

# A new posterior sampler for Bayesian structural vector autoregressive models

## ONLINE APPENDIX

Martin Bruns\*      Michele Piffer<sup>†</sup>

June 23, 2023

---

\*University of East Anglia, School of Economics, Norwich, NR4 7TJ, United Kingdom.

<sup>†</sup>King's Business School, King's College London, Bush House, 30 Aldwych, London, WC2B 4BG, United Kingdom. Corresponding author e-mail: [m.b.piffer@gmail.com](mailto:m.b.piffer@gmail.com)

# Contents

<b>A Likelihood function of the model</b>	<b>1</b>
<b>B The independent NiWU approach</b>	<b>3</b>
B.1 Improper flat & generalized inverse Wishart prior . . . . .	3
B.2 Independent Normal & generalized inverse Wishart prior . . . . .	8
<b>C Np(B) approach proposed in the paper</b>	<b>13</b>
C.1 Derivations . . . . .	13
C.1.1 General case . . . . .	13
C.1.2 Case 1: flat prior on $\boldsymbol{\pi}$ . . . . .	20
C.1.3 Case 2: prior independence . . . . .	22
C.2 Diagnostics for the importance sampler . . . . .	23
C.2.1 Graphical Assessment . . . . .	24
C.2.2 Diagnostic tests . . . . .	24
<b>D The Dynamic Striated Metropolis-Hastings algorithm by Waggoner et al. (2016)</b>	<b>27</b>
D.1 The DSMH sampler . . . . .	27
D.2 Convergence criteria used for the DSMH algorithm . . . . .	31
<b>E Our prior distribution</b>	<b>36</b>
<b>F Additional Tables and Figures</b>	<b>41</b>
<b>References</b>	<b>51</b>

## A Likelihood function of the model

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

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

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

$$Y = \Pi W + BE, \quad (\text{A.2})$$

where  $Y = [\mathbf{y}_1, \dots, \mathbf{y}_t, \dots, \mathbf{y}_T]$  and  $E = [\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_t, \dots, \boldsymbol{\epsilon}_T]$  are  $k \times T$  matrices of data and shocks, and  $W = [\mathbf{w}_1, \dots, \mathbf{w}_t, \dots, \mathbf{w}_T]$  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

$$\tilde{\mathbf{y}} = Z \boldsymbol{\pi} + (I_T \otimes B) \tilde{\boldsymbol{\epsilon}}, \quad \tilde{\boldsymbol{\epsilon}} \sim N(\mathbf{0}, I_{Tk}), \quad (\text{A.3})$$

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

$$Z = (W' \otimes I_k), \quad (\text{A.4})$$

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

$$\tilde{\mathbf{y}} = Z \boldsymbol{\pi} + \tilde{\mathbf{u}}, \quad \tilde{\mathbf{u}} \sim N\left(\mathbf{0}, (I_T \otimes BB')\right), \quad (\text{A.5})$$

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

$$\hat{\Pi}_T = YW'(WW')^{-1}, \quad (\text{A.6})$$

$$\hat{\boldsymbol{\pi}}_T = \text{vec}(\hat{\Pi}_T) = ((WW')^{-1}W \otimes I_k)\tilde{\mathbf{y}}, \quad (\text{A.7})$$

$$\hat{\Sigma}_T = \frac{(Y - \hat{\Pi}_T W)(Y - \hat{\Pi}_T W)'}{T - m}. \quad (\text{A.8})$$

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

$$p(Y|\boldsymbol{\pi}, B) = (2\pi)^{-\frac{kT}{2}} |\det(I_T \otimes BB')|^{-\frac{1}{2}} e^{-\frac{1}{2}(\tilde{\mathbf{y}} - Z\boldsymbol{\pi})'(I_T \otimes BB')^{-1}(\tilde{\mathbf{y}} - Z\boldsymbol{\pi})}, \quad (\text{A.9})$$

$$\propto |\det(B)|^{-T} e^{-\frac{1}{2}(\tilde{\mathbf{y}} - (W' \otimes I_T)\boldsymbol{\pi})'(I_T \otimes (BB')^{-1})(\tilde{\mathbf{y}} - (W' \otimes I_T)\boldsymbol{\pi})}, \quad (\text{A.10})$$

$$\propto |\det(B)|^{-T} e^{-\frac{1}{2}\left\{\tilde{\mathbf{y}}'(I_T \otimes (BB')^{-1})\tilde{\mathbf{y}} + \boldsymbol{\pi}'(WW' \otimes (BB')^{-1})\boldsymbol{\pi} - 2\boldsymbol{\pi}'(W \otimes (BB')^{-1})\tilde{\mathbf{y}}\right\}}, \quad (\text{A.11})$$

$$\propto |\det(B)|^{-T} e^{-\frac{1}{2}\left\{\boldsymbol{\pi}'(WW' \otimes (BB')^{-1})\boldsymbol{\pi} - 2\boldsymbol{\pi}'(WW' \otimes (BB')^{-1})\hat{\boldsymbol{\pi}}_T + \tilde{\mathbf{y}}'(I_T \otimes (BB')^{-1})\tilde{\mathbf{y}}\right\}}, \quad (\text{A.12})$$

where the last step uses

$$2\boldsymbol{\pi}'(W \otimes (BB')^{-1})\tilde{\mathbf{y}} = 2\boldsymbol{\pi}'(WW' \otimes (BB')^{-1})((WW')^{-1}W \otimes I_k)\tilde{\mathbf{y}}, \quad (\text{A.13})$$

$$= 2\boldsymbol{\pi}'(WW' \otimes (BB')^{-1})\hat{\boldsymbol{\pi}}_T. \quad (\text{A.14})$$

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

$$p(Y|\boldsymbol{\pi}, \Sigma) = (2\pi)^{-\frac{kT}{2}} |\det(\Sigma)|^{-\frac{T}{2}} e^{-\frac{T}{2}(\tilde{\mathbf{y}} - Z\boldsymbol{\pi})'(I_T \otimes \Sigma)^{-1}(\tilde{\mathbf{y}} - Z\boldsymbol{\pi})}, \quad (\text{A.15})$$

$$\propto |\det(\Sigma)|^{-\frac{T}{2}} e^{-\frac{1}{2}\left\{\boldsymbol{\pi}'(WW' \otimes \Sigma^{-1})\boldsymbol{\pi} - 2\boldsymbol{\pi}'(WW' \otimes \Sigma^{-1})\hat{\boldsymbol{\pi}}_T + \tilde{\mathbf{y}}'(I_T \otimes \Sigma^{-1})\tilde{\mathbf{y}}\right\}}. \quad (\text{A.16})$$

## B 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  $\Sigma$ . 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, i.e.  $\tilde{p}(\cdot), \tilde{p}_{NiWU,i}(\cdot)$ , or don't introduce sign restrictions, i.e.  $p(\cdot), p_{NiWU,i}(\cdot)$ , as this difference is important for our sampler.

### B.1 Improper flat & generalized inverse Wishart prior

We first consider the prior beliefs

$$\tilde{p}_{NiWU,i}(\boldsymbol{\pi}, \Sigma, Q) \propto \mathbf{I}\{\boldsymbol{\pi}, \Sigma, Q\} \cdot p_{NiWU,i}(\boldsymbol{\pi}, \Sigma, Q), \quad (\text{B.17})$$

$$p_{NiWU,i}(\boldsymbol{\pi}, \Sigma, Q) = p_{NiWU,i}(\boldsymbol{\pi}) \cdot p_{NiWU,i}(\Sigma) \cdot p_{NiWU,i}(Q), \quad (\text{B.18})$$

$$p_{NiWU,i}(\boldsymbol{\pi}) \propto 1, \quad (\text{B.19})$$

$$p_{NiWU,i}(\Sigma) \propto |\det(\Sigma)|^{\frac{c}{2}} \cdot e^{-\frac{1}{2}\text{trace}[\Sigma^{-1}S]}, \quad (\text{B.20})$$

$$p_{NiWU,i}(Q) = U_{O(k)} \propto 1. \quad (\text{B.21})$$

The prior on  $\boldsymbol{\pi}$  is an improper flat prior. The prior on  $\Sigma$  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

$$\tilde{p}_{NiWU,i}(\boldsymbol{\pi}, \Sigma, Q|Y) \propto I\{\boldsymbol{\pi}, \Sigma, Q\} \cdot p_{NiWU,i}(\boldsymbol{\pi}, \Sigma, Q|Y), \quad (\text{B.22})$$

$$p_{NiWU,i}(\boldsymbol{\pi}, \Sigma, Q|Y) = p_{NiWU,i}(\boldsymbol{\pi}, \Sigma|Y) \cdot p_{NiWU,i}(Q|Y, \boldsymbol{\pi}, \Sigma), \quad (\text{B.23})$$

$$p_{NiWU,i}(\boldsymbol{\pi}, \Sigma|Y) \propto |\det(\Sigma)|^{\frac{c}{2}} \cdot e^{-\frac{1}{2}\text{trace}[\Sigma^{-1}S]} \cdot |\det(\Sigma)|^{-\frac{T}{2}} e^{-\frac{1}{2}\{\boldsymbol{\pi}'(WW' \otimes \Sigma^{-1})\boldsymbol{\pi} - 2\boldsymbol{\pi}'(WW' \otimes \Sigma^{-1})\hat{\boldsymbol{\pi}}_T + \tilde{\mathbf{y}}'(I_T \otimes \Sigma^{-1})\tilde{\mathbf{y}}\}}, \quad (\text{B.24})$$

$$p_{NiWU,i}(Q|Y, \boldsymbol{\pi}, \Sigma) = p_{NiWU,i}(Q) = U_{O(k)} \propto 1. \quad (\text{B.25})$$

Since

$$\boldsymbol{\pi}'(WW' \otimes \Sigma^{-1})\boldsymbol{\pi} - 2\boldsymbol{\pi}'(WW' \otimes \Sigma^{-1})\hat{\boldsymbol{\pi}}_T + \tilde{\mathbf{y}}'(I_T \otimes \Sigma^{-1})\tilde{\mathbf{y}}, \quad (\text{B.26})$$

$$= \boldsymbol{\pi}'V_{\pi}^{*-1}\boldsymbol{\pi} - 2\boldsymbol{\pi}'V_{\pi}^{*-1}\boldsymbol{\mu}_{\pi}^* + \tilde{\mathbf{y}}'(I_T \otimes \Sigma^{-1})\tilde{\mathbf{y}}, \quad (\text{B.27})$$

$$= (\boldsymbol{\pi} - \boldsymbol{\mu}_{\pi}^*)'V_{\pi}^{*-1}(\boldsymbol{\pi} - \boldsymbol{\mu}_{\pi}^*) + \tilde{\mathbf{y}}'(I_T \otimes \Sigma^{-1})\tilde{\mathbf{y}} - \boldsymbol{\mu}_{\pi}^{*'}V_{\pi}^{*-1}\boldsymbol{\mu}_{\pi}^*, \quad (\text{B.28})$$

$$= (\boldsymbol{\pi} - \boldsymbol{\mu}_{\pi}^*)'V_{\pi}^{*-1}(\boldsymbol{\pi} - \boldsymbol{\mu}_{\pi}^*) + \tilde{\mathbf{y}}'(I_T \otimes \Sigma^{-1})\tilde{\mathbf{y}} - \hat{\boldsymbol{\pi}}_T'(WW' \otimes \Sigma^{-1})\hat{\boldsymbol{\pi}}_T, \quad (\text{B.29})$$

given

$$V_{\pi}^* = (WW')^{-1} \otimes \Sigma, \quad (\text{B.30})$$

$$\boldsymbol{\mu}_{\pi}^* = \hat{\boldsymbol{\pi}}_T, \quad (\text{B.31})$$

equation (B.24) can be rewritten as

$$p_{NiWU,i}(\boldsymbol{\pi}, \Sigma|Y) \propto |\det(V_{\pi}^*)|^{-\frac{1}{2}} \cdot e^{-\frac{1}{2}(\boldsymbol{\pi} - \boldsymbol{\mu}_{\pi}^*)'V_{\pi}^{*-1}(\boldsymbol{\pi} - \boldsymbol{\mu}_{\pi}^*)} \cdot |\det(V_{\pi}^*)|^{\frac{1}{2}} \cdot |\det(\Sigma)|^{\frac{c}{2}} \cdot |\det(\Sigma)|^{-\frac{T}{2}} \cdot e^{-\frac{1}{2}\{\text{trace}[\Sigma^{-1}S] + \tilde{\mathbf{y}}'(I_T \otimes \Sigma^{-1})\tilde{\mathbf{y}} - \hat{\boldsymbol{\pi}}_T'(WW' \otimes \Sigma^{-1})\hat{\boldsymbol{\pi}}_T\}}. \quad (\text{B.32})$$

