Appendix to Identification of SVAR Models by Combining Sign Restrictions With External Instruments

June 8, 2022

A Prior and posterior distribution of B

In this section of the appendix, we prove that the conjugate prior distribution in equation (2.9) implies

$$p(\tilde{B}; v_0, S_0) \propto p(B; v_0, S_0) p(\Sigma_\eta^{1/2}; v_0, S_0) p(\Phi | B, \Sigma_\eta^{1/2}; v_0, S_0),$$

where:

$$\begin{aligned} p(B; v_0, S_0) &\propto |\det(B)|^{-(v_0+n)} \exp\left(-\frac{1}{2}\text{tr}\left(S_{11}\left(BB'\right)^{-1}\right)\right), \\ p(\Sigma_\eta^{1/2}; v_0, S_0) &\propto |\Sigma_\eta|^{-(v_0+k)/2} \exp\left(-\frac{1}{2}\text{tr}\left(S_{22\cdot 1}\Sigma_\eta^{-1}\right)\right), \\ p(\Phi | B, \Sigma_\eta^{1/2}; v_0, S_0) &\sim \mathcal{MN}(S_{21}S_{11}^{-1}B, \Sigma_\eta, B'S_{11}^{-1}B). \end{aligned}$$

Here, $\Sigma_\eta = \Sigma_\eta^{1/2}(\Sigma_\eta^{1/2})'$, $S_0 = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}$, $S_{22\cdot 1} = S_{22} - S_{21}S_{11}^{-1}S_{12}$, and $X \sim \mathcal{MN}(M, U, V)$ denotes the matrix normal distribution with mean $\text{E}[X] = M$ and variance $\text{Var}[\text{vec}(X)] = V \otimes U$.

Derivation: First, since $\tilde{B}$ is block lower triangular, we have that $|\det(\tilde{B})| = |\det(B)||\det(\Sigma_\eta^{1/2})|$, and hence:

$$p(\tilde{B}; v_0, S_0) \propto |\det(B)|^{-(v_0+\tilde{n})} |\Sigma_\eta|^{\frac{-(v_0+\tilde{n})}{2}} \exp\left(-\frac{1}{2}\text{tr}\left(S_0\left(\tilde{B}\tilde{B}'\right)^{-1}\right)\right).$$

Next, let us focus on the trace term $\text{tr}\left(S_0\left(\tilde{B}\tilde{B}'\right)^{-1}\right) = \text{tr}\left(\tilde{B}^{-1}S_0(\tilde{B}^{-1})'\right)$. Let

$$\tilde{B}^{-1} = \begin{pmatrix} B & 0_{n\times k} \\ \Phi & \Sigma_\eta^{1/2} \end{pmatrix}^{-1} = \begin{pmatrix} B^{-1} & 0_{n\times k} \\ -\Sigma_\eta^{-1/2}\Phi B^{-1} & \Sigma_\eta^{-1/2} \end{pmatrix} = \begin{pmatrix} C_{11} & 0 \\ C_{21} & C_{22} \end{pmatrix}.$$

