Bruns, Martin; Piffer, Michele

Working Paper
Bayesian structural VAR models: A new approach for prior beliefs on impulse responses

DIW Discussion Papers, No. 1796

Provided in Cooperation with:
German Institute for Economic Research (DIW Berlin)

Suggested Citation: Bruns, Martin; Piffer, Michele (2019) : Bayesian structural VAR models: A new approach for prior beliefs on impulse responses, DIW Discussion Papers, No. 1796, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin

This Version is available at:
http://hdl.handle.net/10419/195179

Terms of use:
Documents in EconStor may be saved and copied for your personal and scholarly purposes.
You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.
If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.
Bayesian Structural VAR Models: A New Approach for Prior Beliefs on Impulse Responses

Martin Bruns and Michele Piffer
Opinions expressed in this paper are those of the author(s) and do not necessarily reflect views of the institute.
Bayesian Structural VAR models: a new approach for prior beliefs on impulse responses

Martin Bruns† Michele Piffer‡

March 14, 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.


Keywords: Sign restrictions, Bayesian inference, Oil market.

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

† 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 infinite 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 Normal-inverse-Wishart-Uniform prior (hereafter NiWU) in order to simplify the analysis of the posterior distribution (Uhlig, 2005, Rubio-Ramirez et al., 2010). However, this limits the type of prior information introduced by the researcher to the one that can be modelled by the Normal-inverse-Wishart-Uniform 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. This implies that prior distributions on structural parameters should be selected carefully (Poirier, 1998, Baumeister and Hamilton, 2015). 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 in 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 one implied in the NiWU approach, but the sampling of the corresponding posterior distribution does not become technically more involving 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 show 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, forthcoming). 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 conditionally uniformly distributed. 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 structural 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, potentially leading to new results in applied work. 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 such a 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 restriction is 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 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 very strong 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 demand shocks, we find that 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 helps developing the intuition behind our approach by showing 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 write 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 be used also with prior
beliefs.

2.1 Parametrizations of the structural VAR

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

\[ Ay_t = a_0 + \sum_{l=1}^{p} A_l y_{t-l} + \epsilon_t, \]

\[ = A_+ w_t + \epsilon_t, \quad \epsilon_t \sim N(0, I_k), \tag{1} \]

where \( y_t \) is a \( k \times 1 \) vector of endogenous variables, \( \epsilon_t \) is a \( k \times 1 \) vector of structural shocks, and \( w_t = (1, y'_{t-1}, \ldots, y'_{t-p})' \) is an \( m \times 1 \) vector of the constant and \( p \) lags of the variables, with \( m = kp + 1 \). The \( k \times k \) matrices \( A, A_1, \ldots, A_p \) are gathered in \( A_+ = [a_0, A_1, \ldots, A_p] \), which is of dimension \( k \times m \). We normalize the covariance matrix of \( \epsilon_t \) to the identity matrix.\(^1\) The reduced form representation of the data is

\[ y_t = \pi_0 + \sum_{l=1}^{p} \Pi_l y_{t-l} + u_t, \]

\[ = \Pi w_t + u_t, \quad u_t \sim N(0, \Sigma), \tag{2} \]

where it holds that \( u_t = B \epsilon_t \) and \( \Sigma = BB' \). Matrices \( A, A_+ \) and \( B \) contain structural parameters while \( \Pi \) and \( \Sigma \) represent reduced form parameters. Orthogonal matrices \( Q \), which by construction satisfy \( QQ' = I_k \), allow for the mapping from reduced form to structural parameters, with

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

and \( h(\Sigma) \) a factorization of \( \Sigma \) satisfying \( h(\Sigma)h(\Sigma)' = \Sigma \), for example the Cholesky factorization. Matrix \( A \) in equation (1) highlights the contemporaneous relations among

\(^1\)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).
the variables in the system. Yet, one can premultiply both sides of equation (1) by $B = A^{-1}$ and obtain

$$y_t = \pi_0 + \sum_{l=1}^{p} \Pi_l y_{t-l} + B \epsilon_t,$$

$$= \Pi w_t + B \epsilon_t, \quad \epsilon_t \sim N(0, I_k),$$

(4)

with $\pi_0 = A^{-1} a_0$, $\Pi_l = A^{-1} A_l$ and $\Pi = [\pi_0, \Pi_1, ..., \Pi_p] = A^{-1} A_+$. Matrix $B$ in equation (4) captures the contemporaneous effects of one standard deviation shocks, while future horizons of the impulse responses are calculated using model (4) recursively.\(^2\)

In this paper, we specify the structural VAR as in Uhlig (2005) using model (4) rather than (1) 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.

2.2 The NiWU approach used in the literature

Define $\pi = vec(\Pi)$ as the $km \times 1$ vector that stacks the columns of $\Pi$. The most popular approach for sign restricted SVAR models starts from prior beliefs on the parameters $(\pi, \Sigma, Q)$. In fact, as already known in the literature, when $p(\pi, \Sigma)$ falls within the independent Normal-inverse-Wishart family, the joint posterior distribution

\(^2\)Whether the model is more conveniently expressed as model (1) or (4) (or even as a combined form) depends on whether the identifying restrictions introduced by the researcher are more naturally expressed on $A$ or $B$. 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 (2018). In the terminology used by Lütkepohl (2005), model (1) represents the $A$ form of the SVAR while model (4) represents the $B$ form. 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.
$p(\pi, \Sigma|Y)$ can be conveniently explored using a Gibbs sampler.\footnote{We do not restrict the NiWU approach to the special and 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 $V_\pi$, the distribution $p(\pi)$ in the independent prior specification 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. Extending our approach to conjugate priors only requires minor modifications.} Formally, with prior beliefs
\begin{equation}
 p(\pi, \Sigma, Q) = p(Q|\pi, \Sigma) \cdot p(\pi) \cdot p(\Sigma), \tag{5}
 \end{equation}
given
\begin{align*}
 \pi & \sim N(\mu_\pi, V_\pi), \tag{6} \\
 \Sigma & \sim iW(d, S), \tag{7} \\
 Q|\pi, \Sigma & \sim U, \tag{8}
\end{align*}
it holds that
\begin{equation}
 p(\pi, \Sigma, Q|Y) = p(Q|\pi, \Sigma) \cdot p(\pi, \Sigma|Y), \tag{9}
 \end{equation}
with
\begin{align*}
 \pi|Y, \Sigma & \sim N(\mu^{*}_\pi, V^{*}_\pi), \tag{10} \\
 \Sigma|Y, \Pi & \sim iW(d^{*}, S^{*}), \tag{11} \\
 Q|\pi, \Sigma & \sim U, \tag{12}
\end{align*}
(see Section C of the Appendix for the derivations). Define $Q$ the set of all orthogonal matrices, and $Q_\Sigma$ the subset of $Q$ which implies that the sign restrictions on the structural parameters are satisfied, given a draw of $(\Sigma)$. According to distribution (8), orthogonal matrices in the admissible set $Q_\Sigma$ are treated as equally plausible. The algorithm typically used to conduct posterior analysis with this Normal-inverse-Wishart-Uniform prior specification is summarized in Algorithm 1:
Algorithm 1 (Normal-inverse-Wishart-Uniform approach):

1. run a Gibbs sampler to explore $p(\pi, \Sigma|Y)$ based on the distributions (10) and (11) using $n_1 + n_2$ replications, with $n_1$ the number of burn-in replications and $n_2$ the number of retained replications. Store the retained draws in $\{\pi_d, \Sigma_d\}_{d=1}^{n_2}$;

2. randomly extract $(\pi, \Sigma)$ from $\{\pi_d, \Sigma_d\}_{d=1}^{n_2}$ with replacement;

3. draw an orthogonal matrix $(Q_d)$ using the method by Rubio-Ramirez et al. (2010), which extracts uniformly from the space $Q$, and map $(\pi_d, \Sigma_d, Q_d)$ into the structural parameters of interest;

4a. if $(\pi_d, \Sigma_d, Q_d)$ satisfy the sign restrictions (up to sign and ordering of the shocks), store $(\pi_d, \Sigma_d, Q_d)$ and proceed to Step 5;

4b. if $(\pi_d, \Sigma_d, Q_d)$ do not satisfy the sign restrictions, repeat Step 3 up to $n_3$ times. Stop as soon as $(\pi_d, \Sigma_d, Q_d)$ satisfies the sign restrictions and proceed to Step 5, otherwise discard $(\pi_d, \Sigma_d)$ and move back to Step 2;

5. repeat Steps 2 to 4 until a desired number of draws $n_4$ is obtained.\(^4\)

The convenience of the NiWU approach is that efficient algorithms like Algorithm 1 exist for the sampling of the posterior distribution, and they simplify even further if a conjugate prior is used (see footnote 3). In addition, the possibility of discarding

\(^4\)Algorithm 1 is close in spirit to the algorithms used by Uhlig (2005), Rubio-Ramirez et al. (2010) and Giacomini and Kitagawa (2015). Dismissing draws of $Q$ in Step 4b numerically approximates the extraction from a uniform distribution in $Q_\Sigma$, given that the algorithm by Rubio-Ramirez et al. (2010) draws in $Q$ (an alternative approach is discussed in Amir-Ahmadi and Drautzburg, 2018). For $n_3$ sufficiently large in Step 4b, Algorithm 1 delivers draws $(\pi, \Sigma)$ that approximate $p(\pi, \Sigma|Y)$ from equations (10)-(11) up to an indicator function taking value of one if the reduced form draws imply a non-empty set $Q_\Sigma$. By contrast, when $n_3$ is too small, an additional tilting of these distributions is introduced, since draws $(\pi_1, \Sigma_1)$ and $(\pi_2, \Sigma_2)$ such that $p(\pi_1, \Sigma_1|Y) = p(\pi_2, \Sigma_2|Y)$ are not necessarily retained with equal probability if $Q_{\Sigma_1}$ and $Q_{\Sigma_2}$ differ considerably in size. The algorithm changes if zero restrictions are introduced, see Binning (2013), Arias et al. (2018), and Kilian and Lütkepohl (2017), Chapter 13.
undesired draws allows for the straightforward introduction of sign restrictions not only on contemporaneous impulse responses, but also on future horizons (in which case $Q_\Sigma$ is replaced with $Q_{\pi,\Sigma}$). 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, the implicit prior distribution matters also in a large sample and must be selected carefully (Baumeister and Hamilton, 2015).

The above intuition can be expressed more formally as follows. Consider for simplicity the case of some set of desired sign restrictions of the contemporaneous impulse responses. These 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 infinite probability distributions \{\{p(B)^1, p(B)^2, p(B)^3, \ldots\}\} 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, \ldots\}\} 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). Consider, for example, the contemporaneous impulse responses. Under the NiWU approach the flexibility on $p(B)$ is constrained by the fact that it is expressed indirectly via $p(\Sigma, Q)$, that $p(\Sigma)$ must be the inverse Wishart probability distribution, and that $p(Q|\Sigma)$ is uniform.\footnote{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 C in the Appendix.}

### 2.3 The Np(B) approach proposed in this paper

To overcome the limitations discussed in the previous section, we propose to express 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.
2.3.1 Prior beliefs expressed directly on $(\pi, B)$

We parametrize the model as in equation (4) and express independent prior beliefs

$$p(\pi, B) = p(\pi) \cdot p(B).$$  \hfill (13)

Since $\pi$ is identified, $p(\pi)$ matters less compared to $p(B)$, as long as the sample is sufficiently long. Hence, as also in the NiWU approach from Section 2.2, we restrict $p(\pi)$ to

$$\pi \sim N(\mu_\pi, V_\pi),$$  \hfill (14)

with $\mu_\pi$ and $V_\pi$ not 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.\(^6\)

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

$$p(\pi, B|Y) = p(\pi|B,Y) \cdot p(B|Y),$$  \hfill (15)

$$\pi|B,Y \sim N(\mu_\pi^*, V_\pi^*),$$  \hfill (16)

$$p(B|Y)_{\text{Np}(B)} \propto p(B) \cdot |\text{det}(B)|^{-\frac{T}{2}} \cdot |\text{det}(V_\pi^*)|^{\frac{1}{2}} \cdot e^{-\frac{1}{2} \left\{ \tilde{y}' \left( I_T \otimes (BB')^{-1} \right) \tilde{y} - \mu_\pi^* V_\pi^* \mu_\pi^* \right\}},$$  \hfill (17)

\(^6\)The approach discussed in this section can be extended without difficulty to the case in which $\mu_\pi$ and $V_\pi$ depend on $B$. Indeed, the derivations in Section D of the Appendix consider the general case in which the independence restriction is not introduced. 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 positive number, as in Baumeister and Hamilton (2015) and Baumeister and Hamilton (forthcoming). 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 D.2 of the Appendix.
with

\[ V^*_\pi = [V^{-1}_\pi + (WW' \otimes (BB')^{-1})]^{-1}, \]

\[ \mu^*_\pi = V^*_\pi \cdot [V^{-1}_\pi \mu + W \otimes (BB')^{-1}] \tilde{y}, \]

and \( \tilde{y} = \text{vec}([y_1, \ldots, y_t, \ldots, y_T]) \), \( W = [w_1, \ldots, w_t, \ldots, w_T] \). The suffix \( Np(B) \) indicates that the distribution refers to our approach, rather than to the NiWU approach. 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 \( \pi|B,Y \) 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. Since sign restrictions on impulse responses are frequently introduced contemporaneously rather than on future horizons, we do not view our framework as particularly restrictive. On the other hand, as also the NiWU approach, it makes the analysis more tractable by using a normal prior distribution on \( \pi \), which is not restrictive except in small samples. In addition, by parametrizing the model in \( \Pi \) rather than in \( A_+ = B \Pi \), it makes it straightforward to use the prior by Litterman (1986) (which is applied directly on \( \pi \)), simplifying the analysis compared to Sims and Zha (1998) and Baumeister and Hamilton (2015).\(^7\)

\(^7\)Parametrizing the model in \( \Pi \) rather than in \( A_+ \), equations (1) and (4), simplifies the comparison of the NiWU and the Np(B) approach, as \( \pi = \text{vec}(\Pi) \) appears in both parametrizations. The parametrization in \( \Pi \) rather than \( A_+ \) does not affect the number of parameters for which the algorithm is required, because both \( p(\Pi|B,Y) \) and \( p(A_+|B,Y) \) 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 \( (A, A_+) \), combined with prior independence across the parameters in the different structural equations of the model, allows for posterior simulation equation by equation. In this paper, the presence of \( B \neq I_k \) prevents from breaking the analysis equation by equation.
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 $p(B|Y)_{Np(B)}$ from equation (17). When prior beliefs $p(B)$ take the special case implicit in the NiWU approach, the posterior distribution $p(B|Y)_{Np(B)}$ can be explored using Algorithm 1 from Section 2.2. We now develop an extension of Algorithm 1 to allow for a wider class of prior beliefs on $B$.

We build our sampling procedure on importance sampling techniques. Suppose we are interested in sampling from the target distribution $p(\theta)^{\text{target}}$, and suppose we cannot draw from $p(\theta)^{\text{target}}$ directly, but can evaluate it. In addition, suppose that we can extract proposal draws from the importance function $p(\theta)^{\text{importance}}$. To the extent that the importance function fully covers the support of $p(\theta)^{\text{target}}$, we can obtain draws from $p(\theta)^{\text{target}}$ by resampling with replacement the draws $\{\theta_i\}$ obtained from the importance distribution using weights $w(\theta_i) = \frac{p(\theta=\theta_i)^{\text{target}}}{p(\theta=\theta_i)^{\text{importance}}}$ (see for example Koop, 2003, chapter 4). One can assess the performance of the sampler by computing the effective sample size $ESS = \left(\sum_i (w_i/\sum_i w_i)^2\right)^{-1}$, which captures the effective number of draws used to represent the target probability. 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 importance function. If, instead, we cannot ensure that the importance function gives sufficient mass to the support of the target function, the importance function might need to be changed irrespectively of the effective sample size (see the illustrative example in Section A of the Appendix, and the simulation exercise in Section 3).

In principle, one could use the posterior distribution from the NiWU approach as an importance function to study the posterior distribution associated with our more general approach, setting $p(\theta)^{\text{target}} = p(B|Y)_{Np(B)}$ (equation 17) and $p(\theta)^{\text{importance}} = p(B|Y)_{NiWU}$ (the posterior distribution on $B$ implicit in equations 10-12). 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 $p(B|Y)_{NiWU}$. 

13
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.

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

- $p(\Sigma|Y)_{NiWU}$: posterior distribution on $\Sigma$ corresponding to the NiWU approach;
- $p(\Sigma|Y)_{Np(B)}$: posterior distribution on $\Sigma$ implied by $p(B|Y)_{Np(B)}$ from equation (17), corresponding to our Np(B) approach;
- $p(Q|\Sigma)_{NiWU}$: conditional distribution on $Q$ employed in the NiWU approach, which coincides with a uniform distribution on the space $Q_\Sigma$;
- $p(Q|\Sigma)_{Np(B)}$: conditional distribution on $Q$ implicit in the prior $p(B)$ from our approach.

Then, notice that drawing from $p(B|Y)_{Np(B)}$ is equivalent to drawing from $p(\Sigma|Y)_{Np(B)}$ and mapping $\Sigma$ into $B$ using draws of $Q$ from $p(Q|\Sigma)_{Np(B)}$. Last, consider the following importance sampling procedure. First, explore $p(\Sigma|Y)_{Np(B)}$ using $p(\Sigma|Y)_{NiWU}$ as an importance function. Since $\Sigma$ 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 $Q_\Sigma$, 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 make $p(\Sigma|Y)_{NiWU}$ and $p(\Sigma|Y)_{Np(B)}$ similar, a conjecture that can be verified by computing the effective sample size (see also Section D.2 of the Appendix for a discussion of diagnostic procedures). In the second stage, a low effective sample size only suggests to increase the number of draws from the importance function.
Section D.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 $p(B)$. The sampler can then be implemented using the following algorithm:

Algorithm 2 (algorithm proposed for sign restrictions):

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

1. run a Gibbs sampler to explore $p(\pi, \Sigma|Y)$ based on the distributions (10) and (11) using $m_1 + m_2$ replications, with $m_1$ the number of burn-in replications and $m_2$ the number of retained replications (same as Algorithm 1). Store the retained draws in $\{\Sigma_d\}_{d=1}^{m_2}$.
2. for each $\Sigma_d$ compute weights
   \[
   w_{d}^{\text{stage A}} = \frac{p(\Sigma = \Sigma_d|Y)_{Np(B)}}{p(\Sigma = \Sigma_d|Y)_{NiWU}} \propto |\text{det}(\Sigma_d)|^{\frac{d+K}{2}} \cdot e^{\frac{1}{2}tr[\Sigma_d S]}. \tag{20}
   \]
   Assess if the effective sample size $ESS^A = \left(\sum_d \left(\frac{w_{d}^{\text{stage A}}}{\sum_d (w_{d}^{\text{stage A}})}\right)^2\right)^{-1}$ is sufficiently high;

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

3. randomly select $\Sigma_d$ 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 $m_3$ times. Stop as soon as $(B_d)$ satisfies the sign restrictions and...
proceed to Step 6, otherwise discard $(\Sigma_d)$ and move back to Step 3;

6. repeat Steps 3 to 5 until $m_4$ draws are stored;

7. for all draws $\{B_d, Q_d\}_{d=1}^{m_4}$ compute weights

$$w_{i}^{\text{stage B}} = \frac{p(Q = Q_d | \Sigma_d)_{\text{NiWU}}}{p(Q = Q_d | \Sigma_d)_{\text{NiWU}}} \propto p(B = B_d).$$

Assess if the effective sample size $\text{ESS}^B = \left( \sum_i (w_i^{\text{stage B}} / \sum_i (w_i^{\text{stage B}}))^2 \right)^{-1}$ is larger than a desired minimum number $m_5$. If so, proceed to Step 8, otherwise move back to Steps 3 to 6 and increase $m_4$;

8. generate $\{B_d\}_{d=1}^{\text{ESS}^B}$ by resampling the draws $\{B_d\}_{d=1}^{m_4}$ from Step 7 with replacement using weights $w_{d}^{\text{stage B}}$.

Algorithm 2 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 Section 3, Section 4 and in the Appendix we document that the sampling time of Algorithm 2 is not excessively longer than the computational time from Algorithm 1. In Section D.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, further suggesting that Algorithm 2 can be used in samples frequently used in applied work.

Algorithm 2 offers a way of implementing sign restrictions. Binning (2013) and Arias et al. (2018) extend a variant of Algorithm 1 to include also zero restrictions by offering algorithms that extract $Q$ from the subspace of $Q$ that satisfies the zero restrictions. Only two modifications are required to extend Algorithm 2 to account for also zero restrictions on $B$. First, the computation of $p(\Sigma | Y)_{\text{NiWU}}$ (required for the weights in Stage A, Step 2) must now account for the fact that the mapping from

$^8$Since $\text{ESS}^B$ can be much smaller than $m_4$, we resample $\{B_d\}_{d=1}^{m_4}$ only $\text{ESS}^B$ times (or the closest integer), rather than $m_4$ times. This avoids unnecessary repetitions.
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 B, Step 4), must now be replaced with the methods by either Binning (2013) or Arias et al. (2018). The existing version of Algorithm 2 can then be applied to the case in which zero restrictions are introduced on one structural shock of interest.\(^9\)

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 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 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 E 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 \(p(B)\) 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 \(p(B)\) is challenging, because the literature still provides

\(^9\)In Algorithm 2, extracting Q matrices from the algorithm by Rubio-Ramirez et al. (2010) ensures that the orthogonal parameter space \(Q\) is explored uniformly, and hence is fully explored. 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 (Arias et al., 2016). In this case, the full relevant orthogonal parameter space is still explored, and no modification to the computation of weights in Stage B is required. When zero restrictions are introduced on the effects of more than one shock of interest, the distribution \(p(Q|\Sigma)_{NWU}\) must be evaluated numerically, and the possibility that the relevant part of the orthogonal space is not explored must be addressed.
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 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.