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

$$\tilde{\mathbf{y}}'(I_T \otimes \Sigma^{-1})\tilde{\mathbf{y}} = \text{trace}[Y'\Sigma^{-1}Y] = \text{trace}[\Sigma^{-1}YY'], \quad (\text{B.33})$$

$$\hat{\boldsymbol{\pi}}_T'(WW' \otimes \Sigma^{-1})\hat{\boldsymbol{\pi}}_T = \text{trace}[\hat{\Pi}_T'\Sigma^{-1}\hat{\Pi}_TWW'] = \text{trace}[\Sigma^{-1}\hat{\Pi}_TWW'\hat{\Pi}_T'], \quad (\text{B.34})$$

noticing that

$$YY' - \hat{\Pi}_TWW'\hat{\Pi}_T' = (Y - \hat{\Pi}_TW)(Y - \hat{\Pi}_TW)', \quad (\text{B.35})$$

$$= (Y - \hat{\Pi}_TW)(Y - \hat{\Pi}_TW)' \frac{T-m}{T-m}, \quad (\text{B.36})$$

$$= \hat{\Sigma}_T(T-m), \quad (\text{B.37})$$

and using the result

$$\det(V_\pi^*) = |\det(WW')|^k \cdot |\det(\Sigma)|^m, \quad (\text{B.38})$$

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

$$\begin{aligned} p_{NiWU,i}(\boldsymbol{\pi}, \Sigma|Y) &\propto |\det(V_\pi^*)|^{-\frac{1}{2}} \cdot e^{-\frac{1}{2}(\boldsymbol{\pi}-\boldsymbol{\mu}_\pi^*)'V_\pi^{*-1}(\boldsymbol{\pi}-\boldsymbol{\mu}_\pi^*)} \\ &\quad \cdot |\det(\Sigma)|^{\frac{m}{2}} \cdot |\det(\Sigma)|^{\frac{c}{2}} \cdot |\det(\Sigma)|^{-\frac{T}{2}}. \end{aligned} \quad (\text{B.39})$$

$$\begin{aligned} &\cdot e^{-\frac{1}{2}\text{trace}\left[\Sigma^{-1}\left(S+\hat{\Sigma}_T(T-m)\right)\right]}, \\ &\propto |\det(V_\pi^*)|^{-\frac{1}{2}} \cdot e^{-\frac{1}{2}(\boldsymbol{\pi}-\boldsymbol{\mu}_\pi^*)'V_\pi^{*-1}(\boldsymbol{\pi}-\boldsymbol{\mu}_\pi^*)} \\ &\quad \cdot |\det(\Sigma)|^{-\frac{(T-m-c-k-1)+k+1}{2}} \cdot e^{-\frac{1}{2}\text{trace}\left[\Sigma^{-1}\left(S+\hat{\Sigma}_T(T-m)\right)\right]}. \end{aligned} \quad (\text{B.40})$$

It then follows that the prior beliefs in equations (B.17)-(B.21) imply

$$\tilde{p}_{NiWU,i}(\boldsymbol{\pi}, \Sigma, Q|Y) \propto I\{\boldsymbol{\pi}, \Sigma, Q\} \cdot p_{NiWU,i}(\boldsymbol{\pi}, \Sigma, Q|Y), \quad (\text{B.41})$$

$$p_{NiWU,i}(\boldsymbol{\pi}, \Sigma, Q|Y) = p_{NiWU,i}(\boldsymbol{\pi}|Y, \Sigma) \cdot p_{NiWU,i}(\Sigma|Y) \cdot p_{NiWU,i}(Q|Y, \boldsymbol{\pi}, \Sigma), \quad (\text{B.42})$$

$$p_{NiWU,i}(\Sigma|Y) = iW(S^*, d^*), \quad (\text{B.43})$$

$$p_{NiWU,i}(\boldsymbol{\pi}|Y, \Sigma) = \phi(\boldsymbol{\mu}_\pi^*, V_\pi^*), \quad (\text{B.44})$$

$$p_{NiWU,i}(Q|Y, \boldsymbol{\pi}, \Sigma) = p_{NiWU,i}(Q) = U_{O(k)} \propto 1, \quad (\text{B.45})$$

$$d^* = T - m - c - k - 1, \quad (\text{B.46})$$

$$S^* = S + \hat{\Sigma}_T(T - m), \quad (\text{B.47})$$

$$\boldsymbol{\mu}_\pi^* = \hat{\boldsymbol{\pi}}_T, \quad (\text{B.48})$$

$$V_\pi^* = (WW')^{-1} \otimes \Sigma, \quad (\text{B.49})$$

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

$$d^* \geq k, \quad (\text{B.50})$$

or equivalently,

$$c \leq T - m - 2k - 1, \quad (\text{B.51})$$

$$T \geq c + m + 2k + 1. \quad (\text{B.52})$$



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

$$\log[p_{NiWU,i}(\Sigma|Y)] \propto \frac{T-m-c}{2} \cdot \log(|\det(\Sigma^{-1})|) - \frac{1}{2} \text{trace}[\Sigma^{-1}(S + \hat{\Sigma}_T(T-m))], \quad (\text{B.53})$$

$$\frac{d \log[p_{NiWU,i}(\Sigma|Y)]}{d \Sigma^{-1}} = \frac{T-m-c}{2} \Sigma - \frac{1}{2} (S + \hat{\Sigma}_T(T-m)) = 0, \quad (\text{B.54})$$

$$\Sigma_{mode} = \frac{1}{T-m-c} S + \frac{T-m}{T-m-c} \hat{\Sigma}_T. \quad (\text{B.55})$$

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

1. draw  $\Sigma^{(d)}$  from an  $iW(S^*, d^*)$ , equation (B.43);
2. draw  $\boldsymbol{\pi}^{(d)}$  from a  $N(\boldsymbol{\mu}_\pi^*, V_\pi^*)$ , equation (B.44), using  $V_\pi^*$  associated with  $\Sigma^{(d)}$ ;
3. store  $(\boldsymbol{\pi}^{(d)}, \Sigma^{(d)})$ ;
4. repeat Steps 1-3 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:*  $\tilde{p}_{NiWU,i}(\boldsymbol{\pi}, \Sigma, Q|Y)$

1. draw  $\Sigma^{(d)}$  from an  $iW(S^*, d^*)$ , equation (B.43);

2. draw  $\boldsymbol{\pi}^{(d)}$  from a  $N(\boldsymbol{\mu}_\pi^*, V_\pi^*)$ , equation (B.44), using  $V_\pi^*$  associated with  $\Sigma^{(d)}$ ;
3. draw  $Q^{(d)}$  from the (Haar)uniform distribution using the algorithm by Rubio-Ramirez et al. (2010);
4. compute the structural statistics of interest associated with  $(\boldsymbol{\pi}^{(d)}, \Sigma^{(d)}, Q^{(d)})$ :
  - 4a. if the restrictions are satisfied, store  $(\boldsymbol{\pi}^{(d)}, \Sigma^{(d)}, Q^{(d)})$ ;
  - 4b. if the restrictions are not satisfied, move back to Step 1;
5. repeat Steps 1-4 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 (B.43), and conditional draws for  $\boldsymbol{\pi}$  from the Normal distribution (B.44). 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 *Algorithms 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.

## B.2 Independent Normal & generalized inverse Wishart prior

We now discuss the case of prior beliefs

$$\tilde{p}_{NiWU,i}(\boldsymbol{\pi}, \Sigma, Q) \propto I\{\boldsymbol{\pi}, \Sigma, Q\} \cdot p_{NiWU,i}(\boldsymbol{\pi}, \Sigma, Q), \quad (\text{B.56})$$

$$p_{NiWU,i}(\boldsymbol{\pi}, \Sigma) = p_{NiWU,i}(\boldsymbol{\pi}) \cdot p_{NiWU,i}(\Sigma) \cdot p_{NiWU,i}(Q), \quad (\text{B.57})$$

$$p_{NiWU,i}(\boldsymbol{\pi}) = \phi(\boldsymbol{\mu}_\pi, V_\pi), \quad (\text{B.58})$$

$$p_{NiWU,i}(\Sigma) \propto |\det(\Sigma)|^{\frac{c}{2}} \cdot e^{-\frac{1}{2}\text{trace}[\Sigma^{-1}S]}, \quad (\text{B.59})$$

$$p_{NiWU,i}(Q) = U_{O(k)} \propto 1. \quad (\text{B.60})$$

Compared to [Section B.1](#), the density on  $\boldsymbol{\pi}$  is proper. The above prior implies the joint posterior distribution

$$\tilde{p}_{NiWU,i}(\boldsymbol{\pi}, \Sigma, Q|Y) \propto \mathbf{I}\{\boldsymbol{\pi}, \Sigma, Q\} \cdot p_{NiWU,i}(\boldsymbol{\pi}, \Sigma, Q|Y), \quad (\text{B.61})$$

$$p_{NiWU,i}(\boldsymbol{\pi}, \Sigma, Q|Y) = p_{NiWU,i}(\boldsymbol{\pi}, \Sigma|Y) \cdot p_{NiWU,i}(Q|Y, \boldsymbol{\pi}, \Sigma), \quad (\text{B.62})$$

$$p_{NiWU,i}(\boldsymbol{\pi}, \Sigma|Y) \propto |\det(\Sigma)|^{\frac{c}{2}} \cdot e^{-\frac{1}{2}\text{trace}[\Sigma^{-1}S]} \cdot |\det(V_\pi)|^{-\frac{1}{2}} \cdot e^{-\frac{1}{2}(\boldsymbol{\pi} - \boldsymbol{\mu}_\pi)' V_\pi^{-1} (\boldsymbol{\pi} - \boldsymbol{\mu}_\pi)}. \quad (\text{B.63})$$

$$\begin{aligned} & \cdot |\det(\Sigma)|^{-\frac{T}{2}} \cdot e^{-\frac{1}{2}\left\{\boldsymbol{\pi}'(WW' \otimes \Sigma^{-1})\boldsymbol{\pi} - 2\boldsymbol{\pi}'(WW' \otimes \Sigma^{-1})\hat{\boldsymbol{\pi}}_T + \tilde{\mathbf{y}}'(I_T \otimes \Sigma^{-1})\tilde{\mathbf{y}}\right\}}, \\ & \propto |\det(\Sigma)|^{\frac{c}{2}} \cdot e^{-\frac{1}{2}\text{trace}[\Sigma^{-1}S]} \\ & \cdot e^{-\frac{1}{2}\left\{\boldsymbol{\pi}'V_\pi^{-1}\boldsymbol{\pi} - 2\boldsymbol{\pi}'V_\pi^{-1}\boldsymbol{\mu}_\pi + \boldsymbol{\mu}_\pi'V_\pi^{-1}\boldsymbol{\mu}_\pi\right\}}. \end{aligned} \quad (\text{B.64})$$

$$p_{NiWU,i}(Q|Y, \boldsymbol{\pi}, \Sigma) = p_{NiWU,i}(Q) = U_{O(k)} \propto 1. \quad (\text{B.65})$$

Since

$$\boldsymbol{\pi}'(V_\pi^{-1} + WW' \otimes \Sigma^{-1})\boldsymbol{\pi} - 2\boldsymbol{\pi}'(V_\pi^{-1}\boldsymbol{\mu}_\pi + (WW' \otimes \Sigma^{-1})\hat{\boldsymbol{\pi}}_T) + \tilde{\mathbf{y}}'(I_T \otimes \Sigma^{-1})\tilde{\mathbf{y}} + \boldsymbol{\mu}_\pi'V_\pi^{-1}\boldsymbol{\mu}_\pi, \quad (\text{B.66})$$

$$= \boldsymbol{\pi}'V_\pi^{*-1}\boldsymbol{\pi} - 2\boldsymbol{\pi}'V_\pi^{*-1}\boldsymbol{\mu}_\pi^* + \tilde{\mathbf{y}}'(I_T \otimes \Sigma^{-1})\tilde{\mathbf{y}} + \boldsymbol{\mu}_\pi'V_\pi^{-1}\boldsymbol{\mu}_\pi, \quad (\text{B.67})$$

$$= (\boldsymbol{\pi} - \boldsymbol{\mu}_\pi^*)'V_\pi^{*-1}(\boldsymbol{\pi} - \boldsymbol{\mu}_\pi^*) + \tilde{\mathbf{y}}'(I_T \otimes \Sigma^{-1})\tilde{\mathbf{y}} + \boldsymbol{\mu}_\pi'V_\pi^{-1}\boldsymbol{\mu}_\pi - \boldsymbol{\mu}_\pi^{*'}V_\pi^{*-1}\boldsymbol{\mu}_\pi^*, \quad (\text{B.68})$$

with

$$V_\pi^* = (V_\pi^{-1} + WW' \otimes \Sigma^{-1})^{-1}, \quad (\text{B.69})$$

$$\boldsymbol{\mu}_\pi^* = V_\pi^*(V_\pi^{-1}\boldsymbol{\mu}_\pi + (WW' \otimes \Sigma^{-1})\hat{\boldsymbol{\pi}}_T), \quad (\text{B.70})$$

equation (B.64) can be rewritten as

$$\begin{aligned}
p_{NiWU,i}(\boldsymbol{\pi}, \Sigma | Y) &\propto |\det(V_\pi^*)|^{-\frac{1}{2}} \cdot e^{-\frac{1}{2}(\boldsymbol{\pi} - \boldsymbol{\mu}_\pi^*)' V_\pi^{*-1} (\boldsymbol{\pi} - \boldsymbol{\mu}_\pi^*)} \\
&\cdot |\det(V_\pi^*)|^{\frac{1}{2}} \cdot |\det(\Sigma)|^{\frac{c}{2}} \cdot |\det(\Sigma)|^{-\frac{T}{2}} \cdot \\
&\cdot e^{-\frac{1}{2} \left\{ \text{trace}[\Sigma^{-1} S] + \tilde{\mathbf{y}}' (I_T \otimes \Sigma^{-1}) \tilde{\mathbf{y}} - \boldsymbol{\mu}_\pi^{*'} V_\pi^{*-1} \boldsymbol{\mu}_\pi^* \right\}},
\end{aligned} \tag{B.71}$$

$$\begin{aligned}
&\propto |\det(V_\pi^*)|^{-\frac{1}{2}} \cdot e^{-\frac{1}{2}(\boldsymbol{\pi} - \boldsymbol{\mu}_\pi^*)' V_\pi^{*-1} (\boldsymbol{\pi} - \boldsymbol{\mu}_\pi^*)} \\
&\cdot |\det(V_\pi^*)|^{\frac{1}{2}} \cdot |\det(\Sigma)|^{-\frac{T-c}{2}} \cdot \\
&\cdot e^{-\frac{1}{2} \left\{ \text{trace}[\Sigma^{-1} S] + \tilde{\mathbf{y}}' (I_T \otimes \Sigma^{-1}) \tilde{\mathbf{y}} - \boldsymbol{\mu}_\pi^{*'} V_\pi^{*-1} \boldsymbol{\mu}_\pi^* \right\}}.
\end{aligned} \tag{B.72}$$