Then,

$$\begin{aligned} \text{tr}\left(\tilde{B}^{-1}S_0(\tilde{B}^{-1})'\right) &= \text{tr}\left(\begin{pmatrix} C_{11} & 0 \\ C_{21} & C_{22} \end{pmatrix}\begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}\begin{pmatrix} C_{11}' & C_{21}' \\ 0 & C_{22}' \end{pmatrix}\right) \\ &= \text{tr}\left(\begin{pmatrix} C_{11}S_{11} & C_{11}S_{12} \\ C_{21}S_{11} + C_{22}S_{21} & C_{21}S_{12} + C_{22}S_{22} \end{pmatrix}\begin{pmatrix} C_{11}' & C_{21}' \\ 0 & C_{22}' \end{pmatrix}\right) \\ &= \text{tr}\left[C_{11}S_{11}C_{11}'\right] + \text{tr}\left[(C_{21}S_{11} + C_{22}S_{21})C_{21}' + (C_{21}S_{12} + C_{22}S_{22})C_{22}'\right] \\ &= \text{tr}\left(C_{11}S_{11}C_{11}'\right) + \text{tr}\left((C_{21}S_{11}C_{21}' + C_{22}S_{21}C_{21}' + C_{21}S_{12}C_{22}' + C_{22}S_{22}C_{22}'\right) \\ &= \text{tr}\left(C_{11}S_{11}C_{11}'\right) + \text{tr}(C_{21}S_{11}C_{21}') + 2\text{tr}(C_{22}S_{21}C_{21}') + \text{tr}(C_{22}S_{22}C_{22}') \\ &= \text{tr}\left(B^{-1}S_{11}B^{-1'}\right) + \text{tr}(\Sigma_\eta^{-1}\Phi B^{-1}S_{11}B^{-1'}\Phi') - 2\text{tr}(\Sigma_\eta^{-1}S_{21}B^{-1'}\Phi') + \text{tr}(\Sigma_\eta^{-1}S_{22}). \end{aligned}$$

Adding and subtracting $\text{tr}(\Sigma_\eta^{-1}S_{21}S_{11}^{-1}S_{12})$ yields:

$$\begin{aligned}
\text{tr}\left(\tilde{B}^{-1}S_0\tilde{B}^{-1'}\right) &= \text{tr}\left(B^{-1}S_{11}B^{-1'}\right) + \text{tr}(\Sigma_\eta^{-1}S_{22}) - \text{tr}(\Sigma_\eta^{-1}S_{21}S_{11}^{-1}S_{12}) \\
&\quad + \text{tr}(\Sigma_\eta^{-1}\Phi B^{-1}S_{11}B^{-1'}\Phi') - 2\text{tr}(\Sigma_\eta^{-1}S_{21}B^{-1'}\Phi') + \text{tr}(\Sigma_\eta^{-1}S_{21}S_{11}^{-1}S_{12}) \\
\text{tr}\left(\tilde{B}^{-1}S_0\tilde{B}^{-1'}\right) &= \text{tr}\left(B^{-1}S_{11}B^{-1'}\right) + \text{tr}(\Sigma_\eta^{-1}S_{22\cdot 1}) \\
&\quad + \text{tr}\left(B^{-1}S_{11}B^{-1'}(\Phi - S_{21}S_{11}^{-1}B)'\Sigma_\eta^{-1}(\Phi - S_{21}S_{11}^{-1}B)\right)
\end{aligned}$$

Plugging the trace term back into the prior of $\tilde{B}$ yields:

$$\begin{aligned}
p(\tilde{B}; v_0, S_0) \propto |\det(B)|^{-(v_0+\tilde{n})} \exp\left(-\frac{1}{2}\text{tr}\left(S_{11}\left(BB'\right)^{-1}\right)\right) \cdot |\Sigma_\eta|^{\frac{-v_0+\tilde{n}}{2}} \exp\left(-\frac{1}{2}\text{tr}\left(S_{22\cdot 1}\Sigma_\eta^{-1}\right)\right) \\
\cdot \exp\left(-\frac{1}{2}\text{tr}\left(B^{-1}S_{11}B^{-1'}(\Phi - S_{21}S_{11}^{-1}B)'\Sigma_\eta^{-1}(\Phi - S_{21}S_{11}^{-1}B)\right)\right)
\end{aligned}$$