To overcome this challenge, we propose a prior specification for $p(B)$ that builds on a conventional prior specification used in the literature for $p(\pi)$ known as the Minnesota prior (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 $\sigma_i$ of the residual on univariate AR processes on each variable. Then, Bayesian shrinkage is introduced through a set of hyperparameters that shrink the parameters in $\pi$ 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

$$-\Sigma_{ii}^{0.5} \leq b_{ij} \leq \Sigma_{ii}^{0.5},$$  \hspace{2cm}(22)$$

with $\Sigma_{ii}$ the $i$th element of the diagonal of $\Sigma$.\footnote{Given $\Sigma = BB'$, the equations corresponding to the diagonal elements of $\Sigma$ are $\Sigma_{ii} = b_{i1}^2 + b_{i2}^2 + \ldots + b_{ik}^2$. Since $\Sigma_{ii}$ is nonnegative and since $b_{ij}^2 \geq 0$, each element $b_{ij}$ must satisfy $-\Sigma_{ii}^{0.5} \leq b_{ij} \leq \Sigma_{ii}^{0.5}$. Accordingly, $\gamma_i = \hat{\Sigma}_{ii}^{0.5}$ provides}$

Accordingly, $\gamma_i = \hat{\Sigma}_{ii}^{0.5}$ provides the $i$th moment of $\Sigma_{ii}$.
a candidate assessment of the upperbound for $b_{ij}$, where $\hat{\Sigma}$ is an estimate based on a training sample. We then introduce two hyperparameters $\psi_1$ and $\psi_2$ that control for the location and the spread 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 $(-\psi_2 \gamma_i, \psi_2 \gamma_i)$;

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 truncated on the positive support 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 truncated on the negative support has 95% prior mass in the space $(-\psi_2 \gamma_i, 0)$.

Put differently, $\psi_1$ controls for the first moment of the prior given that 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), while $\psi_2$ 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 $\gamma_i$, and then introduces Bayesian shrinkage explicitly through the hyperparameters $\psi_1$ and $\psi_2$. The hyperparameters $\psi_1$ and $\psi_2$ can be treated hierarchically, further providing flexibility on the prior distribution used. If sign restrictions do not identify all shocks, we suggest to numerically introduce the restriction that the non-identified shocks do not replicate the sign restrictions of the identified shocks. Alternative specifications are also possible.
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 (4) 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 set 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).\footnote{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 $S$ such that $E(\Sigma)$ equals the diagonal matrix displaying the}
For the Np(B) approach, Section 2.3, we specify \( p(B) \) as discussed in Section 2.4, setting \( \psi_1 = 0.8 \) and \( \psi_2 = 1.2 \) for the illustration. We set \( \mu_\pi = 0 \) in equation (6) for all models estimated and use the flat prior specification implied by \( (V_\pi)^{-1} = 0 \). All 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 Algorithm 2. Figure 1 shows the results for the (1, 2) entry of \( \Sigma \) and \( B \) for some of the datasets considered (see Section F of the Appendix for the full illustration). The left column of Figure 1 shows the results for \( \Sigma \) 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 Algorithm 2, 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 are these two empirical distributions, the more Algorithm 2 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 D2 in the Appendix for how we set the tuning parameters required in Algorithm 2, and Table F4 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 \( p(\Sigma|Y)_{NiWU} \) and \( p(\Sigma|Y)_{Np(B)} \) to be similar (dashed and solid lines, respectively). However, as the sample size increases, the likelihood dominates the prior, making \( p(\Sigma|Y)_{NiWU} \) an excellent importance function for \( p(\Sigma|Y)_{Np(B)} \) already.
for $T = 60$. The right column of the figure displays how successful Algorithm 2 is in sampling the posterior $p(B|Y)_{Np(B)}$. As the data reveals information on $\Sigma$, it implicitly defines the bounds $[-\Sigma_{i,j}^{0.5}, +\Sigma_{i,j}^{0.5}]$ within which $b_{i,j}$ must be (see equations (22), as well as Figure F17 in the Appendix). Within these bounds, the marginal distributions $p(b_{i,j}|Y)_{Np(B)}$ reflect a combination of prior beliefs and information on the identified set $\{B : \Sigma = BB'\}$. We find it interesting that while for $T = 30$ the distribution $p(\Sigma|Y)_{Np(B)}$ is still quite different from the distribution generating proposal draws, the associated sampling of $p(B|Y)_{Np(B)}$ is already close to what is detected by the Dynamic Striated Metropolis-Hastings algorithm.

Note: See Figure F4 to Figure F8 in the Appendix for the full illustration.
Table 1: Performance of our algorithm

<table>
<thead>
<tr>
<th>T</th>
<th>T − p</th>
<th>Stage A</th>
<th>Stage B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>proposal</td>
<td>effective</td>
</tr>
<tr>
<td></td>
<td></td>
<td>draws</td>
<td>sample size</td>
</tr>
<tr>
<td></td>
<td></td>
<td>m₂</td>
<td>ESS_A</td>
</tr>
<tr>
<td>30</td>
<td>22</td>
<td>50,000</td>
<td>476</td>
</tr>
<tr>
<td>60</td>
<td>52</td>
<td>50,000</td>
<td>28,846</td>
</tr>
<tr>
<td>120</td>
<td>112</td>
<td>50,000</td>
<td>40,407</td>
</tr>
<tr>
<td>240</td>
<td>232</td>
<td>50,000</td>
<td>45,233</td>
</tr>
<tr>
<td>480</td>
<td>472</td>
<td>50,000</td>
<td>47,692</td>
</tr>
</tbody>
</table>

Table 1 provides additional intuition behind our algorithm by reporting relevant metrics from the sampler. As should be expected, the higher is the sample size of the dataset, the higher is 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 neither expected nor required for the sampler to successfully explore the posterior distribution. As Table 1 shows, approximately 30% of the initial draws are effectively used already for T = 60, a ratio that does not change as the sample size increases.

3.3 Comparison to the NiWU approach

Having discussed the key intuition of the sampler, we now illustrate what drives the differences 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), see Section C.4 in the Appendix for a discussion.

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. As we see from Figure 2, the prior distributions on Σ₁,₂ are quite different, but the

---

12The prior distribution in the NiWU case with the improper prior specification is approximated using \( d = k + 2 \) and \( S = 0.01 \cdot I₂ \), see Figure F10 in the Appendix.
Figure 2: Comparison to the NiWU approach

Note: See Figure F9 to Figure F16 in the Appendix for the full illustration.

associated posterior distributions are very similar already for $T = 60$. By contrast, since $B$ is not identified, differences between the NiWU and the Np(B) approach in prior beliefs on $B$ 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 different parametrizations of the prior under the NiWU approach are quite similar already for $T = 60$. This occurs because, upon learning from the data about $\Sigma$, 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 $p(B|Y)_{NiWU}$
**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 $Q$ an draw from either $p(Q|\Sigma)_{NiWU}$ or $p(Q|\Sigma)_{Np(B)}$. See Section D.2 of the Appendix for further details.

despite the fact that $p(B)_{Np(B)}$ is not tighter 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 $Q_{\Sigma}$.

The intuition behind the differences between the Np(B) and the NiWU approach 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 $\theta$ 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 $\theta$. The top-left plot of Figure 3 shows that indeed the distribution of the angle of
Table 2: Comparison of the computational time

<table>
<thead>
<tr>
<th></th>
<th>Np(B) approach</th>
<th>NiWU approach</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Our algorithm</td>
<td>DSMH algorithm</td>
</tr>
<tr>
<td></td>
<td>Stage A</td>
<td>Stage B</td>
</tr>
<tr>
<td>Stage A</td>
<td>T</td>
<td>h</td>
</tr>
<tr>
<td>30</td>
<td>15</td>
<td>22</td>
</tr>
<tr>
<td>60</td>
<td>19</td>
<td>17</td>
</tr>
<tr>
<td>120</td>
<td>27</td>
<td>17</td>
</tr>
<tr>
<td>240</td>
<td>29</td>
<td>16</td>
</tr>
<tr>
<td>480</td>
<td>38</td>
<td>17</td>
</tr>
</tbody>
</table>

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 rotation matrices that replicate draws of $Q$ from the algorithm by Rubio-Ramirez et al. (2010) is the uniform distribution in the support $[-\pi/2, \pi/2]$. Conditioning on $\Sigma_{true}$, the rotation angles consistent with the sign restrictions are the subset shown in the bottom-left plot of the figure, which correspond to $Q_{\Sigma_{true}}$. While the NiWU approach treats such angles as equally plausible, the Np(B) approach does not, preferring instead to take 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 (22), 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 (equation (33) in Baumeister and Hamilton, 2015 and their Figure 1), while the Np(B) approach implies a distribution that reflects the explicit prior $p(B)$. As the sample size increases, the posterior distributions $p(B|Y)_{NiWU}$ and $p(B|Y)_{Np(B)}$ approach the ones displayed in Figure 3.13

13Figure 3 shows the analysis conditioning on $\Sigma_{true}$. As the sample size increases, both $p(\Sigma|Y)_{NiWU}$ and $p(\Sigma|Y)_{Np(B)}$ collapse to a point mass at $\Sigma_{true}$, making the analysis conditioning on $\Sigma_{true}$ 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 used 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 on the results. It is, instead,
Table 2 compares the computation time of the NiWU and the Np(B) approach, i.e. of Algorithm 1 from Section 2.2, Algorithm 2 from Section 2.3, and the Dynamic Striated Metropolis-Hastings algorithm. All applications of Algorithm 1 and Algorithm 2 take less than a minute to run on Matlab, and Algorithm 2 does not take considerably longer to run compared to Algorithm 1. 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 by Kilian and Murphy (2012). The exercise also illustrates how much 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, and use a flat prior on $\pi$, setting $V_{\pi}^{-1} = 0$ in equation (6).
Table 3: Sign restrictions on the contemporaneous impulse responses

<table>
<thead>
<tr>
<th>A) Sign restrictions used</th>
<th>oil supply shock</th>
<th>aggregate demand shock</th>
<th>oil demand shock</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oil production</td>
<td>–</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>economic activity</td>
<td>–</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>real price of oil</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

B) Prior distributions modelling the sign restrictions

| Prior   | ψ1  | ψ2  | Prob(|b_{i,j}| > γ_i) |
|---------|-----|-----|-----------------------|
| A       | 2   | 4   | 0.83                  |
| B       | 1   | 2   | 0.53                  |
| C       | 0.8 | 1.2 | 0.20                  |

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 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 ψ_1γ_i, while ψ_2 controls for the second moment of the prior by being positively related to the prior mass allocated for |b_{i,j}| ≥ γ_i. We explore the role of prior beliefs by using the three separate specifications for ψ_1 and ψ_2 illustrated in Table 3. Prior A corresponds to a wide prior that attaches approximately 80% probability mass beyond the estimated γ_i. Priors B and C progressively tighten the prior to make it more consistent with the scaling of the variables, giving approximately 50% and 20% prior probability mass beyond γ_i, respectively. We favour Prior C, 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.\textsuperscript{14}

### 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 improper prior specification. We refer to Section G of the Appendix for the analysis of the other variables in the model, as well as for robustness checks and information on the performance of the sampler.

Figure 4 shows the effect of one-standard-deviation shocks on the price of oil. The columns of the figure differ on the structural shock considered, while the rows differ for whether prior specification A, B or C is used to select the prior associated with our 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 then Antolín-Díaz and Rubio-Ramírez (2018) to introduce further restrictions on elasticities and/or shocks and historical decompositions. We find that the wide posterior bands from the NiWU approach can be interpreted as driven by the high variance introduced through the NiWU prior specification. In fact, since the dataset is large enough to imply very similar posteriors distributions for $\Sigma$, the differences in $p(B|Y)_{NiWU}$ and $p(B|Y)_{Np(B)}$ are due to the difference in $p(Q|\Sigma)_{NiWU}$ and $p(Q|\Sigma)_{Np(B)}$ (see Figure G22 and Figure G28-Figure G30 in the Appendix). Prior A from our Np(B) approach replicates the posterior bands from the NiWU approach at the cost of using a prior distribution on $B$ that attaches as much as 80% prior mass to values of $B$ above the estimated reasonable bound $\gamma_i$. We view this prior mass as too wide given the scaling of the variables. As shown with prior C, tightening $p(B)_{Np(B)}$ to make it more consistent with the scaling of the variables tightens the

\textsuperscript{14}We estimate the scale $\gamma_i$ using a training sample on the first 20\% of the available observations, as in Primiceri (2005). The prior distributions, which are shown in Figure G23 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 G10 for a further illustration of the distribution of the probability mass under the prior distribution.
posterior bands considerably. On the short horizon of the response, the 95% credible bands associated with the Np(B) approach are tighter than the 68% credible bands of the NiWU approach. This suggests that introducing explicit information on the scaling of the variables can make inference much sharper.

Focusing on our tighter prior, Figure 4 highlights several results. 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. While we confirm 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 a point-identified model that minimizes the distance between the elasticities implied by the VAR model and external estimates. Yet, 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 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 G43 in the Appendix shows that the posterior distribution 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 decomposition, 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, 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 prior mass
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 G35 to Figure G37 in the Appendix for the full analysis.

As we move our prior from specification \( A \) to \( C \), we find that the unexpected variations in the price of oil are mainly driven by supply shocks and aggregate demand shocks for approximately 30-40\% and 40-60\%, respectively, 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).
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 G40 to Figure G42 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. The top panel of 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 the main 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 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 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 our approach delivers sharper inference. Consistent with Baumeister and Hamilton (forthcoming) 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


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


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

APPENDIX

Martin Bruns* Michele Piffer†

March 14, 2019

Contents

A Illustrative preliminary examples 1
   A.1 The role of prior beliefs for identified and nonidentified parameters . 1
   A.2 Interpreting the effective sample size in importance sampling . . . . 2

B Likelihood function of the model 5

C NiWU approach used in the literature 6
   C.1 Derivations of $p(\Sigma|\pi, Y)$ . . . . . . . . . . . . . . . . . . . . . . 8
   C.2 Derivations of $p(\pi|\Sigma, Y)$ and $p(\Sigma|Y)$ . . . . . . . . . . . . . . 10
   C.3 Derivations of $p(B)$ and $p(B|Y)$ . . . . . . . . . . . . . . . . . . . . . 11
   C.4 More on Algorithm 1 . . . . . . . . . . . . . . . . . . . . . . . . . . . . 12

D Np(B) approach proposed in the paper 13
   D.1 Derivations of $p(B|Y)$ and $p(\pi|B, Y)$ . . . . . . . . . . . . . . . . 13
   D.2 The importance sampler . . . . . . . . . . . . . . . . . . . . . . . . . . . 16
   D.3 Diagnostics for the Importance Sampler . . . . . . . . . . . . . . . . . . 20
      D.3.1 Graphical Assessment . . . . . . . . . . . . . . . . . . . . . . . . . 20
      D.3.2 Diagnostic tests: Wald test, Score test and LR test . . . . . . . 20

E The Dynamic Striated Metropolis-Hastings algorithm by Waggoner et al. (2016) 23
   E.1 The sampler . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 23
   E.2 Convergence criteria . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 27

*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
A Illustrative preliminary examples

In this section we provide two preliminary examples to illustrate some key intuitions recurrently used in the paper. The illustrations are intentionally heuristic and do not aim for a formal analysis, which goes beyond the purpose of this appendix.

A.1 The role of prior beliefs for identified and nonidentified parameters

A fundamental distinction in the paper is the distinction between identified and non-identified parameters. For the former, the prior probability distribution does not matter exidingly except in a finite sample. By contrast, for the latter the prior probability distribution is crucial, because the likelihood function does not fully dominate on the prior, even in a large sample.

As an example for this difference, generate $T$ independent observations from the model $x_t \sim N(\alpha_0, 1)$, setting $\alpha_0 = 3$. The likelihood function of the model is

$$p(\{x_t\}_{t=1}^T|\alpha) \propto e^{-0.5\sum_{t=1}^T(x_t - \alpha)^2},$$

and reaches its maximum at $\hat{\alpha}_{MLE} = \sum_{t=1}^T x_t/T$. Then consider two different prior distributions on $\alpha$, $\alpha \sim N(0, 1)$ and $\alpha \sim N(6, 6)$. Figure A1 compares the prior distributions with the likelihood function and the posterior distributions for $T = 10$ and $T = 30$. Since the parameter $\alpha$ is identified, the likelihood function has a single mode, which approaches $\alpha_0$ as the sample size increases. Accordingly, the divergence in prior beliefs becomes progressively irrelevant, as the likelihood dominates the priors.

Consider now a second example. Generate $T$ independent observations from the model $x_t \sim N(1, \beta_0)$, setting $\beta_0 = -3$. The likelihood function of the model is now

$$p(\{x_t\}_{t=1}^T|\beta) \propto |\beta^2|^{-\frac{1}{2}}e^{-0.5\sum_{t=1}^T(x_t - 1)/\beta^2},$$

and reaches its maximum at $\hat{\beta}_{MLE} = \pm(\sum_{t=1}^T(x_t - 1)^2/T)^{0.5}$. Then consider two different prior distributions on $\beta$, $\beta \sim N(-2, 1)$ and $\beta \sim N(2, 1)$. Figure A2 shows the likelihood function and the posterior distributions for $T = 10$ and $T = 30$. Since the parameter $\beta$ is only identified up to sign, the likelihood function does not have a single mode, but two modes, which approach $\pm\beta_0$ as the sample size increases. Accordingly, the divergence in prior beliefs remains even as the sample size increases, and posterior distributions do not coincide.

We refer the interested reader to Poirier (1998) for a detailed discussion on the conditions under which beliefs on identified and non-identified parameters are updated by the data.
A.2 Interpreting the effective sample size in importance sampling

Another essential intuition behind the analysis in the paper is that in an importance sampler, a low effective sample size (ESS) is not necessarily a bad outcome as long as one can argue that the importance function fully covers the relevant support of the target function.

As an illustration, consider sampling the target probability distribution

\[
p(\theta) \propto e^{-\frac{1}{2}(\theta-3)^2} + e^{-\frac{1}{2}(\theta+2)^2}. \tag{A.1}
\]
Figure A2: Illustrative example: non identified parameter

A) $T = 10$

B) $T = 30$

Note: The model generating the data is $x_t \sim N(1, \beta_0)$, $\beta_0 = -3$.

The top left and right panels of Figure A3 show that the target probability distribution is bi-modal, with peaks at the values $-2$ and $3$. Consider now the case shown in the left column of Figure A3, in which we explore the target function using the distribution $N(1, 100)$ as an importance function. This importance function has a very wide support that fully covers the support of the target function. While the proposal draws are widely spread out, the reweighted draws well approximate the target distribution, capturing the bi-modal nature of such function. In this application, the initial 5,000 draws lead to effectively obtain only 1,365 draws, corresponding to a relatively low ratio of 27.30%. However, this low ratio only suggests that to effectively obtain a desired number of draws from the target distributing we roughly need to extract four times as many draws from the proposal distribution. It does not imply that the importance
Consider instead the case of using the importance distribution $N(3, 1)$. The right column of Figure A3 shows that this distribution has non-negligible probability mass only around the right mode of the target distribution. Accordingly, the sampler only succeeds to capture the right mode of the target distribution. In this application, the initial draws of 5,000 lead to effectively 2,000 draws. The ratio of effective to proposal draws equals $40\%$, which is actually above the one associated with the previous importance function despite the lower performance of the sampler. In the case considered on the right column of the figure, the misrepresentation of the target distribution cannot
be alleviated by increasing the number of proposal draws, as the left mode remains unrepresented.

For an introduction to importance sampling we refer the reader to Koop (2003), chapter 4. See also Robert and Casella (2013), chapter 3, as well as Geweke (1989) and Koopman et al. (2009) for a further discussion.

