Supplement to "A new posterior sampler for Bayesian structural vector autoregressive models"

(*Quantitative Economics*, Vol. 14, No. 4, November 2023, 1221–1250)

Martin Bruns
School of Economics, University of East Anglia

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

S.1. Likelihood function of the model

To derive the likelihood function of the model, start from equation (1) of the paper, which we rewrite here for convenience:

$$\boldsymbol{y}_t = \Pi \boldsymbol{w}_t + B\boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim N(\mathbf{0}, I_k). \tag{S.1}$$

$\boldsymbol{y}_t$ is a $k \times 1$ vector of variables, $\boldsymbol{\epsilon}_t$ is a $k \times 1$ vector of structural shocks, $\boldsymbol{w}_t$ is an $m \times 1$ vector of lagged variables and the constant, with $m = kp + 1$, Π 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

$$Y = \Pi W + BE,$$

where $Y = [\boldsymbol{y}_1, \dots, \boldsymbol{y}_t, \dots, \boldsymbol{y}_T]$ and $E = [\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_t, \dots, \boldsymbol{\epsilon}_T]$ are $k \times T$ matrices of data and shocks, and $W = [\boldsymbol{w}_1, \dots, \boldsymbol{w}_t, \dots, \boldsymbol{w}_T]$ is an $m \times T$ matrix of data. Then make use of the formula $\text{vec}(ABC) = (C' \otimes A) \cdot \text{vec}(B)$ (see Lütkepohl (2005), Mathematical Appendix) and rewrite the model as

$$\tilde{\boldsymbol{y}} = Z\boldsymbol{\pi} + (I_T \otimes B)\tilde{\boldsymbol{\epsilon}}, \quad \tilde{\boldsymbol{\epsilon}} \sim N(\mathbf{0}, I_{Tk}),$$

with $\tilde{\boldsymbol{y}} = \text{vec}(Y)$ and $\tilde{\boldsymbol{\epsilon}} = \text{vec}(E)$ of dimension $kT \times 1$ and

$$Z = \left(W' \otimes I_k\right),$$

of dimension $kT \times mk$. The $mk \times 1$ vector $\boldsymbol{\pi} = \text{vec}(\Pi)$ stacks the columns of Π vertically. Lastly, rewrite the model as

$$\tilde{\boldsymbol{y}} = Z\boldsymbol{\pi} + \tilde{\boldsymbol{u}}, \quad \tilde{\boldsymbol{u}} \sim N\left(\mathbf{0}, \left(I_T \otimes BB'\right)\right),$$

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

Available at http://qeconomics.org. https://doi.org/10.3982/QE2207

with $\tilde{\boldsymbol{u}} = (I_T \otimes B)\tilde{\boldsymbol{\epsilon}}$ the VAR innovations. Define the estimators

$$\begin{aligned}
\hat{\Pi}_T &= YW'\big(WW'\big)^{-1}, \\
\hat{\boldsymbol{\pi}}_T &= \operatorname{vec}(\hat{\Pi}_T) = \big(\big(WW'\big)^{-1}W \otimes I_k\big)\tilde{\boldsymbol{y}}, \\
\hat{\Sigma}_T &= \frac{(Y - \hat{\Pi}_T W)(Y - \hat{\Pi}_T W)'}{T - m}.
\end{aligned}$$

The likelihood function of model (S.1) can be written in $(\boldsymbol{\pi}, B)$ as

$$\begin{aligned}
p(Y|\boldsymbol{\pi}, B) &= (2\pi)^{-\frac{kT}{2}}\big|\det\big(I_T \otimes BB'\big)\big|^{-\frac{1}{2}} e^{-\frac{1}{2}(\tilde{\boldsymbol{y}}-Z\boldsymbol{\pi})'(I_T\otimes BB')^{-1}(\tilde{\boldsymbol{y}}-Z\boldsymbol{\pi})} \\
&\propto \big|\det(B)\big|^{-T} e^{-\frac{1}{2}(\tilde{\boldsymbol{y}}-(W'\otimes I_T)\boldsymbol{\pi})'(I_T\otimes(BB')^{-1})(\tilde{\boldsymbol{y}}-(W'\otimes I_T)\boldsymbol{\pi})} \\
&\propto \big|\det(B)\big|^{-T} e^{-\frac{1}{2}\{\tilde{\boldsymbol{y}}'(I_T\otimes(BB')^{-1})\tilde{\boldsymbol{y}}+\boldsymbol{\pi}'(WW'\otimes(BB')^{-1})\boldsymbol{\pi}-2\boldsymbol{\pi}'(W\otimes(BB')^{-1})\tilde{\boldsymbol{y}}\}} \\
&\propto \big|\det(B)\big|^{-T} e^{-\frac{1}{2}\{\boldsymbol{\pi}'(WW'\otimes(BB')^{-1})\boldsymbol{\pi}-2\boldsymbol{\pi}'(WW'\otimes(BB')^{-1})\hat{\boldsymbol{\pi}}_T+\tilde{\boldsymbol{y}}'(I_T\otimes(BB')^{-1})\tilde{\boldsymbol{y}}\}},
\end{aligned}$$

where the last step uses

$$\begin{aligned}
2\boldsymbol{\pi}'\big(W \otimes \big(BB'\big)^{-1}\big)\tilde{\boldsymbol{y}} &= 2\boldsymbol{\pi}'\big(WW' \otimes \big(BB'\big)^{-1}\big)\big(\big(WW'\big)^{-1}W \otimes I_k\big)\tilde{\boldsymbol{y}} \\
&= 2\boldsymbol{\pi}'\big(WW' \otimes \big(BB'\big)^{-1}\big)\hat{\boldsymbol{\pi}}_T.
\end{aligned}$$

In the reduced form parameters $(\boldsymbol{\pi}, \Sigma)$, the likelihood function is written as

$$\begin{aligned}
p(Y|\boldsymbol{\pi}, \Sigma) &= (2\pi)^{-\frac{kT}{2}}\big|\det(\Sigma)\big|^{-\frac{T}{2}} e^{-\frac{1}{2}(\tilde{\boldsymbol{y}}-Z\boldsymbol{\pi})'(I_T\otimes\Sigma)^{-1}(\tilde{\boldsymbol{y}}-Z\boldsymbol{\pi})} \\
&\propto \big|\det(\Sigma)\big|^{-\frac{T}{2}} e^{-\frac{1}{2}\{\boldsymbol{\pi}'(WW'\otimes\Sigma^{-1})\boldsymbol{\pi}-2\boldsymbol{\pi}'(WW'\otimes\Sigma^{-1})\hat{\boldsymbol{\pi}}_T+\tilde{\boldsymbol{y}}'(I_T\otimes\Sigma^{-1})\tilde{\boldsymbol{y}}\}}.
\end{aligned}$$

S.2. The independent NiWU approach

This Appendix provides the derivations of the posterior distribution under an independent NiWU prior. We report the results for a variety of specifications, including the case of a flat prior on $\boldsymbol{\pi}$ and a generalized improper inverse Wishart prior on Σ. This facilitates the discussion in the paper, in which the exact specification of the NiWU prior for the proposal draws can be selected freely in order to improve the performance of the importance sampler. The derivations are relatively standard in the literature and are provided here for completeness. For a more general discussion, see Canova (2007), Koop and Korobilis (2010), and Kilian and Lütkepohl (2017). As in the paper, we use notation $p(\cdot)$ and $p_{NiWU,i}(\cdot)$ for the densities for our general prior and for the special case of the independent NiWU prior, respectively. We also differentiate between algorithms that introduce sign restrictions, that is, $\tilde{p}(\cdot)$, $\tilde{p}_{NiWU,i}(\cdot)$, or do not introduce sign restrictions, that is, $p(\cdot)$, $p_{NiWU,i}(\cdot)$, as this difference is important for our sampler.

S.2.1 *Improper flat and generalized inverse Wishart prior*

We first consider the prior beliefs

$$
\begin{aligned}
\tilde{p}_{NiWU,i}(\boldsymbol{\pi},\Sigma,Q) &\propto \mathrm{I}\{\boldsymbol{\pi},\Sigma,Q\}\cdot p_{NiWU,i}(\boldsymbol{\pi},\Sigma,Q), \qquad &\text{(S.2)}\\
p_{NiWU,i}(\boldsymbol{\pi},\Sigma,Q) &= p_{NiWU,i}(\boldsymbol{\pi})\cdot p_{NiWU,i}(\Sigma)\cdot p_{NiWU,i}(Q),\\
p_{NiWU,i}(\boldsymbol{\pi}) &\propto 1,\\
p_{NiWU,i}(\Sigma) &\propto \left|\det(\Sigma)\right|^{\frac{c}{2}}\cdot e^{-\frac{1}{2}\operatorname{trace}[\Sigma^{-1}S]},\\
p_{NiWU,i}(Q) &= U_{O(k)}\propto 1. &\text{(S.3)}
\end{aligned}
$$

The prior on $\boldsymbol{\pi}$ is an improper flat prior. The prior on Σ is a generalized version of the standard inverse Wishart distribution in that it is a proper inverse Wishart with $d=-(c+k+1)$ degrees-of-freedom only if $c\leq-(2k+1)$, $p(\Sigma)$. The above prior implies the joint posterior distribution

$$
\begin{aligned}
\tilde{p}_{NiWU,i}(\boldsymbol{\pi},\Sigma,Q|Y) &\propto \mathrm{I}\{\boldsymbol{\pi},\Sigma,Q\}\cdot p_{NiWU,i}(\boldsymbol{\pi},\Sigma,Q|Y),\\
p_{NiWU,i}(\boldsymbol{\pi},\Sigma,Q|Y) &= p_{NiWU,i}(\boldsymbol{\pi},\Sigma|Y)\cdot p_{NiWU,i}(Q|Y,\boldsymbol{\pi},\Sigma),\\
p_{NiWU,i}(\boldsymbol{\pi},\Sigma|Y) &\propto \left|\det(\Sigma)\right|^{\frac{c}{2}}\cdot e^{-\frac{1}{2}\operatorname{trace}[\Sigma^{-1}S]}\left|\det(\Sigma)\right|^{-\frac{T}{2}} \qquad \text{(S.4)}\\
&\quad\cdot e^{-\frac{1}{2}\{\boldsymbol{\pi}'(WW'\otimes\Sigma^{-1})\boldsymbol{\pi}-2\boldsymbol{\pi}'(WW'\otimes\Sigma^{-1})\hat{\boldsymbol{\pi}}_T+\tilde{\mathbf{y}}'(I_T\otimes\Sigma^{-1})\tilde{\mathbf{y}}\}},\\
p_{NiWU,i}(Q|Y,\boldsymbol{\pi},\Sigma) &= p_{NiWU,i}(Q)=U_{O(k)}\propto 1.
\end{aligned}
$$

Since

$$
\begin{aligned}
&\boldsymbol{\pi}'\left(WW'\otimes\Sigma^{-1}\right)\boldsymbol{\pi}-2\boldsymbol{\pi}'\left(WW'\otimes\Sigma^{-1}\right)\hat{\boldsymbol{\pi}}_T+\tilde{\mathbf{y}}'\left(I_T\otimes\Sigma^{-1}\right)\tilde{\mathbf{y}}\\
&\quad=\boldsymbol{\pi}'V_\pi^{*^{-1}}\boldsymbol{\pi}-2\boldsymbol{\pi}'V_\pi^{*^{-1}}\boldsymbol{\mu}_\pi^*+\tilde{\mathbf{y}}'\left(I_T\otimes\Sigma^{-1}\right)\tilde{\mathbf{y}}\\
&\quad=\left(\boldsymbol{\pi}-\boldsymbol{\mu}_\pi^*\right)'V_\pi^{*^{-1}}\left(\boldsymbol{\pi}-\boldsymbol{\mu}_\pi^*\right)+\tilde{\mathbf{y}}'\left(I_T\otimes\Sigma^{-1}\right)\tilde{\mathbf{y}}-\boldsymbol{\mu}_\pi^{*'}V_\pi^{*^{-1}}\boldsymbol{\mu}_\pi^*\\
&\quad=\left(\boldsymbol{\pi}-\boldsymbol{\mu}_\pi^*\right)'V_\pi^{*^{-1}}\left(\boldsymbol{\pi}-\boldsymbol{\mu}_\pi^*\right)+\tilde{\mathbf{y}}'\left(I_T\otimes\Sigma^{-1}\right)\tilde{\mathbf{y}}-\hat{\boldsymbol{\pi}}_T'\left(WW'\otimes\Sigma^{-1}\right)\hat{\boldsymbol{\pi}}_T,
\end{aligned}
$$

given

$$
\begin{aligned}
V_\pi^* &= \left(WW'\right)^{-1}\otimes\Sigma,\\
\boldsymbol{\mu}_\pi^* &= \hat{\boldsymbol{\pi}}_T,
\end{aligned}
$$

equation (S.4) can be rewritten as

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

Using formula $\mathrm{vec}(A)'(D \otimes B)\,\mathrm{vec}(C) = \mathrm{tr}(A'BCD')$ (see Abadir and Magnus (2005, Chapter 10)) to rewrite

$$\tilde{\boldsymbol{y}}'\big(I_T \otimes \Sigma^{-1}\big)\tilde{\boldsymbol{y}} = \mathrm{trace}\big[Y'\Sigma^{-1}Y\big] = \mathrm{trace}\big[\Sigma^{-1}YY'\big],$$
$$\hat{\boldsymbol{\pi}}'_T\big(WW' \otimes \Sigma^{-1}\big)\hat{\boldsymbol{\pi}}_T = \mathrm{trace}\big[\hat{\Pi}'_T\Sigma^{-1}\hat{\Pi}_T WW'\big] = \mathrm{trace}\big[\Sigma^{-1}\hat{\Pi}_T WW'\hat{\Pi}'_T\big],$$

noticing that

$$\begin{aligned} YY' - \hat{\Pi}_T WW'\hat{\Pi}'_T &= (Y - \hat{\Pi}_T W)(Y - \hat{\Pi}_T W)' \\ &= (Y - \hat{\Pi}_T W)(Y - \hat{\Pi}_T W)'\frac{T-m}{T-m} \\ &= \hat{\Sigma}_T(T-m), \end{aligned}$$

and using the result

$$\det\big(V^*_\pi\big) = \big|\det\big(WW'\big)\big|^k \cdot \big|\det(\Sigma)\big|^m,$$

one can simplify $p_{NiWU,i}(\boldsymbol{\pi}, \Sigma|Y)$ as

$$\begin{aligned} p_{NiWU,i}(\boldsymbol{\pi}, \Sigma|Y) \propto& \big|\det\big(V^*_\pi\big)\big|^{-\frac{1}{2}} \cdot e^{-\frac{1}{2}(\boldsymbol{\pi}-\boldsymbol{\mu}^*_\pi)'V^{*-1}_\pi(\boldsymbol{\pi}-\boldsymbol{\mu}^*_\pi)} \\ &\cdot \big|\det(\Sigma)\big|^{\frac{m}{2}} \cdot \big|\det(\Sigma)\big|^{\frac{c}{2}} \cdot \big|\det(\Sigma)\big|^{-\frac{T}{2}} \\ &\cdot e^{-\frac{1}{2}\mathrm{trace}[\Sigma^{-1}(S+\hat{\Sigma}_T(T-m))]} \\ \propto& \big|\det\big(V^*_\pi\big)\big|^{-\frac{1}{2}} \cdot e^{-\frac{1}{2}(\boldsymbol{\pi}-\boldsymbol{\mu}^*_\pi)'V^{*-1}_\pi(\boldsymbol{\pi}-\boldsymbol{\mu}^*_\pi)} \\ &\cdot \big|\det(\Sigma)\big|^{-\frac{(T-m-c-k-1)+k+1}{2}} \cdot e^{-\frac{1}{2}\mathrm{trace}[\Sigma^{-1}(S+\hat{\Sigma}_T(T-m))]}. \end{aligned}$$