Last, we can derive

$$p_{NiWU,i}(\Sigma | Y, \Pi) \propto |\det(\Sigma)|^{\frac{c}{2}} \cdot e^{-\frac{1}{2} \text{trace}[\Sigma^{-1} S]}. \tag{B.73}$$

$$\cdot |\det(\Sigma)|^{-\frac{T}{2}} \cdot e^{-\frac{1}{2} \left\{ \boldsymbol{\pi}' (W W' \otimes \Sigma^{-1}) \boldsymbol{\pi} - 2 \boldsymbol{\pi}' (W W' \otimes \Sigma^{-1}) \tilde{\boldsymbol{\pi}}_T + \tilde{\mathbf{y}}' (I_T \otimes \Sigma^{-1}) \tilde{\mathbf{y}} \right\}},$$

$$\propto |\det(\Sigma)|^{\frac{c}{2}} \cdot e^{-\frac{1}{2} \text{trace}[\Sigma^{-1} S]}. \tag{B.74}$$

$$\cdot |\det(\Sigma)|^{-\frac{T}{2}} \cdot e^{-\frac{1}{2} \left\{ \boldsymbol{\pi}' (W W' \otimes \Sigma^{-1}) \boldsymbol{\pi} - 2 \boldsymbol{\pi}' (W \otimes \Sigma^{-1}) \tilde{\mathbf{y}} + \tilde{\mathbf{y}}' (I_T \otimes \Sigma^{-1}) \tilde{\mathbf{y}} \right\}},$$

$$\propto |\det(\Sigma)|^{-\frac{T-c}{2}} \cdot e^{-\frac{1}{2} \text{trace}[\Sigma^{-1} (S + Y Y' + \Pi W W' \Pi' - 2 Y W' \Pi')]}, \tag{B.75}$$

$$\propto |\det(\Sigma)|^{-\frac{(T-c-k-1)+k+1}{2}} \cdot e^{-\frac{1}{2} \text{trace}[\Sigma^{-1} (S + (Y - \Pi W)(Y - \Pi W)')]} \tag{B.76}$$

It then follows that the prior beliefs in equations (B.56)-(B.60) imply

$$\tilde{p}_{NiWU,i}(\boldsymbol{\pi}, \Sigma, Q|Y) \propto \text{I}\{\boldsymbol{\pi}, \Sigma, Q\} \cdot p_{NiWU,i}(\boldsymbol{\pi}, \Sigma, Q|Y), \quad (\text{B.77})$$

$$p_{NiWU,i}(\boldsymbol{\pi}, \Sigma, Q|Y) = p_{NiWU,i}(\boldsymbol{\pi}|Y, \Sigma) \cdot p_{NiWU,i}(\Sigma|Y) \cdot p_{NiWU,i}(Q|Y, \boldsymbol{\pi}, \Sigma), \quad (\text{B.78})$$

$$p_{NiWU,i}(\boldsymbol{\pi}|Y, \Sigma) = \phi(\boldsymbol{\mu}_\pi^*, V_\pi^*), \quad (\text{B.79})$$

$$p_{NiWU,i}(\Sigma|Y, \Pi) = iW(S^*, d^*), \quad (\text{B.80})$$

$$p_{NiWU,i}(\Sigma|Y) \propto |\det(V_\pi^*)|^{\frac{1}{2}} \cdot |\det(\Sigma)|^{-\frac{T-c}{2}} \cdot e^{-\frac{1}{2} \left\{ \text{trace}[\Sigma^{-1}S] + \tilde{\mathbf{y}}'(I_T \otimes \Sigma^{-1})\tilde{\mathbf{y}} - \boldsymbol{\mu}_\pi^{*'} V_\pi^{*-1} \boldsymbol{\mu}_\pi^* \right\}}, \quad (\text{B.81})$$

$$p_{NiWU,i}(Q|Y, \boldsymbol{\pi}, \Sigma) = p_{NiWU,i}(Q) = U_{O(k)} \propto 1, \quad (\text{B.82})$$

$$V_\pi^* = (V_\pi^{-1} + WW' \otimes \Sigma^{-1})^{-1}, \quad (\text{B.83})$$

$$\boldsymbol{\mu}_\pi^* = V_\pi^* (V_\pi^{-1} \boldsymbol{\mu}_\pi + (WW' \otimes \Sigma^{-1}) \hat{\boldsymbol{\pi}}_T), \quad (\text{B.84})$$

$$d^* = T - c - k - 1, \quad (\text{B.85})$$

$$S^* = S + (Y - \Pi W)(Y - \Pi W)'. \quad (\text{B.86})$$

Contrary to [Section B.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

$$d^* \geq k, \quad (\text{B.87})$$

or equivalently, as long as

$$c \leq T - 2k - 1, \quad (\text{B.88})$$

$$T \geq c + 2k + 1. \quad (\text{B.89})$$

Contrary to [Section B.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 B.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:*  $p_{NiWU,i}(\boldsymbol{\pi}, \Sigma|Y)$

1. select a starting value for  $\Sigma^{(0)}$ ;
2. draw  $\boldsymbol{\pi}^{(d)}$  from a  $N(\boldsymbol{\mu}_\pi^*, V_\pi^*)$ , equation [\(B.79\)](#), using  $(\boldsymbol{\mu}_\pi^*, V_\pi^*)$  associated with  $\Sigma^{(d-1)}$ ;
3. draw  $\Sigma^{(d)}$  from an  $iW(S^*, d^*)$ , equation [\(B.80\)](#), using  $S^*$  associated with  $\boldsymbol{\pi}^{(d)}$ ;
4. if the  $n_2$  burn-in draws have been generated, store  $(\boldsymbol{\pi}^{(d)}, \Sigma^{(d)})$ ;
5. repeat Steps 2-4 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:*  $\tilde{p}_{NiWU,i}(\boldsymbol{\pi}, \Sigma, Q|Y)$

1. select some starting candidate value for  $\Sigma^{(0)}$ ;
2. draw  $\boldsymbol{\pi}^{(c)}$  from a  $N(\boldsymbol{\mu}_\pi^*, V_\pi^*)$ , equation [\(B.79\)](#), using  $(\boldsymbol{\mu}_\pi^*, V_\pi^*)$  associated with  $\Sigma^{(d-1)}$ ;
3. draw  $\Sigma^{(c)}$  from an  $iW(S^*, d^*)$ , equation [\(B.80\)](#), using  $S^*$  associated with  $\boldsymbol{\pi}^{(c)}$ ;
4. draw  $Q^{(c)}$  from the (Haar)uniform distribution using the algorithm by [Rubio-Ramirez et al. \(2010\)](#);
5. compute the structural statistics of interest associated with  $(\boldsymbol{\pi}^{(c)}, \Sigma^{(c)}, Q^{(c)})$ ;

5a. if the restrictions are satisfied, set

$$(\boldsymbol{\pi}^{(d)}, \Sigma^{(d)}, Q^{(d)}) = (\boldsymbol{\pi}^{(c)}, \Sigma^{(c)}, Q^{(c)}); \quad (\text{B.90})$$

5b. if the restrictions are not satisfied, move back to Step 2;

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

7. repeat Steps 2-5 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  $\Sigma$ .

## C 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.

### C.1 Derivations

We first provide the derivations for the more general case of the priors from equations (9a)-(9d) in the paper, and then discuss two common special cases.

#### C.1.1 General case

Start from the prior distribution

$$\tilde{p}(\boldsymbol{\pi}, B) \propto \text{I}\{\boldsymbol{\pi}, B\} \cdot p(\boldsymbol{\pi}, B), \quad (\text{C.91})$$

$$\propto \text{I}\{\boldsymbol{\pi}, B\} \cdot p(\boldsymbol{\pi}|B) \cdot p(B), \quad (\text{C.92})$$

with

$$p(\boldsymbol{\pi}|B) = \phi(\boldsymbol{\mu}_\pi, V_\pi), \quad (\text{C.93})$$

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

$$\tilde{p}(\boldsymbol{\pi}, B|Y) \propto \text{I}\{\boldsymbol{\pi}, B\} \cdot p(\boldsymbol{\pi}, B|Y), \quad (\text{C.94})$$

$$\propto \text{I}\{\boldsymbol{\pi}, B\} \cdot p(\boldsymbol{\pi}, B) \cdot p(Y|\boldsymbol{\pi}, B),$$

$$\propto \text{I}\{\boldsymbol{\pi}, B\} \cdot p(B) \cdot$$

$$\cdot (2\pi)^{-\frac{mk}{2}} |\det(V_\pi)|^{-\frac{1}{2}} e^{-\frac{1}{2}(\boldsymbol{\pi}-\boldsymbol{\mu}_\pi)'V_\pi^{-1}(\boldsymbol{\pi}-\boldsymbol{\mu}_\pi)}. \quad (\text{C.95})$$

$$\cdot (2\pi)^{-\frac{kT}{2}} |\det(B)|^{-T} e^{-\frac{1}{2}(\tilde{\mathbf{y}}-Z\boldsymbol{\pi})'(I_T \otimes BB')^{-1}(\tilde{\mathbf{y}}-Z\boldsymbol{\pi})}.$$

As is standard in the literature, we rewrite equation (C.95) as

$$\tilde{p}(\boldsymbol{\pi}, B|Y) \propto \text{I}\{\boldsymbol{\pi}, B\} \cdot p(\boldsymbol{\pi}, B|Y), \quad (\text{C.96})$$

$$\propto \text{I}\{\boldsymbol{\pi}, B\} \cdot p(\boldsymbol{\pi}|Y, B) \cdot p(B|Y), \quad (\text{C.97})$$