B Likelihood function of the model

To derive the likelihood of the model, start from equation (4) of the paper, which we rewrite here for convenience:

\[ y_t = \Pi w_t + B \epsilon_t, \quad \epsilon_t \sim N(0, I_k). \] (B.2)

\( y_t \) is a \( k \times 1 \) vector of variables, \( \epsilon_t \) is a \( k \times 1 \) vector of structural shocks, \( w_t \) is an \( m \times 1 \) vector of lagged variables and potentially deterministic controls, \( \Pi \) is a \( k \times m \) matrix of reduced form parameters, and \( B \) is a \( k \times k \) matrix of structural parameters. Write the model in compact form as

\[ Y = \Pi W + BE, \] (B.3)

where \( Y = [y_1, ..., y_t, ..., y_T] \) and \( E = [\epsilon_1, ..., \epsilon_t, ..., \epsilon_T] \) are \( k \times T \) matrices of data and shocks, and \( W = [w_1, ..., w_t, ..., w_T] \) is an \( m \times T \) matrix of data. Then, make use of the formula \( \text{vec}(\bar{A}\bar{B}\bar{C}) = (\bar{C}' \otimes \bar{A}) \cdot \text{vec}(\bar{B}) \) (see Lütkepohl, 2005, mathematical appendix) and rewrite the model as

\[ \tilde{y} = Z\pi + (I_T \otimes B)\tilde{\epsilon}, \quad \tilde{\epsilon} \sim N\left(0, (I_T \otimes I_k)\right), \] (B.4)

with \( \tilde{y} = \text{vec}(Y) \) and \( \tilde{\epsilon} = \text{vec}(E) \) of dimension \( kT \times 1 \) and \( Z = (W' \otimes I_k) \) of dimension \( kT \times mk \). The \( mk \times 1 \) vector \( \pi = \text{vec}(\Pi) \) stacks the columns of \( \Pi \) vertically. Last, rewrite the model in reduced form, obtaining

\[ \tilde{y} = Z\pi + \tilde{u}, \quad \tilde{u} \sim N\left(0, \tilde{\Sigma}\right), \] (B.5)

with \( \tilde{u} = (I_T \otimes B)\tilde{\epsilon} \). \( \tilde{\Sigma} \) and \( B \) are related through the equality

\[ \tilde{\Sigma} = (I_T \otimes BB'). \] (B.6)
The likelihood function can now be written in $B$ and $\pi$ as

\[
p(Y|\pi, B) = (2\pi)^{-\frac{kT}{2}} |\det(I_T \otimes BB')|^{-\frac{1}{2}} e^{-\frac{1}{2}(\tilde{y} - Z\pi)'(I_T \otimes BB')^{-1}(\tilde{y} - Z\pi)}, \quad (B.7)
\]

\[
= (2\pi)^{-\frac{kT}{2}} |\det(BB')|^{-\frac{T}{2}} e^{-\frac{1}{2}(\tilde{y} - Z\pi)'(I_T \otimes BB')^{-1}(\tilde{y} - Z\pi)}. \quad (B.8)
\]

C NiWU approach used in the literature

In this section we provide the derivations related to the Normal-inverse-Wishart-Uniform approach discussed in Section 2.2 of the paper. We make the analysis more general by discussing the independent prior specification, rather than the conjugate prior, see footnote 3 of the paper.

Rewrite equation (B.8) in $\Sigma$ rather than $B$ and obtain the likelihood function of the reduced form,

\[
p(Y|\pi, \Sigma) = (2\pi)^{-\frac{kT}{2}} |\det(\Sigma)|^{-\frac{T}{2}} e^{-\frac{1}{2}(\tilde{y} - Z\pi)'(I_T \otimes \Sigma)^{-1}(\tilde{y} - Z\pi)}. \quad (C.9)
\]

With the independent prior specification, prior beliefs on reduced form parameters are given by

\[
p(\pi, \Sigma) = p(\pi)p(\Sigma) \quad (C.10)
\]

with

\[
\pi \sim N(\mu_\pi, V_\pi), \quad (C.11)
\]

\[
\Sigma \sim iW(d, S), \quad (C.12)
\]

i.e.

\[
p(\pi) \propto |\det(V_\pi)|^{-\frac{1}{2}} e^{-\frac{1}{2}(\pi - \mu_\pi)'V_\pi^{-1}(\pi - \mu_\pi)}, \quad (C.13)
\]

\[
p(\Sigma) \propto |\det(\Sigma)|^{-\frac{d+k+1}{2}} e^{-\frac{1}{2}tr[\Sigma^{-1}S]}, \quad (C.14)
\]
The joint posterior distribution is then

\[
p(\pi, \Sigma|Y) \propto |\text{det}(\Sigma)|^{-\frac{d^*+k+1}{2}} \cdot e^{\frac{1}{2}(\tilde{y}-Z\pi)'(I_T \otimes \Sigma)^{-1}(\tilde{y}-Z\pi)} \cdot e^{-\frac{1}{2}(\pi-\mu_\pi)'V_\pi^{-1}(\pi-\mu_\pi)} \cdot |\text{det}(\Sigma)|^{-\frac{d^*+k+1}{2}} \cdot e^{-\frac{1}{2}\text{tr}[\Sigma^{-1}S^*]}.
\]  

(C.15)

Draws of $\Sigma$ are mapped into draws of $B$ using

\[B = h(\Sigma)Q,\]

(C.16)

with $Q$ an orthogonal matrix drawn from

\[Q|\Sigma \sim U\]

(C.17)

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

In the rest of this section we show that equations (C.10), (C.11), (C.12), (C.15) and (C.17) imply

\[
p(\Sigma|\pi, Y) \propto |\text{det}(\Sigma)|^{-\frac{d^*+k+1}{2}} \cdot e^{\frac{1}{2}\text{tr}[\Sigma^{-1}S^*]},
\]  

(C.18)

\[
p(\pi|\Sigma, Y) \propto |\text{det}(V_\pi^*)|^{-\frac{1}{2}} \cdot e^{-\frac{1}{2}(\pi-\mu_\pi)'V_\pi^{-1}(\pi-\mu_\pi)}.
\]  

(C.19)

\[
p(\Sigma|Y) \propto |\text{det}(\Sigma)|^{-\frac{d^*+k+1}{2}} \cdot |\text{det}(V_\pi^*)|^{\frac{1}{2}} \cdot e^{-\frac{1}{2}\left\{\tilde{y}'(I_T \otimes \Sigma^{-1})\tilde{y} - \mu_\pi'V_\pi^{-1}\mu_\pi + \text{tr}[\Sigma^{-1}S]\right\}},
\]  

(C.20)

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

(C.21)

\[
p(B|Y) \propto I\{\text{sign}\} \cdot |\text{det}(B)|^{-(d^*+k)} \cdot |\text{det}(V_\pi^*)|^{\frac{1}{2}} \cdot e^{-\frac{1}{2}\left\{\text{vec}(B^{-1})'(S \otimes I_k)\text{vec}(B^{-1}) + \tilde{y}'(I_T \otimes (BB')^{-1})\tilde{y} - \mu_\pi'V_\pi^{-1}\mu_\pi\right\}},
\]  

(C.22)
with

\[ d^* = d + T; \tag{C.23} \]
\[ S^* = S + (Y - \Pi W)(Y - \Pi W)', \tag{C.24} \]
\[ V^*_\pi = [V^{-1}_\pi + (W W' \otimes (B B')^{-1})]^{-1}, \tag{C.25} \]
\[ \mu^*_\pi = V^*_\pi \cdot [V^{-1}_\pi \mu + [W \otimes (B B')^{-1}] \tilde{y}], \tag{C.26} \]
\[ \mu^*_\pi = V^*_\pi \cdot \left[ V^{-1}_\pi \mu + [W \otimes (BB')^{-1}] \tilde{y} \right], \tag{C.27} \]
\[ = V^*_\pi \cdot \left[ V^{-1}_\pi \mu + (WW' \otimes \Sigma^{-1}) \tilde{y} \right], \tag{C.28} \]

(C.18) and (C.19) are standard results in the literature (see Koop et al. (2010)) and are required to run the Gibbs sampler in Algorithm 1, Section 2.2 of the paper. (C.20) is the analytical expression of the marginal posterior for \( \Sigma \) implicit in the joint posterior (C.15) and is required to run the importance sampler introduced in Section D. (C.21) and (C.22) are the prior and the posterior marginal distributions of \( B \) implicit in the NiWU approach. The derivations of (C.21) build on Arias et al. (2018), which we use to derive equation (C.22).

### C.1 Derivations of \( p(\Sigma | \pi, Y) \)

Rewrite the joint posterior distribution from equation (C.15) as

\[
p(\pi, \Sigma | Y) \propto c(\pi) \cdot |\text{det}(\Sigma)|^{-\frac{d+k+1}{2}} \cdot e^{-\frac{1}{2} tr\left[\Sigma^{-1} S\right]} \cdot |\text{det}(\Sigma)|^{-\frac{T}{2}} \cdot e^{-\frac{1}{2} \left\{ (\tilde{y} - Z \pi)' (I_T \otimes \Sigma)^{-1} (\tilde{y} - Z \pi) \right\}}, \tag{C.29} \]

with \( c(\pi) \) a term that gathers elements which depend on \( \pi \) but not on \( \Sigma \). Since

\[
(\tilde{y} - Z \pi)' (I_T \otimes \Sigma)^{-1} (\tilde{y} - Z \pi) =
\]
\[
= (\tilde{y} - (W' \otimes I_k) \cdot \pi)' (I_T \otimes \Sigma^{-1}) (\tilde{y} - (W' \otimes I_k) \cdot \pi) =
\]
\[
= \tilde{y}' (I_T \otimes \Sigma^{-1}) \tilde{y} + \pi'(W' \otimes I_k)' (I_T \otimes \Sigma^{-1}) (W' \otimes I_k) \pi +
\]
\[
- \tilde{y}' (I_T \otimes \Sigma^{-1}) (W' \otimes I_k) \pi - \pi'(W' \otimes I_k)' (I_T \otimes \Sigma^{-1}) \tilde{y} =
\]
\[
= \tilde{y}' (I_T \otimes \Sigma^{-1}) \tilde{y} + \pi'(WW' \otimes \Sigma^{-1}) \pi +
\]
\[
- \tilde{y}' (W' \otimes \Sigma^{-1}) \pi - \pi'(W \otimes \Sigma^{-1}) \tilde{y},
\]
and since
\[ \text{vec}(A')(I \otimes B)\text{vec}(C) = tr(ABC), \]
\[ \text{vec}(A')(D \otimes B)\text{vec}(C) = tr(A'BCD'), \]
\[ tr(AB) = tr(BA) \]
\[ tr(A + B) = tr(A) + tr(B), \]
it holds that
\[
\begin{align*}
\tilde{y}'(I_T \otimes \Sigma^{-1})\tilde{y} &= tr\{Y'\Sigma^{-1}Y\} = tr\{\Sigma^{-1}YY'\}, \\
\pi'(WW' \otimes \Sigma^{-1})\pi &= tr\{\Pi'\Sigma^{-1}\Pi WW'\} = tr\{\Sigma^{-1}\Pi WW'\Pi'\}, \\
\tilde{y}'(W' \otimes \Sigma^{-1})\tilde{y} &= tr\{Y'\Sigma^{-1}YY'\} = tr\{\Sigma^{-1}YY'\}, \\
\pi'(W \otimes \Sigma^{-1})\pi &= tr\{\Pi'\Sigma^{-1}WW'\} = tr\{\Sigma^{-1}YY'\}. 
\end{align*}
\]
Hence,
\[
\begin{align*}
(y - Z\pi)'(I_T \otimes \Sigma)^{-1}(y - Z\pi) &= tr\{\Sigma^{-1}[YY' + \Pi WW'\Pi' - \Pi YY' - YY'\Pi']\} = \\
&= tr\{\Sigma^{-1}(Y - \Pi W)(Y - \Pi W)'\}. 
\end{align*}
\]
It follows that
\[
p(\Sigma|\pi, Y) \propto c(\pi) \cdot |\text{det}(\Sigma)|^{-\frac{d+k+1}{2}} \cdot e^{-\frac{1}{2}tr\left[\Sigma^{-1}S\right]} \cdot |\text{det}(\Sigma)|^{-\frac{T}{2}} \cdot e^{-\frac{1}{2}tr\left[\Sigma^{-1}(Y - \Pi W)(Y - \Pi W)'\right]} \\
& \propto c(\pi) \cdot |\text{det}(\Sigma)|^{-\frac{(d+T)+k+1}{2}} \cdot e^{-\frac{1}{2}tr\left[\Sigma^{-1}(S + (Y - \Pi W)(Y - \Pi W)')\right]}.
\]
This implies the conditional posterior distribution
\[
\Sigma|\pi, Y \sim iW(S^*, d^*), \quad \text{(C.30)}
\]
with
\[
\begin{align*}
d^* &= d + T, \\
S^* &= S + (Y - \Pi W)(Y - \Pi W)',
\end{align*}
\]
which proves the result from equation (C.18).
C.2 Derivations of $p(\pi|\Sigma, Y)$ and $p(\Sigma|Y)$

Start again from equation (C.15) and consider the terms in its two exponential functions. Since

$$(\pi - \mu_\pi)'V^{-1}_\pi(\pi - \mu_\pi) = \pi'V^{-1}_\pi - 2\pi'V^{-1}_\pi\mu_\pi + \mu_\pi'V^{-1}_\pi\mu_\pi,$$

and since

$$(\bar{y} - Z \cdot \pi)'(I_T \otimes \Sigma)^{-1}(\bar{y} - Z \cdot \pi) =$$

$$(\bar{y} - (W' \otimes I_k) \cdot \pi)'(I_T \otimes \Sigma)^{-1}(\bar{y} - (W' \otimes I_k) \cdot \pi)$$

$$= \bar{y}'(I_T \otimes \Sigma^{-1})\bar{y} + \pi'(W' \otimes I_k)'(I_T \otimes \Sigma^{-1})(W' \otimes I_k)\pi$$

$$- 2\pi'(W' \otimes I_k)'(I_T \otimes \Sigma^{-1})\bar{y}$$

$$= \bar{y}'(I_T \otimes \Sigma^{-1})\bar{y} + \pi'(WW' \otimes \Sigma^{-1})\pi - 2\pi'(W \otimes \Sigma^{-1})\bar{y},$$

the kernel of the joint posterior distribution can be rewritten as

$$p(\pi, \Sigma|Y) \propto e^{-\frac{1}{2}\left\{ \pi'(WW' \otimes \Sigma^{-1})\pi - 2\pi'(W \otimes \Sigma^{-1})\bar{y} + \bar{y}'(I_T \otimes \Sigma^{-1})\bar{y} + \pi'V^{-1}_\pi - 2\pi'V^{-1}_\pi\mu_\pi \right\} \cdot |det(\Sigma)|^{-\frac{1}{2}} \cdot |det(\Sigma)|^{-\frac{d+k+1}{2}} \cdot e^{-\frac{1}{2}tr[\Sigma^{-1}S]}.$$  

Note that we keep all elements that contain $\Sigma$, since we also aim to derive $p(\Sigma|Y)$.

We can rewrite the terms in the exponent of the above equation as

$$\pi'(V^{-1}_\pi + (WW' \otimes \Sigma^{-1}))\pi - 2\pi'(V^{-1}_\pi\mu_\pi + (W \otimes \Sigma^{-1})\bar{y}) + \bar{y}'(I_T \otimes \Sigma^{-1})\bar{y} + tr[\Sigma^{-1}S]$$

$$= \pi'\left(V^{-1}_\pi + (WW' \otimes \Sigma^{-1})\right)\pi +$$

$$- 2\pi'\left(V^{-1}_\pi + (WW' \otimes \Sigma^{-1})\right)(V^{-1}_\pi + (WW' \otimes \Sigma^{-1}))^{-1}(V^{-1}_\pi\mu_\pi + (W \otimes \Sigma^{-1})\bar{y})$$

$$+ \bar{y}'(I_T \otimes \Sigma^{-1})\bar{y} + tr[\Sigma^{-1}S] + \mu_\pi'V^{-1}_\pi\mu_\pi - \mu_\pi'V^{-1}_\pi\mu_\pi,.$$
Hence

\[
p(\pi, \Sigma | Y) \propto \left\{ \frac{\det(V^*_\pi)}{(\pi - \mu^*_\pi)^{d^*+k+1}} \cdot e^{-\frac{1}{2}(\pi - \mu^*_\pi)^\top (\pi - \mu^*_\pi)} \cdot \det(\Sigma) \cdot \det(V^*_\pi) \right\}.
\]

(C.34)

with

\[V^*_\pi = [V^{-1}_\pi + (WW' \otimes \Sigma^{-1})]^{-1},\]
\[\mu^*_\pi = V^*_\pi \cdot [V^{-1}_\pi \mu_\pi + (W \otimes \Sigma^{-1})\tilde{y}]\]

This implies the conditional posterior distribution

\[
\pi | \Sigma, Y \sim N(\mu^*_\pi, V^*_\pi),
\]

(C.35)

which proves the result from equation (C.19). Last, the remaining terms in equation (C.34) imply

\[
p(\Sigma|Y) \propto |\det(\Sigma)|^{-\frac{d^*+k+1}{2}} \cdot |\det(V^*_\pi)|^{\frac{1}{2}} \cdot e^{-\frac{1}{2} \left\{ \tilde{y}'(I_T \otimes \Sigma^{-1})\tilde{y} - \mu^*_\pi V^*_\pi \mu^*_\pi + tr[\Sigma^{-1}S] \right\}}.
\]

(C.36)

with \(d^* = d + T\). Equation (C.36) proves the result from equation (C.20).

**C.3 Derivations of \(p(B)\) and \(p(B|Y)\)**

Given a distribution \(p_\Sigma(\Sigma)\) for \(\Sigma\), a distribution \(p_Q(Q|\Sigma)\) for \(Q\), and \(B\) such that \(B = h(\Sigma)Q\), it holds that

\[
p_B(B) = J(\Sigma \to B) \cdot p_\Sigma(\Sigma) \cdot p_Q(Q|\Sigma),
\]

(C.37)

with \(J(\Sigma \to B)\) the Jacobian of the transformation from \(\Sigma\) to \(B\) (Bibby et al., 1979, Mathai and Haubold, 2008). Equation (C.37) shows that the shape of \(p_B(B)\) on the parameter space of \(B\) depends on the stochastic behaviour of the underlying random variables \(\Sigma\) and \(Q\) and on the curvature that the transformation function introduces.
When no zero restrictions are introduced on $B$, it holds that
\[ J(\Sigma \rightarrow B) = |\det(B)|. \] (C.38)

It follows that the implicit distributions on $B$ associated with equations (C.14), (C.36) and (C.17) are
\[ p(B) \propto \text{I}\{\text{sign}\} \cdot |\det(B)|^{-(d+k)} \cdot e^{-\frac{1}{2} \{\text{vec}(B^{-1})^t (S \otimes I_k) \text{vec}(B^{-1})\}} \] (C.39)
and
\[ p(B|Y) \propto \text{I}\{\text{sign}\} \cdot |\det(B)|^{-(d^*+k)} \cdot |\det(V_*^\pi)|^{\frac{1}{2}} \cdot e^{-\frac{1}{2} \{\text{vec}(B^{-1})^t (S \otimes I_k) \text{vec}(B^{-1}) + \tilde{y}' \left(I_{r} \otimes (BB')^{-1}\right) \tilde{y} - \mu_*^\pi V_*^{-1} \mu_*^\pi\}}, \] (C.40)
with $\text{I}\{\text{sign}\}$ an indicator function taking value of 1 if $B$ satisfies the sign restrictions. Equations (C.39) and (C.40) coincide with equations (C.21) and (C.22), respectively.

Equation (C.39) coincides with the first part of equation (2.8) from Arias et al. (2018), once adjusted for the difference in the notation used (see footnote 1 in this appendix).

If zero restrictions are introduced to $B$, the computation of $J(\Sigma \rightarrow B)$ requires a numerical procedure, see Arias et al. (2018).

\section*{C.4 More on Algorithm 1}

Algorithm 1 (Section 2.2 of the paper), used to explore the posterior under the NiWU approach, runs a Gibbs sampler for $n_1$ burn-in replications and $n_2$ retained replications, and then extracts $n_4$ draws for the structural parameters. By contrast, Algorithm 2 (Section 2.3 of the paper), used to explore the posterior under the Np(B) approach, runs a Gibbs sampler for $m_1$ burn-in replications and $m_2$ retained replications. It then uses the retained draws as \textit{proposal} draws in an importance sampler, leading, in Stage A, to effectively use $\text{ESS}^A$ draws. Similarly, it stores $m_4$ \textit{proposal} draws for the structural parameters, which are then reweighted, leading to effectively $\text{ESS}^B$ draws.

To make sure that the computational time of the two approaches are comparable in the applications of the paper, we first run Algorithm 2, and then run Algorithm 1.
to ensure that it generates a number of draws close to the ones effectively obtained for Algorithm 2, see Table C1. For example, in the simulation exercise of Section 3, for $T = 120$ Algorithm 1 effectively generates 40,407 draws from Stage A and 28,443 draws from Stage B. We then run Algorithm 1 setting $n_2 = n_4 = 28,443$.

Table C1: Tuning parameters used for Algorithm 1 in the paper

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Correspondence</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_1$</td>
<td>burn-in draws</td>
<td>equal to $ESS^A$ in Algorithm 2</td>
</tr>
<tr>
<td>$n_2$</td>
<td>retained draws</td>
<td>equal to $ESS^B$ in algorithm 2</td>
</tr>
<tr>
<td>$n_3$</td>
<td>maximum number of draws of Q</td>
<td>equal to $m_3$ in Algorithm 2</td>
</tr>
<tr>
<td>$n_4$</td>
<td>desired number of draws</td>
<td>equal to $ESS^B$ in algorithm 2</td>
</tr>
</tbody>
</table>

D  Np(B) approach proposed in the paper

In this section we provide the derivations of the posterior distribution for the approach proposed in the paper, and discuss our importance sampler. We then discuss the diagnostic procedures that we use to ensure that the weights in the importance sampler have a finite variance. To make the analysis more general, we provide the derivations for the posterior in the case in which prior beliefs on $\pi$ can depend on $B$. As discussed in the paper, footnote 6, the applications in the paper use the special case of independence between $B$ and $\pi$.