It then follows that the prior beliefs in equations (S.2)–(S.3) imply

$$\begin{aligned} \tilde{p}_{NiWU,i}(\boldsymbol{\pi}, \Sigma, Q|Y) &\propto \mathrm{I}\{\boldsymbol{\pi}, \Sigma, Q\} \cdot p_{NiWU,i}(\boldsymbol{\pi}, \Sigma, Q|Y), \\ p_{NiWU,i}(\boldsymbol{\pi}, \Sigma, Q|Y) &= p_{NiWU,i}(\boldsymbol{\pi}|Y, \Sigma) \cdot p_{NiWU,i}(\Sigma|Y) \cdot p_{NiWU,i}(Q|Y, \boldsymbol{\pi}, \Sigma), \\ p_{NiWU,i}(\Sigma|Y) &= iW\big(S^*, d^*\big), \qquad \text{(S.5)} \\ p_{NiWU,i}(\boldsymbol{\pi}|Y, \Sigma) &= \phi\big(\boldsymbol{\mu}^*_\pi, V^*_\pi\big), \qquad \text{(S.6)} \\ p_{NiWU,i}(Q|Y, \boldsymbol{\pi}, \Sigma) &= p_{NiWU,i}(Q) = U_{O(k)} \propto 1, \\ d^* &= T - m - c - k - 1, \\ S^* &= S + \hat{\Sigma}_T(T-m), \\ \boldsymbol{\mu}^*_\pi &= \hat{\boldsymbol{\pi}}_T, \\ V^*_\pi &= \big(WW'\big)^{-1} \otimes \Sigma, \end{aligned}$$

with $iW(\cdot,\cdot)$ and $\phi(\cdot,\cdot)$ the probability density functions of the inverse Wishart and the Normal distribution, respectively. While in the independent NiWU prior it is usually $p_{NiWU,i}(\Sigma|Y,\boldsymbol{\pi})$ rather than $p_{NiWU,i}(\Sigma|Y)$ to be inverse Wishart, it is shown above that $p_{NiWU,i}(\Sigma|Y)$ is inverse Wishart if $p_{NiWU,i}(\boldsymbol{\pi})\propto 1$. $p_{NiWU,i}(\Sigma|Y)$ is a proper inverse Wishart distribution as long as

$$d^*\geq k,$$

or equivalently,

$$c\leq T-m-2k-1,$$
$$T\geq c+m+2k+1.$$

The mode of $p_{NiWU,i}(\Sigma|Y)$ can be computed as follows:

$$\log\big[p_{NiWU,i}(\Sigma|Y)\big]\propto\frac{T-m-c}{2}\cdot\log\big(\big|\det\big(\Sigma^{-1}\big)\big|\big)-\frac{1}{2}\operatorname{trace}\big[\Sigma^{-1}\big(S+\hat{\Sigma}_T(T-m)\big)\big],$$
$$\frac{d\log\big[p_{NiWU,i}(\Sigma|Y)\big]}{d\Sigma^{-1}}=\frac{T-m-c}{2}\Sigma-\frac{1}{2}\big(S+\hat{\Sigma}_T(T-m)\big)=0,$$
$$\Sigma_{\text{mode}}=\frac{1}{T-m-c}S+\frac{T-m}{T-m-c}\hat{\Sigma}_T.$$

In footnote 2 of the paper, we use this result for $c=-(d+k+1)$.

Different algorithms can be used to explore the joint posterior distribution, depending on whether the analysis is done only for the reduced-form parameters without sign restrictions on the corresponding structural parameters ($p_{NiWU,i}(\boldsymbol{\pi},\Sigma|Y)$), or on structural parameters under sign restrictions ($\tilde{p}_{NiWU,i}(\boldsymbol{\pi},\Sigma,Q|Y)$). In the former case, the algorithm used to explore $p_{NiWU,i}(\boldsymbol{\pi},\Sigma|Y)$ is

Algorithm A: $p_{NiWU,i}(\boldsymbol{\pi},\Sigma|Y)$

(i) draw $\Sigma^{(d)}$ from an iW(S^*,d^*), equation (S.5);

(ii) draw $\boldsymbol{\pi}^{(d)}$ from a N$(\boldsymbol{\mu}_\pi^*,V_\pi^*)$, equation (S.6), using V_π^* associated with $\Sigma^{(d)}$;

(iii) store $(\boldsymbol{\pi}^{(d)},\Sigma^{(d)})$;

(iv) repeat steps (i)–(iii) until a desired number of draws n_1 is stored.

In the latter case, the algorithm used to explore $\tilde{p}_{NiWU,i}(\boldsymbol{\pi},\Sigma,Q|Y)$ is

Algorithm B: $\tilde{p}_{NiWU,i}(\boldsymbol{\pi},\Sigma,Q|Y)$

(i) draw $\Sigma^{(d)}$ from an iW(S^*,d^*), equation (S.5);

(ii) draw $\boldsymbol{\pi}^{(d)}$ from a N$(\boldsymbol{\mu}_\pi^*,V_\pi^*)$, equation (S.6), using V_π^* associated with $\Sigma^{(d)}$;

(iii) draw $Q^{(d)}$ from the (Haar) uniform distribution using the algorithm by Rubio-Ramirez, Waggoner, and Zha (2010);

(iv) compute the structural statistics of interest associated with $(\boldsymbol{\pi}^{(d)},\Sigma^{(d)},Q^{(d)})$:

(iv,a) if the restrictions are satisfied, store $(\boldsymbol{\pi}^{(d)}, \Sigma^{(d)}, Q^{(d)})$;

(iv,b) if the restrictions are not satisfied, move back to step 1;

1. repeat steps (i)–(iv) until a desired number of draws n_1 is stored.

Both algorithms rely on direct sampling. By construction, *Algorithm A* delivers marginal draws from $p_{NiWU,i}(\Sigma|Y)$ in equation (S.5) and conditional draws for $\boldsymbol{\pi}$ from the Normal distribution (S.6). The same is not true for *Algorithm B*, because the accept/reject part of the algorithm potentially tilts the Normal inverse Wishart distribution for $(\boldsymbol{\pi}, \Sigma)$. In the paper, we use *Algorithm A* to generate proposal draws for our sampler because we must be in a position to evaluate the distribution of the proposal draws in the importance sampler. We use *Algorithm B* as a term of comparison to the NiWU approach in Section 4.3 of the paper.

S.2.2 *Independent normal and generalized inverse Wishart prior*

We now discuss the case of prior beliefs

$$\begin{aligned}
\tilde{p}_{NiWU,i}(\boldsymbol{\pi}, \Sigma, Q) &\propto \mathrm{I}\{\boldsymbol{\pi}, \Sigma, Q\} \cdot p_{NiWU,i}(\boldsymbol{\pi}, \Sigma, Q), && \text{(S.7)}\\
p_{NiWU,i}(\boldsymbol{\pi}, \Sigma) &= p_{NiWU,i}(\boldsymbol{\pi}) \cdot p_{NiWU,i}(\Sigma) \cdot p_{NiWU,i}(Q),\\
p_{NiWU,i}(\boldsymbol{\pi}) &= \phi(\boldsymbol{\mu}_\pi, V_\pi),\\
p_{NiWU,i}(\Sigma) &\propto \left|\det(\Sigma)\right|^{\frac{c}{2}} \cdot e^{-\frac{1}{2}\operatorname{trace}[\Sigma^{-1}S]},\\
p_{NiWU,i}(Q) &= U_{O(k)} \propto 1. && \text{(S.8)}
\end{aligned}$$

Compared to Section S.2.1, the density on $\boldsymbol{\pi}$ is proper. The above prior implies the joint posterior distribution

$$\begin{aligned}
\tilde{p}_{NiWU,i}(\boldsymbol{\pi}, \Sigma, Q|Y) &\propto \mathrm{I}\{\boldsymbol{\pi}, \Sigma, Q\} \cdot p_{NiWU,i}(\boldsymbol{\pi}, \Sigma, Q|Y),\\
p_{NiWU,i}(\boldsymbol{\pi}, \Sigma, Q|Y) &= p_{NiWU,i}(\boldsymbol{\pi}, \Sigma|Y) \cdot p_{NiWU,i}(Q|Y, \boldsymbol{\pi}, \Sigma),\\
p_{NiWU,i}(\boldsymbol{\pi}, \Sigma|Y) &\propto \left|\det(\Sigma)\right|^{\frac{c}{2}} \cdot e^{-\frac{1}{2}\operatorname{trace}[\Sigma^{-1}S]}\\
&\qquad \cdot \left|\det(V_\pi)\right|^{-\frac{1}{2}} \cdot e^{-\frac{1}{2}(\boldsymbol{\pi}-\boldsymbol{\mu}_\pi)'V_\pi^{-1}(\boldsymbol{\pi}-\boldsymbol{\mu}_\pi)}\left|\det(\Sigma)\right|^{-\frac{T}{2}}\\
&\qquad \cdot e^{-\frac{1}{2}\{\boldsymbol{\pi}'(WW'\otimes\Sigma^{-1})\boldsymbol{\pi}-2\boldsymbol{\pi}'(WW'\otimes\Sigma^{-1})\hat{\boldsymbol{\pi}}_T+\tilde{\mathbf{y}}'(I_T\otimes\Sigma^{-1})\tilde{\mathbf{y}}\}} && \text{(S.9)}\\
&\propto \left|\det(\Sigma)\right|^{\frac{c}{2}} \cdot e^{-\frac{1}{2}\operatorname{trace}[\Sigma^{-1}S]}\\
&\qquad \cdot e^{-\frac{1}{2}\{\boldsymbol{\pi}'V_\pi^{-1}\boldsymbol{\pi}-2\boldsymbol{\pi}'V_\pi^{-1}\boldsymbol{\mu}_\pi+\boldsymbol{\mu}_\pi'V_\pi^{-1}\boldsymbol{\mu}_\pi\}} \cdot \left|\det(\Sigma)\right|^{-\frac{T}{2}}\\
&\qquad \cdot e^{-\frac{1}{2}\{\boldsymbol{\pi}'(WW'\otimes\Sigma^{-1})\boldsymbol{\pi}-2\boldsymbol{\pi}'(WW'\otimes\Sigma^{-1})\hat{\boldsymbol{\pi}}_T+\tilde{\mathbf{y}}'(I_T\otimes\Sigma^{-1})\tilde{\mathbf{y}}\}},\\
p_{NiWU,i}(Q|Y, \boldsymbol{\pi}, \Sigma) &= p_{NiWU,i}(Q) = U_{O(k)} \propto 1.
\end{aligned}$$

Since

$$\begin{aligned}
&\boldsymbol{\pi}'\big(V_\pi^{-1} + WW' \otimes \Sigma^{-1}\big)\boldsymbol{\pi} - 2\boldsymbol{\pi}'\big(V_\pi^{-1}\boldsymbol{\mu}_\pi + \big(WW' \otimes \Sigma^{-1}\big)\hat{\boldsymbol{\pi}}_T\big) + \tilde{\mathbf{y}}'\big(I_T \otimes \Sigma^{-1}\big)\tilde{\mathbf{y}} + \boldsymbol{\mu}_\pi' V_\pi^{-1}\boldsymbol{\mu}_\pi \\
&\quad = \boldsymbol{\pi}' V_\pi^{*^{-1}} \boldsymbol{\pi} - 2\boldsymbol{\pi}' V_\pi^{*^{-1}} \boldsymbol{\mu}_\pi^* + \tilde{\mathbf{y}}'\big(I_T \otimes \Sigma^{-1}\big)\tilde{\mathbf{y}} + \boldsymbol{\mu}_\pi' V_\pi^{-1}\boldsymbol{\mu}_\pi \\
&\quad = \big(\boldsymbol{\pi} - \boldsymbol{\mu}_\pi^*\big)' V_\pi^{*^{-1}} \big(\boldsymbol{\pi} - \boldsymbol{\mu}_\pi^*\big) + \tilde{\mathbf{y}}'\big(I_T \otimes \Sigma^{-1}\big)\tilde{\mathbf{y}} + \boldsymbol{\mu}_\pi' V_\pi^{-1}\boldsymbol{\mu}_\pi - \boldsymbol{\mu}_\pi^{*'} V_\pi^{*^{-1}} \boldsymbol{\mu}_\pi^*,
\end{aligned}$$

with

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

equation (S.9) can be rewritten as

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

Lastly, we can derive

$$\begin{aligned}
p_{NiWU,i}(\Sigma|Y, \Pi) \propto{}& \big|\det(\Sigma)\big|^{\frac{c}{2}} \cdot e^{-\frac{1}{2}\,\mathrm{trace}[\Sigma^{-1}S]} \\
&\cdot \big|\det(\Sigma)\big|^{-\frac{T}{2}} \cdot e^{-\frac{1}{2}\{\boldsymbol{\pi}'(WW'\otimes\Sigma^{-1})\boldsymbol{\pi} - 2\boldsymbol{\pi}'(WW'\otimes\Sigma^{-1})\hat{\boldsymbol{\pi}}_T + \tilde{\mathbf{y}}'(I_T\otimes\Sigma^{-1})\tilde{\mathbf{y}}\}} \\
\propto{}& \big|\det(\Sigma)\big|^{\frac{c}{2}} \cdot e^{-\frac{1}{2}\,\mathrm{trace}[\Sigma^{-1}S]} \\
&\cdot \big|\det(\Sigma)\big|^{-\frac{T}{2}} \cdot e^{-\frac{1}{2}\{\boldsymbol{\pi}'(WW'\otimes\Sigma^{-1})\boldsymbol{\pi} - 2\boldsymbol{\pi}'(W\otimes\Sigma^{-1})\tilde{\mathbf{y}} + \tilde{\mathbf{y}}'(I_T\otimes\Sigma^{-1})\tilde{\mathbf{y}}\}} \\
\propto{}& \big|\det(\Sigma)\big|^{-\frac{T-c}{2}} \cdot e^{-\frac{1}{2}\,\mathrm{trace}[\Sigma^{-1}(S + YY' + \Pi WW'\Pi' - 2YW'\Pi')]} \\
\propto{}& \big|\det(\Sigma)\big|^{-\frac{(T-c-k-1)+k+1}{2}} \cdot e^{-\frac{1}{2}\,\mathrm{trace}[\Sigma^{-1}(S + (Y-\Pi W)(Y-\Pi W)')]}.
\end{aligned}$$