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

$$\Psi(B) = (I_T \otimes BB'), \quad (\text{C.98})$$

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

$$p(\boldsymbol{\pi}, B|Y) = \kappa \cdot e^{-\frac{1}{2}[(\boldsymbol{\pi}-\boldsymbol{\mu}_\pi)'V_\pi^{-1}(\boldsymbol{\pi}-\boldsymbol{\mu}_\pi) + (\tilde{\mathbf{y}}-Z\boldsymbol{\pi})'\Psi(B)^{-1}(\tilde{\mathbf{y}}-Z\boldsymbol{\pi})]}, \quad (\text{C.99})$$



with  $\kappa$  a term that includes elements which are not a function of  $\boldsymbol{\pi}$ . As done also in [Section B.2](#), factorize the terms in the exponent of (C.99) as

$$\begin{aligned}
& (\boldsymbol{\pi} - \boldsymbol{\mu}_\pi)' V_\pi^{-1} (\boldsymbol{\pi} - \boldsymbol{\mu}_\pi) + (\tilde{\mathbf{y}} - Z\boldsymbol{\pi})' \Psi(B)^{-1} (\tilde{\mathbf{y}} - Z\boldsymbol{\pi}) = \\
& = \boldsymbol{\pi}' V_\pi^{-1} \boldsymbol{\pi} - 2\boldsymbol{\pi}' V_\pi^{-1} \boldsymbol{\mu}_\pi + \boldsymbol{\mu}_\pi' V_\pi^{-1} \boldsymbol{\mu}_\pi + \\
& + \tilde{\mathbf{y}}' \Psi(B)^{-1} \tilde{\mathbf{y}} - 2\boldsymbol{\pi}' Z' \Psi(B)^{-1} \tilde{\mathbf{y}} + \boldsymbol{\pi}' Z' \Psi(B)^{-1} Z \boldsymbol{\pi} = \quad (\text{C.100}) \\
& = \boldsymbol{\pi}' [V_\pi^{-1} + Z' \Psi(B)^{-1} Z] \boldsymbol{\pi} - 2\boldsymbol{\pi}' [V_\pi^{-1} \boldsymbol{\mu}_\pi + Z' \Psi(B)^{-1} \tilde{\mathbf{y}}] + \\
& + \tilde{\mathbf{y}}' \Psi(B)^{-1} \tilde{\mathbf{y}} + \boldsymbol{\mu}_\pi' V_\pi^{-1} \boldsymbol{\mu}_\pi = \\
& = (\boldsymbol{\pi} - \boldsymbol{\mu}_\pi^*)' V_\pi^{*-1} (\boldsymbol{\pi} - \boldsymbol{\mu}_\pi^*) + \tilde{\mathbf{y}}' \Psi(B)^{-1} \tilde{\mathbf{y}} + \boldsymbol{\mu}_\pi' V_\pi^{-1} \boldsymbol{\mu}_\pi - \boldsymbol{\mu}_\pi^{*'} V_\pi^{*-1} \boldsymbol{\mu}_\pi^*,
\end{aligned}$$

with

$$V_\pi^* = [V_\pi^{-1} + Z' \Psi(B)^{-1} Z]^{-1}, \quad (\text{C.101})$$

$$= \left( V_\pi^{-1} + [WW' \otimes (BB')^{-1}] \right)^{-1}, \quad (\text{C.102})$$

$$\boldsymbol{\mu}_\pi^* = V_\pi^* \cdot [V_\pi^{-1} \boldsymbol{\mu}_\pi + Z' \Psi(B)^{-1} \tilde{\mathbf{y}}], \quad (\text{C.103})$$

$$= V_\pi^* \cdot \left( V_\pi^{-1} \boldsymbol{\mu}_\pi + [W \otimes (BB')^{-1}] \tilde{\mathbf{y}} \right), \quad (\text{C.104})$$

$$= V_\pi^* \cdot \left( V_\pi^{-1} \boldsymbol{\mu}_\pi + [WW' \otimes (BB')^{-1}] \hat{\boldsymbol{\pi}}_T \right). \quad (\text{C.105})$$

The joint posterior distribution can now be written as

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

It follows that

$$\tilde{p}(\boldsymbol{\pi}, B|Y) \propto \mathbf{I}\{\boldsymbol{\pi}, B\} p(\boldsymbol{\pi}|Y, B) \cdot p(B|Y), \quad (\text{C.106})$$

$$p(\boldsymbol{\pi}|B, Y) = \phi(\boldsymbol{\mu}_\pi^*, V_\pi^*), \quad (\text{C.107})$$

$$\begin{aligned} p(B|Y) &\propto p(B) \cdot |\det(B)|^{-T} \cdot |\det(V_\pi)|^{-\frac{1}{2}} \cdot |\det(V_\pi^*)|^{\frac{1}{2}} \\ &\cdot e^{-\frac{1}{2} \left\{ \tilde{\mathbf{y}}' (I_T \otimes (BB')^{-1}) \tilde{\mathbf{y}} - \boldsymbol{\mu}_\pi^{*'} V_\pi^{*-1} \boldsymbol{\mu}_\pi^* + \boldsymbol{\mu}_\pi' V_\pi^{-1} \boldsymbol{\mu}_\pi \right\}}. \end{aligned} \quad (\text{C.108})$$

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

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

$$B = f_1(\Sigma, Q) = h(\Sigma)Q, \quad (\text{C.109})$$

$$\Sigma = f_2(B) = BB', \quad (\text{C.110})$$

$$Q = f_3(B) = h(f_2(B))^{-1}B, \quad (\text{C.111})$$

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

$$p_{\Sigma, Q}(\Sigma, Q) = v_{\{B \rightarrow \Sigma, Q\}} \cdot p_B(f_1(\Sigma, Q)), \quad (\text{C.112})$$

$$p_B(B) = v_{\{\Sigma, Q \rightarrow B\}} \cdot p_{\Sigma, Q}(f_2(B), f_3(B)), \quad (\text{C.113})$$

with  $v_{\{\cdot \rightarrow \cdot\}}$  the volume elements in the mapping from  $B$  to  $(\Sigma, Q)$  or vice versa (see [Casella and Berger, 2021](#) for the definition of the volume element). We compute  $v_{\{B \rightarrow \Sigma, Q\}}$  and  $v_{\{\Sigma, Q \rightarrow B\}}$  using the results in [Bibby et al. \(1979\)](#) (Chapter 2) and [Mathai](#)

and Haubold (2008) (Chapter 11):<sup>1</sup>

$$v_{\{B \rightarrow \Sigma, Q\}} = |\det(\Sigma)|^{-\frac{1}{2}}, \quad (\text{C.114})$$

$$v_{\{\Sigma, Q \rightarrow B\}} = |\det(B)|. \quad (\text{C.115})$$

We can then compute the joint and marginal posterior distributions on  $(\Sigma, Q)$  implied by equation (C.108).

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

$$p_{\Sigma, Q}(\Sigma, Q|Y) \propto v_{\{B \rightarrow \Sigma, Q\}} \cdot p_B(f_1(\Sigma, Q)|Y), \quad (\text{C.116})$$

$$\propto |\det(\Sigma)|^{-\frac{1}{2}} \cdot |\det(\Sigma)|^{-\frac{T}{2}} \cdot |\det(V_\pi)|^{-\frac{1}{2}} \cdot |\det(V_\pi^*)|^{\frac{1}{2}}. \quad (\text{C.117})$$

$$\cdot e^{-\frac{1}{2} \left\{ \tilde{\mathbf{y}}' (I_T \otimes \Sigma^{-1}) \tilde{\mathbf{y}} - \boldsymbol{\mu}_\pi^{*'} V_\pi^{*-1} \boldsymbol{\mu}_\pi^* + \boldsymbol{\mu}_\pi' V_\pi^{-1} \boldsymbol{\mu}_\pi \right\}} \cdot p_B(f_1(\Sigma, Q)),$$

$$\propto |\det(\Sigma)|^{-\frac{T+1}{2}} \cdot |\det(V_\pi)|^{-\frac{1}{2}} \cdot |\det(V_\pi^*)|^{\frac{1}{2}}. \quad (\text{C.118})$$

$$\cdot e^{-\frac{1}{2} \left\{ \tilde{\mathbf{y}}' (I_T \otimes \Sigma^{-1}) \tilde{\mathbf{y}} - \boldsymbol{\mu}_\pi^{*'} V_\pi^{*-1} \boldsymbol{\mu}_\pi^* + \boldsymbol{\mu}_\pi' V_\pi^{-1} \boldsymbol{\mu}_\pi \right\}} \cdot p_B(h(\Sigma)Q),$$

$$p_\Sigma(\Sigma|Y) \propto \int_{O(k)} p_{\Sigma, Q}(\Sigma, Q|Y) dQ, \quad (\text{C.119})$$

$$\propto |\det(\Sigma)|^{-\frac{T+1}{2}} \cdot |\det(V_\pi)|^{-\frac{1}{2}} \cdot |\det(V_\pi^*)|^{\frac{1}{2}} \cdot e^{-\frac{1}{2} \left\{ \tilde{\mathbf{y}}' (I_T \otimes \Sigma^{-1}) \tilde{\mathbf{y}} - \boldsymbol{\mu}_\pi^{*'} V_\pi^{*-1} \boldsymbol{\mu}_\pi^* + \boldsymbol{\mu}_\pi' V_\pi^{-1} \boldsymbol{\mu}_\pi \right\}}. \quad (\text{C.120})$$

$$\cdot \int_{O(k)} p_B(h(\Sigma)Q) dQ,$$

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  $\Sigma$  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  $\Sigma$ . For a given value of  $\Sigma_d$ , we approximate the integral using the following simulation-based method:

---

<sup>1</sup>When zero restrictions are introduced, the volume elements must be evaluated numerically, see Arias et al. (2018).

*Algorithm E:*

1. draw a matrix  $Q_c$  from the algorithm by [Rubio-Ramirez et al. \(2010\)](#);
2. compute  $B_c = h(\Sigma_d)Q_c$ ;
  - 2a. if  $B_c$  satisfies the sign restrictions on  $B$ , proceed to Step 3;
  - 2b. if  $B_c$  does not satisfy the sign restrictions on  $B$ , move back to Step 1;
3. store  $Q_i = Q_c$ ;
4. repeat Steps 1-4 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.
5. compute

$$\int_{O(k)} p_B(h(\Sigma_d)Q) dQ \approx \frac{\sum_{i=1}^{m_2} p_B(h(\Sigma_d)Q_i)}{m_3(\Sigma_d)}. \quad (\text{C.121})$$

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

$$p_{\Sigma,Q}(\Sigma, Q) = v_{\{B \rightarrow \Sigma, Q\}} \cdot p_B(h(\Sigma)Q), \quad (\text{C.122})$$

$$p_{\Sigma}(\Sigma) = \int_{O(k)} v_{\{B \rightarrow \Sigma, Q\}} \cdot p_B(h(\Sigma)Q) dQ, \quad (\text{C.123})$$

$$p_{Q|\Sigma}(Q|\Sigma) = \frac{p_{\Sigma,Q}(\Sigma, Q)}{\tilde{p}_{\Sigma}(\Sigma)}, \quad (\text{C.124})$$

$$= \frac{v_{\{B \rightarrow \Sigma, Q\}} \cdot p_B(h(\Sigma)Q)}{\int_{O(k)} v_{\{B \rightarrow \Sigma, Q\}} \cdot p_B(h(\Sigma)Q) dQ}, \quad (\text{C.125})$$

$$= \frac{p_B(h(\Sigma)Q)}{\int_{O(k)} p_B(h(\Sigma)Q) dQ}, \quad (\text{C.126})$$

and the same holds for the posterior,

$$p_{Q|Y,\Sigma}(Q|Y, \Sigma) = \frac{p_{\Sigma,Q|Y}(\Sigma, Q|Y)}{p_{\Sigma|Y}(\Sigma|Y)}, \quad (\text{C.127})$$

$$= \frac{v_{\{B \rightarrow \Sigma, Q\}} \cdot p_B(h(\Sigma)Q)}{\int_{O(k)} v_{\{B \rightarrow \Sigma, Q\}} \cdot p_B(h(\Sigma)Q) dQ}, \quad (\text{C.128})$$

$$= \frac{p_B(h(\Sigma)Q)}{\int_{O(k)} p_B(h(\Sigma)Q) dQ}, \quad (\text{C.129})$$

$$= p_{Q|\Sigma}(Q|\Sigma). \quad (\text{C.130})$$

To appreciate the usefulness of this result (C.130) for our sampler, consider two matrices  $(B_1, B_2)$  that imply the same matrix  $\Sigma$ ,

$$B_1 = h(\Sigma)Q_1, \quad (\text{C.131})$$

$$B_2 = h(\Sigma)Q_2. \quad (\text{C.132})$$

Equations (C.126) and (C.129) imply that

$$\frac{p_{Q|\Sigma}(Q_1|\Sigma)}{p_{Q|\Sigma}(Q_2|\Sigma)} = \frac{p_{Q|Y,\Sigma}(Q_1|Y, \Sigma)}{p_{Q|Y,\Sigma}(Q_2|Y, \Sigma)} = \frac{p_B(h(\Sigma)Q_1)}{p_B(h(\Sigma)Q_2)} = \frac{p(B_1)}{p(B_2)}. \quad (\text{C.133})$$

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  $\Sigma$  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  $\Sigma$ , if the posterior favours  $B_1$  over  $B_2$  defined in equations (C.131)-(C.132), it is because the prior does the same by the same proportion, i.e.  $\frac{p(B_2|Y)}{p(B_1|Y)} = \frac{p(B_2)}{p(B_1)}$ . 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:*

1. select the conditioning value  $\bar{\Sigma}$  and compute the Cholesky decomposition  $B_c = h(\bar{\Sigma})$ ;
2. draw  $Q^{(d)}$  from the (Haar)uniform distribution using the algorithm by [Rubio-Ramirez et al. \(2010\)](#);
3. 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;
  - 3a. if the sign restrictions are satisfied, proceed to Step 4;
  - 3b. if the sign restrictions are not satisfied, move back to Step 2;
4. compute  $\theta^{(d)} = \text{atan}\left(\frac{Q_{21}^{(d)}}{Q_{11}^{(d)}}\right)$ , with  $Q_{ij}^{(d)}$  the  $(i, j)$  entry of  $Q_{11}^{(d)}$  and with  $\text{atan}$  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

$$Q(\theta) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}; \quad (\text{C.134})$$

5. compute  $w^{(d)}$  as the prior  $p(B)$  evaluated in  $B^{(d)}$ ;
6. store  $(\theta^{(d)}, w^{(d)})$ ;
7. repeat Steps 2-6 until a desired number of draws  $n_4$  is stored;
8. resample  $\{\theta^{(d)}\}_{d=1}^{n_4}$  with replacement using weights  $\{w^{(d)}\}_{d=1}^{n_4}$ .