D.1 Derivations of $p(B|Y)$ and $p(\pi|B,Y)$

Start from the prior distribution

$$p(\pi, B) = p(\pi|B) \cdot p(B),$$

(D.41)

with

$$p(\pi|B) = (2\pi)^{-\frac{k}{2}} |det(V_\pi)|^{-\frac{1}{2}} e^{-\frac{1}{2}(\pi - \mu_\pi)^t V_\pi^{-1}(\pi - \mu_\pi)},$$

(D.42)
where $\mu_\pi$ and $V_\pi$ can be a function of $B$. The joint posterior distribution then equals

$$p(\pi, B|Y) = p(\pi, B)\frac{p(Y|\pi, B)}{p(Y)} = p(B) \cdot (2\pi)^{-\frac{k}{2}}|\text{det}(V_\pi)|^{-\frac{1}{2}} e^{-\frac{1}{2}(\pi - \mu_\pi)'V_\pi^{-1}(\pi - \mu_\pi)},$$

(D.43)

$$= (2\pi)^{-\frac{kT}{2}}|\text{det}(B)|^{-\frac{1}{2}} e^{-\frac{1}{2}(\tilde{y} - Z\pi)'(I_T \otimes BB')^{-1}(\tilde{y} - Z\pi)p(Y)^{-1}},$$

with $p(Y) = \int_\pi \int_B p(Y|\pi, B)p(\pi, B)d\pi dB$. We aim to rewrite equation (D.43) as

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

(D.44)

and to exploit analytical results for $p(\pi|B, Y)$. To do so, define

$$\Psi(B) = (I_T \otimes BB'),$$

(D.45)

and rewrite the joint posterior distribution as

$$p(\pi, B|Y) = c \cdot e^{-\frac{1}{2}\left[(\pi - \mu_\pi)'V_\pi^{-1}(\pi - \mu_\pi) + (\tilde{y} - Z\pi)'\Psi(B)^{-1}(\tilde{y} - Z\pi)\right]},$$

(D.46)

with $c$ a term that includes elements which are not a function of $\pi$. Factorize the terms in the exponent of (D.46) as

$$(\pi - \mu_\pi)'V_\pi^{-1}(\pi - \mu_\pi) + (\tilde{y} - Z\pi)'\Psi(B)^{-1}(\tilde{y} - Z\pi) =$$

$$= \pi'V_\pi^{-1}\pi - 2\pi'V_\pi^{-1}\mu_\pi + \mu_\pi'V_\pi^{-1}\mu_\pi +$$

$$+ \tilde{y}'\Psi(B)^{-1}\tilde{y} - 2\pi'Z\Psi(B)^{-1}Z\pi + \pi'Z\Psi(B)^{-1}Z\pi =$$

$$= \pi'[V_\pi^{-1} + Z'\Psi(B)^{-1}Z]\pi - 2\pi'[V_\pi^{-1}\mu_\pi + Z'\Psi(B)^{-1}\tilde{y}] +$$

$$+ \tilde{y}'\Psi(B)^{-1}\tilde{y} + \mu_\pi'V_\pi^{-1}\mu_\pi =$$

$$= (\pi - \mu_\pi)^*V_\pi^{-1}(\pi - \mu_\pi^*) + \tilde{y}'\Psi(B)^{-1}\tilde{y} + \mu_\pi'V_\pi^{-1}\mu_\pi - \mu_\pi'^*V_\pi^{-1}\mu_\pi^*,$$
with

\[ V_\pi^* = [V_\pi^{-1} + Z'\Psi(B)^{-1}Z]^{-1}, \]  
\[ = \left[ V_\pi^{-1} + [WW' \otimes (BB')^{-1}] \right]^{-1}, \]  
\[ \mu_\pi^* = V_\pi^* \cdot \left[ V_\pi^{-1} \mu_\pi + Z'\Psi(B)^{-1}\tilde{y} \right], \]  
\[ = V_\pi^* \cdot \left[ V_\pi^{-1} \mu_\pi + [W \otimes (BB')^{-1}]\tilde{y} \right]. \]

(D.48) \hspace{1cm} (D.49) \hspace{1cm} (D.50) \hspace{1cm} (D.51)

The joint posterior distribution can now be written as

\[ p(\pi, B|Y) = (2\pi)^{-\frac{k}{2}} \cdot |\text{det}(V_\pi^*)|^{-\frac{1}{2}} \cdot e^{-\frac{1}{2}(\pi - \mu_\pi^*)'V_\pi^{-1}(\pi - \mu_\pi^*)} \cdot p(\pi|B,Y) \cdot |\text{det}(B)|^{-T} \cdot |\text{det}(V_\pi^*)|^{\frac{1}{2}} \cdot e^{-\frac{1}{2}(\tilde{y}'(I_T \otimes (BB')^{-1})\tilde{y} - \mu_\pi^*V_\pi^{-1}\mu_\pi + \mu_\pi^*V_\pi^{-1}\mu_\pi')} \cdot e^{-\frac{1}{2} \left\{ \tilde{y}'(I_T \otimes (BB')^{-1})\tilde{y} - \mu_\pi^*V_\pi^{-1}\mu_\pi + \mu_\pi^*V_\pi^{-1}\mu_\pi \right\}}. \]

It follows that

\[ \pi|B,Y \sim N(\mu_\pi^*, V_\pi^*), \]  
\[ p(B|Y) \propto p(B) \cdot \ldots \cdot |\text{det}(B)|^{-T} \cdot |\text{det}(V_\pi)|^{\frac{1}{2}} \cdot e^{-\frac{1}{2} \left\{ \tilde{y}'(I_T \otimes (BB')^{-1})\tilde{y} - \mu_\pi^*V_\pi^{-1}\mu_\pi + \mu_\pi^*V_\pi^{-1}\mu_\pi \right\}}, \]

which proves equations (16) and (17) in Section 2.3 of the paper.

The algorithm discussed in the next section requires being able to evaluate the function \( p(\Sigma|Y) \), which is the distribution of \( \Sigma \) implied by equation (D.53). To illustrate how to compute this function, consider first the case of the distribution \( p(\Sigma) \) implicit in the distribution \( p(B) \). Given \( p(\Sigma) \) generating \( \Sigma \) and a distribution \( p(Q|\Sigma) \) generating \( Q \), the distribution \( p(B) \) evaluated at \( \tilde{B} \) with \( \tilde{B}\tilde{B}' = \Sigma \) and \( \tilde{Q} = \tilde{B}h(\Sigma)^{-1} \) is

\[ p(B = \tilde{B}) \propto J(\Sigma \rightarrow B) \cdot p(\Sigma = \tilde{\Sigma}) \cdot p(Q = \tilde{Q}|\tilde{\Sigma}). \]

(D.54)

It follows that

\[ p(\Sigma = \tilde{\Sigma}) \propto J(\Sigma \rightarrow B)^{-1} \cdot p(B = \tilde{B}) \cdot p(Q = \tilde{Q}|\tilde{\Sigma})^{-1}. \]

(D.55)
In addition, if we can evaluate \( p(B) \) we can then evaluate \( p(Q|\Sigma) \), since
\[
p(Q = \tilde{Q}|\tilde{\Sigma}) \propto p(B = \tilde{B}).
\] (D.56)

Applied to the posterior distribution \( p(B|Y) \), this implies
\[
p(\Sigma|Y) \propto J(\Sigma \to B)^{-1} \cdot p(B|Y) \cdot p(Q|\Sigma)^{-1}
\propto |\text{det}(\Sigma)|^{-\frac{1}{2}} \cdot p(B|Y) \cdot p(B)^{-1}
\propto |\text{det}(\Sigma)|^{-\frac{T+1}{2}} \cdot |\text{det}(V_\pi)|^{-\frac{1}{2}} \cdot |\text{det}(V_\ast)|^\frac{1}{2} \cdot e^{-\frac{1}{2}\left\{ \tilde{y}^\top (I_T \otimes \Sigma^{-1}) \tilde{y} - \mu_\pi^\top V_\ast^{-1} \mu_\pi + \mu_\pi^\top V_\ast^{-1} \mu_\pi \right\}},
\] (D.57)

\[ \begin{align*}
\text{D.2 The importance sampler}
\end{align*} \]

Before discussing the importance sampler, we clarify the notation used so far by noticing that several distributions have been derived twice: once for the special case associated with the NiWU approach, and once for the less restrictive \( \text{Np}(B) \) approach associated with the general prior distribution \( p(B) \). By construction, these sets of functions coincide when the distribution \( p(B) \) from the \( \text{Np}(B) \) approach coincides with the distribution implicit in the NiWU approach, equation (C.21). To avoid confusion, the discussion in this section and in the paper uses the following notation:

\( \text{Np}(B) \) approach:
- \( p(B)_{\text{Np}(B)} \): prior distribution on \( B \) used in the \( \text{Np}(B) \) approach;
- \( p(\Sigma)_{\text{Np}(B)} \): prior distribution on \( \Sigma \) associated with the \( \text{Np}(B) \) approach;
- \( p(B|Y)_{\text{Np}(B)} \): posterior distribution on \( B \) associated with \( p(B)_{\text{Np}(B)} \) and derived analytically in equation (D.53);
- \( p(\Sigma|Y)_{\text{Np}(B)} \): posterior distribution on \( \Sigma \) implicit in the \( \text{Np}(B) \) approach and derived analytically in equation (D.57);
- \( p(Q|\Sigma)_{\text{Np}(B)} \): distribution on \( Q \) implicit in the prior \( p(B)_{\text{Np}(B)} \) conditioning on the space \( \Sigma = BB' \), evaluated as in equation (D.56);

\[ ^2 \text{Given some } \tilde{\Sigma}, \tilde{\Sigma} = \tilde{B}\tilde{B}' \text{ and } \tilde{Q} = h(\tilde{\Sigma})^{-1} \tilde{B}, \text{ it holds that } \frac{p(B = \tilde{B}|Y)_{\text{NiWU}}}{p(B = \tilde{B}|Y)_{\text{Np}(B)}} = \frac{p(Q = \tilde{Q}|\Sigma)_{\text{NiWU}}}{p(Q = \tilde{Q}|\Sigma)_{\text{Np}(B)}}, \text{ since the data do not update beliefs on } Q \text{ and differences along the identified set are only driven by the prior.} \]
NiWU approach:

- \( p(B)_{NiWU} \): prior distribution on \( B \) implied by the NiWU approach and derived analytically in equation (C.39);
- \( p(\Sigma)_{NiWU} \): prior distribution on \( \Sigma \) explicitly introduced in the NiWU approach, which is the inverse Wishart distribution;
- \( p(B|Y)_{NiWU} \): posterior distribution on \( B \) implied by the NiWU approach and derived analytically in equation (C.40);
- \( p(\Sigma|Y)_{NiWU} \): marginal posterior distribution on \( \Sigma \) implicit in the NiWU approach and derived analytically in equation (C.36);
- \( p(Q|\Sigma)_{NiWU} \): distribution on \( Q \) explicitly used in the NiWU approach, which is uniform in the space \( Q_{\Sigma} \) of orthogonal matrices that satisfy the identifying restrictions, given \( \Sigma \);

Our importance sampling procedure aims to draw from \( p(B|Y)_{NP(B)} \) by drawing from \( p(\Sigma|Y)_{NP(B)} \) and \( p(Q|\Sigma)_{NP(B)} \). To this aim, Stage A of Algorithm 2 in the paper uses \( p(\Sigma|Y)_{NiWU} \) as an importance function for \( p(\Sigma|Y)_{NP(B)} \). While the prior distributions \( p(\Sigma)_{NP(B)} \) and \( p(\Sigma)_{NiWU} \) may well differ, the corresponding posterior distributions \( p(\Sigma|Y)_{NP(B)} \) and \( p(\Sigma|Y)_{NiWU} \) are closer to each other since \( \Sigma \) is identified and the likelihood function dominates the prior distribution. Accordingly, \( p(\Sigma)_{NiWU} \), which can be explored conveniently with a Gibbs sampler, can be used as an importance function for \( p(\Sigma|Y)_{NP(B)} \). When no zero restrictions are introduced, the weights for Stage A are computed as

\[
    w^\text{stage A}_d = \frac{p(\Sigma|Y)_{NP(B)}}{p(\Sigma|Y)_{NiWU}} \propto \left| \det(\Sigma_d) \right|^{\frac{d+1}{2}} \cdot e^{\frac{1}{2}tr[S_d]\Sigma_d^{-1}S_d}. \tag{D.58}
\]

which follows from equations (D.57) and (C.36). If instead zero restrictions are introduced on \( B \), the evaluation of \( p(\Sigma|Y)_{NP(B)} \) must compute the Jacobian transformation \( J(B \rightarrow \Sigma) \) numerically, see Section D.

Having draws from \( p(\Sigma|Y)_{NP(B)} \), it remains to draw from \( p(Q|\Sigma)_{NP(B)} \) and map draws of \( \Sigma \) and \( Q \) into \( B \). Extracting from \( p(Q|\Sigma)_{NP(B)} \) is not straightforward, especially if \( p(B) \) has not been specified yet. However, the algorithm developed by Rubio-Ramirez et al. (2010) extracts orthogonal matrices uniformly from the full orthogonal space \( \mathcal{Q} \), and hence also from the general subset in which the target function
\( p(Q|\Sigma)_{Np(B)} \) has mass. Hence, the distribution \( p(Q|\Sigma)_{NiWU} \) can be used as an importance function for \( p(Q|\Sigma)_{Np(B)} \). This requires being able to evaluate \( p(Q|\Sigma)_{NiWU} \) rather than draw from \( p(Q|\Sigma)_{Np(B)} \). This requires being able to evaluate \( p(Q|\Sigma)_{NiWU} \) in correspondence to the proposal extractions, which can be done using \( p(Q|\Sigma)_{Np(B)} \propto p(B = \tilde{B}) \). Last, the uniformity of the importance distribution for Stage B implies \( p(Q|\Sigma)_{NiWU} \propto 1 \). All in all, it follows that the weights for Stage B can be computed as

\[
w_{\text{stage B}}^{d} = \frac{p(Q|\Sigma)_{Np(B)}}{p(Q|\Sigma)_{NiWU}} \propto p(B)_{Np(B)}. \tag{D.59}
\]

Implementing Algorithm 2 requires setting the hyperparameters \( \mu_\pi, V_\pi, d, S \) for the NiWU approach run as an importance function, Step 1 of Algorithm 2. We propose to set \( \mu_\pi, V_\pi \) equal to the values used for the prior specification of \( \pi \), equation (14) in the paper, and to set \( d, S \) to heuristically minimize the distance between \( p(\Sigma|Y)_{Np(B)} \) and \( p(\Sigma|Y)_{NiWU} \). We set \( d = k + 2 \) and \( S \) such that the expected value of \( \Sigma \) associated with the prior distribution implicit in the importance function equals the pointwise mean of \( \Sigma \) associated with \( p(B)_{Np(B)} \). This procedure requires to

1. specify \( p(B) \);
2. extract \( m_0 \) draws from \( p(B) \);
3. compute the corresponding draws of \( \Sigma = BB' \);
4. set \( S = (d - k - 1) \cdot \frac{\sum_{d=1}^{m_0} \Sigma_d}{m_0} \).

Other parametrizations are also possible.

The simulation exercise in Section 3 of the paper discusses posterior distributions in the hypothetical case in which the researcher has access to an infinitely large dataset. Algorithm 2 can be modified to explore such a scenario by running only Stage B using \( \Sigma = \Sigma_{true} \) in Step 3, with \( \Sigma_{true} \) the true value of \( \Sigma \) used in the data generating process. The same simulation exercise also explores the distribution of the rotation angle \( \theta \) implicit in the orthogonal matrices extracted from \( p(Q|\Sigma)_{Np(B)} \) and \( p(Q|\Sigma)_{NiWU} \). In the bivariate case, since orthogonal matrices can be viewed as Givens rotations

\[
Q(\theta) = \begin{pmatrix}
\cos(\theta) & -\sin(\theta) \\
\sin(\theta) & \cos(\theta)
\end{pmatrix}, \tag{D.60}
\]
one can explore the distribution on \( \theta \) implicit in \( p(Q|\Sigma)_{NP(B)} \) or \( p(Q|\Sigma)_{NiWU} \) by computing \( \theta = \text{atan} \left( \frac{Q_{2,1}}{Q_{1,1}} \right) \) for each extraction \( Q \) from \( p(Q|\Sigma)_{NP(B)} \) or \( p(Q|\Sigma)_{NiWU} \), with \( \text{atan} \) the inverse tangent function evaluated at \( \frac{Q_{2,1}}{Q_{1,1}} \), given \( \bar{Q}_{i,j} \) the \((i, j)\) entry of \( \bar{Q} \). \( Q(\theta) = \bar{Q} \) up to sign and ordering of the columns.

Stage A of our importance sampler relies on the convergence of \( p(\Sigma|Y)_{NiWU} \) and \( p(\Sigma|Y)_{NP(B)} \) towards a point mass at \( \Sigma_{\text{true}} \) to justify the use of the former distribution as an importance function for the latter. How large the dataset must be to ensure that \( p(\Sigma|Y)_{NiWU} \) and \( p(\Sigma|Y)_{NP(B)} \) are sufficiently close is ultimately an empirical question, and depends not only on the available dataset but also on the prior distributions \( p(\Sigma)_{NiWU} \) and \( p(B)_{NP(B)} \) used. The simulations on the bivariate VAR model from Section 3 find that the ratio of effective draws to proposal draws equals 0.50 (the threshold value used by Herbst and Schorfheide, 2014 to judge whether to change importance function) already for \( \tilde{T} = T - p = 52 \). In the application to the oil market from Section 4 of the paper, the dataset effectively has \( \tilde{T} = T - p = 405 \) observations and delivers a ratio of effective to proposal draws between 0.78 and 0.88 (Table G11). As a heuristic, back-of-the-envelop calculation, we kept the training sample constant and used the first \( \tau \) observations from the estimation sample, increasing \( \tau \) from 85 to \( T = 405 \). The results, reported in Figure G44, show that the ratio of effective to proposal draws increases fairly rapidly as the size of the dataset increases. The threshold value of 0.50 is reached for a number of observations between \( \tau = 115 \) (for the wider prior) and \( \tau = 185 \) (for the tighter prior). Most of the applications of Bayesian structural VAR model use larger datasets.

<table>
<thead>
<tr>
<th>Table D2: Tuning parameters used for Algorithm 2 in the paper</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stage A</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Stage B</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
D.3 Diagnostics for the Importance Sampler

Stage B of the importance sampler uses an importance function that features positive mass everywhere in the support of the target function. Accordingly, the favourable condition highlighted by Geweke (1989) in his equation (5) holds, ensuring that the variance of the estimators constructed on the parameters of interest is finite (see also Robert and Casella (2013), chapter 3). However, this condition does not hold for Stage A of the algorithm, which hence requires ensuring that the non-normalised importance weights \( \{w_i\}_{i=1}^N \) of size \( N \) have a finite variance. As pointed out by Geweke (1989), the finite variance condition is not necessary to ensure consistency of the estimators based on the weighted draws, but it reduces the risk that the sampler performs poorly (see also Robert and Casella (2013), chapter 3). In order to assess whether the variance of the weights in Stage A is bounded, we employ a graphical procedure and three diagnostic tests proposed by Koopman et al. (2009).

D.3.1 Graphical Assessment

A measure frequently used to assess the quantitative importance of outliers in importance weights (and hence concerns about the finiteness of the variance) is the recursively estimated variance of the weights. Define this estimated variance as \( \{v_i\}_{i=1}^N \), where \( v_i = Var(w_{1:i}) \) is the variance of the first \( i \) weights. If outliers do not raise quantitatively relevant concerns, \( \{v_i\}_{i=1}^N \) should converge smoothly towards a constant. If, instead, individual outliers dominate the recursive variance, the plot will reveal large jumps. Jumps in \( \{v_i\}_{i=1}^N \) are indicative of an unbounded variance of the weights, which raises concerns on their use in the re-sampling procedure. We use this graphical procedure to assess the performance of the weights in our sampler.

D.3.2 Diagnostic tests: Wald test, Score test and LR test

In addition to using graphical illustrations, we employ the more formal testing procedures by Koopman et al. (2009) to formally assess the quantitative relevance of outliers. They propose a Wald test, a score test and a Likelihood Ratio test to test

\[ H_0 : \text{the variance is finite} \quad \text{(D.61)} \]

against

\[ H_1 : \text{the variance is unbounded.} \quad \text{(D.62)} \]
To set the stage, note that the problematic part of $\{w_i\}_{i=1}^N$ are the excessively large weights. It is therefore natural to consider only those weights that are larger than a certain threshold $u$. After specifying $u$, generate a new random variable

$$Z_i = w_i - u \quad \text{if} \quad w_i > u.$$  \hfill (D.63)

Then, if $\{w_i\}_{i=1}^N$ are i.i.d. random draws from the same random variable, then, as shown by Pickands et al. (1975), for large $N$ and $u$ the new sequence of random variables $\{Z_i\}_{i=1}^n$ approximates a generalised Pareto distribution with density function

$$f(z|\xi, \beta) = \frac{1}{\beta} \left(1 + \frac{\xi z}{\beta}\right)^{-\frac{1}{\xi} - 1}. \hfill (D.64)$$

The attractive feature of this distribution is that the finiteness of the variance can directly be assessed: If $\xi \leq 0.5$, the variance of $Z_i$ exists, otherwise it is unbounded.

In practice, the threshold $u$ plays a crucial role. We follow Arias et al. (2018) and use five different threshold values, $v_1 = 0.5N$, $v_2 = 0.6N$, $v_3 = 0.7N$, $v_4 = 0.9N$ and $v_5 = 0.99N$ and set $u_j = w(v_j)$, where $w(i)$ are the ordered weights, i.e. $w_1 \leq w(2) \leq \ldots \leq w(N)$.