It then follows that the prior beliefs in equations (S.7)–(S.8) imply

$$\begin{aligned}
\tilde{p}_{NiWU,i}(\boldsymbol{\pi}, \Sigma, Q|Y) &\propto \mathrm{I}\{\boldsymbol{\pi}, \Sigma, Q\} \cdot p_{NiWU,i}(\boldsymbol{\pi}, \Sigma, Q|Y), \\
p_{NiWU,i}(\boldsymbol{\pi}, \Sigma, Q|Y) &= p_{NiWU,i}(\boldsymbol{\pi}|Y, \Sigma) \cdot p_{NiWU,i}(\Sigma|Y) \cdot p_{NiWU,i}(Q|Y, \boldsymbol{\pi}, \Sigma), \\
p_{NiWU,i}(\boldsymbol{\pi}|Y, \Sigma) &= \phi\big(\boldsymbol{\mu}_\pi^*, V_\pi^*\big), \qquad \text{(S.10)} \\
p_{NiWU,i}(\Sigma|Y, \Pi) &= iW\big(S^*, d^*\big), \qquad \text{(S.11)}
\end{aligned}$$

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

$$
\begin{aligned}
p_{NiWU,i}(Q|Y,\boldsymbol{\pi},\Sigma) &= p_{NiWU,i}(Q) = U_{O(k)} \propto 1, \\
V_\pi^* &= \left(V_\pi^{-1} + WW' \otimes \Sigma^{-1}\right)^{-1}, \\
\boldsymbol{\mu}_\pi^* &= V_\pi^*\left(V_\pi^{-1}\boldsymbol{\mu}_\pi + \left(WW' \otimes \Sigma^{-1}\right)\hat{\boldsymbol{\pi}}_T\right), \\
d^* &= T - c - k - 1, \\
S^* &= S + (Y - \Pi W)(Y - \Pi W)'.
\end{aligned}
$$

Contrary to Section S.2.1, S^* is now a function of $\boldsymbol{\pi}$, although this is not made explicit in the notation. Note that $p_{NiWU,i}(\Sigma|Y,\Pi)$ is a proper inverse Wishart distribution as long as

$$d^* \geq k,$$

or equivalently, as long as

$$
\begin{aligned}
c &\leq T - 2k - 1, \\
T &\geq c + 2k + 1.
\end{aligned}
$$

Contrary to Section S.2.1, drawing directly from $p_{NiWU,i}(\Sigma|Y)$ is now not possible and a Gibbs sampler is needed, but $p_{NiWU,i}(\Sigma|Y)$ can still be evaluated.

As in Section S.2.1, different algorithms can be used to explore the joint posterior distribution, depending on whether the analysis is done only on reduced form parameters without accounting for the indirect restrictions from sign restrictions on the corresponding structural parameters ($p_{NiWU,i}(\boldsymbol{\pi},\Sigma|Y)$) or by also considering sign restrictions ($\tilde{p}_{NiWU,i}(\boldsymbol{\pi},\Sigma,Q|Y)$). In the former case, the algorithm used to explore $p_{NiWU,i}(\boldsymbol{\pi},\Sigma|Y)$ is

Algorithm C: $p_{NiWU,i}(\boldsymbol{\pi},\Sigma|Y)$

(i) select a starting value for $\Sigma^{(0)}$;

(ii) draw $\boldsymbol{\pi}^{(d)}$ from a $\mathrm{N}(\boldsymbol{\mu}_\pi^*, V_\pi^*)$, equation (S.10), using $(\boldsymbol{\mu}_\pi^*, V_\pi^*)$ associated with $\Sigma^{(d-1)}$;

(iii) draw $\Sigma^{(d)}$ from an $\mathrm{iW}(S^*, d^*)$, equation (S.11), using S^* associated with $\boldsymbol{\pi}^{(d)}$;

(iv) if the n_2 burn-in draws have been generated, store $(\boldsymbol{\pi}^{(d)}, \Sigma^{(d)})$;

(v) repeat steps (ii)–(iv) until a desired number of draws n_3 is stored.

In the latter case, the algorithm used to explore $\tilde{p}_{NiWU,i}(\boldsymbol{\pi},\Sigma,Q|Y)$ is

Algorithm D: $\tilde{p}_{NiWU,i}(\boldsymbol{\pi},\Sigma,Q|Y)$

(i) select some starting candidate value for $\Sigma^{(0)}$;

(ii) draw $\boldsymbol{\pi}^{(c)}$ from a $\mathrm{N}(\boldsymbol{\mu}_\pi^*, V_\pi^*)$, equation (S.10), using $(\boldsymbol{\mu}_\pi^*, V_\pi^*)$ associated with $\Sigma^{(d-1)}$;

(iii) draw $\Sigma^{(c)}$ from an $\mathrm{iW}(S^*, d^*)$, equation (S.11), using S^* associated with $\boldsymbol{\pi}^{(c)}$;

(iv) draw $Q^{(c)}$ from the (Haar)uniform distribution using the algorithm by Rubio-Ramirez, Waggoner, and Zha (2010);

(v) compute the structural statistics of interest associated with $(\boldsymbol{\pi}^{(c)}, \Sigma^{(c)}, Q^{(c)})$:

(v a) if the restrictions are satisfied, set

$$\left(\boldsymbol{\pi}^{(d)}, \Sigma^{(d)}, Q^{(d)}\right) = \left(\boldsymbol{\pi}^{(c)}, \Sigma^{(c)}, Q^{(c)}\right);$$

(v b) if the restrictions are not satisfied, move back to step 2;

(vi) if the n_2 burn-in draws have been passed already, store $(\boldsymbol{\pi}^{(d)}, \Sigma^{(d)}, Q^{(d)})$;

(vii) repeat steps (ii)–(v) until a desired number of draws n_3 is stored.

Both algorithms rely on Gibbs sampling. As for *Algorithms A* and *B*, we use *Algorithm C* to generate proposal draws for our sampler, because we must be in a position to evaluate the marginal distribution associated with the proposal draws for Σ.

S.3. 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. We then discuss the diagnostics procedures we use to assess the performance of our importance sampler.

S.3.1 *Derivations*

We first provide the derivations for the more general case of the priors from equations (19a)–(19d) in the paper, and then discuss two common special cases.

S.3.1.1 *General case* Start from the prior distribution

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

with

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

where $\phi(\cdot,\cdot)$ is the probability density function of a Normal distribution and $\boldsymbol{\mu}_\pi$ and V_π can be a function of B. The joint posterior distribution then equals

$$\begin{aligned}\tilde{p}(\boldsymbol{\pi}, B|Y) &\propto \mathrm{I}\{\boldsymbol{\pi}, B\} \cdot p(\boldsymbol{\pi}, B|Y)\\ &\propto \mathrm{I}\{\boldsymbol{\pi}, B\} \cdot p(\boldsymbol{\pi}, B) \cdot p(Y|\boldsymbol{\pi}, B)\\ &\propto \mathrm{I}\{\boldsymbol{\pi}, B\} \cdot p(B)\\ &\quad \cdot (2\pi)^{-\frac{mk}{2}} \left|\det(V_\pi)\right|^{-\frac{1}{2}} e^{-\frac{1}{2}(\boldsymbol{\pi}-\boldsymbol{\mu}_\pi)' V_\pi^{-1}(\boldsymbol{\pi}-\boldsymbol{\mu}_\pi)}\\ &\quad \cdot (2\pi)^{-\frac{kT}{2}} \left|\det(B)\right|^{-T} e^{-\frac{1}{2}(\tilde{y}-Z\boldsymbol{\pi})'(I_T \otimes BB')^{-1}(\tilde{y}-Z\boldsymbol{\pi})}. \end{aligned} \tag{S.12}$$

As is standard in the literature, we rewrite equation (S.12) as

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

and exploit analytical results for $p(\boldsymbol{\pi}|Y, B)$. Define

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

and rewrite $p(\boldsymbol{\pi}, B|Y)$ as

$$p(\boldsymbol{\pi}, B|Y) = \kappa \cdot e^{-\frac{1}{2}[(\boldsymbol{\pi}-\boldsymbol{\mu}_\pi)'V_\pi^{-1}(\boldsymbol{\pi}-\boldsymbol{\mu}_\pi)+(\tilde{\mathbf{y}}-Z\boldsymbol{\pi})'\Psi(B)^{-1}(\tilde{\mathbf{y}}-Z\boldsymbol{\pi})]}, \tag{S.13}$$

with κ a term that includes elements which are not a function of $\boldsymbol{\pi}$. As done also in Section S.2.2, factorize the terms in the exponent of (S.13) as

$$\begin{aligned}&(\boldsymbol{\pi} - \boldsymbol{\mu}_\pi)'V_\pi^{-1}(\boldsymbol{\pi} - \boldsymbol{\mu}_\pi) + (\tilde{\mathbf{y}} - Z\boldsymbol{\pi})'\Psi(B)^{-1}(\tilde{\mathbf{y}} - Z\boldsymbol{\pi})\\ &\quad= \boldsymbol{\pi}'V_\pi^{-1}\boldsymbol{\pi} - 2\boldsymbol{\pi}'V_\pi^{-1}\boldsymbol{\mu}_\pi + \boldsymbol{\mu}_\pi'V_\pi^{-1}\boldsymbol{\mu}_\pi\\ &\qquad+ \tilde{\mathbf{y}}'\Psi(B)^{-1}\tilde{\mathbf{y}} - 2\boldsymbol{\pi}'Z'\Psi(B)^{-1}\tilde{\mathbf{y}} + \boldsymbol{\pi}'Z'\Psi(B)^{-1}Z\boldsymbol{\pi}\\ &\quad= \boldsymbol{\pi}'\left[V_\pi^{-1} + Z'\Psi(B)^{-1}Z\right]\boldsymbol{\pi} - 2\boldsymbol{\pi}'\left[V_\pi^{-1}\boldsymbol{\mu}_\pi + Z'\Psi(B)^{-1}\tilde{\mathbf{y}}\right]\\ &\qquad+ \tilde{\mathbf{y}}'\Psi(B)^{-1}\tilde{\mathbf{y}} + \boldsymbol{\mu}_\pi'V_\pi^{-1}\boldsymbol{\mu}_\pi\\ &\quad= \left(\boldsymbol{\pi} - \boldsymbol{\mu}_\pi^*\right)'V_\pi^{*^{-1}}\left(\boldsymbol{\pi} - \boldsymbol{\mu}_\pi^*\right) + \tilde{\mathbf{y}}'\Psi(B)^{-1}\tilde{\mathbf{y}} + \boldsymbol{\mu}_\pi'V_\pi^{-1}\boldsymbol{\mu}_\pi - \boldsymbol{\mu}_\pi^{*'}V_\pi^{*^{-1}}\boldsymbol{\mu}_\pi^*,\end{aligned}$$

with

$$\begin{aligned}V_\pi^* &= \left[V_\pi^{-1} + Z'\Psi(B)^{-1}Z\right]^{-1}\\ &= \left(V_\pi^{-1} + \left[WW' \otimes \left(BB'\right)^{-1}\right]\right)^{-1},\\ \boldsymbol{\mu}_\pi^* &= V_\pi^* \cdot \left[V_\pi^{-1}\boldsymbol{\mu}_\pi + Z'\Psi(B)^{-1}\tilde{\mathbf{y}}\right]\\ &= V_\pi^* \cdot \left(V_\pi^{-1}\boldsymbol{\mu}_\pi + \left[W \otimes \left(BB'\right)^{-1}\right]\tilde{\mathbf{y}}\right)\\ &= V_\pi^* \cdot \left(V_\pi^{-1}\boldsymbol{\mu}_\pi + \left[WW' \otimes \left(BB'\right)^{-1}\right]\hat{\boldsymbol{\pi}}_T\right).\end{aligned}$$

The joint posterior distribution can now be written as

$$\begin{aligned}\tilde{p}(\boldsymbol{\pi}, B|Y) \propto \mathrm{I}(\boldsymbol{\pi}, B) \cdot &\underbrace{\left|\det\left(V_\pi^*\right)\right|^{-\frac{1}{2}} \cdot e^{-\frac{1}{2}(\boldsymbol{\pi}-\boldsymbol{\mu}_\pi^*)'V_\pi^{*^{-1}}(\boldsymbol{\pi}-\boldsymbol{\mu}_\pi^*)}}_{p(\boldsymbol{\pi}|Y,B)=\phi(\boldsymbol{\mu}_\pi^*, V_\pi^*)}\\ &\cdot p(B) \cdot \left|\det(B)\right|^{-T} \cdot \left|\det\left(V_\pi^*\right)\right|^{\frac{1}{2}} \cdot \left|\det(V_\pi)\right|^{-\frac{1}{2}}\\ &\cdot e^{-\frac{1}{2}\{\tilde{\mathbf{y}}'(I_T\otimes(BB')^{-1})\tilde{\mathbf{y}}-\boldsymbol{\mu}_\pi^{*'}V_\pi^{*^{-1}}\boldsymbol{\mu}_\pi^*+\boldsymbol{\mu}_\pi'V_\pi^{-1}\boldsymbol{\mu}_\pi\}}.\end{aligned}$$

It follows that

$$
\begin{aligned}
&\tilde{p}(\boldsymbol{\pi}, B|Y) \propto \mathrm{I}\{\boldsymbol{\pi}, B\} p(\boldsymbol{\pi}|Y, B) \cdot p(B|Y), \\
&p(\boldsymbol{\pi}|B, Y) = \phi\big(\boldsymbol{\mu}_\pi^*, V_\pi^*\big), \\
&\quad p(B|Y) \propto p(B) \cdot \big|\det(B)\big|^{-T} \cdot \big|\det(V_\pi)\big|^{-\frac{1}{2}} \cdot \big|\det\big(V_\pi^*\big)\big|^{\frac{1}{2}} \\
&\qquad\qquad \cdot e^{-\frac{1}{2}\{\tilde{\mathbf{y}}'(I_T \otimes (BB')^{-1})\tilde{\mathbf{y}} - \boldsymbol{\mu}_\pi^{*\prime} V_\pi^{*-1} \boldsymbol{\mu}_\pi^* + \boldsymbol{\mu}_\pi' V_\pi^{-1} \boldsymbol{\mu}_\pi\}}.
\end{aligned}
\tag{S.14}
$$

While not explicit in the notation, $\boldsymbol{\mu}_\pi^*$ and V_π^* are a function of B.