### C.1.2 Case 1: flat prior on $\pi$

The first special case used in the paper (equation 16a) consists of the limit case in which the prior distribution  $p(\pi|B)$  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 et al., 1984](#) [Del Negro and Schorfheide, 2011](#)). Under the prior

$$\tilde{p}(\boldsymbol{\pi}, B) = \mathbb{I}\{\boldsymbol{\pi}, B\} \cdot p(\boldsymbol{\pi}, B), \quad (\text{C.135})$$

$$p(\boldsymbol{\pi}, B) = p(\boldsymbol{\pi}) \cdot p(B), \quad (\text{C.136})$$

$$p(\boldsymbol{\pi}) \propto 1, \quad (\text{C.137})$$

$$p(B), \quad (\text{C.138})$$

the key results from [Section C.1.1](#) simplify to

$$\tilde{p}(\boldsymbol{\pi}, B|Y) \propto \mathbb{I}\{\boldsymbol{\pi}, B\} \cdot p(\boldsymbol{\pi}|B, Y) \cdot p(B|Y), \quad (\text{C.139})$$

$$p(\boldsymbol{\pi}|B, Y) = \phi(\boldsymbol{\mu}_\pi^*, V_\pi^*), \quad (\text{C.140})$$

$$p(B|Y) \propto p(B) \cdot |\det(B)|^{-(T-m)} \cdot e^{-\frac{1}{2}\text{trace}[(BB')^{-1}\hat{\Sigma}_T(T-m)]}, \quad (\text{C.141})$$

$$p(\Sigma|Y) \propto v_{\{B \rightarrow \Sigma, Q\}} \cdot |\det(\Sigma)|^{-\frac{T-m}{2}} \cdot e^{-\frac{1}{2}\text{trace}[\Sigma^{-1}\hat{\Sigma}_T(T-m)]} \cdot \int_{O(k)} p_B(h(\Sigma)Q) dQ, \quad (\text{C.142})$$

$$\propto |\det(\Sigma)|^{-\frac{T-m+1}{2}} \cdot e^{-\frac{1}{2}\text{trace}[\Sigma^{-1}\hat{\Sigma}_T(T-m)]} \cdot \int_{O(k)} p_B(h(\Sigma)Q) dQ, \quad (\text{C.143})$$

$$V_\pi^* = (WW')^{-1} \otimes BB', \quad (\text{C.144})$$

$$\boldsymbol{\mu}_\pi^* = \hat{\boldsymbol{\pi}}_T. \quad (\text{C.145})$$

We use equation [\(C.143\)](#) to compute weights  $w(\Sigma^{(d)})^{\text{stage A}}$  in the Appendix of the paper.

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

$$\begin{aligned} \log[p(\Sigma|Y)] \propto & \frac{T-m+1}{2} \cdot \log(|\det(\Sigma^{-1})|) - \frac{1}{2} \text{trace}[\Sigma^{-1} \hat{\Sigma}_T(T-m)] + \\ & + \log\left(\int_{O(k)} p_B(h(\Sigma)Q) dQ\right), \end{aligned} \quad (\text{C.146})$$

$$\frac{d \log[p(\Sigma|Y)]}{d \Sigma^{-1}} = \frac{T-m+1}{2} \Sigma - \frac{1}{2} \hat{\Sigma}_T(T-m) + \frac{d}{d \Sigma^{-1}} \log\left(\int_{O(k)} p_B(h(\Sigma)Q) dQ\right) = 0, \quad (\text{C.147})$$

hence the mode is implicitly defined by

$$\Sigma = \frac{T-m}{T-m+1} \hat{\Sigma}_T - \frac{2}{T-m+1} \frac{d}{d \Sigma^{-1}} \log\left(\int_{O(k)} p_B(h(\Sigma)Q) dQ\right), \quad (\text{C.148})$$

a result used in footnote 2 of the paper.

### C.1.3 Case 2: prior independence

The second special case used in the paper (equation 16b) 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 modelled through dummy variables.

Under the prior

$$\tilde{p}(\boldsymbol{\pi}, B) \propto \mathbb{I}\{\boldsymbol{\pi}, B\} \cdot p(\boldsymbol{\pi}, B), \quad (\text{C.149})$$

$$p(\boldsymbol{\pi}, B) = p(\boldsymbol{\pi}) \cdot p(B), \quad (\text{C.150})$$

$$p(\boldsymbol{\pi}) = \phi(\boldsymbol{\mu}_\pi, V_\pi), \quad (\text{C.151})$$

$$p(B), \quad (\text{C.152})$$



the key results from [Section C.1.1](#) simplify to

$$\tilde{p}(\boldsymbol{\pi}, B|Y) \propto p(\boldsymbol{\pi}|B, Y) \cdot p(B|Y) \cdot \mathbf{I}(\boldsymbol{\pi}, B), \quad (\text{C.153})$$

$$p(\boldsymbol{\pi}|B, Y) = \phi(\boldsymbol{\mu}_\pi^*, V_\pi^*), \quad (\text{C.154})$$

$$p(B|Y) \propto p(B) \cdot |\det(B)|^{-T} \cdot |\det(V_\pi^*)|^{\frac{1}{2}} \cdot e^{-\frac{1}{2} \left\{ \tilde{\mathbf{y}}' (I_T \otimes (BB')^{-1}) \tilde{\mathbf{y}} - \boldsymbol{\mu}_\pi^{*\prime} V_\pi^{*-1} \boldsymbol{\mu}_\pi^* \right\}}, \quad (\text{C.155})$$

$$p(\Sigma|Y) \propto v_{\{B \rightarrow \Sigma, Q\}} \cdot |\det(\Sigma)|^{-\frac{T}{2}} \cdot |\det(V_\pi^*)|^{\frac{1}{2}} \cdot e^{-\frac{1}{2} \left\{ \tilde{\mathbf{y}}' (I_T \otimes \Sigma^{-1}) \tilde{\mathbf{y}} - \boldsymbol{\mu}_\pi^{*\prime} V_\pi^{*-1} \boldsymbol{\mu}_\pi^* \right\}} \cdot \int_{O(k)} p_B(h(\Sigma)Q) dQ, \quad (\text{C.156})$$

$$\propto |\det(\Sigma)|^{-\frac{T+1}{2}} \cdot |\det(V_\pi^*)|^{\frac{1}{2}} \cdot e^{-\frac{1}{2} \left\{ \tilde{\mathbf{y}}' (I_T \otimes \Sigma^{-1}) \tilde{\mathbf{y}} - \boldsymbol{\mu}_\pi^{*\prime} V_\pi^{*-1} \boldsymbol{\mu}_\pi^* \right\}} \cdot \int_{O(k)} p_B(h(\Sigma)Q) dQ, \quad (\text{C.157})$$

$$V_\pi^* = (V_\pi^{-1} + [WW' \otimes (BB')^{-1}])^{-1}, \quad (\text{C.158})$$

$$\boldsymbol{\mu}_\pi^* = V_\pi^* \cdot (V_\pi^{-1} \boldsymbol{\mu}_\pi + [WW' \otimes (BB')^{-1}] \hat{\boldsymbol{\pi}}_T). \quad (\text{C.159})$$

We use equation [\(C.157\)](#) to compute weights  $w(\Sigma^{(d)})^{\text{stage A}}$  in the Appendix of the paper.

## C.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 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. 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 et al. \(2009\)](#). We discuss these approaches in the rest of this section, and refer to [Figure F3](#) and [Table F6](#) for the results of the diagnostics associated with the application in the paper.

### C.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 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.

### C.2.2 Diagnostic tests

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

$$H_0 : \text{the variance is finite} \tag{C.160}$$

against

$$H_1 : \text{the variance is unbounded.} \tag{C.161}$$

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

$$Z_i = w_i - u \quad \text{if} \quad w_i > u. \tag{C.162}$$

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

$$f(z|\xi, \beta) = \frac{1}{\beta} \left(1 + \xi \frac{z}{\beta}\right)^{-\frac{1}{\xi}-1}. \quad (\text{C.163})$$

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.  $w_1 \leq w_{(2)} \leq \dots \leq w_{(N)}$ . To make the tests operational we use the null and alternative hypotheses

$$H_0 : \xi = 0.5, \quad (\text{C.164})$$

$$H_1 : \xi > 0.5. \quad (\text{C.165})$$

### Wald test

The log likelihood function based on the generalised Pareto distribution in [\(C.163\)](#) equals

$$\log f(z|\beta, \xi) = -n \cdot \log(\beta) - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^n \log(1 + \xi \beta^{-1} z_i). \quad (\text{C.166})$$

To construct the test statistic, we need to numerically maximise [\(C.166\)](#) 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 [\(C.166\)](#) to obtain  $\beta^{MLE}$  and  $\xi^{MLE}$ ;

2. construct the test statistic as

$$t = \sqrt{\frac{n}{3\hat{\beta}^{MLE}}} \left( \hat{\xi}^{MLE} - \frac{1}{2} \right). \quad (\text{C.167})$$

The statistic  $t$  has an approximate standard normal distribution for  $n \rightarrow \infty$ . 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 (C.166) under the null hypothesis of  $\xi = 0.5$ . We follow these steps:

1. numerically maximise (C.166) 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 (C.166) with respect to  $\xi$  and set  $\xi = 0.5$  and  $\beta = \beta^{MLE,r}$ . This gives <sup>2</sup>

$$\hat{s}_r^\xi = 4 \sum_{i=1}^n \log\left(1 + \frac{z_i}{2\beta^{MLE,r}}\right) - 6 \sum_{i=1}^n \frac{z_i}{2\beta^{MLE,r} + z_i}; \quad (\text{C.168})$$

3. compute

$$s_*^\xi = \frac{1}{\sqrt{2n}} \hat{s}_r^\xi. \quad (\text{C.169})$$

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

---

<sup>2</sup>There is a typing error in Koopman (page 4) in the corresponding equation.

## D The Dynamic Striated Metropolis-Hastings algorithm by Waggoner et al. (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 et al. (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 (11d) 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 D.1 briefly discusses the DSMH sampler, while Section D.2 discusses the convergence criteria that we use to assess the performance of the posterior chains.

### D.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

$$f_\lambda(\boldsymbol{\theta}) = f^s(\boldsymbol{\theta}) \cdot \left(f^i(\boldsymbol{\theta})\right)^\lambda \cdot \left(f^d(\boldsymbol{\theta})\right)^{1-\lambda}, \quad (\text{D.170})$$

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  $\lambda$  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 et al. (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, one could depart from Waggoner et al. (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.<sup>3</sup> 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 (C.141). One could sample  $p(B|Y)$  by setting

$$f^s(\boldsymbol{\theta}) = p(B), \quad (\text{D.171})$$

$$f^i(\boldsymbol{\theta}) = |\det(B)|^{-(T-m)} \cdot e^{-\frac{1}{2}[(BB')^{-1}\hat{\Sigma}_T(T-m)]}, \quad (\text{D.172})$$

$$f^d(\boldsymbol{\theta}) = 1. \quad (\text{D.173})$$

This specification follows Waggoner et al. (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 B.1 due to the accept-reject part of the algorithm, the following modification of *Algorithm A*,

*Algorithm A-bis:*

1. draw  $\Sigma^{(d)}$  from  $iW(S^*, d^*)$ , equation (B.43);

---

<sup>3</sup>A similar idea is explored by Mlikota and Schorfheide (2022), who, like us, implement sequential sampling starting from a posterior distribution of a convenient form rather than from the prior.