The final step is to multiply the term by $\frac{|B'S_{11}^{-1}B|^{k/2}}{|B'S_{11}^{-1}B|^{k/2}} \cdot \frac{|\Sigma_\eta|^{n/2}}{|\Sigma_\eta|^{n/2}}$, yielding the result:

$$\begin{aligned}
p(\tilde{B}; v_0, S_0) &\propto |\det(B)|^{-(v_0+\tilde{n})+k} \exp\left(-\frac{1}{2}\text{tr}\left(S_{11}\left(BB'\right)^{-1}\right)\right) \cdot |\Sigma_\eta|^{\frac{-(v_0+\tilde{n})+n}{2}} \exp\left(-\frac{1}{2}\text{tr}\left(S_{22\cdot 1}\Sigma_\eta^{-1}\right)\right) \\
&\quad \frac{1}{|B'S_{11}^{-1}B|^{k/2}|\Sigma_\eta|^{n/2}} \exp\left(-\frac{1}{2}\text{tr}\left(B^{-1}S_{11}B^{-1'}(\Phi - S_{21}S_{11}^{-1}B)'\Sigma_\eta^{-1}(\Phi - S_{21}S_{11}^{-1}B)\right)\right) \\
&= |\det(B)|^{-(v_0+n)} \exp\left(-\frac{1}{2}\text{tr}\left(S_{11}\left(BB'\right)^{-1}\right)\right) \cdot |\Sigma_\eta|^{\frac{-(v_0+k)}{2}} \exp\left(-\frac{1}{2}\text{tr}\left(S_{22\cdot 1}\Sigma_\eta^{-1}\right)\right) \\
&\quad \frac{1}{|B'S_{11}^{-1}B|^{k/2}|\Sigma_\eta|^{n/2}} \exp\left(-\frac{1}{2}\text{tr}\left(B^{-1}S_{11}B^{-1'}(\Phi - S_{21}S_{11}^{-1}B)'\Sigma_\eta^{-1}(\Phi - S_{21}S_{11}^{-1}B)\right)\right)
\end{aligned}$$

B Posterior inference

We start with some notation. Let S_a and S_b be full rank selection matrices of zeros and ones such that $\alpha = S_a\text{vec}(\tilde{A})$ and $\beta = S_b\text{vec}(\tilde{B})$ are the nonzero free elements in $\tilde{A}$ and $\tilde{B}$. Denote by $\theta = \{\alpha, \beta\}$ the set of augmented SVAR parameters, and by θ_{-x} the set of parameters excluding x. Setting arbitrary initial values $\theta^{(0)} = \{\alpha^{(0)}, \beta^{(0)}\}$, we propose a MCMC that generates draws $\theta^{(i)}, i = 1, \ldots, M$ from the posterior, by iteratively drawing from the following conditional distributions:

1. Draw $\alpha^{(i)}$ from $p\left(\alpha|\theta_{-\alpha}, \tilde{Y}\right) \sim \mathcal{N}(\bar{\alpha}, \bar{V}_\alpha)$ where mean and variance are:

$$\begin{aligned}
\bar{V}_\alpha^{-1} &= V_\alpha^{-1} + S_a((\tilde{B}\tilde{B}')^{-1} \otimes X'X)S_a', \\
\bar{\alpha} &= \bar{V}_\alpha\left(V_\alpha^{-1} + S_a\text{vec}(X'\tilde{Y}(\tilde{B}\tilde{B}')^{-1})\right).
\end{aligned}$$

2. Draw $\beta^{(i)}$ from $p\left(\beta|\theta_{-\beta},\tilde{Y}\right) \propto p(\tilde{Y}|\alpha,\beta)p(\beta)$. Since the conditional distribution is of no known form, we rely on an Accept Reject Metropolis Hastings (AR-MH) step (Tierney 1994, Chib & Greenberg 1995). For a given proposal distribution $p^{\star}\left(\beta|\theta_{-\beta},\tilde{Y}\right)$, which we discuss at a later point, the AR-MH algorithm involves two steps:

   (a) *Accept-reject step*: Generate a candidate $\beta^{\star} \sim p^{\star}\left(\beta|\theta_{-\beta},\tilde{Y}\right)$ and accept it with probability

$$\alpha_{\text{AR}}\left(\beta^{\star}\right) = \min\left\{1, \frac{p\left(\beta^{\star}|\theta_{-\beta},\tilde{Y}\right)}{c_{AR} \times p^{\star}\left(\beta^{\star}|\theta_{-\beta},\tilde{Y}\right)}\right\},$$

   which is repeated until a draw is accepted.