Our algorithm requires being able to evaluate the probability distribution on (Σ, Q) implied by a distribution on B. To compute such transformations through the change of a variable theorem, define functions $f_1(\cdot)$, $f_2(\cdot)$, $f_3(\cdot)$ as

$$
\begin{aligned}
B &= f_1(\Sigma, Q) = h(\Sigma)Q, \\
\Sigma &= f_2(B) = BB', \\
Q &= f_3(B) = h\big(f_2(B)\big)^{-1} B,
\end{aligned}
$$

with $h(\Sigma)$ the Cholesky decomposition of Σ, or any other unique decomposition. Referring now more rigorously to probability distributions on B and (Σ, Q) as $p_B(B)$ and $p_{\Sigma,Q}(\Sigma, Q)$, we obtain

$$
\begin{aligned}
p_{\Sigma,Q}(\Sigma, Q) &= v_{\{B\to\Sigma,Q\}} \cdot p_B\big(f_1(\Sigma, Q)\big), \\
p_B(B) &= v_{\{\Sigma,Q\to B\}} \cdot p_{\Sigma,Q}\big(f_2(B), f_3(B)\big),
\end{aligned}
$$

with $v_{\{\cdot\to\cdot\}}$ the volume elements in the mapping from B to (Σ, Q) or vice versa (see Casella and Berger (2021) for the definition of the volume element). We compute $v_{\{B\to\Sigma,Q\}}$ and $v_{\{\Sigma,Q\to B\}}$ using the results in Bibby, Kent, and Mardia (1979) (Chapter 2) and Mathai and Haubold (2008) (Chapter 11):[1]

$$
\begin{aligned}
v_{\{B\to\Sigma,Q\}} &= \big|\det(\Sigma)\big|^{-\frac{1}{2}}, \\
v_{\{\Sigma,Q\to B\}} &= \big|\det(B)\big|.
\end{aligned}
$$

We can then compute the joint and marginal posterior distributions on (Σ, Q) implied by equation (S.14).

Consider first $p_{\Sigma,Q}(\Sigma, Q|Y)$. It holds that

$$
\begin{aligned}
p_{\Sigma,Q}(\Sigma, Q|Y) &\propto v_{\{B\to\Sigma,Q\}} \cdot p_B\big(f_1(\Sigma, Q)|Y\big) \\
&\propto \big|\det(\Sigma)\big|^{-\frac{1}{2}} \cdot \big|\det(\Sigma)\big|^{-\frac{T}{2}} \cdot \big|\det(V_\pi)\big|^{-\frac{1}{2}} \cdot \big|\det\big(V_\pi^*\big)\big|^{\frac{1}{2}} \\
&\quad \cdot e^{-\frac{1}{2}\{\tilde{\mathbf{y}}'(I_T \otimes \Sigma^{-1})\tilde{\mathbf{y}} - \boldsymbol{\mu}_\pi^{*\prime} V_\pi^{*-1} \boldsymbol{\mu}_\pi^* + \boldsymbol{\mu}_\pi' V_\pi^{-1} \boldsymbol{\mu}_\pi\}} \cdot p_B\big(f_1(\Sigma, Q)\big)
\end{aligned}
$$

[1]When zero restrictions are introduced, the volume elements must be evaluated numerically; see Arias, Rubio-Ramírez, and Waggoner (2018).

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

with $O(k)$ the space of orthogonal matrices of dimensions $k \times k$. The integral $\int_{O(k)} p_B(h(\Sigma)Q)\, dQ$ is a function of Σ and coincides with integrating the prior distribution $p(B)$ along the subspace $\mathcal{B}(\Sigma)$ defined as the parameter space of B such that BB' equals a constant value Σ. For a given value of Σ_d, we approximate the integral using the following simulation-based method:

Algorithm E:

(i) draw a matrix Q_c from the algorithm by Rubio-Ramirez, Waggoner, and Zha (2010);

(ii) compute $B_c = h(\Sigma_d)Q_c$;

(ii,a) if B_c satisfies the sign restrictions on B, proceed to step 3;

(ii,b) if B_c does not satisfy the sign restrictions on B, move back to step 1;

(iii) store $Q_i = Q_c$;

(iv) repeat steps (i)–(iv) until m_2 draws $\{Q_i\}_{i=1}^{m_2}$ are stored, and save the number of generated draws $m_3(\Sigma_d)$ required to obtain the necessary m_2 draws.

(v) compute

$$
\int_{O(k)} p_B\left(h(\Sigma_d)Q\right) dQ \approx \frac{\sum_{i=1}^{m_2} p_B\left(h(\Sigma_d)Q_i\right)}{m_3(\Sigma_d)}.
$$

Our algorithm also requires evaluating the conditional distribution $p(Q|\Sigma)$ implicit in $p(B)$. Note that

$$
\begin{aligned}
p_{\Sigma,Q}(\Sigma, Q) &= v_{\{B\to\Sigma,Q\}} \cdot p_B\left(h(\Sigma)Q\right), \\
p_\Sigma(\Sigma) &= \int_{O(k)} v_{\{B\to\Sigma,Q\}} \cdot p_B\left(h(\Sigma)Q\right) dQ, \\
p_{Q|\Sigma}(Q|\Sigma) &= \frac{p_{\Sigma,Q}(\Sigma, Q)}{\tilde{p}_\Sigma(\Sigma)}
\end{aligned}
\tag{S.15}
$$

$$= \frac{v_{\{B \to \Sigma, Q\}} \cdot p_B\big(h(\Sigma)Q\big)}{\int_{O(k)} v_{\{B \to \Sigma, Q\}} \cdot p_B\big(h(\Sigma)Q\big)\, dQ}$$
$$= \frac{p_B\big(h(\Sigma)Q\big)}{\int_{O(k)} p_B\big(h(\Sigma)Q\big)\, dQ},$$

and the same holds for the posterior,

$$\begin{aligned}
p_{Q|Y,\Sigma}(Q|Y,\Sigma) &= \frac{p_{\Sigma,Q|Y}(\Sigma, Q|Y)}{p_{\Sigma|Y}(\Sigma|Y)} \\
&= \frac{v_{\{B \to \Sigma, Q\}} \cdot p_B\big(h(\Sigma)Q\big)}{\int_{O(k)} v_{\{B \to \Sigma, Q\}} \cdot p_B\big(h(\Sigma)Q\big)\, dQ} \\
&= \frac{p_B\big(h(\Sigma)Q\big)}{\int_{O(k)} p_B\big(h(\Sigma)Q\big)\, dQ} \qquad \text{(S.16)} \\
&= p_{Q|\Sigma}(Q|\Sigma). \qquad \text{(S.17)}
\end{aligned}$$

To appreciate the usefulness of this result (S.17) for our sampler, consider two matrices (B_1, B_2) that imply the same matrix Σ,

$$B_1 = h(\Sigma)Q_1, \qquad \text{(S.18)}$$
$$B_2 = h(\Sigma)Q_2. \qquad \text{(S.19)}$$

Equations (S.15) and (S.16) imply that

$$\frac{p_{Q|\Sigma}(Q_1|\Sigma)}{p_{Q|\Sigma}(Q_2|\Sigma)} = \frac{p_{Q|Y,\Sigma}(Q_1|Y,\Sigma)}{p_{Q|Y,\Sigma}(Q_2|Y,\Sigma)} = \frac{p_B\big(h(\Sigma)Q_1\big)}{p_B\big(h(\Sigma)Q_2\big)} = \frac{p(B_1)}{p(B_2)}.$$

Hence, evaluating $p_{Q|\Sigma}(Q|\Sigma)$ and $p_{Q|Y,\Sigma}(Q|Y,\Sigma)$ along different values of Q and conditioning on the same Σ only requires evaluating the prior distribution $p_B(h(\Sigma)Q) = p(B)$. We use this result in Stage B of our algorithm to map draws from $p(\Sigma|Y)$ into draws from $p(B|Y)$. The above result is intuitive. Conditioning on Σ, if the posterior favors B_1 over B_2 defined in equations (S.18)–(S.19), it is because the prior does the same by the same proportion, that is, $\frac{p(B_2|Y)}{p(B_1|Y)} = \frac{p(B_2)}{p(B_1)}$. Hence, in our algorithm, mapping draws from $p(\Sigma|Y)$ into draws from $p(B|Y)$ only requires evaluating $p(B)$. For completeness, the paper uses notation $p(Q|Y,\Sigma)$ rather than the equivalent $p(Q|\Sigma)$.

The above intuition is used in the analysis of Figure 4 of the paper, which is generated as follows:

Algorithm F:

(i) select the conditioning value $\bar{\Sigma}$ and compute the Cholesky decomposition $B_c = h(\bar{\Sigma})$;

(ii) draw $Q^{(d)}$ from the (Haar) uniform distribution using the algorithm by Rubio-Ramirez, Waggoner, and Zha (2010);

(iii) compute $B^{(d)} = B_c \cdot Q^{(d)}$, verify if the sign restrictions are satisfied, checking up to sign and permutations of the columns of $B^{(d)}$ by appropriately adjusting the sign and ordering of the columns of $Q^{(d)}$, if necessary;

(iii,a) if the sign restrictions are satisfied, proceed to step 4;

(iii,b) if the sign restrictions are not satisfied, move back to step 2;

(iv) compute $\theta^{(d)} = \text{atan}(\frac{Q_{21}^{(d)}}{Q_{11}^{(d)}})$, with $Q_{ij}^{(d)}$ the (i,j) entry of $Q_{11}^{(d)}$ and with *atan* the inverse tangent function evaluated at $\frac{Q_{2,1}}{Q_{1,1}}$, so that $Q(\theta) = Q^{(d)}$, with the $Q(\theta)$ the Givens rotation matrix

$$Q(\theta) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix};$$

(v) compute $w^{(d)}$ as the prior $p(B)$ evaluated in $B^{(d)}$;

(vi) store $(\theta^{(d)}, w^{(d)})$;

(vii) repeat steps (ii)–(vi) until a desired number of draws n_4 is stored;

(viii) resample $\{\theta^{(d)}\}_{d=1}^{n_4}$ with replacement using weights $\{w^{(d)}\}_{d=1}^{n_4}$.

S.3.1.2 *Case 1: Flat prior on* $\boldsymbol{\pi}$ The first special case used in the paper (equation (27)) consists of the limit case in which the prior distribution $p(\boldsymbol{\pi}|B)$ becomes flat and improper ($V_\pi^{-1} = 0$). Such prior distribution is used in the literature either to reflect uninformative priors, or to introduce a Minnesota prior through dummy observations (see Doan, Litterman, and Sims (1984), Del Negro and Schorfheide (2011)). Under the prior,

$$\begin{aligned}
&\tilde{p}(\boldsymbol{\pi}, B) = \mathrm{I}\{\boldsymbol{\pi}, B\} \cdot p(\boldsymbol{\pi}, B), \\
&p(\boldsymbol{\pi}, B) = p(\boldsymbol{\pi}) \cdot p(B), \\
&p(\boldsymbol{\pi}) \propto 1, \\
&p(B),
\end{aligned}$$

the key results from Section S.3.1.1 simplify to

$$\begin{aligned}
&\tilde{p}(\boldsymbol{\pi}, B|Y) \propto \mathrm{I}\{\boldsymbol{\pi}, B\} \cdot p(\boldsymbol{\pi}|B, Y) \cdot p(B|Y), \\
&p(\boldsymbol{\pi}|B, Y) = \phi\big(\boldsymbol{\mu}_\pi^*, V_\pi^*\big),
\end{aligned}$$

$$p(B|Y) \propto p(B) \cdot \big|\det(B)\big|^{-(T-m)} \cdot e^{-\frac{1}{2}\,\text{trace}[(BB')^{-1}\hat{\Sigma}_T(T-m)]}, \tag{S.20}$$

$$\begin{aligned}
p(\Sigma|Y) &\propto v_{\{B\to\Sigma,Q\}} \cdot \big|\det(\Sigma)\big|^{-\frac{T-m}{2}} \cdot e^{-\frac{1}{2}\,\text{trace}[\Sigma^{-1}\hat{\Sigma}_T(T-m)]} \cdot \int_{O(k)} p_B\big(h(\Sigma)Q\big)\,dQ \\
&\propto \big|\det(\Sigma)\big|^{-\frac{T-m+1}{2}} \cdot e^{-\frac{1}{2}\,\text{trace}[\Sigma^{-1}\hat{\Sigma}_T(T-m)]} \cdot \int_{O(k)} p_B\big(h(\Sigma)Q\big)\,dQ,
\end{aligned} \tag{S.21}$$

$$V_\pi^* = \left(WW'\right)^{-1} \otimes BB',$$
$$\boldsymbol{\mu}_\pi^* = \hat{\boldsymbol{\pi}}_T.$$

We use equation (S.21) to compute weights $w(\Sigma^{(d)})^{\text{stage A}}$ in the Appendix of the paper.

The mode of $p(\Sigma|Y)$ is computed as

$$\begin{aligned}
\log\left[p(\Sigma|Y)\right] &\propto \frac{T-m+1}{2} \cdot \log\left(\left|\det\left(\Sigma^{-1}\right)\right|\right) - \frac{1}{2}\operatorname{trace}\left[\Sigma^{-1}\hat{\Sigma}_T(T-m)\right] \\
&\quad + \log\left(\int_{O(k)} p_B\left(h(\Sigma)Q\right) dQ\right), \\
\frac{d\log\left[p(\Sigma|Y)\right]}{d\Sigma^{-1}} &= \frac{T-m+1}{2}\Sigma - \frac{1}{2}\hat{\Sigma}_T(T-m) + \frac{d}{d\Sigma^{-1}}\log\left(\int_{O(k)} p_B\left(h(\Sigma)Q\right) dQ\right) = 0;
\end{aligned}$$

hence, the mode is implicitly defined by

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

a result used in footnote 2 of the paper.