To make the tests operational we use the null and alternative hypotheses

$$H_0 : \xi = 0.5, \hfill (D.65)$$

$$H_1 : \xi > 0.5. \hfill (D.66)$$

Koopman et al. (2009) propose testing procedures based on a Wald test, a score test, and a LR test.

Wald test

The log likelihood function based on the generalised Pareto distribution in (D.64) equals

$$\log f(z|\beta, \xi) = -n \cdot \log(\beta) - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^n \log(1 + \xi \beta^{-1} z_i). \hfill (D.67)$$

To construct the test statistic, we need to numerically maximise (D.67) with respect to $\beta$ and $\xi$ to obtain $\beta^{MLE}$ and $\xi^{MLE}$, and then construct a test statistic to test for $\xi = 0.5$. In practice, we follow these steps:

1. numerically maximise the unrestricted log likelihood function given in (D.67) to
obtain $\beta^{MLE}$ and $\xi^{MLE}$;

2. construct the test statistic as

$$t = \sqrt{\frac{n}{3\beta^{MLE}}} \left(\xi^{MLE} - \frac{1}{2}\right).$$  \hfill (D.68)

The statistic $t$ has an approximate standard normal distribution for $n \to \infty$. Large values indicate that $\xi$ exceeds 0.5, which is indicative of unbounded weight variance. Therefore, we reject $H_0$ of finite variance of the weights if $t$ exceeds the critical value obtained from the standard normal distribution.

**Score Test**

In order to construct a score test, we need to maximise (D.67) under the null hypothesis of $\xi = 0.5$. Follow these steps:

1. numerically maximise (D.67) with respect to $\beta$ under the restriction $\xi = 0.5$ to obtain $\beta^{MLE,r}$;

2. since the test statistic is based on the score function of $\xi$, differentiate (D.67) with respect to $\xi$ and set $\xi = 0.5$ and $\beta = \beta^{MLE,r}$. This gives \footnote{Please note that there is a typo in Koopman (page 4) in the corresponding equation.}

$$\hat{s}^\xi = 4 \sum_{i=1}^{n} \log(1 + \frac{z_i}{2\beta^{MLE,r}}) - 6 \sum_{i=1}^{n} \frac{z_i}{2\beta^{MLE,r} + z_i};$$  \hfill (D.69)

3. compute

$$s^\xi_* = \frac{1}{\sqrt{2n}} \hat{s}^\xi.$$  \hfill (D.70)

Since $s^\xi_* \to N(0, 1)$ for $n \to \infty$, we reject $H_0$ of finite weight variances if $s^\xi_*$ exceeds the critical values obtained from a standard normal distribution.

**LR Test**

In order to construct a likelihood ratio test statistic, follow these steps:

1. Estimate the model under the restriction implying infinite weight variance, i.e. maximise (D.67) under the restrictions $\xi \geq 0.5$ to obtain $\tilde{\beta}$ and $\tilde{\xi}$. Compute $\log \tilde{f} = f(z|\tilde{\beta}, \tilde{\xi});$

\footnote{Please note that there is a typo in Koopman (page 4) in the corresponding equation.}
2. The restricted MLE estimator for $\beta$, $\beta^{MLE,r}$, is computed in the previous section. Compute $\log f^r = f(z|\beta^{MLE,r}, \xi = 0.5)$;

3. The test statistic is $LR = 2(\log \tilde{f} - \log f^r)$.

The null hypothesis of finite weight variance is rejected for large values compared to a $\chi(1)$ distribution.

Figure F18 and Figure G45 report the results of the graphical assessment for the applications developed in the paper, while Table F4 and Table G13 report the test statistics.

E  The Dynamic Striated Metropolis-Hastings algorithm by Waggoner et al. (2016)

To assess whether our sampling procedure successfully explores the posterior distributions of the applications in the paper, we sample the same posterior distributions using the sampler developed by Waggoner et al. (2016). We do so because such sampler was intentionally designed to explore posterior distributions that are potentially irregularly shaped and challenging to explore, and hence is appropriate for the application of the paper. While being more computationally demanding, the Dynamic Striated Metropolis-Hastings sampler offers a benchmark against which to compare the extractions of the posterior distribution from our sampler. Section E.1 briefly discusses the functioning of the sampler, while Section E.2 discusses the convergence criteria that we use to assess the performance of the posterior chains.

E.1  The sampler

The key intuition behind the Dynamic Striated Metropolis-Hastings sampler is that one does not immediately sample $p(B|Y)_{Np(B)}$, which might feature an irregular shape and multiple peaks, but a simpler function. The draws from this starting function are then progressively used to draw from $p(B|Y)_{Np(B)}$. Define $\theta$ the vector including the parameters of interest and the function $f_\lambda(\theta)$ as

$$
    f_\lambda(\theta) = f^s(\theta) \cdot \left( f^i(\theta) \right)^{\lambda} \cdot \left( f^d(\theta) \right)^{1-\lambda},
$$

(E.71)
with tempering parameter $\lambda \in [0,1]$. $f^s(\theta)$, $f^i(\theta)$ and $f^d(\theta)$ are selected such that $f_{\lambda=0}(\theta)$ is the kernel of a starting distribution and $f_{\lambda=1}(\theta)$ coincides with the kernel of the probability distribution that one ultimately wants to explore. As $\lambda$ increases from 0 to 1, $f_{\lambda}(\theta)$ progressively introduces the elements of $f^i(\theta)$ and drops the elements of $f^d(\theta)$.

Waggoner et al. (2016) specify $f^s(\theta)$, $f^i(\theta)$ and $f^d(\theta)$ in order to initialize the algorithm from the prior distribution. We depart from Waggoner et al. (2016) and specify $f^s(\theta)$, $f^i(\theta)$ and $f^d(\theta)$ to initialize the algorithm from $p(B|Y)_{NiWU}$. Initializing the algorithm in $p(B|Y)_{NiWU}$ rather than $p(B)_{Np(B)}$ speeds up the posterior sampler, because the starting kernel already introduces information from the data. Then, we progressively convert $p(B|Y)_{NiWU}$ into $p(B|Y)_{Np(B)}$ using the sampler by Waggoner et al. (2016). In our application, the sampler is required for nine parameters, which is well within the range of parameters in which the Dynamic Striated Metropolis-Hastings sampler performs efficiently.

More precisely, in our applications $\theta$ contains the entries of $B$, excluding the ones restricted to zero, if any. In sampling $p(B|Y)_{Np(B)}$ from equation (17), one could set

\begin{align*}
f^s(\theta) &= p(B)_{Np(B)}, \quad \text{(E.72)} \\
f^i(\theta) &= |\det(B)|^{-T} \cdot |\det(V^*_\pi)|^{\frac{1}{2}} \cdot e^{-\frac{1}{2}\left\{ \tilde{y}^{\top}(I_T \otimes (BB')^{-1}) \tilde{y} - \mu^*_x V^*_x^{-1} \mu^*_x \right\}}, \quad \text{(E.73)} \\
f^d(\theta) &= 1. \quad \text{(E.74)}
\end{align*}

This specification follows Waggoner et al. (2016) in initializing the algorithm at the prior distribution. We propose, instead, to start from a function that allows exploiting the computational advantage of the NiWU approach from Section 2.2 of the paper. Having derived the posterior distribution corresponding to the NiWU approach in equation (C.22), we use this function as the starting point of the algorithm and set

\begin{align*}
f^s(\theta) &= |\det(B)|^{-T} \cdot |\det(V^*_\pi)|^{\frac{1}{2}} \cdot e^{-\frac{1}{2}\left\{ \tilde{y}^{\top}(I_T \otimes (BB')^{-1}) \tilde{y} - \mu^*_x V^*_x^{-1} \mu^*_x \right\}}, \quad \text{(E.75)} \\
f^i(\theta) &= p(B)_{Np(B)} \cdot |\det(V^*_\pi)|^{-\frac{1}{2}} \cdot e^{-\frac{1}{2}\left\{ \mu^*_x V^*_x^{-1} \mu^*_x \right\}}, \quad \text{(E.76)} \\
f^d(\theta) &= \left|\det(B)\right|^{-(d+k)} e^{-\frac{1}{2}\left\{ \text{vec}(B^{-1})(S \otimes I_k)\text{vec}(B^{-1}) \right\}}, \quad \text{(E.77)}
\end{align*}

We set $d$ and $S$ as in Algorithm 2, see Section D.2. If zero restrictions on $B$ are introduced, either equation (E.77) is modified to account for the volume element featuring also zero restrictions or the algorithm is initialized in $p(B)_{Np(B)}$. 

24
Once the functions $f^s(\theta)$, $f^i(\theta)$ and $f^d(\theta)$ are selected, define the target function

$$
\log[f_{\lambda h}(\theta)] = \log[f^s(\theta)] + \lambda_h \cdot \log[f^i(\theta)] + (1 - \lambda_h) \cdot \log[f^d(\theta)],
$$

(E.78)

as the logarithm of the tempered function $f_{\lambda h}(\theta)$ at stage $h$, with $h = 1, \ldots, H$ and $H$ the total number of stages. To the extent that, among other requirements, a sufficient number of stages is used, the target function at stage $h - 1$ is sufficiently close to the target function at stage $h$. This makes the draws representative of the target function at stage $h - 1$ a useful point of departure to numerically explore the target distribution at stage $h$.

Within this sequential approach, Waggoner et al. (2016) propose sampling $\log[f_{\lambda h}(\theta)]$ using a modified Metropolis-Hastings algorithm. In the general stage $h$, the algorithm can be summarized in the following steps:

1. start stage $h$ with $N \cdot G$ draws $\{\theta^{(h-1)}_d\}_{d=1}^{N \cdot G}$, which are representative of the target distribution at stage $h - 1$. If $h = 1$, then $\{\theta^{(0)}_d\}_{d=1}^{N \cdot G}$ are drawn from the starting function, otherwise, they are computed at the end of stage $h - 1$;

2. evaluate the function $\log[f^i(\theta)]$ at each $\{\theta^{(h-1)}_d\}_{d=1}^{N \cdot G}$ and group draws $\{\theta^{(h-1)}_d\}_{d=1}^{N \cdot G}$ into $M$ ‘striations’ (subsets), depending on the corresponding value of $\log[f^i(\theta)]$;

3. for each $\{\theta^{(h-1)}_d\}_{d=1}^{N \cdot G}$ compute weights $\tilde{\omega}_d = \frac{f_{\lambda h}(\theta^{(h-1)}_d)}{f_{\lambda h-1}(\theta^{(h-1)}_d)}$. As with importance sampling techniques, $\omega_d = \frac{\tilde{\omega}_d}{\sum_{d=1}^{N \cdot G} \tilde{\omega}_d}$, $d = 1, \ldots, N \cdot G$ allow reweighing the draws from the previous stage such that they become representative of the target distribution of the current stage, provided that the effective sample size does not shrink excessively;

4. use $\{\theta^{(h-1)}_d, \omega_d\}_{d=1}^{N \cdot G}$ to compute numerically the variance $\Omega_h$ of the target function at stage $h$;

5. explore the target function $\log[f_{\lambda h}(\theta)]$ as follows. For each group $g$, set the initial draw $\theta_{old}$ to a random draw from $\{\theta^{(h-1)}_d\}_{d=1}^{N \cdot G}$, extracted with replacement using $\{\omega_d\}_{d=1}^{N \cdot G}$. Then, with probability $p$, set $\theta_{new} = \theta_{old} + \theta_{shock}$ with $\theta_{shock}$ a multivariate zero-mean normal random variable with variance $c_h \cdot \Omega_h$, while with probability $1 - p$ set $\theta_{new}$ to a randomly extracted draw from the subset of

---

See, for example, Herbst and Schorfheide (2014) for an alternative approach to sequential Monte Carlo samplers.
\{\theta_d^{(h-1)}\}_{d=1}^{N \cdot G}$ from the striation associated with function $\log[f_i(\theta)]$ evaluated at $\theta_{\text{old}}$. Accept $\theta_{\text{new}}$ with probability $\min\left\{1, \frac{f_{\lambda_h}(\theta_{\text{new}})}{f_{\lambda_h}(\theta_{\text{old}})}\right\}$ if $\theta_{\text{new}}$ was generated from the random walk extraction, and with probability $\min\left\{1, \frac{f_{\lambda_h}(\theta_{\text{new}})}{f_{\lambda_h-1}(\theta_{\text{old}})}\right\}$ if $\theta_{\text{new}}$ was randomly selected from the strations at the previous stage. Continue for $N \cdot \tau$ iterations;

6. store one every $\tau$ draws for each group and collect the $N \cdot G$ draws $\{\theta_d^{(h)}\}_{d=1}^{N \cdot G}$. Use $\{\theta_d^{(h)}\}_{d=1}^{N \cdot G}$ to initialize the next stage $h + 1$. If $h = H$, the last stage has been reached, and $\{\theta_d^{(h)}\}_{d=1}^{N \cdot G}$ are interpreted as posterior draws from the desired distribution.

**Table E3:** Tuning parameters used for the DSMH algorithm

<table>
<thead>
<tr>
<th></th>
<th>Section 3</th>
<th>Section 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H$</td>
<td>number of stages</td>
<td>6</td>
</tr>
<tr>
<td>$G$</td>
<td>number of groups</td>
<td>8</td>
</tr>
<tr>
<td>$N$</td>
<td>number of MH iterations per group</td>
<td>8,000</td>
</tr>
<tr>
<td>$K$</td>
<td>number of iterations to calibrate $c$</td>
<td>1,000</td>
</tr>
<tr>
<td>$\tau$</td>
<td>frequency of storage of iterations</td>
<td>1</td>
</tr>
<tr>
<td>$M$</td>
<td>number of striations</td>
<td>10</td>
</tr>
</tbody>
</table>

To make the above algorithm operational, we need to set several tuning parameters. Following Waggoner et al. (2016), we set $p = 1/(10\tau)$. We then set the parameter $c_h$ at each stage following the guidance discussed by Waggoner et al. (2016) in Appendix A, hence ensuring, within each stage, an acceptance ratio between 0.20 and 0.30 from a preliminary Metropolis-Hastings algorithm with $K$ iterations. We set the progression of the tempering parameter as in Herbst and Schorfheide (2014), using $\lambda_h = \left(\frac{h-1}{H-1}\right)^2$. It remains to calibrate the number of stages $H$, the number of groups $G$, the effective number of iterations $N$ within each group, the number $K$ affecting the number of iterations $K \cdot G$ for the calibration of the parameter $c_h$, the frequency $\tau$ at which draws stored, and the number of striations $M$. These parameters affect the time required for the algorithm to run. For the applications in this paper, we speed up posterior sampling by setting these parameters to relatively low values, under the following conditions:

1. that the effective sample size computed for the importance sampling from one stage to the next one never falls below 80%. If this occurs a higher number of stages $H$ is used;
2. that the convergence criteria discussed in the next section of this appendix confirm that the draws at stage $h$ have converged to the distribution of the target function at stage $h$. If this is not the case, a higher number of effective draws $N$ and/or groups $G$ is used.

Table E3 reports the values of the tuning parameters used in the applications, as well as the running time.

**E.2 Convergence criteria**

Consider the chain \( \{ \theta_d \}_{d=1}^{N \cdot G} \), with \( \theta_d = (\theta_{1,d}, ..., \theta_{j,d}, ..., \theta_{\kappa,d})' \) of dimension \( \kappa \times 1 \). We employ four convergence criteria in order to assess if the chain has converged in distribution, namely the convergence criteria developed by Geweke (1992), by Raftery and Lewis (1992), by Gelman and Rubin (1992) and by Brooks and Gelman (1998). Intuitively, the criteria used operate by assessing whether the series has excessive auto-dependence (which indicates that the draws are not from a stationary distribution) and whether it depends on the starting point (which indicates that the chain is not long enough).

The convergence criteria that we use can be classified according to two main features. First, whether the convergence of each parameter is assessed in isolation from the remaining \( \kappa - 1 \) parameters or jointly (i.e. whether the object of interest is \( \theta_{j,d} \) or \( \theta_d \)). Second, whether the series of \( N \cdot G \) draws are considered in a long chain from a single starting value or in multiple chains from multiple starting points (i.e. whether the object of interest is \( \{ \theta_{j,d} \}_{d=1}^{N \cdot G} \) or \( \{ \theta_d \}_{d=1}^{N \cdot G} \) for univariate chains, and \( \{ \theta_{j,d} \}_{d=1}^{N \cdot G} \) or \( \{ \theta_d \}_{d=1}^{N \cdot G} \) for multivariate chains). The statistic by Brooks and Gelman (1998) considers the multidimensional objects, while the remaining criteria consider univariate objects. The criteria by Geweke (1992) and Raftery and Lewis (1992) consider single chains, while the criteria by Gelman and Rubin (1992) and Brooks and Gelman (1998) consider multiple chains.

For a detailed comparative review of convergence criteria for Markov Chain Monte Carlo mechanisms see Cowles and Carlin (1996) and Brooks and Roberts (1998).

**E.2.1 Geweke (1992)**

The univariate approach by Geweke (1992) assesses the convergence of each parameter of the series in isolation, using the series \( \{ \theta_{j,d} \}_{d=1}^{N \cdot G} \) for each parameter \( j \). The assessment is based on a comparison of means across different parts of the chain. If the means are close to each other, the procedure detects convergence.
To run the test we proceed in four steps:

1. extract the first 10% and the last 40% of the draws of \( \{ \theta_{j,d} \}_{d=1}^{N \cdot G} \), i.e. \( \{ \theta_{j,d} \}_{d=0.10 \cdot N \cdot G}^{0.10 \cdot N \cdot G} \) and \( \{ \theta_{j,d} \}_{d=0.60 \cdot N \cdot G}^{0.60 \cdot N \cdot G} \);

2. for each subseries, compute the mean and the standard deviation and call them \( \hat{\mu}_{\text{first}}, \hat{\mu}_{\text{last}}, \hat{\sigma}_{\text{first}} \) and \( \hat{\sigma}_{\text{last}} \);

3. compute the test statistic

\[
CD = \frac{\hat{\mu}_{\text{first}} - \hat{\mu}_{\text{last}}}{\sqrt{\frac{\hat{\sigma}_{\text{first}}}{0.10 NG} + \frac{\hat{\sigma}_{\text{last}}}{0.40 NG}}},
\]  

(E.79)

Under the conditions mentioned in Geweke (1992), \( CD \) has an asymptotic standard normal distribution;

4. compute the \( p \)-value.

The final statistic of the test is the \( p \)-value associated with the statistic \( CD \). A \( p \)-value below the significant level indicates that the null hypothesis of convergence, captured by the equality of means across the chain, can be rejected, and hence that the series has not converged.

### E.2.2 Raftery and Lewis (1992)

The approach by Raftery and Lewis (1992), like the one by Geweke (1992), investigates one long univariate chain of draws for one parameter in isolation, \( \{ \theta_{j,d} \}_{d=1}^{N \cdot G} \). The main objects of interest are the quantiles of the probability distribution for the parameter \( j \).

The method assesses if the chain is long enough to get precise estimates of quantiles of this distribution.

To define the notion of closeness, three values have to be specified by the user: \( s, q \) and \( r \). If the interest lies in \( q_{j,0.025} \), the 0.025 quantile of the posterior of a parameter \( \theta_j \), then \( s = 0.25 \). If one exerts 95% of the posterior draws to lie in an interval of +/- 0.0125 around the true 0.025 quantile, then \( s = 0.95 \) and \( r = 0.0125 \). These specifications are standard for output from an MCMC chain. The implementation of the algorithm for each parameter \( j \) proceeds in 4 steps:
1. transform \( \{\theta_{j,d}\}_{d=1}^{NG} \) into a dichotomous random variable \( Z_d \):

\[
Z_d = \begin{cases} 
1 & \text{if } \theta_{j,d} < q_{0.025}, \\
0 & \text{otherwise}; 
\end{cases}
\]

(E.80)

2. write the matrix of transition probabilities for \( Z_d \) conditioning on the previous state,

\[
P = \begin{bmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{bmatrix},
\]

with \( \alpha = P(Z_{d+1} = 1|Z_d = 0) \) and \( \beta = P(Z_{d+1} = 0|Z_d = 1) \). The unconditional probabilities of being in one state or another are

\[
\pi_0 = P(\theta_{j,d} < q_{0.025}) = P(Z_d = 0) = \frac{\beta}{\alpha + \beta} \quad (E.81)
\]

\[
\pi_1 = P(\theta_{j,d} \geq q_{0.025}) = P(Z_d = 1) = 1 - \pi_0 = \frac{\alpha}{\alpha + \beta} \quad (E.82)
\]

3. approximate the probability that a draw of the parameter is smaller than the quantile of interest as

\[
P(\theta_{j,d} < q_{0.025}) \approx \bar{Z}_{NG,j} = \frac{1}{NG} \sum_{d=1}^{NG} Z_d. \quad (E.83)
\]

As shown by Raftery and Lewis (1992), \( \bar{Z}_{NG} \) is approximately normally distributed with mean \( q_{0.025} \) and variance \( \frac{1}{NG} \frac{(2-\alpha-\beta)\alpha\beta}{(\alpha+\beta)^3} \).