2. draw  $\boldsymbol{\pi}^{(d)}$  from  $N(\boldsymbol{\mu}_\pi^*, V_\pi^*)$ , equation (B.44), using  $V_\pi^*$  associated with  $\Sigma^{(d)}$ ;
3. draw  $Q^{(d)}$  using the algorithm by Rubio-Ramirez et al. (2010);
4. store  $(\boldsymbol{\pi}^{(d)}, \Sigma^{(d)}, Q^{(d)})$ ;
5. repeat Steps 1-4 until a desired number of draws  $n_1$  is stored,

delivers draws from the joint distribution

$$\begin{aligned}
p(B|Y) &\propto v_{\Sigma, Q \rightarrow B} \cdot p(\Sigma|Y) \cdot 1, \\
&\propto |\det(B)| \cdot |\det(BB')|^{-\frac{T-m-c}{2}} \cdot e^{-\frac{1}{2}[(BB')^{-1}(S+\hat{\Sigma}_T(T-m))]}, \\
&\propto |\det(B)|^{-(T-m-c-1)} \cdot e^{-\frac{1}{2}[(BB')^{-1}(S+\hat{\Sigma}_T(T-m))]}. \tag{D.174}
\end{aligned}$$

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 (D.174) is very similar to equation (D.172), suggesting to start the DSMH from draws from *Algorithm A-bis* and then setting

$$f^s(\boldsymbol{\theta}) = |\det(B)|^{-(T-m)} \cdot e^{-\frac{1}{2}[(BB')^{-1}\hat{\Sigma}_T(T-m)]}, \tag{D.175}$$

$$f^i(\boldsymbol{\theta}) = p(B), \tag{D.176}$$

$$f^d(\boldsymbol{\theta}) = 1. \tag{D.177}$$

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

$$\log[f_{\lambda_h}(\boldsymbol{\theta})] = \log[f^s(\boldsymbol{\theta})] + \lambda_h \cdot \log[f^i(\boldsymbol{\theta})] + (1 - \lambda_h) \cdot \log[f^d(\boldsymbol{\theta})], \tag{D.178}$$

as the logarithm of the tempered function  $f_{\lambda_h}(\boldsymbol{\theta})$  at stage  $h$ , with  $h = 1, \dots, H$  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 et al. \(2016\)](#) propose sampling  $\log[f_{\lambda_h}(\boldsymbol{\theta})]$  using a modified Metropolis-Hastings algorithm.<sup>4</sup> In the general stage  $h$ , the algorithm can be summarized in the following steps:

1. 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  $h - 1$ ;
2. 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  $\log[f^i(\boldsymbol{\theta})]$ ;
3. 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,  $\omega_d = \frac{\tilde{\omega}_d}{\sum_{d=1}^{N \cdot G} \tilde{\omega}_d}$ ,  $d = 1, \dots, N \cdot G$  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;
4. use  $\{\boldsymbol{\theta}_d^{(h-1)}, \omega_d\}_{d=1}^{N \cdot G}$  to compute numerically the variance  $\Omega_h$  of the target function at stage  $h$ ;
5. explore the target function  $\log[f_{\lambda_h}(\boldsymbol{\theta})]$  as follows. For each group  $g$ , set the initial draw  $\boldsymbol{\theta}_{old}$  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  $\boldsymbol{\theta}_{new} = \boldsymbol{\theta}_{old} + \boldsymbol{\theta}_{shock}$  with  $\boldsymbol{\theta}_{shock}$  a multivariate zero-mean normal random variable with variance  $c_h \cdot \Omega_h$ , while with probability  $1 - p$  set  $\boldsymbol{\theta}_{new}$  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  $\boldsymbol{\theta}_{old}$ . Accept  $\boldsymbol{\theta}_{new}$  with probability  $\min\left\{1, \frac{f_{\lambda_h}(\boldsymbol{\theta}_{new})}{f_{\lambda_h}(\boldsymbol{\theta}_{old})}\right\}$  if  $\boldsymbol{\theta}_{new}$  was generated from the random walk extraction, and with probability  $\min\left\{1, \frac{f_{\lambda_h}(\boldsymbol{\theta}_{new})}{f_{\lambda_{h-1}}(\boldsymbol{\theta}_{old})} \frac{f_{\lambda_h}(\boldsymbol{\theta}_{old})}{f_{\lambda_{h-1}}(\boldsymbol{\theta}_{new})}\right\}$

---

<sup>4</sup>See, for example, [Herbst and Schorfheide \(2014\)](#) for an alternative approach to sequential Monte Carlo samplers.



if  $\theta_{new}$  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.

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, 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 et al. (2016) and set the number of stages  $H$  at 50, using their suggested progression for the tempering parameter  $\lambda_h$ . 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  $\tau$  at which draws stored, and the number of striations  $M$ . We set  $(G, K, M, \tau)$  as indicated in Table F1. 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  $\theta$ . For each application, multiple convergence diagnostics are used in each of the 50 stages of the sampler. The diagnostics results are not reported.

## D.2 Convergence criteria used for the DSMH algorithm

Consider the chain  $\{\theta_d\}_{d=1}^{N \cdot G}$ , with  $\theta_d = (\theta_{1,d}, \dots, \theta_{j,d}, \dots, \theta_{\kappa,d})'$  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 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  $\boldsymbol{\theta}_d$ ). 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  $\{\theta_{j,d}\}_{d=1}^N, \{\theta_{j,d}\}_{d=N+1}^{2N}, \dots, \{\theta_{j,d}\}_{d=(G-1)N+1}^{N \cdot G}$  for univariate chains, and  $\{\boldsymbol{\theta}_d\}_{d=1}^{N \cdot G}$  or  $\{\boldsymbol{\theta}_d\}_{d=1}^N, \{\boldsymbol{\theta}_d\}_{d=N+1}^{2N}, \dots, \{\boldsymbol{\theta}_d\}_{d=(G-1)N+1}^{N \cdot G}$  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:

1. extract the first 10% and the last 40% of the draws of  $\{\theta_{j,d}\}_{d=1}^{N \cdot G}$ , i.e.  $\{\theta_{j,d}\}_{d=1}^{0.10 \cdot N \cdot G}$  and  $\{\theta_{j,d}\}_{d=0.60 \cdot N \cdot G}^{N \cdot G}$ ;
2. for each subseries, compute the mean and the standard deviation and call them  $\hat{\mu}_{first}$ ,  $\hat{\mu}_{last}$ ,  $\hat{\Sigma}_{first}$  and  $\hat{\sigma}_{last}$ ;

3. compute the test statistic

$$CD = \frac{\hat{\mu}_{first} - \hat{\mu}_{last}}{\frac{\hat{\sigma}_{first}}{\sqrt{0.1NG}} + \frac{\hat{\sigma}_{last}}{\sqrt{0.4NG}}}. \quad (\text{D.179})$$

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.

### **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  $q = 0.025$ . If one expects 95% of the posterior draws to lie in an interval of  $\pm 0.0125$  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  $\{\theta_{j,d}\}_{d=1}^{N \cdot G}$  into a dichotomous random variable  $Z_d$ :

$$Z_d = \begin{cases} 1 & \text{if } \theta_{j,d} < q_{0.025}, \\ 0 & \text{otherwise;} \end{cases} \quad (\text{D.180})$$

2. 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

$$\pi_0 = P(\theta_{j,d} < q_{0.025}) = P(Z_d = 0) = \frac{\beta}{\alpha + \beta}, \quad (\text{D.181})$$

$$\pi_1 = P(\theta_{j,d} \geq q_{0.025}) = P(Z_d = 1) = 1 - \pi_0 = \frac{\alpha}{\alpha + \beta}; \quad (\text{D.182})$$

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

$$P(\theta_{j,d} < q_{0.025}) \approx \bar{Z}_{NG,j} = \frac{1}{NG} \sum_{d=1}^{NG} Z_d. \quad (\text{D.183})$$

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

$$n^* = \frac{(2 - \alpha - \beta)\alpha\beta}{(\alpha + \beta)^3} \left\{ \frac{\Phi^{-1}(\frac{1}{2}(s+1))}{r} \right\}^2. \quad (\text{D.184})$$

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,  $\{\theta_{j,d}\}_{d=1}^N, \{\theta_{j,d}\}_{d=N+1}^{2N}, \dots, \{\theta_{j,d}\}_{d=(G-1)N+1}^{N \cdot G}$ . 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

$$B = \frac{1}{G-1} \sum_{g=1}^G (\bar{\theta}_{j,\cdot,g} - \bar{\theta}_j)^2, \quad (\text{D.185})$$

where  $\bar{\theta}_{j,\cdot,g}$  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

$$W = \frac{1}{G(N-1)} \sum_{g=1}^G \sum_{n=1}^N (\theta_{j,n,g} - \bar{\theta}_{j,\cdot,g})^2; \quad (\text{D.186})$$

3. estimate the overall variance as

$$\hat{V} = \hat{\sigma}_+^2 + \frac{B}{G}, \quad (\text{D.187})$$

with

$$\hat{\sigma}_+^2 = \frac{N-1}{N} W + B; \quad (\text{D.188})$$

4. compute the statistic

$$\hat{R} = \frac{\hat{V}}{W}. \quad (\text{D.189})$$

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

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

$$D = \frac{1}{G-1} \sum_{g=1}^G (\bar{\boldsymbol{\theta}}_g - \bar{\boldsymbol{\theta}})(\bar{\boldsymbol{\theta}}_g - \bar{\boldsymbol{\theta}})'; \quad (\text{D.190})$$

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

$$W = \frac{1}{G(N-1)} \sum_{g=1}^G \sum_{n=1}^N (\boldsymbol{\theta}_{n,g} - \bar{\boldsymbol{\theta}}_g)(\boldsymbol{\theta}_{n,g} - \bar{\boldsymbol{\theta}}_g)'; \quad (\text{D.191})$$

3. estimate the overall variance as

$$\hat{V} = \frac{N-1}{N} W + \left(1 + \frac{1}{G}\right) D; \quad (\text{D.192})$$

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

$$\hat{R}^{mult} = \frac{N}{N-1} + \frac{G+1}{G} \lambda_1, \quad (\text{D.193})$$

where  $\lambda_1$  is the largest eigenvalue of the matrix  $W^{-1}D$ .

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

## E Our prior distribution

In this section we show that our prior can be viewed as a generalization of the Generalized Normal prior by [Arias et al. \(2018\)](#). Consider first the case of no sign restrictions as in Section (2.5) of [Arias et al. \(2018\)](#). 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 et al., 2018](#)) or in its independent form. Define  $A = B^{-1}$ . The joint inverse Wishart Uniform distribution,

$$p_{NiWU}(\Sigma, Q) = p_{NiWU}(\Sigma) \cdot p_{NiWU}(Q), \quad (\text{E.194})$$

$$\propto |\det(\Sigma)|^{-\frac{d+k+1}{2}} \cdot e^{-\frac{1}{2}\text{trace}[\Sigma^{-1}S]} \cdot 1, \quad (\text{E.195})$$

implies the following distributions for B and for A:

$$p_{NiWU}(B) \propto v_{\Sigma, Q \rightarrow B} \cdot |\det(BB')|^{-\frac{d+k+1}{2}} \cdot e^{-\frac{1}{2}\text{trace}[(BB')^{-1}S]}, \quad (\text{E.196})$$

$$\propto |\det(B)| \cdot |\det(B)|^{-(d+k+1)} \cdot e^{-\frac{1}{2}\text{trace}[(BB')^{-1}S]}, \quad (\text{E.197})$$

$$\propto |\det(B)|^{-(d+k)} \cdot e^{-\frac{1}{2}\text{trace}[B'^{-1}B^{-1}S]}, \quad (\text{E.198})$$

$$\propto |\det(B)|^{-(d+k)} \cdot e^{-\frac{1}{2}\text{trace}[B^{-1}SB'^{-1}]}, \quad (\text{E.199})$$

$$\propto |\det(B)|^{-(d+k)} \cdot e^{-\frac{1}{2}\text{vec}(B^{-1'})'(I_k \otimes S)\text{vec}(B^{-1'})}, \quad (\text{E.200})$$

$$p_{NiWU}(A) \propto v_{B \rightarrow A} \cdot |\det(A^{-1})|^{-(d+k)} \cdot e^{-\frac{1}{2}\text{vec}(A')'(I_k \otimes S)\text{vec}(A')}, \quad (\text{E.201})$$

$$\propto |\det(A)|^{-2k} \cdot |\det(A)|^{d+k} \cdot e^{-\frac{1}{2}\text{vec}(A')'(I_k \otimes S)\text{vec}(A')}, \quad (\text{E.202})$$

$$\propto |\det(A)|^{d-k} \cdot e^{-\frac{1}{2}\text{vec}(A')'(I_k \otimes S)\text{vec}(A')}. \quad (\text{E.203})$$

Equation (E.203) is the Generalized Normal distribution from [Arias et al. \(2018\)](#) and coincides with their equation (2.8), adjusted for the different notation (i.e.  $A = A'$ ). It is constant in  $Q$ , i.e.  $p_{NiWU}(A_1) = p_{NiWU}(A_2)$  for  $A_2 = QA_1$ . This holds immediately when inspecting equation (E.195) jointly with the fact that  $v_{\Sigma, Q \rightarrow B}$  does not depend on  $Q$ , or noticing that

$$p_{NiWU}(QA) \propto |\det(QA)|^{d-k} \cdot e^{-\frac{1}{2}\text{vec}((QA)')'(I_k \otimes S)\text{vec}((QA)')}, \quad (\text{E.204})$$

$$\propto |\det(A)|^{d-k} \cdot e^{-\frac{1}{2}\text{trace}[(QA)S(QA)']}, \quad (\text{E.205})$$

$$\propto |\det(A)|^{d-k} \cdot e^{-\frac{1}{2}\text{trace}[QASA'Q']}, \quad (\text{E.206})$$

$$\propto |\det(A)|^{d-k} \cdot e^{-\frac{1}{2}\text{trace}[A'Q'QAS]}, \quad (\text{E.207})$$

$$\propto |\det(A)|^{d-k} \cdot e^{-\frac{1}{2}\text{trace}[A'AS]} = p_{NiWU}(A). \quad (\text{E.208})$$

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  $\Sigma$ .