S.3.1.3 *Case 2: Prior independence* The second special case used in the paper (equation (28)) consists of a Normal distribution for $\boldsymbol{\pi}$ that features independence from B. It is frequently used to introduce the flexible Minnesota prior, compared to the Minnesota prior modeled through dummy variables.

Under the prior,

$$\begin{aligned}
\tilde{p}(\boldsymbol{\pi}, B) &\propto \mathrm{I}\{\boldsymbol{\pi}, B\} \cdot p(\boldsymbol{\pi}, B), \\
p(\boldsymbol{\pi}, B) &= p(\boldsymbol{\pi}) \cdot p(B), \\
p(\boldsymbol{\pi}) &= \phi(\boldsymbol{\mu}_\pi, V_\pi), \\
p(B)&,
\end{aligned}$$

the key results from Section S.3.1.1 simplify to

$$\begin{aligned}
\tilde{p}(\boldsymbol{\pi}, B|Y) &\propto p(\boldsymbol{\pi}|B, Y) \cdot p(B|Y) \cdot \mathrm{I}(\boldsymbol{\pi}, B), \\
p(\boldsymbol{\pi}|B, Y) &= \phi\left(\boldsymbol{\mu}_\pi^*, V_\pi^*\right), \\
p(B|Y) &\propto p(B) \cdot \left|\det(B)\right|^{-T} \cdot \left|\det\left(V_\pi^*\right)\right|^{\frac{1}{2}} \cdot e^{-\frac{1}{2}\{\tilde{\mathbf{y}}'(I_T \otimes (BB')^{-1})\tilde{\mathbf{y}} - \boldsymbol{\mu}_\pi^{*\prime} V_\pi^{*-1} \boldsymbol{\mu}_\pi^*\}}, \\
p(\Sigma|Y) &\propto v_{\{B\to\Sigma,Q\}} \cdot \left|\det(\Sigma)\right|^{-\frac{T}{2}} \cdot \left|\det\left(V_\pi^*\right)\right|^{\frac{1}{2}} \cdot e^{-\frac{1}{2}\{\tilde{\mathbf{y}}'(I_T \otimes \Sigma^{-1})\tilde{\mathbf{y}} - \boldsymbol{\mu}_\pi^{*\prime} V_\pi^{*-1} \boldsymbol{\mu}_\pi^*\}} \\
&\quad \cdot \int_{O(k)} p_B\left(h(\Sigma)Q\right) dQ \qquad \text{(S.22)} \\
&\propto \left|\det(\Sigma)\right|^{-\frac{T+1}{2}} \cdot \left|\det\left(V_\pi^*\right)\right|^{\frac{1}{2}} \cdot e^{-\frac{1}{2}\{\tilde{\mathbf{y}}'(I_T \otimes \Sigma^{-1})\tilde{\mathbf{y}} - \boldsymbol{\mu}_\pi^{*\prime} V_\pi^{*-1} \boldsymbol{\mu}_\pi^*\}} \\
&\quad \cdot \int_{O(k)} p_B\left(h(\Sigma)Q\right) dQ,
\end{aligned}$$

$$V_\pi^* = \left(V_\pi^{-1} + \left[WW' \otimes (BB')^{-1}\right]\right)^{-1},$$
$$\boldsymbol{\mu}_\pi^* = V_\pi^* \cdot \left(V_\pi^{-1}\boldsymbol{\mu}_\pi + \left[WW' \otimes (BB')^{-1}\right]\hat{\boldsymbol{\pi}}_T\right).$$

We use equation (S.22) to compute weights $w(\Sigma^{(d)})^{\text{stage A}}$ in the Appendix of the paper.

S.3.2 *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 $p(Q|\Sigma)$. Accordingly, the favorable 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 nonnormalized importance weights $\{w_i\}_{i=1}^N$ of size N have a finite variance. In order to assess whether the variance of the weights in Stage A is bounded, we employ a graphical procedure and two diagnostic tests proposed by Koopman, Shephard, and Creal (2009). We discuss these approaches in the rest of this section, and refer to Figure S.3 and Table S.6 for the results of the diagnostics associated with the application in the paper.

S.3.2.1 *Graphical assessment* A measure frequently used to assess the quantitative importance of outliers in importance weights (and hence possible 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 $v_i = \text{Var}(w_{1:i})$ is the variance of the first i weights. If outliers do not raise quantitatively relevant concerns, $\{v_i\}_{i=1}^N$ should converge smoothly toward 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.

S.3.2.2 *Diagnostic tests* In addition to using graphical illustrations, we employ the more formal testing procedures by Koopman, Shephard, and Creal (2009) to assess the quantitative relevance of outliers. Koopman, Shephard, and Creal (2009) propose a Wald test and a score test to test

$$H_0 : \text{the variance is finite}$$

against

$$H_1 : \text{the variance is unbounded.}$$

To set the stage, note that the problematic part of $\{w_i\}_{i=1}^N$ is 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

$$Z_i = w_i - u \quad \text{if } w_i > u.$$

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

approximates a generalized Pareto distribution with density function

$$f(z|\xi,\beta)=\frac{1}{\beta}\left(1+\xi\frac{z}{\beta}\right)^{-\frac{1}{\xi}-1}. \tag{S.23}$$

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, Rubio-Ramírez, and Waggoner (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, that is, $w_1 \leq w_{(2)} \leq \cdots \leq w_{(N)}$. To make the tests operational, we use the null and alternative hypotheses

$$H_0: \xi = 0.5,$$
$$H_1: \xi > 0.5.$$

Wald test The log likelihood function based on the generalized Pareto distribution in (S.23) equals

$$\log f(z|\beta,\xi) = -n\cdot\log(\beta) - \left(1+\frac{1}{\xi}\right)\sum_{i=1}^{n}\log\bigl(1+\xi\beta^{-1}z_i\bigr). \tag{S.24}$$

To construct the test statistic, we need to numerically maximize (S.24) with respect to β and ξ to obtain β^{MLE} and ξ^{MLE}, and then construct a test statistic to test for $\xi = 0.5$. In practice, we follow these steps:

1. numerically maximize the unrestricted log likelihood function given in (S.24) to obtain β^{MLE} and ξ^{MLE};

2. construct the test statistic as

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

The statistic t has an approximate standard normal distribution for $n \to \infty$. Large values indicate that ξ 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 maximize (S.24) under the null hypothesis of $\xi = 0.5$. We follow these steps:

1. Numerically maximize (S.24) with respect to β under the restriction $\xi = 0.5$ to obtain $\beta^{\mathrm{MLE},r}$;

2. since the test statistic is based on the score function of ξ, differentiate (S.24) with respect to ξ and set $\xi = 0.5$ and $\beta = \beta^{\text{MLE},r}$. This gives [2]

$$\hat{s}_r^{\xi} = 4\sum_{i=1}^{n} \log\left(1 + \frac{z_i}{2\beta^{\text{MLE},r}}\right) - 6\sum_{i=1}^{n} \frac{z_i}{2\beta^{\text{MLE},r} + z_i};$$

3. compute

$$s_*^{\xi} = \frac{1}{\sqrt{2n}} \hat{s}_r^{\xi}.$$

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

S.4. The dynamic striated Metropolis–Hastings algorithm by Waggoner, Wu, and Zha (2016)

We assess the performance of our sampler by exploring the posterior distribution not only through our sampler, but also using the Dynamic Striated Metropolis–Hastings (DSMH) algorithm developed by Waggoner, Wu, and Zha (2016). While being more demanding, the DSMH algorithm is designed to explore distributions that are potentially irregularly shaped, and hence, offers a valid benchmark against which to compare the draws from our sampler. We use the DSMH algorithm to explore $p(B|Y)$ from equation (22) in the paper, and then use the accept/reject steps 7–8 from our algorithm to convert draws from $p(B|Y)$ into draws from the ultimate distribution of interest $\tilde{p}(B|Y) = \int \tilde{p}(\boldsymbol{\pi}, B|Y)\, d\boldsymbol{\pi}$. Section S.4.1 briefly discusses the DSMH sampler, while Section S.4.2 discusses the convergence criteria that we use to assess the performance of the posterior chains.

S.4.1 *The DSMH sampler*

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

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

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

Waggoner, Wu, and Zha (2016) specify $f^s(\boldsymbol{\theta})$, $f^i(\boldsymbol{\theta})$, and $f^d(\boldsymbol{\theta})$ in order to initialize the algorithm from the prior distribution behind their posterior of interest. Alternatively,

[2] There is a typing error in Koopman (p. 4) in the corresponding equation.

one could depart from Waggoner, Wu, and Zha (2016) and specify $f^s(\boldsymbol{\theta})$, $f^i(\boldsymbol{\theta})$, and $f^d(\boldsymbol{\theta})$ to initialize the algorithm from $p_{NiWU,i}(B|Y)$, which we defined in the paper as the posterior distribution on B associated with the NiWU approach.[3] To illustrate that both approaches are feasible, we initialize the bivariate simulation exercise in Section 3 at the prior while using the distribution associated with the NiWU approach for the monetary policy application in Section 4. In our applications, the sampler is required for either four or twenty-five parameters, which is well within the range of parameters in which the Dynamic Striated Metropolis–Hastings sampler performs efficiently.

More precisely, in our applications $\boldsymbol{\theta}$ contains the entries of B. Consider for simplicity *Case 1*, which leads to the posterior of interest $p(B|Y)$ from equation (S.20). One could sample $p(B|Y)$ by setting

$$\begin{aligned} f^s(\boldsymbol{\theta}) &= p(B), \\ f^i(\boldsymbol{\theta}) &= \left|\det(B)\right|^{-(T-m)} \cdot e^{-\frac{1}{2}[(BB')^{-1}\hat{\Sigma}_T(T-m)]}, \\ f^d(\boldsymbol{\theta}) &= 1. \end{aligned} \tag{S.25}$$

This specification follows Waggoner, Wu, and Zha (2016) in initializing the algorithm at the prior distribution behind the posterior distribution of interest. Alternatively, one can start from a function that allows exploiting the computational advantage of the NiWU approach. While it is hard to analytically derive the posterior distribution of the draws delivered by *Algorithm B* in Section S.2.1 due to the accept/reject part of the algorithm, the following modification of *Algorithm A*:

Algorithm A-bis:

(i) draw $\Sigma^{(d)}$ from iW(S^*, d^*), equation (S.5);

(ii) draw $\boldsymbol{\pi}^{(d)}$ from N$(\boldsymbol{\mu}_\pi^*, V_\pi^*)$, equation (S.6), using V_π^* associated with $\Sigma^{(d)}$;

(iii) draw $Q^{(d)}$ using the algorithm by Rubio-Ramirez, Waggoner, and Zha (2010);

(iv) store $(\boldsymbol{\pi}^{(d)}, \Sigma^{(d)}, Q^{(d)})$;

(v) repeat steps 1–4 until a desired number of draws n_1 is stored,

delivers draws from the joint distribution

$$\begin{aligned} p(B|Y) &\propto v_{\Sigma,Q\to B} \cdot p(\Sigma|Y) \cdot 1 \\ &\propto \left|\det(B)\right| \cdot \left|\det\bigl(BB'\bigr)\right|^{-\frac{T-m-c}{2}} \cdot e^{-\frac{1}{2}[(BB')^{-1}(S+\hat{\Sigma}_T(T-m))]} \\ &\propto \left|\det(B)\right|^{-(T-m-c-1)} \cdot e^{-\frac{1}{2}[(BB')^{-1}(S+\hat{\Sigma}_T(T-m))]}. \end{aligned} \tag{S.26}$$

For $c = -1$ and $S = 0$ (the same values used for the proposal draws for *Case 1* in our algorithm, see the Appendix of the paper), equation (S.26) is very similar to equation

[3]A similar idea is explored by Mlikota and Schorfheide (forthcoming), who, like us, implement sequential sampling starting from a posterior distribution of a convenient form rather than from the prior.

(S.25), suggesting to start the DSMH from draws from *Algorithm A-bis* and then setting

$$f^s(\boldsymbol{\theta}) = \left|\det(B)\right|^{-(T-m)} \cdot e^{-\frac{1}{2}[(BB')^{-1}\hat{\Sigma}_T(T-m)]},$$
$$f^i(\boldsymbol{\theta}) = p(B),$$
$$f^d(\boldsymbol{\theta}) = 1.$$

Once the functions $f^s(\boldsymbol{\theta})$, $f^i(\boldsymbol{\theta})$, and $f^d(\boldsymbol{\theta})$ are specified, define the target function

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

as the logarithm of the tempered function $f_{\lambda_h}(\boldsymbol{\theta})$ 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 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, Wu, and Zha (2016) propose sampling $\log[f_{\lambda_h}(\boldsymbol{\theta})]$ using a modified Metropolis–Hastings algorithm.[4] In the general stage h, the algorithm can be summarized in the following steps:

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

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

(iii) for each $\{\boldsymbol{\theta}_d^{(h-1)}\}_{d=1}^{N\cdot G}$, compute weights $\tilde{\omega}_d = \frac{f_{\lambda_h}(\boldsymbol{\theta}_d^{(h-1)})}{f_{\lambda_{h-1}}(\boldsymbol{\theta}_d^{(h-1)})}$. 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 reweighting 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;

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

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

[4]See, for example, Herbst and Schorfheide (2014) for an alternative approach to sequential Monte Carlo samplers.

with probability $\min\{1, \frac{f_{\lambda_h}(\boldsymbol{\theta}_{\text{new}})}{f_{\lambda_h}(\boldsymbol{\theta}_{\text{old}})}\}$ if $\boldsymbol{\theta}_{\text{new}}$ was generated from the random walk extraction, and with probability $\min\{1, \frac{f_{\lambda_h}(\boldsymbol{\theta}_{\text{new}})}{f_{\lambda_{h-1}}(\boldsymbol{\theta}_{\text{old}})} \frac{f_{\lambda_h}(\boldsymbol{\theta}_{\text{old}})}{f_{\lambda_{h-1}}(\boldsymbol{\theta}_{\text{new}})}\}$ if $\boldsymbol{\theta}_{\text{new}}$ was randomly selected from the striations at the previous stage. Continue for $N \cdot \tau$ iterations;

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

Following Waggoner, Wu, and Zha (2016), we set $p = 1/(10\tau)$. We then set the parameter c_h at each stage following the guidance discussed by Waggoner, Wu, and Zha (2016) in Appendix A, ensuring, within each stage, an acceptance ratio between 0.20 and 0.30 from a preliminary Metropolis–Hastings algorithm with K iterations. We follow Waggoner, Wu, and Zha (2016) and set the number of stages H at 50, using their suggested progression for the tempering parameter λ_h. It remains to calibrate the number of groups G, the effective number of iterations N within each group, the number K affecting the number of iterations KG for the calibration of the parameter c_h, the frequency τ at which draws stored, and the number of striations M. We set (G, K, M, τ) as indicated in Table S.1. We then set an initial value for N and progressively increase it within each stage if this is needed to ensure that the convergence criteria discussed in the next section confirm that the draws at stage h have converged to the distribution of the target function at stage h for at least 90% of the parameters in $\boldsymbol{\theta}$. For each application, multiple convergence diagnostics are used in each of the 50 stages of the sampler. The diagnostics results are not reported.