4. compute the optimal length of the chain as the length that ensures \( P(q - r \leq \bar{Z}_{NG} \leq q + r) \) using

\[
n^* = \frac{(2 - \alpha - \beta)\alpha\beta}{(\alpha + \beta)^3} \left\{ \frac{\Phi^{-1}\left(\frac{1}{2}(s + 1)\right)}{r} \right\}^2 \quad (E.84)
\]

The key statistic of the test is \( n^* \), which has an intuitive interpretation: it is the minimum number of draws we need for the desired level of accuracy of the quantile \( q \) (given by \( r \) and \( s \)). If \( N \cdot G \) is lower than \( n^* \), this suggests that the chain length needs to be increased.
E.2.3 Gelman and Rubin (1992)

The convergence diagnostic by Gelman and Rubin (1992) uses multiple univariate chains, \( \{\theta_{j,d}\}_{d=1}^{N} \), \( \{\theta_{j,d}\}_{d=N+1}^{2N} \), \( \ldots \), \( \{\theta_{j,d}\}_{d=(G-1)N+1}^{N\cdot G} \). If the chains have converged, then they should not depend on starting values any more. The convergence statistic is based on a comparison of between-sequence variance and within sequence variance.

The procedure consists of four steps:

1. compute the variance of the mean of each sequence ("between-sequence variance") as

\[
B = \frac{1}{G-1} \sum_{g=1}^{G} (\bar{\theta}_{j,d,g} - \bar{\theta}_{j})^2, \tag{E.85}
\]

where \( \bar{\theta}_{j,d,g} \) is the mean of \( \theta_{j,d} \) within the \( g \)-th chain and \( \bar{\theta}_{j} \) is the mean across all chains for parameter \( j \).

2. compute the mean across sequences of the variances within sequence (the "within-sequence variance") as

\[
W = \frac{1}{G(N-1)} \sum_{g=1}^{G} \sum_{n=1}^{N} (\theta_{j,n,g} - \bar{\theta}_{j,d,g})^2; \tag{E.86}
\]

3. estimate the overall variance as

\[
\hat{V} = \hat{\sigma}^2 + \frac{B}{G}, \tag{E.87}
\]

with

\[
\hat{\sigma}^2 = \frac{N-1}{N} W + B; \tag{E.88}
\]

4. compute the statistic

\[
\hat{R} = \frac{\hat{V}}{W}. \tag{E.89}
\]

The key statistic of the test is \( \hat{R} \). As a rule-of-thumb, \( \hat{R} \) should be below 1.2 to assert that the chain has converged.

E.2.4 Brooks and Gelman (1998)

The statistic by Brooks and Gelman (1998) is a multivariate extension of Gelman and Rubin (1992) and requires different chains of a multivariate series, \( \{\theta_{d}\}_{d=1}^{N} \), \( \{\theta_{d}\}_{d=N+1}^{2N} \), \( \ldots \), \( \{\theta_{d}\}_{d=(G-1)N+1}^{N\cdot G} \).
\( \theta_d \) \( d = (G-1)N+1 \). Intuitively, as in the test by Gelman and Rubin (1992), the approach by Brooks and Gelman (1998) builds the analysis by comparing the between-chain and within-chain variances. The test builds on the multivariate extension of the steps used for the approach by Gelman and Rubin (1992):

1. compute the variance of the mean of each sequence (“between-sequence variance”) as
\[
B = \frac{1}{G-1} \sum_{g=1}^{G} (\bar{\theta}_g - \bar{\theta})(\bar{\theta}_g - \bar{\theta})';
\] (E.90)

2. compute the mean across sequences of the variances within sequence (the “within-sequence variance”) as
\[
W = \frac{1}{G(N-1)} \sum_{g=1}^{G} \sum_{n=1}^{N} (\theta_{n,g} - \bar{\theta}_g)(\theta_{n,g} - \bar{\theta}_g)';
\] (E.91)

3. estimate the overall variance as
\[
\hat{V} = \frac{N-1}{N} W + \left( 1 + \frac{1}{G} \right) B;
\] (E.92)

4. compute the (scalar) distance measure between these two matrices as
\[
\hat{R}^{\text{mult}} = \frac{N}{N-1} + \frac{G+1}{G} \lambda_1,
\] (E.93)

where \( \lambda_1 \) is the largest eigenvalue of the matrix \( W^{-1}B \).

The final statistic of the test is \( \hat{R}^{\text{mult}} \). As for the approach by Gelman and Rubin (1992), the rule-of-thumb prescribes that \( \hat{R}^{\text{mult}} \) is below 1.2 in order to assert convergence.

Table F5 to Table F9 and Table G14 to Table G16 report the convergence diagnostics for the two applications developed in the paper.

**F Additional tables/figures for Section 3 of the paper**
Figure F4: Illustration of our algorithm, $T = 30$

Stage A, $p(\Sigma|Y)$

Note: In the upper panel, the proposal draws refer to the draws obtained in Step 1 of Algorithm 2, while the reweighted draws are obtained using the weights from Step 2. In the lower panel, the proposal draws refer to draws obtained from Step 6 of Algorithm 2, while the reweighted draws are obtained in Step 8.
Figure F5: Illustration of our algorithm, $T = 60$

Stage A, $p(\Sigma | Y)$

Stage B, $p(B|Y)$

Note: In the upper panel, the proposal draws refer to the draws obtained in Step 1 of Algorithm 2, while the reweighted draws are obtained using the weights from Step 2. In the lower panel, the proposal draws refer to draws obtained from Step 6 of Algorithm 2, while the reweighted draws are obtained in Step 8.
Figure F6: Illustration of our algorithm, $T = 120$

Stage A, $p(\Sigma|Y)$

Stage B, $p(B|Y)$

Note: In the upper panel, the proposal draws refer to the draws obtained in Step 1 of Algorithm 2, while the reweighted draws are obtained using the weights from Step 2. In the lower panel, the proposal draws refer to draws obtained from Step 6 of Algorithm 2, while the reweighted draws are obtained in Step 8.
Figure F7: Illustration of our algorithm, $T = 240$

Stage A, $p(\Sigma|Y)$

Note: In the upper panel, the proposal draws refer to the draws obtained in Step 1 of Algorithm 2, while the reweighted draws are obtained using the weights from Step 2. In the lower panel, the proposal draws refer to draws obtained from Step 6 of Algorithm 2, while the reweighted draws are obtained in Step 8.
Figure F8: Illustration of our algorithm, $T = 480$

*Stage A, $p(\Sigma|Y)$*

![Stage A, proposal draws](image1)
![Stage A, reweighted draws](image2)
![True (DSMH algorithm)](image3)

*Stage B, $p(B|Y)$*

![Demand shock](image4)
![Supply shock](image5)

Note: In the upper panel, the proposal draws refer to the draws obtained in Step 1 of Algorithm 2, while the reweighted draws are obtained using the weights from Step 2. In the lower panel, the proposal draws refer to draws obtained from Step 6 of Algorithm 2, while the reweighted draws are obtained in Step 8.
Figure F9: Comparing our approach to the NiWU approach: priors

Stage A, $p(\Sigma)$

Note: Since the prior distribution for the NiWU approach with $d = 0$, $S = 0$ is an improper function, in this figure we approximate it by setting $d = k + 2$, $S = 0.01I_k$. See also Figure F10.
Figure F10: More on the approximate improper prior implicit in the NiWU approach

Stage A, \( p(\Sigma) \)

Stage B, \( p(B) \)
Figure F11: Comparing our approach to the NiWU approach, posteriors for $T = 30$

Stage A, $p(\Sigma|Y)$

```
\begin{figure}
\centering
\begin{subfigure}{0.4\textwidth}
\centering
\includegraphics[width=\textwidth]{wage.png}
\caption{Wage}
\end{subfigure}
\begin{subfigure}{0.4\textwidth}
\centering
\includegraphics[width=\textwidth]{employment.png}
\caption{Employment}
\end{subfigure}
\end{figure}
```

Stage B, $p(B|Y)$

```
\begin{figure}
\centering
\begin{subfigure}{0.4\textwidth}
\centering
\includegraphics[width=\textwidth]{demand.png}
\caption{Demand shock}
\end{subfigure}
\begin{subfigure}{0.4\textwidth}
\centering
\includegraphics[width=\textwidth]{supply.png}
\caption{Supply shock}
\end{subfigure}
\end{figure}
```
Figure F12: Comparing our approach to the NiWU approach, posteriors for $T = 60$

*Stage A, $p(\Sigma|Y)$*

![Graph showing posteriors for $\Sigma_{1,1}$, $\Sigma_{2,1}$, and $\Sigma_{2,2}$ for Wage and Employment.]

*Stage B, $p(B|Y)$*

![Graph showing posteriors for $B_{1,1}$, $B_{1,2}$, $B_{2,1}$, and $B_{2,2}$ for Demand shock and Supply shock.]

Legend:
- Np(B) approach (this paper)
- NiWU approach, improper prior
- NiWU approach, KK(1997) prior
- $\Sigma_{i,j,\text{true}}$
Figure F13: Comparing our approach to the NiWU approach, posteriors for $T = 120$

Stage A, $p(\Sigma|Y)$

$\Sigma_{1,1}$

Stage B, $p(B|Y)$

Demand shock

Supply shock
Figure F14: Comparing our approach to the NiWU approach, posteriors for $T = 240$

Stage A, $p(\Sigma | Y)$

![Graph showing Wage, Employment, and posterior distributions for $\Sigma_{1,1}$, $\Sigma_{2,1}$, and $\Sigma_{2,2}$ for Stage A.

Stage B, $p(B | Y)$

![Graph showing Wage and Employment distributions for demand and supply shocks for $B_{1,1}$, $B_{1,2}$, $B_{2,1}$, and $B_{2,2}$.]

Legend:
- Np(B) approach (this paper)
- NiWU approach, improper prior
- NiWU approach, KK(1997) prior
- $\Sigma_{ij}$, true

42
Figure F15: Comparing our approach to the NiWU approach, posteriors for $T = 480$

Stage A, $p(\Sigma|Y)$

Stage B, $p(B|Y)$
Figure F16: Comparing our approach to the NiWU approach, posteriors asymptotically

*Stage A, \( p(\Sigma | Y) \)*

*Stage B, \( p(B | Y) \)*
**Figure F17**: Posterior distribution of the estimated bounds $\pm \Sigma_{i,i}^{0.5}$ associated with $p(\Sigma | Y)_{NWU}$ and $p(\Sigma | Y)_{NP(B)}$, and posterior distributions $p(B | Y)_{NWU}$ and $p(B | Y)_{NP(B)}$

**A) $T = 30$**

- **Demand shock**
  - Wage
  - Employment
- **Supply shock**
  - Wage
  - Employment

**B) $T = 60$**

- **Demand shock**
  - Wage
  - Employment
- **Supply shock**
  - Wage
  - Employment
C) $T = 120$

Demand shock

Supply shock

D) $T = 240$

Demand shock

Supply shock
Note: The grey solid line refers to the distribution $p(B|Y)_{Np(B)}$, the dashed blue line refers to the distribution $p(B|Y)_{NiWU}$, the blue and grey dotted lines refer to the distributions of the upper bounds $\pm \Sigma_{i,j}^{0.5}$ corresponding to $p(\Sigma|Y)_{Np(B)}$ and $p(\Sigma|Y)_{NiWU}$, respectively. The distributions associated with the NiWU approach refer to the improper prior specification.
### Table F4: Diagnostics on the importance weights, tests

<table>
<thead>
<tr>
<th>$T$</th>
<th>u</th>
<th>0.5N</th>
<th>0.6N</th>
<th>0.7N</th>
<th>0.9N</th>
<th>0.99N</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>Wald</td>
<td>1415.72</td>
<td>985.61</td>
<td>630.74</td>
<td>120.39</td>
<td>3.00</td>
</tr>
<tr>
<td></td>
<td>Score</td>
<td>26.52</td>
<td>23.11</td>
<td>19.47</td>
<td>10.32</td>
<td>2.07</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>NaN</td>
<td>NaN</td>
<td>NaN</td>
<td>NaN</td>
<td>NaN</td>
</tr>
<tr>
<td></td>
<td>Score</td>
<td>-12.48</td>
<td>-10.81</td>
<td>-9.09</td>
<td>-5.18</td>
<td>-1.53</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>120</td>
<td>Wald</td>
<td>-47.72</td>
<td>-42.27</td>
<td>-35.98</td>
<td>-19.11</td>
<td>-5.41</td>
</tr>
<tr>
<td></td>
<td>Score</td>
<td>-18.47</td>
<td>-15.91</td>
<td>-13.30</td>
<td>-6.97</td>
<td>-2.18</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>240</td>
<td>Wald</td>
<td>-52.92</td>
<td>-48.63</td>
<td>-43.02</td>
<td>-25.48</td>
<td>-8.20</td>
</tr>
<tr>
<td></td>
<td>Score</td>
<td>-21.82</td>
<td>-18.72</td>
<td>-15.87</td>
<td>-8.29</td>
<td>-2.76</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>480</td>
<td>Wald</td>
<td>-57.16</td>
<td>-53.74</td>
<td>-48.59</td>
<td>-30.96</td>
<td>-11.11</td>
</tr>
<tr>
<td></td>
<td>Score</td>
<td>-23.47</td>
<td>-20.04</td>
<td>-16.92</td>
<td>-9.05</td>
<td>-2.67</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Note: Reported are the test statistics. The null hypothesis implies finite weight variance. The corresponding critical values above which the null hypothesis is rejected are 1.65 for the Wald test, 1.65 for the score test and 2.68 for the LR test (5% significance level). The corresponding p-values are close to 0 for $T = 30$ and close to 1 for all other cases.
Figure F18: Diagnostics on the importance weights, graphical assessment

Note: The graph shows the recursive variance \( \{v_i\}_{i=1}^N \), where \( v_i = \text{var}(w_{1:i}) \) computed using de-meaned and standardized weights.
**Table F5**: Convergence diagnostics for the application in Section 3 of the paper, $T = 30$

<table>
<thead>
<tr>
<th>parameter</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>all</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geweke (1992), $p$-value. Convergence found if $p$-value $\geq 0.01$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 1</td>
<td>0.20</td>
<td>0.97</td>
<td>0.80</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td>stage 2</td>
<td>0.73</td>
<td>0.17</td>
<td>0.52</td>
<td>0.62</td>
<td></td>
</tr>
<tr>
<td>stage 3</td>
<td>0.37</td>
<td>0.96</td>
<td>0.71</td>
<td>0.31</td>
<td></td>
</tr>
<tr>
<td>stage 4</td>
<td>0.22</td>
<td>0.73</td>
<td>0.98</td>
<td>0.60</td>
<td></td>
</tr>
<tr>
<td>stage 5</td>
<td>0.43</td>
<td>0.80</td>
<td>0.16</td>
<td>0.36</td>
<td></td>
</tr>
<tr>
<td>stage 6</td>
<td>0.44</td>
<td>0.18</td>
<td>0.57</td>
<td>0.89</td>
<td></td>
</tr>
<tr>
<td>Raftery and Lewis (1992), $n^<em>$. Convergence found if $n^</em> \leq N \cdot G$ (64,000)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 1</td>
<td>27,996</td>
<td>29,797</td>
<td>27,655</td>
<td>29,681</td>
<td></td>
</tr>
<tr>
<td>stage 2</td>
<td>26,520</td>
<td>27,536</td>
<td>23,243</td>
<td>26,926</td>
<td></td>
</tr>
<tr>
<td>stage 3</td>
<td>24,413</td>
<td>25,273</td>
<td>21,471</td>
<td>23,524</td>
<td></td>
</tr>
<tr>
<td>stage 4</td>
<td>20,577</td>
<td>22,874</td>
<td>18,598</td>
<td>22,693</td>
<td></td>
</tr>
<tr>
<td>stage 5</td>
<td>25,423</td>
<td>22,634</td>
<td>21,363</td>
<td>24,069</td>
<td></td>
</tr>
<tr>
<td>stage 6</td>
<td>20,655</td>
<td>23,180</td>
<td>21,118</td>
<td>21,405</td>
<td></td>
</tr>
<tr>
<td>Gelman and Rubin (1992), $\hat{R}$. Convergence found if $\hat{R} \leq 1.2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 1</td>
<td>1.002</td>
<td>1.004</td>
<td>1.004</td>
<td>1.003</td>
<td></td>
</tr>
<tr>
<td>stage 2</td>
<td>1.003</td>
<td>1.002</td>
<td>1.001</td>
<td>1.001</td>
<td></td>
</tr>
<tr>
<td>stage 3</td>
<td>1.002</td>
<td>1.001</td>
<td>1.001</td>
<td>1.003</td>
<td></td>
</tr>
<tr>
<td>stage 4</td>
<td>1.002</td>
<td>1.000</td>
<td>1.001</td>
<td>1.002</td>
<td></td>
</tr>
<tr>
<td>stage 5</td>
<td>1.001</td>
<td>1.002</td>
<td>1.001</td>
<td>1.001</td>
<td></td>
</tr>
<tr>
<td>stage 6</td>
<td>1.001</td>
<td>1.002</td>
<td>1.001</td>
<td>1.001</td>
<td></td>
</tr>
<tr>
<td>Brooks and Gelman (1998), $\hat{R}^{mult}$. Convergence found if $\hat{R}^{mult} \leq 1.2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 1</td>
<td>1.006</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 2</td>
<td>1.006</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 3</td>
<td>1.004</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 4</td>
<td>1.003</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 5</td>
<td>1.003</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 6</td>
<td>1.002</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table F6: Convergence diagnostics for the application in Section 3 of the paper, $T = 60$

<table>
<thead>
<tr>
<th>parameter</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>all</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geweke (1992), $p$-value. Convergence found if $p$-value $\geq 0.01$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 1</td>
<td>0.88</td>
<td>0.53</td>
<td>0.34</td>
<td>0.48</td>
<td></td>
</tr>
<tr>
<td>stage 2</td>
<td>0.67</td>
<td>0.01</td>
<td>0.61</td>
<td>0.63</td>
<td></td>
</tr>
<tr>
<td>stage 3</td>
<td>0.73</td>
<td>0.74</td>
<td>0.86</td>
<td>0.87</td>
<td></td>
</tr>
<tr>
<td>stage 4</td>
<td>0.97</td>
<td>0.99</td>
<td>0.84</td>
<td>0.87</td>
<td></td>
</tr>
<tr>
<td>stage 5</td>
<td>0.45</td>
<td>0.75</td>
<td>0.29</td>
<td>0.83</td>
<td></td>
</tr>
<tr>
<td>stage 6</td>
<td>0.07</td>
<td>0.99</td>
<td>0.66</td>
<td>0.81</td>
<td></td>
</tr>
<tr>
<td>Raftery and Lewis (1992), $n^<em>$. Convergence found if $n^</em> \leq N \cdot G (64,000)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 1</td>
<td>29,590</td>
<td>30,642</td>
<td>19,815</td>
<td>26,467</td>
<td></td>
</tr>
<tr>
<td>stage 2</td>
<td>31,074</td>
<td>34,613</td>
<td>21,118</td>
<td>27,324</td>
<td></td>
</tr>
<tr>
<td>stage 3</td>
<td>25,240</td>
<td>27,007</td>
<td>23,181</td>
<td>26,194</td>
<td></td>
</tr>
<tr>
<td>stage 4</td>
<td>24,078</td>
<td>25,324</td>
<td>20,228</td>
<td>24,278</td>
<td></td>
</tr>
<tr>
<td>stage 5</td>
<td>23,460</td>
<td>21,716</td>
<td>21,914</td>
<td>23,460</td>
<td></td>
</tr>
<tr>
<td>stage 6</td>
<td>22,397</td>
<td>23,273</td>
<td>20,856</td>
<td>20,757</td>
<td></td>
</tr>
<tr>
<td>Gelman and Rubin (1992), $\hat{R}$. Convergence found if $\hat{R} \leq 1.2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 1</td>
<td>1.001</td>
<td>1.000</td>
<td>1.001</td>
<td>1.001</td>
<td></td>
</tr>
<tr>
<td>stage 2</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td></td>
</tr>
<tr>
<td>stage 3</td>
<td>1.001</td>
<td>1.002</td>
<td>1.003</td>
<td>1.002</td>
<td></td>
</tr>
<tr>
<td>stage 4</td>
<td>1.000</td>
<td>1.001</td>
<td>1.001</td>
<td>1.002</td>
<td></td>
</tr>
<tr>
<td>stage 5</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>stage 6</td>
<td>1.001</td>
<td>1.000</td>
<td>1.001</td>
<td>1.001</td>
<td></td>
</tr>
<tr>
<td>Brooks and Gelman (1998), $\hat{R}^{mult}$. Convergence found if $\hat{R}^{mult} \leq 1.2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 1</td>
<td>1.002</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 2</td>
<td>1.004</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 3</td>
<td>1.003</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 4</td>
<td>1.002</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 5</td>
<td>1.004</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 6</td>
<td>1.002</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table F7: Convergence diagnostics for the application in Section 3 of the paper, $T = 120$