   (b) *Metropolis-Hastings step*: Accept the proposal $\beta^{\star}$ with probability $\alpha_{\text{MH}}(\beta^{(i-1)}|\beta^{\star})$. Let $\mathcal{D}\left(\beta\right) = \left\{\beta : p\left(\beta|\theta_{-\beta},\tilde{Y}\right) \leq c_{AR} \times p^{\star}\left(\beta|\theta_{-\beta},\tilde{Y}\right)\right\}$ and $\mathcal{D}^{C}\left(\beta\right)$ its complement. Then:

$$\alpha_{\text{MH}}(\beta^{(i-1)}|\beta^{\star}) = \begin{cases} 1 & \text{if } \beta^{(i-1)} \in \mathcal{D}\left(\beta\right) \\ \dfrac{c_{AR} \times p^{\star}\left(\beta^{\star}|\theta_{-\beta},\tilde{Y}\right)}{p\left(\beta^{\star}|\theta_{-\beta},\tilde{Y}\right)} & \text{if } \beta^{(i-1)} \in \mathcal{D}^{C}\left(\beta\right), \beta^{\star} \in \mathcal{D}\left(\beta\right) \\ \dfrac{p\left(\beta^{\star}|\theta_{-\beta},\tilde{Y}\right) p^{\star}\left(\beta^{(i-1)}|\theta_{-\beta},\tilde{Y}\right)}{p\left(\beta^{(i-1)}|\theta_{-\beta},\tilde{Y}\right) p^{\star}\left(\beta^{\star}|\theta_{-\beta},\tilde{Y}\right)} & \text{if } \beta^{(i-1)}, \beta^{\star} \in \mathcal{D}^{C}\left(\beta\right) \end{cases}$$

The constant c_{AR} in the AR-MH step can be tuned to trade off the efficiency of the AR step against the acceptance probability in the MH step. To see this, note that for increasing values of c_{AR}, the MH acceptance probability eventually approaches one given that any $\beta^{\star} \in \mathcal{D}$. However, at the same time the performance of the AR step deteriorates, as more and more draws are necessary until a draw is accepted. We iteratively tune this constant over a preliminary run of the MCMC as to capture twice the average ratio between target and proposal distribution. For the applications considered in this paper, this resulted in a reasonable trade-off between AR and MH steps, yielding acceptance probabilities of the latter in the range of 85%-99%.

The success of the AR-MH step depends critically on the design of the proposal distribution $p^{\star}\left(\beta|\theta_{-\beta},\tilde{Y}\right)$. In Appendix B.1, we outline in detail a proposal distribution which relies on the methodology developed in Arias et al. (2018, 2021) to efficiently explore the conditional distribution of the set-identified parameters in $\tilde{B}$. Briefly summarized, the proposal involves drawing a candidate $\beta^{\star} = S_b\text{vec}(\tilde{B}^{\star})$ for $\tilde{B}^{\star} = \text{chol}(\Sigma)Q$ by drawing $\Sigma \sim i\mathcal{W}(v,S)$ from an inverse Wishart with shape parameter S and degrees of freedom v, and $Q = \text{diag}(Q_1, Q_2)$ from a uniform distribution of Q_1 and Q_2 subject to the zero and sign restrictions discussed in Section 2.2 and 2.3. In order to capture the shape of the

conditional distribution, we set $v = v_0 + T$ and $S = S_0 + \tilde{U}\tilde{U}'$. To evaluate the importance density of a candidate draw $\beta^\star$, we use numerical derivatives which account for the change of variables underlying the transformation of random variables Σ, Q to β. After some burn in period, the algorithm is used to generate a large number of draws of the posterior distribution of θ. Those draws are then used in a standard fashion to summarize posterior quantities numerically.