S.4.2 *Convergence criteria used for the DSMH algorithm*

Consider the chain $\{\boldsymbol{\theta}_d\}_{d=1}^{N\cdot G}$, with $\boldsymbol{\theta}_d = (\theta_{1,d}, \ldots, \theta_{j,d}, \ldots, \theta_{\kappa,d})'$ of dimension $\kappa \times 1$. We employ four convergence criteria in order to assess if the chain has converged in distribution, namely the converge 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 autodependence (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 $\boldsymbol{\theta}_d$). 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 $\{\theta_{j,d}\}_{d=1}^{N}, \{\theta_{j,d}\}_{d=N+1}^{2N}, \ldots, \{\theta_{j,d}\}_{d=(G-1)N+1}^{N\cdot G}$ for univariate chains, and $\{\boldsymbol{\theta}_d\}_{d=1}^{N\cdot G}$ or $\{\boldsymbol{\theta}_d\}_{d=1}^{N}, \{\boldsymbol{\theta}_d\}_{d=N+1}^{2N}, \ldots, \{\boldsymbol{\theta}_d\}_{d=(G-1)N+1}^{N\cdot G}$ 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).

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:

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

(ii) 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 $\hat{\sigma}_{\text{last}}$;

(iii) compute the test statistic

$$CD = \frac{\hat{\mu}_{\text{first}} - \hat{\mu}_{\text{last}}}{\frac{\hat{\sigma}_{\text{first}}}{\sqrt{0.1NG}} + \frac{\hat{\sigma}_{\text{last}}}{\sqrt{0.4NG}}}.$$

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

(iv) 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.

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 θ_j, then $q = 0.025$. 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 a MCMC chain. The implementation of the algorithm for each parameter j proceeds in 4 steps:

(i) transform $\{\theta_{j,d}\}_{d=1}^{N\cdot G}$ into a dichotomous random variable Z_d:

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

(ii) write the matrix of transition probabilities for Z_d conditioning on the previous state,

$$\mathcal{P} = \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix},$$

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

$$\pi_0 = P(\theta_{j,d} < q_{0.025}) = P(Z_d = 0) = \frac{\beta}{\alpha+\beta},$$
$$\pi_1 = P(\theta_{j,d} \geq q_{0.025}) = P(Z_d = 1) = 1 - \pi_0 = \frac{\alpha}{\alpha+\beta};$$

(iii) approximate the probability that a draw of the parameter is smaller than the quantile of interest as

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

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

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

$$n^* = \frac{(2-\alpha-\beta)\alpha\beta}{(\alpha+\beta)^3}\left\{\frac{\Phi^{-1}\left(\frac{1}{2}(s+1)\right)}{r}\right\}^2.$$

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.

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:

(i) compute the variance of the mean of each sequence (“between-sequence variance”) as

$$B = \frac{1}{G-1}\sum_{g=1}^{G}(\bar{\theta}_{j,\cdot,g} - \bar{\theta}_j)^2,$$

where $\bar{\theta}_{j,\cdot,g}$ is the mean of $\theta_{j,d}$ within the gth chain and $\bar{\theta}_j$ is the mean across all chains for parameter j.

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

$$W = \frac{1}{G(N-1)} \sum_{g=1}^{G} \sum_{n=1}^{N} (\theta_{j,n,g} - \bar{\theta}_{j,\cdot,g})^2;$$

(iii) estimate the overall variance as

$$\hat{V} = \hat{\sigma}_+^2 + \frac{B}{G},$$

with

$$\hat{\sigma}_+^2 = \frac{N-1}{N} W + B;$$

(iv) compute the statistic

$$\hat{R} = \frac{\hat{V}}{W}.$$

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.

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, $\{\boldsymbol{\theta}_d\}_{d=1}^{N}, \{\boldsymbol{\theta}_d\}_{d=N+1}^{2N}, \ldots, \{\boldsymbol{\theta}_d\}_{d=(G-1)N+1}^{N\cdot G}$. 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):

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

$$D = \frac{1}{G-1} \sum_{g=1}^{G} (\bar{\boldsymbol{\theta}}_g - \bar{\boldsymbol{\theta}})(\bar{\boldsymbol{\theta}}_g - \bar{\boldsymbol{\theta}})';$$

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

$$W = \frac{1}{G(N-1)} \sum_{g=1}^{G} \sum_{n=1}^{N} (\boldsymbol{\theta}_{n,g} - \bar{\boldsymbol{\theta}}_g)(\boldsymbol{\theta}_{n,g} - \bar{\boldsymbol{\theta}}_g)';$$

(iii) estimate the overall variance as

$$\hat{V} = \frac{N-1}{N} W + \left(1 + \frac{1}{G}\right) D;$$

(iv) compute the (scalar) distance measure between these two matrices as

$$\hat{R}^{\text{mult}} = \frac{N}{N-1} + \frac{G+1}{G}\lambda_1,$$

where λ_1 is the largest eigenvalue of the matrix $W^{-1}D$.

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.

S.5. Our prior distribution

In this section, we show that our prior can be viewed as a generalization of the generalized Normal prior by Arias, Rubio-Ramírez, and Waggoner (2018). Consider first the case of no sign restrictions as in Section (2.5) of Arias, Rubio-Ramírez, and Waggoner (2018). We use notation $p_{NiWU}(\cdot)$ to indicate that the derivations are the same irrespectively of whether the NiWU prior is specified in its conjugate form (as in Arias, Rubio-Ramírez, and Waggoner (2018)) or in its independent form. Define $A = B^{-1}$. The joint inverse Wishart Uniform distribution,

$$\begin{aligned} p_{NiWU}(\Sigma, Q) &= p_{NiWU}(\Sigma) \cdot p_{NiWU}(Q) \\ &\propto \left|\det(\Sigma)\right|^{-\frac{d+k+1}{2}} \cdot e^{-\frac{1}{2}\operatorname{trace}[\Sigma^{-1}S]} \cdot 1, \end{aligned} \tag{S.27}$$

implies the following distributions for B and for A:

$$\begin{aligned} p_{NiWU}(B) &\propto v_{\Sigma,Q\to B} \cdot \left|\det\big(BB'\big)\right|^{-\frac{d+k+1}{2}} \cdot e^{-\frac{1}{2}\operatorname{trace}[(BB')^{-1}S]} \\ &\propto \left|\det(B)\right| \cdot \left|\det(B)\right|^{-(d+k+1)} \cdot e^{-\frac{1}{2}\operatorname{trace}[(BB')^{-1}S]} \\ &\propto \left|\det(B)\right|^{-(d+k)} \cdot e^{-\frac{1}{2}\operatorname{trace}[B'^{-1}B^{-1}S]} \\ &\propto \left|\det(B)\right|^{-(d+k)} \cdot e^{-\frac{1}{2}\operatorname{trace}[B^{-1}SB'^{-1}]} \\ &\propto \left|\det(B)\right|^{-(d+k)} \cdot e^{-\frac{1}{2}\operatorname{vec}(B^{-1'})'(I_k\otimes S)\operatorname{vec}(B^{-1'})}, \\ p_{NiWU}(A) &\propto v_{B\to A} \cdot \left|\det\big(A^{-1}\big)\right|^{-(d+k)} \cdot e^{-\frac{1}{2}\operatorname{vec}(A')'(I_k\otimes S)\operatorname{vec}(A')} \\ &\propto \left|\det(A)\right|^{-2k} \cdot \left|\det(A)\right|^{d+k} \cdot e^{-\frac{1}{2}\operatorname{vec}(A')'(I_k\otimes S)\operatorname{vec}(A')} \\ &\propto \left|\det(A)\right|^{d-k} \cdot e^{-\frac{1}{2}\operatorname{vec}(A')'(I_k\otimes S)\operatorname{vec}(A')}. \end{aligned} \tag{S.28}$$

Equation (S.28) is the generalized Normal distribution from Arias, Rubio-Ramírez, and Waggoner (2018) and coincides with their equation (2.8), adjusted for the different notation (i.e., $A = A'$). It is constant in Q, that is, $p_{NiWU}(A_1) = p_{NiWU}(A_2)$ for $A_2 = QA_1$. This holds immediately when inspecting equation (S.27) jointly with the fact that $v_{\Sigma,Q\to B}$ does not depend on Q, or noticing that

$$\begin{aligned} p_{NiWU}(QA) &\propto \left|\det(QA)\right|^{d-k} \cdot e^{-\frac{1}{2}\operatorname{vec}((QA)')'(I_k\otimes S)\operatorname{vec}((QA)')} \\ &\propto \left|\det(A)\right|^{d-k} \cdot e^{-\frac{1}{2}\operatorname{trace}[(QA)S(QA)']} \end{aligned}$$

$$
\begin{aligned}
&\propto \left|\det(A)\right|^{d-k} \cdot e^{-\frac{1}{2}\operatorname{trace}[QASA'Q']} \\
&\propto \left|\det(A)\right|^{d-k} \cdot e^{-\frac{1}{2}\operatorname{trace}[A'Q'QAS]} \\
&\propto \left|\det(A)\right|^{d-k} \cdot e^{-\frac{1}{2}\operatorname{trace}[A'AS]} = p_{NiWU}(A).
\end{aligned}
$$

Hence, $p_{NiWU}(Q|\Sigma)$ is constant in the entire space $O(k)$. When instead sign restrictions are introduced, it is constant in a subspace of $O(k)$ that depends on Σ.

Consider now our prior distribution. Consider first the case in which no sign restrictions are introduced. Define

$$
\begin{aligned}
V_b &= \left(\frac{\psi_2^2}{1.96^2} \cdot I_k \otimes \operatorname{diag}\left([\gamma_1^2, \ldots, \gamma_i^2, \ldots, \gamma_k^2]'\right)\right) \\
&= C^{-1} \otimes D^{-1}, \\
C &= \frac{1.96^2}{\psi_2^2} I_k, \\
D &= \operatorname{diag}\left(\left[\frac{1}{\gamma_1^2}, \ldots, \frac{1}{\gamma_i^2}, \ldots, \frac{1}{\gamma_k^2}\right]'\right).
\end{aligned}
$$

Under no sign restrictions, the prior density $\tilde{p}(B)$ discussed in Section 2.4 of the paper equals

$$
\begin{aligned}
\tilde{p}(B) &= p(B) \\
&= \prod_i \prod_j p(b_{ij}|\mu_{ij}, \sigma_{ij}) \\
&= \prod_i \prod_j \phi\left(0, \frac{\psi_2^2}{1.96^2} \cdot \gamma_i^2\right) \\
&\propto \left|\det(V_b)\right|^{-\frac{1}{2}} \cdot e^{-\frac{1}{2}\operatorname{vec}(B)' V_b^{-1} \operatorname{vec}(B)} \\
&\propto \left|\det(C \otimes D)\right|^{\frac{1}{2}} \cdot e^{-\frac{1}{2}\operatorname{vec}(B)'(C\otimes D)\operatorname{vec}(B)} \\
&\propto \left|\det(C)\right|^{\frac{k}{2}} \cdot \left|\det(D)\right|^{\frac{k}{2}} \cdot e^{-\frac{1}{2}\operatorname{vec}(B)'(C\otimes D)\operatorname{vec}(B)} \\
&\propto \left(\frac{1.96}{\psi_2}\right)^{\frac{k}{2}} \cdot \prod_{i=1}^{k} \gamma_i^{-\frac{k}{2}} \cdot e^{-\frac{1}{2}\operatorname{trace}[B'DBC']} \\
&\propto \left(\frac{1.96}{\psi_2}\right)^{\frac{k}{2}} \cdot \prod_{i=1}^{k} \gamma_i^{-\frac{k}{2}} \cdot e^{-\frac{1}{2}\operatorname{trace}[B'DB\frac{1.96^2}{\psi_2^2}]} \\
&\propto \left(\frac{1.96}{\psi_2}\right)^{\frac{k}{2}} \cdot \prod_{i=1}^{k} \gamma_i^{-\frac{k}{2}} \cdot e^{-\frac{1}{2}\frac{1.96^2}{\psi_2^2}\operatorname{trace}[DBB']},
\end{aligned}
$$

which implies $p(B) = p(BQ)$, that is, $p(Q|\Sigma) = \tilde{p}(Q|\Sigma)$ are flat in Q, as in the NiWU case.