<table>
<thead>
<tr>
<th>parameter</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>all</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geweke (1992), $p$-value. Convergence found if $p$-value $\geq 0.01$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 1</td>
<td>0.24</td>
<td>0.34</td>
<td>0.11</td>
<td>0.30</td>
<td></td>
</tr>
<tr>
<td>stage 2</td>
<td>0.44</td>
<td>1.00</td>
<td>0.97</td>
<td>0.65</td>
<td></td>
</tr>
<tr>
<td>stage 3</td>
<td>0.58</td>
<td>0.33</td>
<td>0.78</td>
<td>0.83</td>
<td></td>
</tr>
<tr>
<td>stage 4</td>
<td>0.78</td>
<td>0.89</td>
<td>0.35</td>
<td>0.85</td>
<td></td>
</tr>
<tr>
<td>stage 5</td>
<td>0.08</td>
<td>0.00</td>
<td>0.01</td>
<td>0.21</td>
<td></td>
</tr>
<tr>
<td>stage 6</td>
<td>0.29</td>
<td>0.80</td>
<td>0.98</td>
<td>0.60</td>
<td></td>
</tr>
<tr>
<td>Raftery and Lewis (1992), $n^<em>$. Convergence found if $n^</em> \leq N \cdot G$ (64,000)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 1</td>
<td>36,972</td>
<td>36,049</td>
<td>24,406</td>
<td>34,613</td>
<td></td>
</tr>
<tr>
<td>stage 2</td>
<td>36,368</td>
<td>34,740</td>
<td>22,195</td>
<td>37,209</td>
<td></td>
</tr>
<tr>
<td>stage 3</td>
<td>31,218</td>
<td>34,292</td>
<td>23,584</td>
<td>26,926</td>
<td></td>
</tr>
<tr>
<td>stage 4</td>
<td>28,648</td>
<td>27,584</td>
<td>21,856</td>
<td>28,375</td>
<td></td>
</tr>
<tr>
<td>stage 5</td>
<td>26,746</td>
<td>25,469</td>
<td>21,689</td>
<td>25,676</td>
<td></td>
</tr>
<tr>
<td>stage 6</td>
<td>23,335</td>
<td>23,273</td>
<td>21,224</td>
<td>21,468</td>
<td></td>
</tr>
<tr>
<td>Gelman and Rubin (1992), $\hat{R}$. Convergence found if $\hat{R} \leq 1.2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 1</td>
<td>1.002</td>
<td>1.002</td>
<td>1.002</td>
<td>1.001</td>
<td></td>
</tr>
<tr>
<td>stage 2</td>
<td>1.001</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>stage 3</td>
<td>1.001</td>
<td>1.003</td>
<td>1.002</td>
<td>1.001</td>
<td></td>
</tr>
<tr>
<td>stage 4</td>
<td>1.001</td>
<td>1.002</td>
<td>1.001</td>
<td>1.002</td>
<td></td>
</tr>
<tr>
<td>stage 5</td>
<td>1.002</td>
<td>1.002</td>
<td>1.002</td>
<td>1.001</td>
<td></td>
</tr>
<tr>
<td>stage 6</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td></td>
</tr>
<tr>
<td>Brooks and Gelman (1998), $\hat{R}^{mult}$. Convergence found if $\hat{R}^{mult} \leq 1.2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 1</td>
<td>1.004</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 2</td>
<td>1.003</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 3</td>
<td>1.004</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 4</td>
<td>1.004</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 5</td>
<td>1.002</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 6</td>
<td>1.004</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table F8: Convergence diagnostics for the application in Section 3 of the paper, $T = 240$

<table>
<thead>
<tr>
<th>parameter</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>all</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geweke (1992), p-value. Convergence found if $p$-value $\geq 0.01$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 1</td>
<td>0.14</td>
<td>0.09</td>
<td>0.10</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td>stage 2</td>
<td>0.45</td>
<td>0.79</td>
<td>0.88</td>
<td>0.51</td>
<td></td>
</tr>
<tr>
<td>stage 3</td>
<td>0.30</td>
<td>0.64</td>
<td>0.62</td>
<td>0.35</td>
<td></td>
</tr>
<tr>
<td>stage 4</td>
<td>0.00</td>
<td>0.10</td>
<td>0.06</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>stage 5</td>
<td>0.19</td>
<td>0.35</td>
<td>0.20</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>stage 6</td>
<td>0.76</td>
<td>0.50</td>
<td>0.59</td>
<td>0.67</td>
<td></td>
</tr>
<tr>
<td>Raftery and Lewis (1992), $n^\ast$. Convergence found if $n^\ast \leq N \cdot G (64,000)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 1</td>
<td>40,778</td>
<td>50,738</td>
<td>30,664</td>
<td>41,426</td>
<td></td>
</tr>
<tr>
<td>stage 2</td>
<td>44,744</td>
<td>42,822</td>
<td>28,279</td>
<td>40,519</td>
<td></td>
</tr>
<tr>
<td>stage 3</td>
<td>39,909</td>
<td>39,909</td>
<td>26,128</td>
<td>37,019</td>
<td></td>
</tr>
<tr>
<td>stage 4</td>
<td>35,180</td>
<td>37,161</td>
<td>24,506</td>
<td>38,437</td>
<td></td>
</tr>
<tr>
<td>stage 5</td>
<td>27,668</td>
<td>32,870</td>
<td>22,286</td>
<td>28,625</td>
<td></td>
</tr>
<tr>
<td>stage 6</td>
<td>25,832</td>
<td>24,072</td>
<td>20,651</td>
<td>26,152</td>
<td></td>
</tr>
<tr>
<td>Gelman and Rubin (1992), $\hat{R}$. Convergence found if $\hat{R} \leq 1.2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 1</td>
<td>1.003</td>
<td>1.002</td>
<td>1.002</td>
<td>1.003</td>
<td></td>
</tr>
<tr>
<td>stage 2</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td></td>
</tr>
<tr>
<td>stage 3</td>
<td>1.003</td>
<td>1.003</td>
<td>1.004</td>
<td>1.004</td>
<td></td>
</tr>
<tr>
<td>stage 4</td>
<td>1.001</td>
<td>1.002</td>
<td>1.001</td>
<td>1.002</td>
<td></td>
</tr>
<tr>
<td>stage 5</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td></td>
</tr>
<tr>
<td>stage 6</td>
<td>1.001</td>
<td>1.002</td>
<td>1.001</td>
<td>1.001</td>
<td></td>
</tr>
<tr>
<td>Brooks and Gelman (1998), $\hat{R}^{\text{mult}}$. Convergence found if $\hat{R}^{\text{mult}} \leq 1.2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 1</td>
<td>1.003</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 2</td>
<td>1.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 3</td>
<td>1.004</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 4</td>
<td>1.004</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 5</td>
<td>1.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 6</td>
<td>1.003</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table F9: Convergence diagnostics for the application in Section 3 of the paper, \( T = 480 \)

<table>
<thead>
<tr>
<th>parameter</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>all</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geweke (1992), ( p )-value. Convergence found if ( p )-value ( \geq 0.01 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 1</td>
<td>0.23</td>
<td>0.26</td>
<td>0.17</td>
<td>0.24</td>
<td></td>
</tr>
<tr>
<td>stage 2</td>
<td>0.95</td>
<td>0.77</td>
<td>0.74</td>
<td>0.87</td>
<td></td>
</tr>
<tr>
<td>stage 3</td>
<td>0.85</td>
<td>0.42</td>
<td>0.76</td>
<td>0.89</td>
<td></td>
</tr>
<tr>
<td>stage 4</td>
<td>0.09</td>
<td>0.01</td>
<td>0.03</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>stage 5</td>
<td>0.75</td>
<td>0.47</td>
<td>0.96</td>
<td>0.80</td>
<td></td>
</tr>
<tr>
<td>stage 6</td>
<td>0.07</td>
<td>0.02</td>
<td>0.03</td>
<td>0.08</td>
<td></td>
</tr>
</tbody>
</table>

| Raftery and Lewis (1992), \( n^* \). Convergence found if \( n^* \leq N \cdot G (64,000) \) |    |    |    |    |     |
| stage 1   | 50,928 | 40,175 | 34,822 | 47,481 |     |
| stage 2   | 49,372 | 43,116 | 30,363 | 46,902 |     |
| stage 3   | 51,247 | 47,285 | 30,858 | 53,787 |     |
| stage 4   | 42,552 | 35,474 | 27,897 | 42,262 |     |
| stage 5   | 36,416 | 34,673 | 23,651 | 34,829 |     |
| stage 6   | 27,789 | 33,590 | 22,108 | 28,220 |     |

| Gelman and Rubin (1992), \( \hat{R} \). Convergence found if \( \hat{R} \leq 1.2 \) |    |    |    |    |     |
| stage 1   | 1.002 | 1.001 | 1.001 | 1.002 |     |
| stage 2   | 1.002 | 1.001 | 1.001 | 1.001 |     |
| stage 3   | 1.002 | 1.004 | 1.004 | 1.002 |     |
| stage 4   | 1.001 | 1.002 | 1.001 | 1.001 |     |
| stage 5   | 1.001 | 1.001 | 1.001 | 1.001 |     |
| stage 6   | 1.001 | 1.001 | 1.001 | 1.001 |     |

| Brooks and Gelman (1998), \( \hat{R}^{mult} \). Convergence found if \( \hat{R}^{mult} \leq 1.2 \) |    |    |    |    |     |
| stage 1   | 1.003 |    |    |    |     |
| stage 2   | 1.005 |    |    |    |     |
| stage 3   | 1.006 |    |    |    |     |
| stage 4   | 1.003 |    |    |    |     |
| stage 5   | 1.002 |    |    |    |     |
| stage 6   | 1.002 |    |    |    |     |
Figure F19: Marginal distribution on the rotation angle $\theta$ implicit in the marginal prior and posterior distributions $p(Q)$ and $p(Q|Y)$

A) $T = 30$

B) $T = 60$
C) $T = 120$

Np(B) approach

\[ \theta, \text{ radians} \]

NiWU approach

D) $T = 240$

Np(B) approach

\[ \theta, \text{ radians} \]

NiWU approach
Note: The NiWU approach uses the distribution $p(Q|\Sigma)_{NiWU}$, which is the uniform distribution in the matrix parameter space $Q_{\Sigma}$. The Np(B) approach uses the distribution $p(Q|\Sigma)_{Np(B)}$ that is implied by the prior $p(B)_{Np(B)}$. The marginal distributions $p(Q)_{NiWU}$, $p(Q|Y)_{NiWU}$, $p(Q)_{Np(B)}$ and $p(Q|Y)_{Np(B)}$ are obtained by marginalizing out $\Sigma$ using $p(\Sigma)_{NiWU}$, $p(\Sigma|Y)_{NiWU}$, $p(\Sigma|Y)_{Np(B)}$ and $p(\Sigma|Y)_{Np(B)}$, respectively. The figure shows the distribution on the rotation angle $\theta$ associated with the distributions $p(Q)_{NiWU}$, $p(Q|Y)_{NiWU}$, $p(Q)_{Np(B)}$ and $p(Q|Y)_{Np(B)}$. See also the discussion in footnote 13 of the paper and Section D.2 of the Appendix.
G Additional tables/figures for Section 4 of the paper
Figure G20: Data, as it enters the model
Figure G21: $p(\Sigma)_{NWU}$ and $p(\Sigma)_{Np(B)}$

Note: The figure shows the prior distribution $p(\Sigma)$ associated with our approach and with the prior specification by Kadiyala and Karlsson (1997) of the NiWU approach.
Figure G22: $p(\Sigma|Y)_{NiWU}$ and $p(\Sigma|Y)_{Np(B)}$

Note: The figure shows the posterior distribution $p(\Sigma|Y)$ associated with our approach and with the NiWU approach.
Figure G23: $p(B)_{Np(B)}$ comparing prior specifications for our approach

Note: The figure shows the three alternative prior specifications we use for the prior distribution in our Np(B) approach. See also Table 3 in the paper.
Table G10: Allocation of prior probability mass in our Np(B) approach

Estimated bounds $\gamma_i$ used to specify the prior $p(B)_{Np(B)}$

$i = 1$: Oil production growth \quad 0.978  
$i = 2$: Economic activity index \quad 2.668  
$i = 3$: Real oil price \quad 2.953  

Prior probability mass of the absolute value of $B_{i,j}$ beyond $\gamma_i$

<table>
<thead>
<tr>
<th></th>
<th>prior A</th>
<th></th>
<th></th>
<th>prior B</th>
<th></th>
<th></th>
<th>prior C</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_i$</td>
<td>0.84</td>
<td>0.83</td>
<td>0.83</td>
<td>0.52</td>
<td>0.54</td>
<td>0.53</td>
<td>0.20</td>
<td>0.21</td>
<td>0.21</td>
</tr>
<tr>
<td>$\gamma_i$</td>
<td>0.84</td>
<td>0.84</td>
<td>0.84</td>
<td>0.54</td>
<td>0.53</td>
<td>0.53</td>
<td>0.21</td>
<td>0.22</td>
<td>0.21</td>
</tr>
<tr>
<td>$\gamma_i$</td>
<td>0.84</td>
<td>0.84</td>
<td>0.84</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.21</td>
<td>0.20</td>
<td>0.19</td>
</tr>
</tbody>
</table>

97.5th percentile of the marginal distributions $p(B_{i,j})_{Np(B)}$

<table>
<thead>
<tr>
<th></th>
<th>prior A</th>
<th></th>
<th></th>
<th>prior B</th>
<th></th>
<th></th>
<th>prior C</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_i$</td>
<td>-4.331</td>
<td>4.279</td>
<td>4.352</td>
<td>-2.158</td>
<td>2.149</td>
<td>2.135</td>
<td>-1.246</td>
<td>1.259</td>
<td>1.253</td>
</tr>
</tbody>
</table>

63
Figure G24: $p(B|Y)_{Np(B)}$, comparing algorithms (prior $A$)

Note: The figure shows the posterior distribution $p(B|Y)_{Np(B)}$ associated with prior specification $A$ (Table 3 in the paper), sampled using either our algorithm or the Dynamic Strated Metropolis-Hastings algorithm by Waggoner et al. (2016).
Figure G25: $p(B|Y)_{NP(B)}$, comparing algorithms (prior $B$)

Note: The figure shows the posterior distribution $p(B|Y)_{NP(B)}$ associated with prior specification $B$ (Table 3 in the paper), sampled using either our algorithm or the Dynamic Stratified Metropolis-Hastings algorithm by Waggoner et al. (2016).
Figure G26: $p(B|Y)_{Np(B)}$, comparing algorithms (prior C)

Note: The figure shows the posterior distribution $p(B|Y)_{Np(B)}$ associated with prior specification C (Table 3 in the paper), sampled using either our algorithm or the Dynamic Strated Metropolis-Hastings algorithm by Waggoner et al. (2016).
Figure G27: $p(B|Y)_{NiWU}$, comparing NiWU approach with improper prior and with prior from Kadiyala and Karlsson (1997)

Note: The figure shows the posterior distribution $p(B|Y)_{NiWU}$ when parametrizing the underlying inverse Wishart either with the improper prior or with the prior specification by Kadiyala and Karlsson (1997).
Figure G28: \( p(B|Y) \), comparing our approach to the NiWU approach (prior A)

Note: The figure shows the posterior distribution \( p(B|Y)_{Np(B)} \) associated with prior specification A (Table 3 in the paper) and the posterior distribution \( p(B|Y)_{NiWU} \) associated with the improper prior specification. It then shows the distributions of the upper bounds \( \pm \Sigma_{i,j}^{0.5} \) (equation (22) in the paper) associated with the posterior distributions \( p(\Sigma|Y)_{Np(B)} \) and \( p(\Sigma|Y)_{NiWU} \).
Figure G29: $p(B|Y)$, comparing our approach to the NiWU approach (prior $B$)

Note: The figure shows the posterior distribution $p(B|Y)_{Np(B)}$ associated with prior specification $B$ (Table 3 in the paper) and the posterior distribution $p(B|Y)_{NiWU}$ associated with the improper prior specification. It then shows the distributions of the upper bounds $\pm \Sigma_{i,j}^0.5$ (equation (22) in the paper) associated with the posterior distributions $p(\Sigma|Y)_{Np(B)}$ and $p(\Sigma|Y)_{NiWU}$. 
Figure G30: $p(B|Y)$, comparing our approach to the NiWU approach (prior C)

Note: The figure shows the posterior distribution $p(B|Y)_{N_p(B)}$ associated with prior specification C (Table 3 in the paper) and the posterior distribution $p(B|Y)_{NiWU}$ associated with the improper prior specification. It then shows the distributions of the upper bounds $\pm \Sigma_{i,j}^{0.5}$ (equation (22) in the paper) associated with the posterior distributions $p(\Sigma|Y)_{N_p(B)}$ and $p(\Sigma|Y)_{NiWU}$.
Figure G31: IRFs, comparing NiWU approach with improper prior and with prior from Kadiyala and Karlsson (1997)

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 inverse Wishart distribution. The x-ed, dashed and solid lines show the same statistics associated with the prior specification of the inverse Wishart by Kadiyala and Karlsson (1997). As in Antolín-Díaz and Rubio-Ramírez (2018), the response of the oil production has been accumulated to the level.
Figure G32: IRFs, comparing NiWU and Np(B) (prior A)

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 under prior specification A. As in Antolín-Díaz and Rubio-Ramírez (2018), the response of the oil production has been accumulated to the level.
Figure G33: IRFs, comparing NiWU and Np(B) (prior 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 under prior specification B. As in Antolín-Díaz and Rubio-Ramírez (2018), the response of the oil production has been accumulated to the level.
**Figure G34:** IRFs, comparing NiWU and Np(B) (prior $C$)

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 under prior specification $C$. As in Antolín-Díaz and Rubio-Ramírez (2018), the response of the oil production has been accumulated to the level.
Figure G35: FEVD, comparing NiWU and Np(B) (prior $A$)

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 under prior specification $A$. 


Figure G36: FEVD, comparing NiWU and Np(B) (prior $B$)

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 under prior specification $B$. 
Figure G37: FEVD, comparing NiWU and Np(B) (prior C)

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 under prior specification C.
Figure G38: Sensitivity analysis for Figure 4 in the paper. Posterior impulse responses for the real oil price, comparing NiWU and \(N_p(B)\), estimating \(\gamma_i\) with the univariate AR process used in the Minnesota prior.

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 \(N_p(B)\) approach. The rows of the figure differ for the parametrization used for the prior distribution \(p(B)_{N_p(B)}\), as from Table 3.
Figure G39: Sensitivity analysis for Figure 4 in the paper. Posterior impulse responses for all shocks and all variables, exploring $p(B|Y)_{Np(B)}$ for Prior C setting $m_5 = 2,000$ rather than 1,000.

Note: The dotted lines and the shaded areas show the pointwise median, 68 and 95% credible bands. The figure shows the case posterior associated with Prior C, explored either setting $m_5 = 1,000$ in Algorithm 2 (baseline parametrization) or $m_5 = 2,000$. As in Antolín-Díaz and Rubio-Ramírez (2018), the response of the oil production has been accumulated to the level.
Figure G40: Historical decomposition, June – December 1990 (prior A)

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 June 1990.
Figure G41: Historical decomposition, June – December 1990 (prior B)

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 June 1990.
Figure G42: Historical decomposition, June – December 1990 (prior $C$)

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 June 1990.
Figure G43: Price and demand elasticities

Np(B) approach, prior

Price elasticity of oil supply

Price elasticity of oil demand

Np(B) approach, posterior

Price elasticity of oil supply

Price elasticity of oil demand

NiWU approach, posterior

Price elasticity of oil supply

Price elasticity of oil demand

Note: Elasticities are computed from the A form of the structural VAR, equation (1) in the paper. The figure also shows the posterior medians estimated by Baumeister and Hamilton (forthcoming) and by Caldara et al. (2018).
Figure G44: Behaviour of the relevant effective sample size in our algorithm when the size of the dataset increases

Note: The figure reports the ratio of the effective sample size relative to the proposal draws from Stage A of our algorithm. The training sample used to set the prior beliefs is the same as in the application in the paper. We then use the first $p + \tau$ observations of the estimation sample and progressively increase $\tau$, starting from $\tau = 85$. The ratio reaches approximately 0.50 for $\tau = 115$ (prior A), $\tau = 165$ (prior B) and $\tau = 185$ (prior C). 50,000 proposal draws were initially extracted for all cases considered.
### Table G11: Performance of our algorithm

<table>
<thead>
<tr>
<th></th>
<th>Stage A</th>
<th></th>
<th>Stage B</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>proposal effective sample size</td>
<td>relative effective sample size</td>
<td>proposal effective sample size</td>
<td>relative effective sample size</td>
</tr>
<tr>
<td></td>
<td>$m_2$</td>
<td>$ESS_A$</td>
<td>$ESS_A^{m_2}$</td>
<td>$m_4$</td>
</tr>
<tr>
<td>Prior A</td>
<td>50,000</td>
<td>44,278</td>
<td>0.8856</td>
<td>80,000</td>
</tr>
<tr>
<td>Prior B</td>
<td>50,000</td>
<td>41,044</td>
<td>0.8209</td>
<td>80,000</td>
</tr>
<tr>
<td>Prior C</td>
<td>50,000</td>
<td>39,212</td>
<td>0.7842</td>
<td>880,000</td>
</tr>
</tbody>
</table>

Note: See Figure G39 for the sensitivity analysis for Prior C when ensuring a higher number of effective draws in Stage B.