Consider now our prior distribution. Consider first the case in which no sign



restrictions are introduced. Define

$$V_b = \left( \frac{\psi_2^2}{1.96^2} \cdot I_k \otimes \text{diag}([\gamma_1^2, \dots, \gamma_i^2, \dots, \gamma_k^2]') \right), \quad (\text{E.209})$$

$$= C^{-1} \otimes D^{-1}, \quad (\text{E.210})$$

$$C = \frac{1.96^2}{\psi_2^2} I_k, \quad (\text{E.211})$$

$$D = \text{diag}\left(\left[\frac{1}{\gamma_1^2}, \dots, \frac{1}{\gamma_i^2}, \dots, \frac{1}{\gamma_k^2}\right]'\right). \quad (\text{E.212})$$

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

$$\tilde{p}(B) = p(B), \quad (\text{E.213})$$

$$= \prod_i \prod_j p(b_{ij} | \mu_{ij}, \sigma_{ij}), \quad (\text{E.214})$$

$$= \prod_i \prod_j \phi\left(0, \frac{\psi_2^2}{1.96^2} \cdot \gamma_i^2\right), \quad (\text{E.215})$$

$$\propto |\det(V_b)|^{-\frac{1}{2}} \cdot e^{-\frac{1}{2} \text{vec}(B)' V_b^{-1} \text{vec}(B)}, \quad (\text{E.216})$$

$$\propto |\det(C \otimes D)|^{\frac{1}{2}} \cdot e^{-\frac{1}{2} \text{vec}(B)' (C \otimes D) \text{vec}(B)}, \quad (\text{E.217})$$

$$\propto |\det(C)|^{\frac{k}{2}} \cdot |\det(D)|^{\frac{k}{2}} \cdot e^{-\frac{1}{2} \text{vec}(B)' (C \otimes D) \text{vec}(B)}, \quad (\text{E.218})$$

$$\propto \left(\frac{1.96}{\psi_2}\right)^{\frac{k}{2}} \cdot \prod_{i=1}^k \gamma_i^{-\frac{k}{2}} \cdot e^{-\frac{1}{2} \text{trace}[B' D B C']}, \quad (\text{E.219})$$

$$\propto \left(\frac{1.96}{\psi_2}\right)^{\frac{k}{2}} \cdot \prod_{i=1}^k \gamma_i^{-\frac{k}{2}} \cdot e^{-\frac{1}{2} \text{trace}[B' D B \frac{1.96^2}{\psi_2^2}]}, \quad (\text{E.220})$$

$$\propto \left(\frac{1.96}{\psi_2}\right)^{\frac{k}{2}} \cdot \prod_{i=1}^k \gamma_i^{-\frac{k}{2}} \cdot e^{-\frac{1}{2} \frac{1.96^2}{\psi_2^2} \text{trace}[D B B']}, \quad (\text{E.221})$$

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

Consider now the case with sign restrictions. Define

$$V_b = \left( I_k \otimes \text{diag}([s_1^2, \dots, s_i^2, \dots, s_k^2]') \right), \quad (\text{E.222})$$

$$= I_k \otimes D^{-1}, \quad (\text{E.223})$$

$$D = \text{diag}\left(\left[\frac{1}{s_1^2}, \dots, \frac{1}{s_i^2}, \dots, \frac{1}{s_k^2}\right]'\right), \quad (\text{E.224})$$

$$\mathbf{e} = \psi_1 \cdot (\boldsymbol{\iota}_k \otimes [\gamma_1, \dots, \gamma_i, \dots, \gamma_k]') \cdot \text{vec}(\text{sign}(B)), \quad (\text{E.225})$$

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

$$\tilde{p}(B) \propto \mathbf{I}\{B\} \cdot p(B), \quad (\text{E.226})$$

$$\tilde{p}(B) = \prod_i \prod_j \tilde{p}(b_{ij} | \mu_{ij}, \sigma_{ij}), \quad (\text{E.227})$$

$$= \prod_i \prod_j \mathbf{I}\{b_{ij}\} \phi\left(\psi_1 \gamma_i \cdot \text{sign}(b_{ij}), s_i^2\right), \quad (\text{E.228})$$

$$\propto \mathbf{I}\{B\} \cdot |\det(V_b)|^{-\frac{1}{2}} \cdot e^{-\frac{1}{2}(\text{vec}(B) - \mathbf{e})' V_b^{-1} (\text{vec}(B) - \mathbf{e})}, \quad (\text{E.229})$$

$$\propto \mathbf{I}\{B\} \cdot |\det(I_k \otimes D)|^{\frac{1}{2}} \cdot e^{-\frac{1}{2}(\text{vec}(B) - \mathbf{e})' (I_k \otimes D) (\text{vec}(B) - \mathbf{e})}, \quad (\text{E.230})$$

$$\propto \mathbf{I}\{B\} \cdot |\det(D)|^{\frac{k}{2}} \cdot e^{-\frac{1}{2}(\text{vec}(B) - \mathbf{e})' (I_k \otimes D) (\text{vec}(B) - \mathbf{e})}, \quad (\text{E.231})$$

$$\propto \mathbf{I}\{B\} \cdot |\det(D)|^{\frac{k}{2}} \cdot e^{-\frac{1}{2} \left\{ \text{vec}(B)' (I_k \otimes D) \text{vec}(B) - 2 \text{vec}(B)' (I_k \otimes D) \mathbf{e} + \mathbf{e}' (I_k \otimes D) \mathbf{e} \right\}}, \quad (\text{E.232})$$

$$\propto \mathbf{I}\{B\} \cdot |\det(D)|^{\frac{k}{2}} \cdot e^{-\frac{1}{2} \text{vec}(B)' (I_k \otimes D) \text{vec}(B)} \cdot e^{\text{vec}(B)' (I_k \otimes D) \mathbf{e}}, \quad (\text{E.233})$$

$$\propto \mathbf{I}\{B\} \cdot |\det(D)|^{\frac{k}{2}} \cdot e^{-\frac{1}{2} \text{trace}[D B B']} \cdot e^{\text{vec}(B)' (I_k \otimes D) \mathbf{e}}, \quad (\text{E.234})$$

which implies  $p(B) \neq p(BQ)$ , unless  $\psi_1 = 0$ , under which  $\text{vec}(B)' (I_k \otimes D) \mathbf{e} = 0$  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.

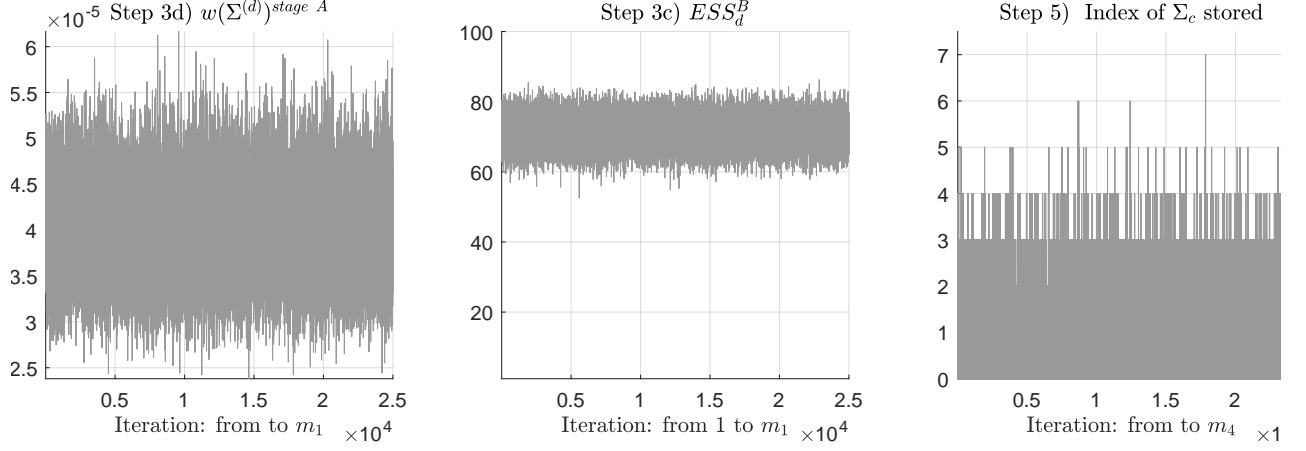
## F Additional Tables and Figures

**Table F1:** Tuning parameters used in the algorithms

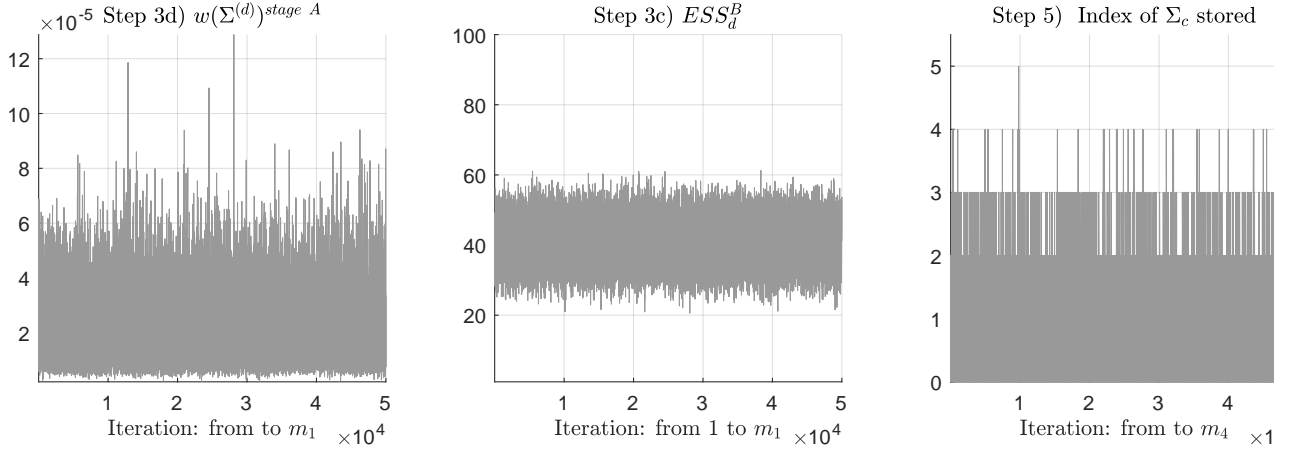
Algorithm	Discussion of the algorithm		Figure 2	Figure 3 Table 1 Figure 5	Figure 4	Figure 6 Figure 7 Figure 8 Table 2 Table 3
A	Section B.1	$n_1$		$m_1$		$m_1$
B	Section B.1	$n_1$	20,000			20,000
C	Section B.2	$n_2$ $n_3$				
D	Section B.2	$n_2$ $n_3$				
Our	Section 2.3	$m_1$ $m_2$ $m_4$ $m_5$		25,000 100 $m_1$ 20,000		50,000 100 $m_1$ 20,000
F	Section C.1.1	$n_4$			50,000	
DSMH	Section D	$H$ $G$ initial $N$ $K$ $M$ $\tau$		50 10 20,000 100 10 10		50 10 50,000 100 10 10

**Figure F1:** Performance of our algorithm

Simulation exercise from [Section 3](#),  $T = 120$ , single pseudo dataset

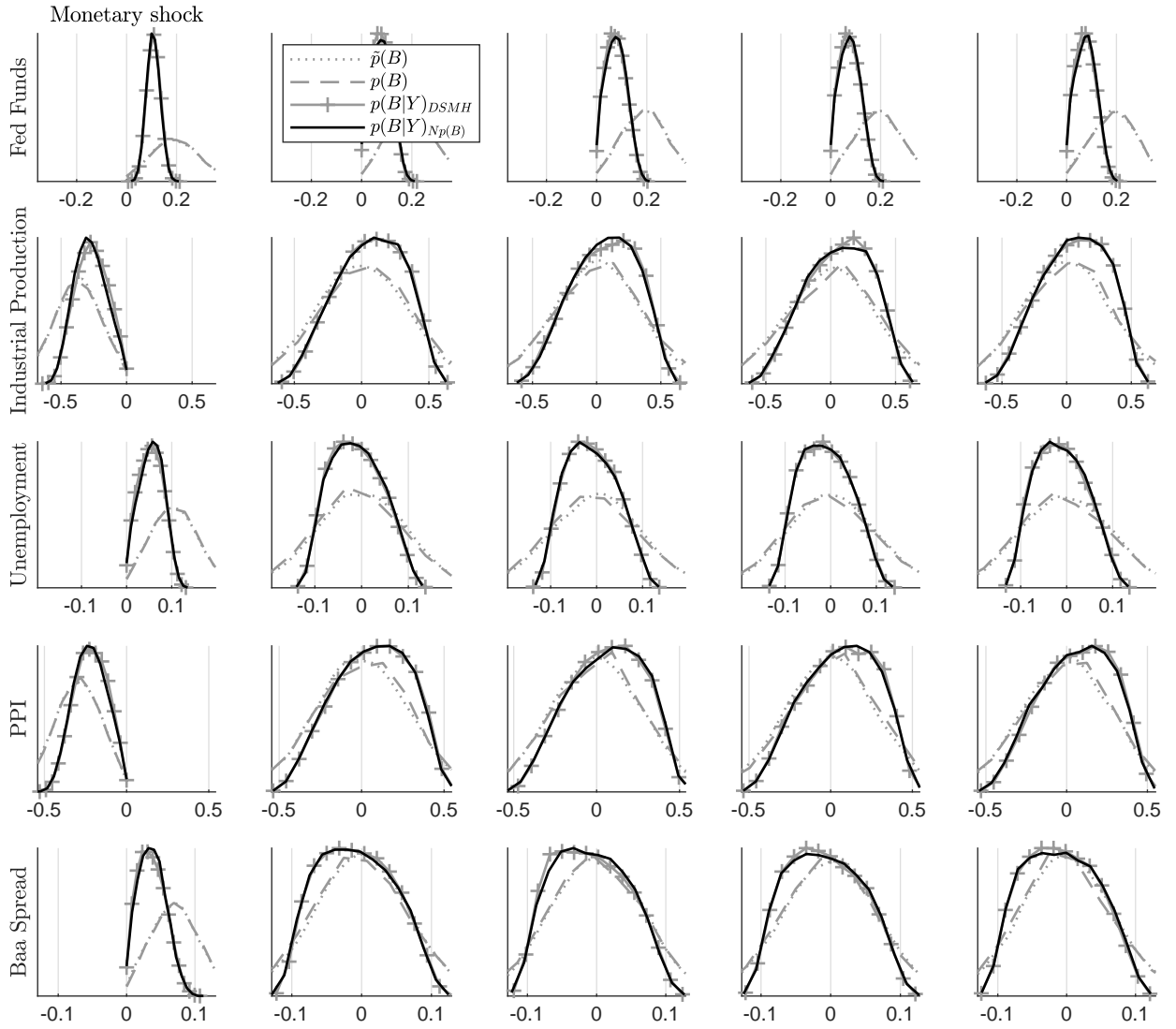


Application to monetary policy shocks, [Section 4](#)



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  $ESS_d^B = \left( \sum_i (w_i(\Sigma^{(d)})^{\text{stage B}})^2 \right)^{-1}$  computed, for each of the  $m_1$  draws of  $\Sigma$ , 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  $\Sigma$ .