Consider now the case with sign restrictions. Define

$$\begin{aligned}
V_b &= \left(I_k \otimes \operatorname{diag}\left([s_1^2, \ldots, s_i^2, \ldots, s_k^2]'\right)\right) \\
&= I_k \otimes D^{-1}, \\
D &= \operatorname{diag}\left(\left[\frac{1}{s_1^2}, \ldots, \frac{1}{s_i^2}, \ldots, \frac{1}{s_k^2}\right]'\right), \\
\boldsymbol{e} &= \psi_1 \cdot \left(\boldsymbol{\iota}_k \otimes [\gamma_1, \ldots, \gamma_i, \ldots, \gamma_k]'\right) \cdot \operatorname{vec}\big(\operatorname{sign}(B)\big),
\end{aligned}$$

with s_i calibrated as explained in Section 2.4 of the paper and $\operatorname{sign}(B)$ a matrix whose ij entry equals $\operatorname{sign}(b_{ij})$, the sign restriction on b_{ij}. The prior density $\tilde{p}(B)$ equals

$$\begin{aligned}
\tilde{p}(B) &\propto \mathrm{I}\{B\} \cdot p(B), \\
\tilde{p}(B) &= \prod_i \prod_j \tilde{p}(b_{ij}|\mu_{ij}, \sigma_{ij}) \\
&= \prod_i \prod_j \mathrm{I}\{b_{ij}\} \phi\left(\psi_1 \gamma_i \cdot \operatorname{sign}(b_{ij}), s_i^2\right) \\
&\propto \mathrm{I}\{B\} \cdot \left|\det(V_b)\right|^{-\frac{1}{2}} \cdot e^{-\frac{1}{2}(\operatorname{vec}(B)-\boldsymbol{e})' V_b^{-1} (\operatorname{vec}(B)-\boldsymbol{e})} \\
&\propto \mathrm{I}\{B\} \cdot \left|\det(I_k \otimes D)\right|^{\frac{1}{2}} \cdot e^{-\frac{1}{2}(\operatorname{vec}(B)-\boldsymbol{e})'(I_k \otimes D)(\operatorname{vec}(B)-\boldsymbol{e})} \\
&\propto \mathrm{I}\{B\} \cdot \left|\det(D)\right|^{\frac{k}{2}} \cdot e^{-\frac{1}{2}(\operatorname{vec}(B)-\boldsymbol{e})'(I_k \otimes D)(\operatorname{vec}(B)-\boldsymbol{e})} \\
&\propto \mathrm{I}\{B\} \cdot \left|\det(D)\right|^{\frac{k}{2}} \cdot e^{-\frac{1}{2}\{\operatorname{vec}(B)'(I_k \otimes D)\operatorname{vec}(B) - 2\operatorname{vec}(B)'(I_k \otimes D)\boldsymbol{e} + \boldsymbol{e}'(I_k \otimes D)\boldsymbol{e}\}} \\
&\propto \mathrm{I}\{B\} \cdot \left|\det(D)\right|^{\frac{k}{2}} \cdot e^{-\frac{1}{2}\operatorname{vec}(B)'(I_k \otimes D)\operatorname{vec}(B)} \cdot e^{\operatorname{vec}(B)'(I_k \otimes D)\boldsymbol{e}} \\
&\propto \mathrm{I}\{B\} \cdot \left|\det(D)\right|^{\frac{k}{2}} \cdot e^{-\frac{1}{2}\operatorname{trace}[DBB']} \cdot e^{\operatorname{vec}(B)'(I_k \otimes D)\boldsymbol{e}},
\end{aligned}$$

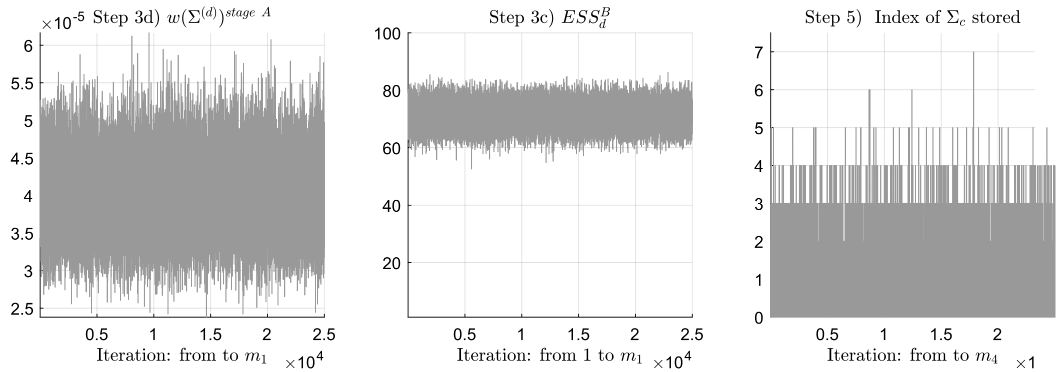
which implies $p(B) \neq p(BQ)$, unless $\psi_1 = 0$, under which $\operatorname{vec}(B)'(I_k \otimes D)\boldsymbol{e} = 0$ and $p(BQ) = p(B)$, $\forall Q$. Hence, in general, our prior implies distributions $p(Q|\Sigma)$, $\tilde{p}(Q|\Sigma)$ that are not flat in the parameter space of orthogonal matrices.

S.6. Additional tables and figures

Table S.1. Tuning parameters used in the algorithms.

Algorithm	Discussion of the algorithm		Figure 2	Figure 3 Table 1 Figure 5	Figure 4	Figure 6 Figure 7 Figure 8 Table 2 Table 3
A	Section S.2.1	n_1		m_1		m_1
B	Section S.2.1	n_1	20,000			20,000
C	Section S.2.2	n_2 n_3				
D	Section S.2.2	n_2 n_3				
Our	Section 2.3	m_1 m_2 m_4 m_5		25,000 100 m_1 20,000		50,000 100 m_1 20,000
F	Section S.3.1.1	n_4			50,000	
DSMH	Section S.4	H G initial N K M τ		50 10 20,000 100 10 10		50 10 50,000 100 10 10

Simulation exercise from section 3, $T = 120$, single pseudo dataset



Application to monetary policy shocks, section 4

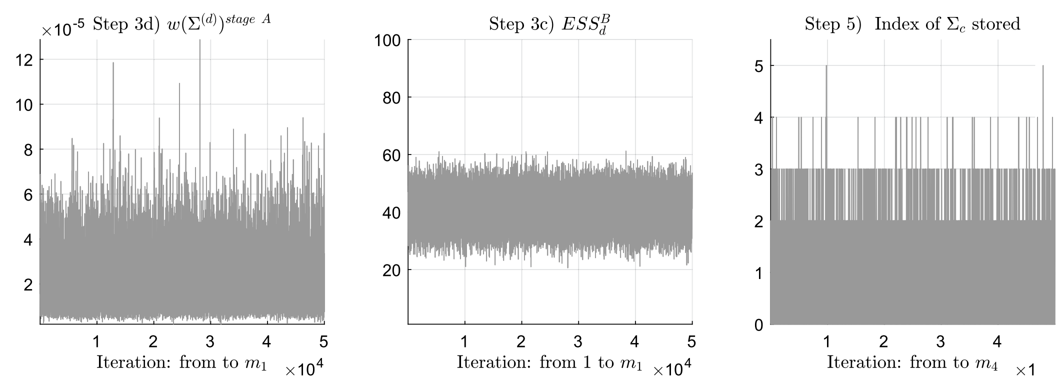


FIGURE S.1. Performance of our algorithm. *Note*: For each subfigure, the plot to the left shows the weights $w(\Sigma^{(d)})^{\text{stage A}}$ from Stage A of the sampler. The middle plot shows the effective sample size $\text{ESS}_d^B = (\sum_i (w_i(\Sigma^{(d)})^{\text{stage B}})^2)^{-1}$ computed, for each of the m_1 draws of Σ, out of the m_2 generated and stored draws for (Q, B). The plot to the right shows for each of the m_4 draws generated in step 4 how many times it was stored in step 8. Overall, the weights from Stage A are very balanced and correspond to a relative effective sample size of 0.9849 and 0.8013. The effective number of draws used in Stage B is relatively large, and the resampling from Stage B does not excessively use any specific draw of Σ.

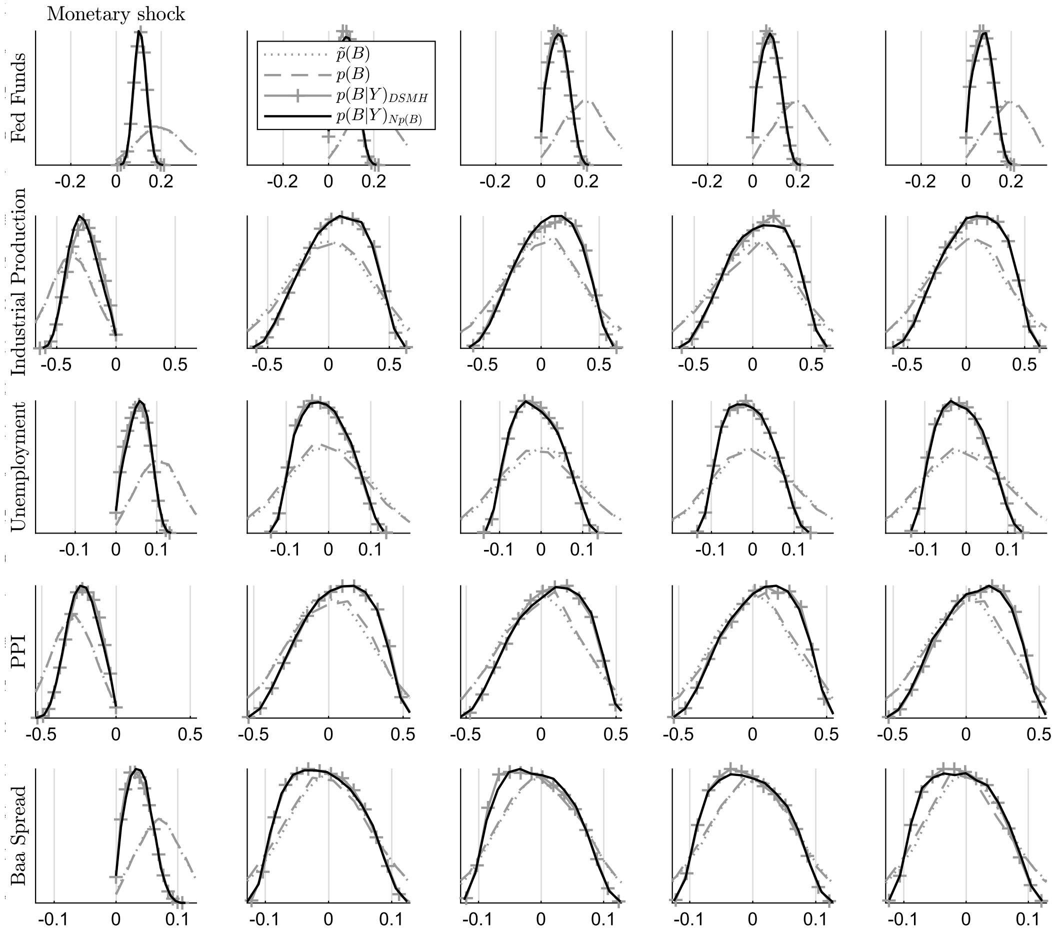


FIGURE S.2. Update on B.

TABLE S.2. Comparison of impulse responses (pointwise median). Robustness for $\psi_1 = 1$, $\psi_2 = 2$.

	0	1	2	3	4	5	6	7	8	9	10	11	12
							Using the NiWU prior, i_0						
Fed F.	25.00	17.78	13.04	9.23	5.96	3.04	0.34	−2.14	−4.37	−6.49	−8.29	−10.00	−11.47
IP	−0.44	−0.45	−0.48	−0.51	−0.55	−0.58	−0.60	−0.63	−0.66	−0.68	−0.70	−0.71	−0.73
Unempl.	8.62	9.14	9.78	10.35	10.87	11.32	11.78	12.19	12.56	12.86	13.13	13.40	13.59
PPI	−0.36	−0.37	−0.38	−0.39	−0.41	−0.43	−0.44	−0.46	−0.47	−0.48	−0.49	−0.50	−0.50
Baa S.	6.04	8.34	8.88	8.79	8.55	8.33	8.08	7.75	7.49	7.20	6.86	6.54	6.25
							Using our prior, i_1						
Fed F.	25.00	16.56	11.12	6.83	3.18	−0.07	−3.05	−5.71	−8.18	−10.44	−12.48	−14.28	−15.93
IP	−0.60	−0.61	−0.64	−0.68	−0.72	−0.75	−0.78	−0.81	−0.83	−0.86	−0.88	−0.89	−0.91
Unempl.	11.56	12.16	12.88	13.44	14.00	14.52	14.95	15.37	15.71	16.05	16.30	16.50	16.68
PPI	−0.48	−0.49	−0.50	−0.52	−0.54	−0.55	−0.57	−0.58	−0.59	−0.61	−0.62	−0.63	−0.63
Baa S.	7.93	10.65	11.22	11.06	10.67	10.29	9.91	9.55	9.14	8.72	8.31	7.90	7.49
							Ratio of our prior to NiWU prior, $100 \cdot (i_1 / i_0 - 1)$ (percent)						
Fed F.	0	−7	−15	−26	−47	−102	−991	167	87	61	51	43	39
IP	38	35	34	32	32	30	29	28	27	26	26	25	25
Unempl.	34	33	32	30	29	28	27	26	25	25	24	23	23
PPI	34	33	32	31	30	29	28	27	27	26	26	26	26
Baa S.	31	28	26	26	25	24	23	23	22	21	21	21	20

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

TABLE S.3. Comparison of impulse responses (pointwise median). Robustness for $\psi_1 = 0.6$, $\psi_2 = 1.2$.

	0	1	2	3	4	5	6	7	8	9	10	11	12
						Using the NiWU prior, i_0							
Fed F.	25.00	17.65	12.91	9.08	5.79	2.86	0.18	−2.25	−4.51	−6.61	−8.42	−10.10	−11.61
IP	−0.43	−0.45	−0.48	−0.52	−0.55	−0.58	−0.60	−0.63	−0.65	−0.67	−0.69	−0.71	−0.72
Unempl.	8.86	9.40	10.05	10.63	11.13	11.63	12.05	12.42	12.79	13.12	13.36	13.59	13.80
PPI	−0.36	−0.37	−0.38	−0.39	−0.41	−0.43	−0.44	−0.46	−0.47	−0.48	−0.49	−0.50	−0.51
Baa S.	6.16	8.37	8.87	8.79	8.53	8.25	7.98	7.70	7.42	7.13	6.84	6.54	6.21
						Using our prior, i_1							
Fed F.	25.00	16.36	10.81	6.41	2.63	−0.68	−3.70	−6.43	−8.96	−11.24	−13.29	−15.17	−16.86
IP	−0.60	−0.61	−0.64	−0.68	−0.72	−0.75	−0.78	−0.81	−0.83	−0.85	−0.87	−0.89	−0.90
Unempl.	12.68	13.09	13.83	14.42	15.00	15.46	15.91	16.27	16.63	16.90	17.11	17.34	17.47
PPI	−0.48	−0.49	−0.49	−0.51	−0.53	−0.54	−0.56	−0.57	−0.59	−0.60	−0.61	−0.61	−0.62
Baa S.	8.40	11.18	11.71	11.47	11.10	10.66	10.28	9.87	9.46	9.02	8.58	8.13	7.66
						Ratio of our prior to NiWU prior, $100 \cdot (i_1 / i_0 - 1)$ (percent)							
Fed F.	0	−7	−16	−29	−55	−124	−2105	185	99	70	58	50	45
IP	39	35	34	32	31	30	29	28	27	27	26	26	26
Unempl.	43	39	38	36	35	33	32	31	30	29	28	28	27
PPI	34	32	31	29	29	27	26	25	25	24	24	23	22
Baa S.	36	34	32	30	30	29	29	28	28	26	26	24	23

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

TABLE S.4. Comparison of forecast error variance decompositions (pointwise median). Robustness for $\psi_1 = 1$, $\psi_2 = 2$.