### Table G12: Comparison of the computational time

<table>
<thead>
<tr>
<th></th>
<th>Np(B) approach</th>
<th>NiWU approach</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Our algorithm</td>
<td>DSMH algorithm</td>
</tr>
<tr>
<td></td>
<td>Stage A</td>
<td>Stage B</td>
</tr>
<tr>
<td>Prior A</td>
<td>4</td>
<td>32</td>
</tr>
<tr>
<td>Prior B</td>
<td>4</td>
<td>32</td>
</tr>
<tr>
<td>Prior C</td>
<td>4</td>
<td>30</td>
</tr>
</tbody>
</table>

Note: All codes are run on Matlab, except for the Dynamic Striated Metropolis-Hastings algorithm, which we coded on Fortran to reduce computational time.
### Table G13: Diagnostics on the importance weights, tests

<table>
<thead>
<tr>
<th></th>
<th>u</th>
<th>0.5N</th>
<th>0.6N</th>
<th>0.7N</th>
<th>0.9N</th>
<th>0.99N</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>using prior A</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wald</td>
<td>-51.78</td>
<td>-47.14</td>
<td>-41.34</td>
<td>-23.86</td>
<td>-6.95</td>
<td></td>
</tr>
<tr>
<td>Score</td>
<td>-21.11</td>
<td>-18.16</td>
<td>-15.23</td>
<td>-8.04</td>
<td>-2.54</td>
<td></td>
</tr>
<tr>
<td>LR</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td><strong>using prior B</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wald</td>
<td>-49.87</td>
<td>-44.92</td>
<td>-38.93</td>
<td>-21.76</td>
<td>-6.27</td>
<td></td>
</tr>
<tr>
<td>Score</td>
<td>-20.12</td>
<td>-17.17</td>
<td>-14.71</td>
<td>-7.93</td>
<td>-2.44</td>
<td></td>
</tr>
<tr>
<td>LR</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td><strong>using prior C</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wald</td>
<td>-48.62</td>
<td>-43.48</td>
<td>-37.35</td>
<td>-20.57</td>
<td>-5.85</td>
<td></td>
</tr>
<tr>
<td>LR</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
</tbody>
</table>

Note: Reported are the test statistics. The null hypothesis implies finite weight variance. The corresponding critical values above which the null hypothesis is rejected are 1.65 for the Wald test, 1.65 for the score test and 2.68 for the LR test (5% significance level). The corresponding p-values are close to 1 in all cases.
Figure G45: Diagnostics on the importance weights, graphical assessment

Note: The graph shows the recursive variance $\{v_i\}_{i=1}^N$, where $v_i = \text{var}(w_{1,i})$ computed using de-meaned and standardized weights.
Table G14: Convergence for the application in Section 4 of the paper (prior A)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geweke (1992), ( p )-value. Convergence found if ( p )-value ( \geq ) 0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage 1</td>
<td>0.89</td>
<td>0.99</td>
<td>0.76</td>
<td>0.52</td>
<td>0.97</td>
<td>0.81</td>
<td>0.89</td>
<td>0.93</td>
<td>0.60</td>
<td></td>
</tr>
<tr>
<td>Stage 2</td>
<td>0.34</td>
<td>0.10</td>
<td>0.91</td>
<td>0.70</td>
<td>0.45</td>
<td>0.53</td>
<td>0.50</td>
<td>0.95</td>
<td>0.32</td>
<td></td>
</tr>
<tr>
<td>Stage 3</td>
<td>0.97</td>
<td>0.00</td>
<td>0.44</td>
<td>0.38</td>
<td>0.73</td>
<td>0.11</td>
<td>0.20</td>
<td>0.84</td>
<td>0.77</td>
<td></td>
</tr>
<tr>
<td>Stage 4</td>
<td>0.34</td>
<td>0.13</td>
<td>0.07</td>
<td>0.10</td>
<td>0.02</td>
<td>0.00</td>
<td>0.53</td>
<td>0.11</td>
<td>0.19</td>
<td></td>
</tr>
<tr>
<td>Stage 5</td>
<td>0.04</td>
<td>0.47</td>
<td>0.30</td>
<td>0.80</td>
<td>0.43</td>
<td>0.37</td>
<td>0.09</td>
<td>0.20</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>Stage 6</td>
<td>0.51</td>
<td>0.22</td>
<td>0.73</td>
<td>0.89</td>
<td>0.13</td>
<td>0.84</td>
<td>0.13</td>
<td>0.48</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>Stage 7</td>
<td>0.24</td>
<td>0.09</td>
<td>0.91</td>
<td>0.42</td>
<td>0.54</td>
<td>0.88</td>
<td>0.13</td>
<td>0.10</td>
<td>0.67</td>
<td></td>
</tr>
<tr>
<td>Stage 8</td>
<td>0.71</td>
<td>0.87</td>
<td>0.35</td>
<td>0.67</td>
<td>0.53</td>
<td>0.84</td>
<td>0.82</td>
<td>0.83</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td>Raftery and Lewis (1992), ( n^* ). Convergence found if ( n^* \leq N \cdot G ) (96,000)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage 1</td>
<td>41,344</td>
<td>46,534</td>
<td>49,785</td>
<td>35,215</td>
<td>51,856</td>
<td>51,204</td>
<td>54,525</td>
<td>38,755</td>
<td>35,627</td>
<td></td>
</tr>
<tr>
<td>Stage 2</td>
<td>37,434</td>
<td>50,066</td>
<td>53,377</td>
<td>36,224</td>
<td>54,932</td>
<td>46,233</td>
<td>56,568</td>
<td>36,924</td>
<td>37,273</td>
<td></td>
</tr>
<tr>
<td>Stage 3</td>
<td>38,730</td>
<td>44,984</td>
<td>56,174</td>
<td>34,828</td>
<td>50,566</td>
<td>53,771</td>
<td>57,904</td>
<td>39,664</td>
<td>35,354</td>
<td></td>
</tr>
<tr>
<td>Stage 4</td>
<td>34,979</td>
<td>48,661</td>
<td>47,399</td>
<td>31,605</td>
<td>57,975</td>
<td>46,849</td>
<td>50,863</td>
<td>39,008</td>
<td>34,591</td>
<td></td>
</tr>
<tr>
<td>Stage 5</td>
<td>33,342</td>
<td>54,051</td>
<td>48,941</td>
<td>29,899</td>
<td>45,119</td>
<td>48,333</td>
<td>48,579</td>
<td>36,675</td>
<td>31,792</td>
<td></td>
</tr>
<tr>
<td>Stage 6</td>
<td>32,761</td>
<td>44,036</td>
<td>44,523</td>
<td>28,301</td>
<td>48,893</td>
<td>51,204</td>
<td>46,194</td>
<td>34,675</td>
<td>31,982</td>
<td></td>
</tr>
<tr>
<td>Stage 7</td>
<td>32,849</td>
<td>42,128</td>
<td>50,356</td>
<td>28,183</td>
<td>46,572</td>
<td>47,749</td>
<td>49,348</td>
<td>36,461</td>
<td>29,521</td>
<td></td>
</tr>
<tr>
<td>Stage 8</td>
<td>31,937</td>
<td>42,577</td>
<td>47,819</td>
<td>28,189</td>
<td>46,003</td>
<td>40,744</td>
<td>44,222</td>
<td>32,982</td>
<td>27,444</td>
<td></td>
</tr>
<tr>
<td>Gelman and Rubin (1992), ( \hat{R} ). Convergence found if ( \hat{R} \leq 1.2 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage 1</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td></td>
</tr>
<tr>
<td>Stage 2</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td></td>
</tr>
<tr>
<td>Stage 3</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td></td>
</tr>
<tr>
<td>Stage 4</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td></td>
</tr>
<tr>
<td>Stage 5</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td></td>
</tr>
<tr>
<td>Stage 6</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td></td>
</tr>
<tr>
<td>Stage 7</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td></td>
</tr>
<tr>
<td>Stage 8</td>
<td>1.006</td>
<td>1.001</td>
<td>1.005</td>
<td>1.002</td>
<td>1.002</td>
<td>1.004</td>
<td>1.002</td>
<td>1.002</td>
<td>1.003</td>
<td></td>
</tr>
<tr>
<td>Brooks and Gelman (1998), ( \hat{R}^{mult} ). Convergence found if ( \hat{R}^{mult} \leq 1.2 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.005</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.006</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.006</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.005</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.005</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage 7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.006</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage 8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.009</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table G15: Convergence diagnostics for the application in Section 4 of the paper (prior $B$)

<table>
<thead>
<tr>
<th>parameter</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>all</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geweke (1992), $p$-value. Convergence found if $p$-value $\geq 0.01$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 1</td>
<td>0.89</td>
<td>0.99</td>
<td>0.91</td>
<td>0.77</td>
<td>0.76</td>
<td>0.63</td>
<td>0.99</td>
<td>0.97</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td>stage 2</td>
<td>0.43</td>
<td>0.83</td>
<td>0.69</td>
<td>0.51</td>
<td>0.74</td>
<td>0.38</td>
<td>0.64</td>
<td>0.52</td>
<td>0.92</td>
<td></td>
</tr>
<tr>
<td>stage 3</td>
<td>0.43</td>
<td>0.22</td>
<td>0.96</td>
<td>0.16</td>
<td>0.72</td>
<td>0.64</td>
<td>0.30</td>
<td>0.19</td>
<td>0.97</td>
<td></td>
</tr>
<tr>
<td>stage 4</td>
<td>0.83</td>
<td>0.46</td>
<td>0.47</td>
<td>0.10</td>
<td>0.39</td>
<td>0.12</td>
<td>0.60</td>
<td>0.47</td>
<td>0.87</td>
<td></td>
</tr>
<tr>
<td>stage 5</td>
<td>0.00</td>
<td>0.06</td>
<td>0.04</td>
<td>0.23</td>
<td>0.87</td>
<td>0.40</td>
<td>0.00</td>
<td>0.00</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td>stage 6</td>
<td>0.47</td>
<td>0.92</td>
<td>0.92</td>
<td>0.74</td>
<td>0.80</td>
<td>0.72</td>
<td>0.39</td>
<td>0.92</td>
<td>0.66</td>
<td></td>
</tr>
<tr>
<td>stage 7</td>
<td>0.40</td>
<td>0.03</td>
<td>0.18</td>
<td>0.28</td>
<td>0.89</td>
<td>0.20</td>
<td>0.11</td>
<td>0.32</td>
<td>0.43</td>
<td></td>
</tr>
<tr>
<td>stage 8</td>
<td>0.01</td>
<td>0.89</td>
<td>0.09</td>
<td>0.78</td>
<td>0.01</td>
<td>0.03</td>
<td>0.07</td>
<td>0.02</td>
<td>0.33</td>
<td></td>
</tr>
<tr>
<td>Raftery and Lewis (1992), $n^<em>$. Convergence found if $n^</em> \leq N \cdot G (96,000)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 1</td>
<td>38,119</td>
<td>47,370</td>
<td>45,685</td>
<td>33,838</td>
<td>48,710</td>
<td>54,814</td>
<td>50,787</td>
<td>38,627</td>
<td>35,620</td>
<td></td>
</tr>
<tr>
<td>stage 2</td>
<td>36,111</td>
<td>51,075</td>
<td>45,272</td>
<td>36,428</td>
<td>53,673</td>
<td>52,694</td>
<td>50,161</td>
<td>38,833</td>
<td>36,033</td>
<td></td>
</tr>
<tr>
<td>stage 3</td>
<td>32,480</td>
<td>52,883</td>
<td>47,749</td>
<td>32,248</td>
<td>47,931</td>
<td>47,439</td>
<td>45,531</td>
<td>36,582</td>
<td>34,444</td>
<td></td>
</tr>
<tr>
<td>stage 4</td>
<td>34,444</td>
<td>45,531</td>
<td>44,619</td>
<td>31,061</td>
<td>46,337</td>
<td>47,739</td>
<td>46,519</td>
<td>36,616</td>
<td>29,441</td>
<td></td>
</tr>
<tr>
<td>stage 5</td>
<td>32,019</td>
<td>43,502</td>
<td>46,992</td>
<td>30,289</td>
<td>47,548</td>
<td>44,677</td>
<td>46,233</td>
<td>34,586</td>
<td>31,792</td>
<td></td>
</tr>
<tr>
<td>stage 6</td>
<td>35,469</td>
<td>43,375</td>
<td>41,654</td>
<td>29,404</td>
<td>40,952</td>
<td>47,123</td>
<td>50,108</td>
<td>32,027</td>
<td>29,643</td>
<td></td>
</tr>
<tr>
<td>stage 7</td>
<td>31,382</td>
<td>41,270</td>
<td>43,832</td>
<td>30,067</td>
<td>41,937</td>
<td>45,093</td>
<td>47,313</td>
<td>32,065</td>
<td>28,585</td>
<td></td>
</tr>
<tr>
<td>stage 8</td>
<td>30,502</td>
<td>38,932</td>
<td>43,630</td>
<td>31,056</td>
<td>44,353</td>
<td>44,017</td>
<td>40,908</td>
<td>34,958</td>
<td>28,010</td>
<td></td>
</tr>
<tr>
<td>Gelman and Rubin (1992), $\hat{R}$. Convergence found if $\hat{R} \leq 1.2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 1</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td></td>
</tr>
<tr>
<td>stage 2</td>
<td>1.002</td>
<td>1.000</td>
<td>1.002</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td></td>
</tr>
<tr>
<td>stage 3</td>
<td>1.002</td>
<td>1.003</td>
<td>1.001</td>
<td>1.003</td>
<td>1.002</td>
<td>1.003</td>
<td>1.004</td>
<td>1.004</td>
<td>1.001</td>
<td></td>
</tr>
<tr>
<td>stage 4</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.002</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.002</td>
<td>1.001</td>
<td></td>
</tr>
<tr>
<td>stage 5</td>
<td>1.003</td>
<td>1.003</td>
<td>1.004</td>
<td>1.004</td>
<td>1.001</td>
<td>1.003</td>
<td>1.003</td>
<td>1.002</td>
<td>1.003</td>
<td></td>
</tr>
<tr>
<td>stage 6</td>
<td>1.003</td>
<td>1.001</td>
<td>1.002</td>
<td>1.002</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.003</td>
<td></td>
</tr>
<tr>
<td>stage 7</td>
<td>1.002</td>
<td>1.002</td>
<td>1.002</td>
<td>1.001</td>
<td>1.001</td>
<td>1.003</td>
<td>1.002</td>
<td>1.001</td>
<td>1.002</td>
<td></td>
</tr>
<tr>
<td>stage 8</td>
<td>1.001</td>
<td>1.002</td>
<td>1.001</td>
<td>1.000</td>
<td>1.002</td>
<td>1.002</td>
<td>1.001</td>
<td>1.002</td>
<td>1.001</td>
<td></td>
</tr>
<tr>
<td>Brooks and Gelman (1998), $\hat{R}^{mult}$. Convergence found if $\hat{R}^{mult} \leq 1.2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 1</td>
<td>1.004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 2</td>
<td>1.005</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 3</td>
<td>1.006</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 4</td>
<td>1.006</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 5</td>
<td>1.007</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 6</td>
<td>1.006</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 7</td>
<td>1.007</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 8</td>
<td>1.004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>parameter</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>all</td>
</tr>
<tr>
<td>-----------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>Geweke (1992), p-value. Convergence found if ( p )-value ( \geq 0.01 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 1</td>
<td>0.02</td>
<td>0.68</td>
<td>0.09</td>
<td>0.14</td>
<td>0.76</td>
<td>0.31</td>
<td>0.06</td>
<td>0.93</td>
<td>0.28</td>
<td></td>
</tr>
<tr>
<td>stage 2</td>
<td>0.93</td>
<td>0.93</td>
<td>0.46</td>
<td>0.87</td>
<td>0.28</td>
<td>0.86</td>
<td>0.44</td>
<td>0.34</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td>stage 3</td>
<td>0.13</td>
<td>0.36</td>
<td>0.84</td>
<td>0.60</td>
<td>0.52</td>
<td>0.75</td>
<td>0.28</td>
<td>0.69</td>
<td>0.20</td>
<td></td>
</tr>
<tr>
<td>stage 4</td>
<td>0.51</td>
<td>0.96</td>
<td>0.45</td>
<td>0.50</td>
<td>0.89</td>
<td>0.78</td>
<td>0.70</td>
<td>0.85</td>
<td>0.69</td>
<td></td>
</tr>
<tr>
<td>stage 5</td>
<td>0.59</td>
<td>0.46</td>
<td>0.55</td>
<td>0.88</td>
<td>0.23</td>
<td>0.39</td>
<td>0.96</td>
<td>0.69</td>
<td>0.90</td>
<td></td>
</tr>
<tr>
<td>stage 6</td>
<td>0.83</td>
<td>0.98</td>
<td>0.23</td>
<td>0.56</td>
<td>0.74</td>
<td>0.66</td>
<td>0.31</td>
<td>0.84</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>stage 7</td>
<td>0.28</td>
<td>0.01</td>
<td>0.50</td>
<td>0.74</td>
<td>0.00</td>
<td>0.23</td>
<td>0.15</td>
<td>0.49</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>stage 8</td>
<td>0.50</td>
<td>0.27</td>
<td>0.94</td>
<td>0.17</td>
<td>0.52</td>
<td>0.29</td>
<td>0.23</td>
<td>0.11</td>
<td>0.84</td>
<td></td>
</tr>
<tr>
<td>Raftery and Lewis (1992), ( n^* ). Convergence found if ( n^* \leq N \cdot G ) (96,000)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 1</td>
<td>35,121</td>
<td>50,821</td>
<td>52,477</td>
<td>36,431</td>
<td>49,224</td>
<td>55,566</td>
<td>52,166</td>
<td>38,020</td>
<td>38,813</td>
<td></td>
</tr>
<tr>
<td>stage 2</td>
<td>33,157</td>
<td>44,446</td>
<td>48,859</td>
<td>31,199</td>
<td>51,033</td>
<td>46,716</td>
<td>51,724</td>
<td>38,660</td>
<td>33,370</td>
<td></td>
</tr>
<tr>
<td>stage 3</td>
<td>33,031</td>
<td>46,236</td>
<td>44,946</td>
<td>31,490</td>
<td>46,066</td>
<td>45,493</td>
<td>47,971</td>
<td>35,633</td>
<td>35,043</td>
<td></td>
</tr>
<tr>
<td>stage 4</td>
<td>30,094</td>
<td>39,710</td>
<td>44,581</td>
<td>29,695</td>
<td>38,689</td>
<td>39,670</td>
<td>39,937</td>
<td>29,771</td>
<td>30,855</td>
<td></td>
</tr>
<tr>
<td>stage 5</td>
<td>31,738</td>
<td>34,611</td>
<td>34,979</td>
<td>30,146</td>
<td>34,503</td>
<td>38,252</td>
<td>37,792</td>
<td>30,211</td>
<td>27,660</td>
<td></td>
</tr>
<tr>
<td>stage 6</td>
<td>28,836</td>
<td>31,296</td>
<td>34,924</td>
<td>28,604</td>
<td>33,818</td>
<td>36,831</td>
<td>33,774</td>
<td>30,490</td>
<td>28,408</td>
<td></td>
</tr>
<tr>
<td>stage 7</td>
<td>28,922</td>
<td>30,369</td>
<td>31,034</td>
<td>26,680</td>
<td>31,453</td>
<td>32,818</td>
<td>31,605</td>
<td>26,367</td>
<td>27,115</td>
<td></td>
</tr>
<tr>
<td>stage 8</td>
<td>29,268</td>
<td>28,358</td>
<td>32,257</td>
<td>27,054</td>
<td>31,453</td>
<td>29,453</td>
<td>29,229</td>
<td>26,862</td>
<td>28,670</td>
<td></td>
</tr>
<tr>
<td>Gelman and Rubin (1992), ( \hat{R} ). Convergence found if ( \hat{R} \leq 1.2 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 1</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.002</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.003</td>
<td>1.003</td>
</tr>
<tr>
<td>stage 2</td>
<td>1.001</td>
<td>1.001</td>
<td>1.000</td>
<td>1.001</td>
<td>1.001</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.001</td>
<td>1.001</td>
</tr>
<tr>
<td>stage 3</td>
<td>1.001</td>
<td>1.001</td>
<td>1.000</td>
<td>1.001</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.001</td>
<td>1.001</td>
</tr>
<tr>
<td>stage 4</td>
<td>1.001</td>
<td>1.001</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
</tr>
<tr>
<td>stage 5</td>
<td>1.001</td>
<td>1.001</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.001</td>
<td>1.002</td>
</tr>
<tr>
<td>stage 6</td>
<td>1.001</td>
<td>1.001</td>
<td>1.002</td>
<td>1.002</td>
<td>1.000</td>
<td>1.002</td>
<td>1.000</td>
<td>1.000</td>
<td>1.001</td>
<td>1.002</td>
</tr>
<tr>
<td>stage 7</td>
<td>1.003</td>
<td>1.001</td>
<td>1.003</td>
<td>1.002</td>
<td>1.002</td>
<td>1.001</td>
<td>1.002</td>
<td>1.001</td>
<td>1.001</td>
<td>1.002</td>
</tr>
<tr>
<td>stage 8</td>
<td>1.002</td>
<td>1.001</td>
<td>1.001</td>
<td>1.000</td>
<td>1.001</td>
<td>1.002</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
<td>1.001</td>
</tr>
<tr>
<td>Brooks and Gelman (1998), ( \hat{R}^{mult} ). Convergence found if ( \hat{R}^{mult} \leq 1.2 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage 1</td>
<td></td>
<td>1.009</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.009</td>
</tr>
<tr>
<td>stage 2</td>
<td></td>
<td>1.005</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.005</td>
</tr>
<tr>
<td>stage 3</td>
<td></td>
<td>1.006</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.006</td>
</tr>
<tr>
<td>stage 4</td>
<td></td>
<td>1.004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.004</td>
</tr>
<tr>
<td>stage 5</td>
<td></td>
<td>1.004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.004</td>
</tr>
<tr>
<td>stage 6</td>
<td></td>
<td>1.005</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.005</td>
</tr>
<tr>
<td>stage 7</td>
<td></td>
<td>1.007</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.007</td>
</tr>
<tr>
<td>stage 8</td>
<td></td>
<td>1.005</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.005</td>
</tr>
</tbody>
</table>
References


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