**Figure F2: Update on  $B$**



**Table F2:** 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
<i>Using the NiWU prior, <math>i_0</math></i>													
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
<i>Using our prior, <math>i_1</math></i>													
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
<i>Ratio of our prior to NiWU prior, <math>100 \cdot (i_1/i_0 - 1)</math> (percent)</i>													
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 F3:** 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
<i>Using the NiWU prior, <math>i_0</math></i>													
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
<i>Using our prior, <math>i_1</math></i>													
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
<i>Ratio of our prior to NiWU prior, <math>100 \cdot (i_1/i_0 - 1)</math> (percent)</i>													
Fed F.	0	-7	-16	-29	-55	-124	-210	5	185	99	70	58	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 F4:** 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
<i>Using the NiWU prior, <math>f_0</math></i>													
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
<i>Using our prior, <math>f_1</math></i>													
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
<i>Ratio of our prior to NiWU prior, <math>100 \cdot (f_1/f_0 - 1)</math> (percent)</i>													
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).



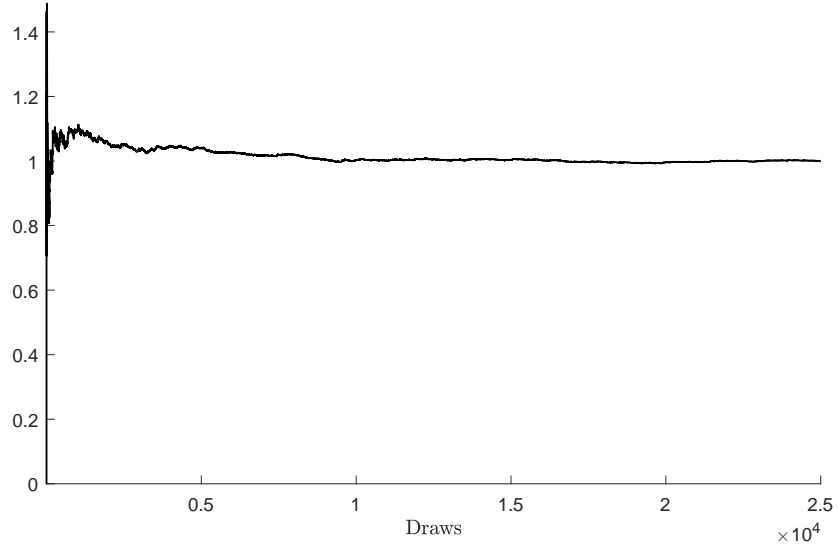
**Table F5:** Comparison of Forecast Error Variance Decompositions (pointwise median)  
Robustness for  $\psi_1 = 0.6$ ,  $\psi_2 = 1.2$

	0	1	2	3	4	5	6	7	8	9	10	11	12
<i>Using the NiWU prior, <math>f_0</math></i>													
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
<i>Using our prior, <math>f_1</math></i>													
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
<i>Ratio of our prior to NiWU prior, <math>100 \cdot (f_1/f_0 - 1)</math> (percent)</i>													
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

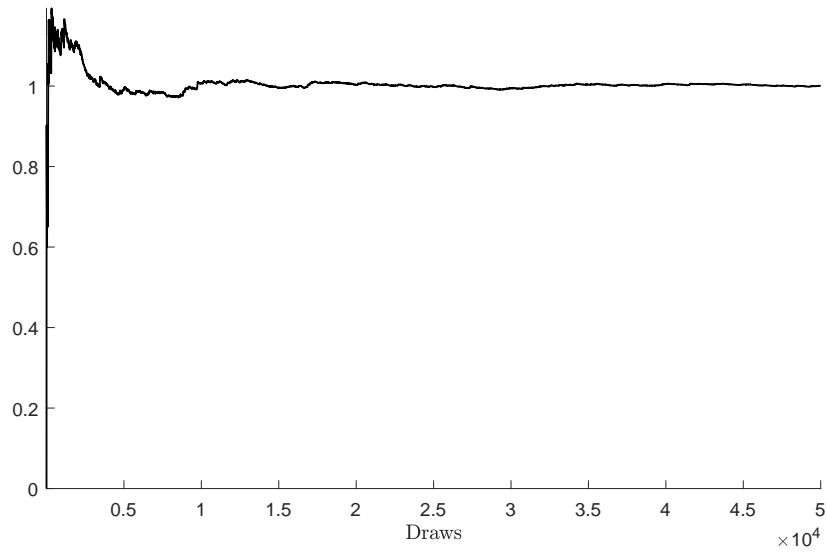
Note: Pointwise median forecast error decompositions using the NiWU prior (top panel) or our prior (middle panel), together with the percent difference between the two (bottom panel).

**Figure F3:** Diagnostics on the importance weights, graphical assessment

Simulation exercise from [Section 3](#),  $T = 120$ , single pseudo dataset



Application to monetary shocks, [Section 4](#)



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 F6:** Diagnostics on the importance weights in Stage A of our sampler: tests

---

Simulation exercise from [Section 3](#),  $T = 120$ , single pseudo dataset

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

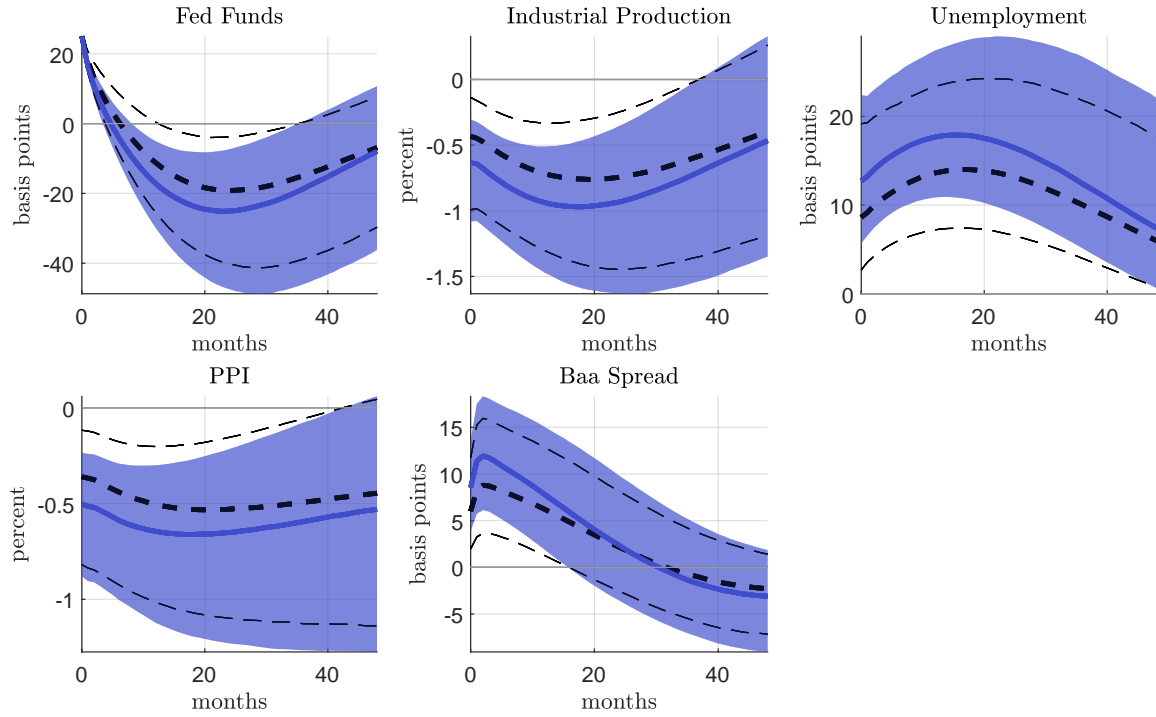
Application to monetary policy shocks, [Section 4](#)

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 F4:** 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, K. M. and Magnus, J. R. (2005), *Matrix Algebra*, Vol. 1, Cambridge University Press.
- Arias, J. E., Rubio-Ramírez, J. F. and Waggoner, D. F. (2018), ‘Inference based on Structural Vector Autoregressions identified with sign and zero restrictions: Theory and applications’, *Econometrica* **86**(2), 685–720.
- Bibby, J., Kent, J. and Mardia, K. (1979), *Multivariate analysis*, Academic Press, London.
- Brooks, S. P. and Gelman, A. (1998), ‘General methods for monitoring convergence of iterative simulations’, *Journal of computational and graphical statistics* **7**(4), 434–455.
- Brooks, S. P. and Roberts, G. O. (1998), ‘Assessing convergence of Markov Chain Monte Carlo algorithms’, *Statistics and Computing* **8**(4), 319–335.
- Canova, F. (2007), *Methods for applied macroeconomic research*, Vol. 13, Princeton University Press.
- Casella, G. and Berger, R. L. (2021), *Statistical inference*, Cengage Learning.
- Cowles, M. K. and Carlin, B. P. (1996), ‘Markov Chain Monte Carlo convergence diagnostics: a comparative review’, *Journal of the American Statistical Association* **91**(434), 883–904.
- Del Negro, M. and Schorfheide, F. (2011), Bayesian Macroeconometrics, *in* G. K. Herman van Dijk and J. Geweke, eds, ‘The Oxford Handbook of Bayesian Econometrics’, Oxford University Press, pp. 293–389.
- Doan, T., Litterman, R. and Sims, C. (1984), ‘Forecasting and conditional projection using realistic prior distributions’, *Econometric Reviews* **3**(1), 1–100.

- Gelman, A. and Rubin, D. B. (1992), ‘Inference from iterative simulation using multiple sequences’, *Statistical Science* pp. 457–472.
- Geweke, J. (1989), ‘Bayesian inference in econometric models using Monte Carlo integration’, *Econometrica* pp. 1317–1339.
- Geweke, J. (1992), Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments, *in* ‘Bayesian Statistics’.
- Herbst, E. and Schorfheide, F. (2014), ‘Sequential Monte Carlo sampling for DSGE models’, *Journal of Applied Econometrics* **29**(7), 1073–1098.
- Kilian, L. and Lütkepohl, H. (2017), *Structural vector autoregressive analysis*, Cambridge University Press.
- Koop, G. and Korobilis, D. (2010), ‘Bayesian multivariate time series methods for empirical macroeconomics’, *Foundations and Trends in Econometrics* **3**(4), 267–358.
- Koopman, S. J., Shephard, N. and Creal, D. (2009), ‘Testing the assumptions behind importance sampling’, *Journal of Econometrics* **149**(1), 2–11.
- Lütkepohl, H. (2005), *New introduction to multiple time series analysis*, Springer Science & Business Media.
- Mathai, A. M. and Haubold, H. J. (2008), *Special functions for applied scientists*, Vol. 4, Springer.
- Mlikota, M. and Schorfheide, F. (2022), ‘Sequential monte carlo with model tempering’, *arXiv preprint arXiv:2202.07070* .
- Pickands, J. et al. (1975), ‘Statistical inference using extreme order statistics’, *the Annals of Statistics* **3**(1), 119–131.
- Raftery, A. and Lewis, S. (1992), ‘How many iterations in the Gibbs sampler?’, *Bayesian Statistics* **4**, 763–773.

Robert, C. and Casella, G. (2013), *Monte Carlo statistical methods*, Springer Science & Business Media.

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

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