	0	1	2	3	4	5	6	7	8	9	10	11	12
							Using the NiWU prior, f_0						
Fed F.	0.36	0.28	0.22	0.17	0.14	0.12	0.10	0.09	0.09	0.08	0.08	0.08	0.08
IP	0.13	0.14	0.15	0.16	0.17	0.18	0.19	0.19	0.20	0.21	0.21	0.22	0.22
Unempl.	0.12	0.13	0.14	0.15	0.17	0.18	0.18	0.19	0.20	0.21	0.22	0.22	0.23
PPI	0.12	0.12	0.12	0.13	0.13	0.14	0.14	0.14	0.15	0.15	0.16	0.16	0.17
Baa S.	0.06	0.07	0.07	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08
							Using our prior, f_1						
Fed F.	0.28	0.21	0.16	0.12	0.10	0.08	0.07	0.07	0.07	0.06	0.06	0.07	0.07
IP	0.21	0.22	0.23	0.24	0.25	0.26	0.27	0.28	0.28	0.29	0.29	0.30	0.30
Unempl.	0.17	0.19	0.21	0.22	0.23	0.24	0.25	0.26	0.27	0.28	0.29	0.30	0.30
PPI	0.18	0.17	0.18	0.18	0.18	0.19	0.19	0.20	0.20	0.21	0.21	0.22	0.22
Baa S.	0.09	0.10	0.10	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11
							Ratio of our prior to NiWU prior, $100 \cdot (f_1/f_0 - 1)$ (percent)						
Fed F.	-22	-25	-27	-29	-29	-29	-27	-24	-22	-20	-18	-16	-14
IP	55	54	53	51	49	47	45	43	42	40	38	36	35
Unempl.	42	43	43	42	40	39	38	37	35	34	33	32	31
PPI	48	47	45	43	42	40	39	37	37	36	34	33	32
Baa S.	45	42	41	39	38	38	36	35	34	34	34	33	32

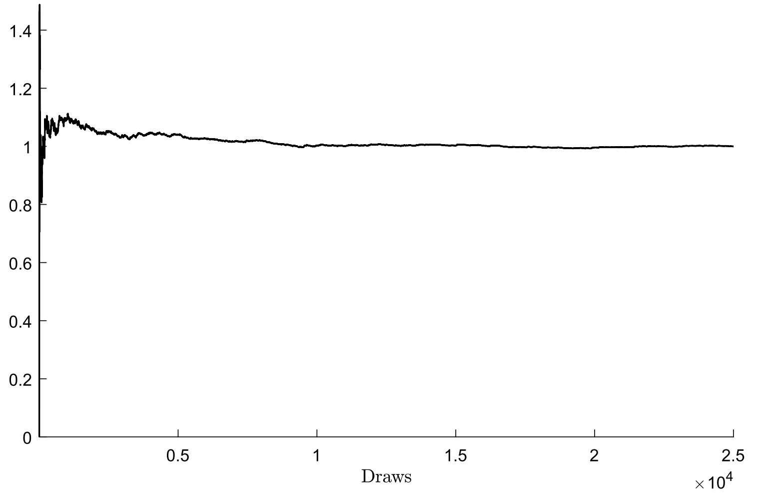
Note: Pointwise median forecast error decompositions using the NiWU prior (top panel) or our prior (middle panel), together with the percent difference between the two (bottom panel).

Table S.5. Comparison of forecast error variance decompositions (pointwise median). Robustness for $\psi_1 = 0.6$, $\psi_2 = 1.2$.

	0	1	2	3	4	5	6	7	8	9	10	11	12
Fed F.	0.36	0.27	0.21	0.17	0.14	0.11	0.10	0.09	0.08	0.08	0.08	0.08	0.08
IP	0.13	0.15	0.15	0.16	0.17	0.18	0.19	0.20	0.20	0.21	0.21	0.22	0.22
Unempl.	0.12	0.13	0.14	0.16	0.17	0.18	0.19	0.20	0.20	0.21	0.22	0.22	0.23
PPI	0.12	0.12	0.12	0.12	0.13	0.13	0.14	0.14	0.15	0.15	0.15	0.16	0.16
Baa S.	0.06	0.07	0.07	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08
						Using our prior, f_1							
Fed F.	0.28	0.21	0.16	0.12	0.10	0.08	0.07	0.07	0.07	0.07	0.07	0.07	0.07
IP	0.21	0.22	0.23	0.24	0.25	0.26	0.27	0.28	0.29	0.29	0.29	0.30	0.30
Unempl.	0.20	0.22	0.24	0.26	0.27	0.28	0.29	0.31	0.31	0.32	0.33	0.34	0.34
PPI	0.17	0.17	0.17	0.18	0.18	0.19	0.19	0.20	0.20	0.21	0.21	0.21	0.22
Baa S.	0.10	0.11	0.12	0.12	0.12	0.12	0.13	0.13	0.13	0.13	0.13	0.13	0.13
						Ratio of our prior to NiWU prior, $100 \cdot (f_1/f_0 - 1)$ (percent)							
Fed F.	−21	−25	−27	−28	−28	−28	−26	−24	−21	−19	−16	−14	−11
IP	57	53	51	49	47	45	43	42	41	39	38	37	35
Unempl.	73	69	66	64	62	61	58	56	54	52	51	49	49
PPI	47	46	45	43	42	41	41	39	38	37	36	35	34
Baa S.	71	65	63	61	59	58	57	56	55	53	51	50	49

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

Simulation exercise from section 3, $T = 120$, single pseudo dataset



Application to monetary shocks, section 4

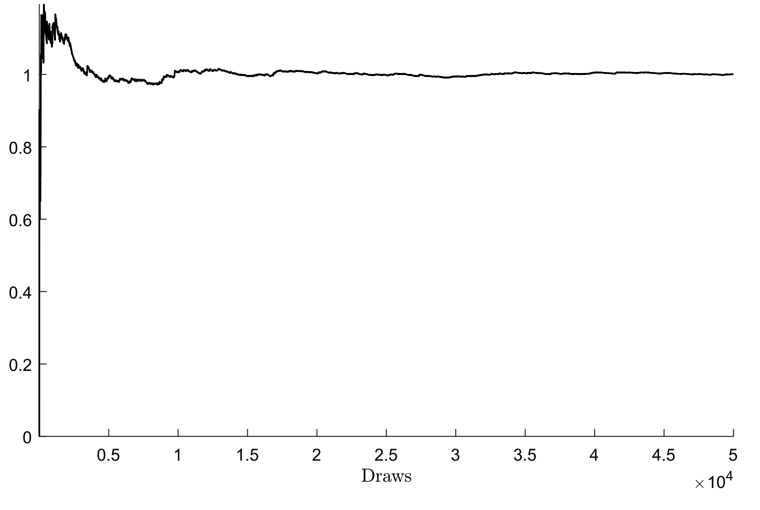


FIGURE S.3. Diagnostics on the importance weights, graphical assessment. *Note*: The graph shows the recursive variance $\{v_i\}_{i=1}^{N}$, where $v_i = \mathrm{Var}(w_{1:i})$ computed using demeaned and standardized weights.

TABLE S.6. Diagnostics on the importance weights in Stage A of our sampler: tests.

Simulation exercise from Section 3, $T = 120$, single pseudo data set					
u	0.5N	0.6N	0.7N	0.9N	0.99N
Wald	−46.3176	−45.0236	−42.5017	−30.7455	−13.6695
Score	−18.0855	−15.5888	−13.0071	−7.35923	−2.42388
Application to monetary policy shocks, Section 4					
u	0.5N	0.6N	0.7N	0.9N	0.99N
Wald	−50.6431	−45.959	−40.0677	−23.1004	−6.99473
Score	−20.62	−17.5873	−15.0084	−8.09104	−2.62551

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 and 1.65 for the score test. The corresponding p-values are close to 1 in all cases.

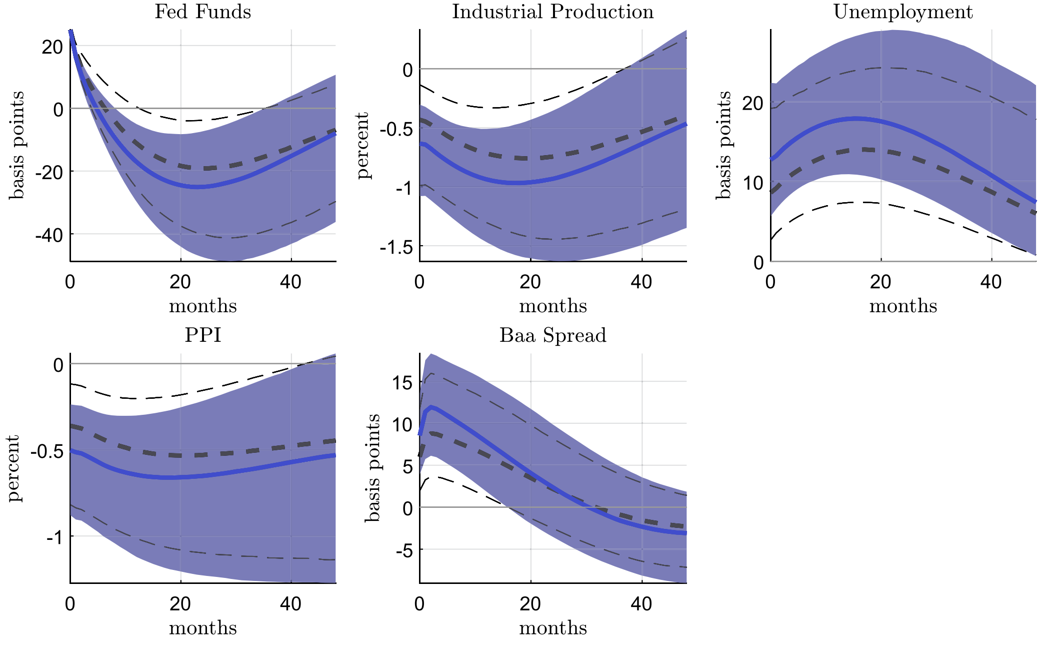


FIGURE S.4. Monetary policy shocks: Impulse responses, comparison to posterior associated with NiWU prior. *Note*: Pointwise median and 68% credible bands. The monetary shock is normalized to imply an impact 25 basis points increase in the policy rate. Dashed lines report the case under the NiWU prior.

REFERENCES

Abadir, Karim M. and Jan R. Magnus (2005), *Matrix Algebra*, Vol. 1. Cambridge University Press. [4]

Arias, Jonas E., Juan F. Rubio-Ramírez, and Daniel F. Waggoner (2018), "Inference based on structural vector autoregressions identified with sign and zero restrictions: Theory and applications." *Econometrica*, 86 (2), 685–720. [11, 17, 25]

Bibby, John M., John Kent, and Kantilal Mardia (1979), *Multivariate Analysis*. Academic Press, London. [11]

Brooks, Stephen P. and Andrew Gelman (1998), "General methods for monitoring convergence of iterative simulations." *Journal of Computational and Graphical Statistics*, 7 (4), 434–455. [21, 24]

Brooks, Stephen P. and Gareth O. Roberts (1998), "Assessing convergence of Markov chain Monte Carlo algorithms." *Statistics and Computing*, 8 (4), 319–335. [22]

Canova, Fabio (2007), *Methods for Applied Macroeconomic Research*, Vol. 13. Princeton University Press. [2]

Casella, George and Roger L. Berger (2021), *Statistical Inference*. Cengage Learning. [11]

Cowles, Mary Kathryn and Bradley P. Carlin (1996), "Markov chain Monte Carlo convergence diagnostics: A comparative review." *Journal of the American Statistical Association*, 91 (434), 883–904. [22]

Del Negro, Marco and Frank Schorfheide (2011), "Bayesian macroeconometrics." In *The Oxford Handbook of Bayesian Econometrics* (Gary Koop, Herman van Dijk, and John Geweke, eds.), 293–389, Oxford University Press. [14]

Doan, Thomas, Robert Litterman, and Christopher Sims (1984), "Forecasting and conditional projection using realistic prior distributions." *Econometric Reviews*, 3 (1), 1–100. [14]

Gelman, Andrew and Donald B. Rubin (1992), "Inference from iterative simulation using multiple sequences." *Statistical Science*, 7 (4), 457–472. [21, 23, 24, 25]

Geweke, John (1989), "Bayesian inference in econometric models using Monte Carlo integration." *Econometrica*, 57 (6), 1317–1339. [16]

Geweke, John (1992), "Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments." In *Bayesian Statistics*. [21, 22]

Herbst, Edward and Frank Schorfheide (2014), "Sequential Monte Carlo sampling for DSGE models." *Journal of Applied Econometrics*, 29 (7), 1073–1098. [20]

Kilian, Lutz and Helmut Lütkepohl (2017), *Structural Vector Autoregressive Analysis*. Cambridge University Press. [2]

Koop, Gary and Dimitris Korobilis (2010), "Bayesian multivariate time series methods for empirical macroeconomics." *Foundations and Trends in Econometrics*, 3 (4), 267–358. [2]

Koopman, Siem Jan, Neil Shephard, and Drew Creal (2009), "Testing the assumptions behind importance sampling." *Journal of Econometrics*, 149 (1), 2–11. [16]

Lütkepohl, Helmut (2005), *New Introduction to Multiple Time Series Analysis*. Springer Science & Business Media. [1]

Mathai, Arakaparampil M. and Hans J. Haubold (2008), *Special Functions for Applied Scientists*, Vol. 4. Springer. [11]

Mlikota, Marko and Frank Schorfheide (forthcoming), "Sequential Monte Carlo with model tempering." *Studies in Nonlinear Dynamics & Econometrics*. [19]

Pickands, James (1975), "Statistical inference using extreme order statistics." *The Annals of Statistics*, 3 (1), 119–131. [16]

Raftery, Adrian E. and Steven Lewis (1992), "How many iterations in the Gibbs sampler?" *Bayesian Statistics*, 4, 763–773. [21, 22, 23]

Robert, Christian and George Casella (2013), *Monte Carlo Statistical Methods*. Springer Science & Business Media. [16]

Rubio-Ramirez, Juan F., Daniel F. Waggoner, and Tao Zha (2010), "Structural vector autoregressions: Theory of identification and algorithms for inference." *The Review of Economic Studies*, 77 (2), 665–696. [5, 9, 12, 14, 19]

Waggoner, Daniel F., Hongwei Wu, and Tao Zha (2016), "Striated Metropolis–Hastings sampler for high-dimensional models." *Journal of Econometrics*, 192 (2), 406–420. [18, 19, 20, 21]

Co-editor James D. Hamilton handled this manuscript.

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