B.1 Proposal distribution used in the AR-MH algorithms

We describe the proposal distribution $p^\star(\beta; v, S)$ used by the AR-MH algorithm in Section 2.4 and 2.5 in more detail. We use the following notation. Let $\tilde{n} = n + k$, $e_{\tilde{n},j}$ be the jth column of $I_{\tilde{n}}$, $Q = \text{diag}(Q_1, Q_2)$ be a $\tilde{n} \times \tilde{n}$ orthogonal block diagonal matrix where Q_1 is orthogonal of size $n \times n$ and Q_2 orthogonal of size $k \times k$. Furthermore, Σ is a symmetric positive definite matrix dimension $\tilde{n}$. As mentioned in the main part of the paper, the structural impact matrix of the proxy-augmented SVAR, $\tilde{B}$, is parameterized as $\tilde{B} = \text{chol}(\Sigma)Q = PQ$ where $\text{chol}(\cdot)$ is the lower triangular Cholesky decomposition. If the external variable is assumed to be a valid instrument, we have specified zero restriction on $\tilde{B}$ as discussed in Section 2.2. We follow Arias et al. (2021) and denote the restrictions as:

$$J\tilde{B}e_{\tilde{n},j} = 0_{k\times 1} \text{ for } 1 \leq j \leq n-k, \tag{B.1}$$

$$JPQe_{\tilde{n},j} = JPL'Q_1e_{n,j} = 0_{k\times 1} \text{ for } 1 \leq j \leq n-k, \tag{B.2}$$

where $J = [0_{k\times n} : I_k]$ and $L = [I_n : 0_{n\times k}]$. That is, the exogeneity restrictions can be written as linear constraints on either $\tilde{B}$ or Q. Denote by $\tilde{z}_j$ the number of restrictions on the jth column of Q_1, which is k for $1 \leq j \leq n-k$ if the exogeneity constraints are imposed and 0 otherwise. Then, the proposal distribution in the AR-MH algorithm draws $\beta^\star$ by the following algorithm:

1. Draw $P = \text{chol}(\Sigma)$ where $\Sigma \sim i\mathcal{W}(v, S)$.

2. Generate $Q = \text{diag}(Q_1, Q_2)$ from a uniform distribution, subject to zero and sign restrictions, as in Arias et al. (2021):

   (a) For $1 \leq j \leq n$, draw $w_{1,j} = x_{1,j}/||x_{1,j}||$ with $x_{1,j} \sim \mathcal{N}(0, I_{n+1-j-\tilde{z}_j})$

   (b) For $1 \leq j \leq k$, draw $w_{2,j} = x_{2,j}/||x_{2,j}||$ with $x_{2,j} \sim \mathcal{N}(0, I_{k+1-j})$

   (c) Compute $Q_1 = [q_{1,1} : \cdots : q_{1,n}]$ recursively by setting $q_{1,j} = K_{1,j}w_{1,j}$, where $K_{1,j}$ is such that it forms a null space of the matrix $M_{1,j} = [q_{1,1} : \cdots : q_{1,j-1} : G(P)']'$ with $G(P) := JPL'$ and for $1 \leq j \leq n-k$. For $n-k+1 \leq j \leq n$, set $M_{1,j} = [q_{1,1} : \cdots : q_{1,j-1}]'$. This captures the exogeneity restrictions as in Section 2.2. If they do not hold (as discussed in Section 2.3), simply use $M_{1,j} = [q_{1,1} : \cdots : q_{1,j-1}]'$ for $1 \leq j \leq n$.

   (d) Compute $Q_2 = [q_{2,1} : \cdots : q_{2,n}]$ recursively by setting $q_{2,j} = K_{2,j}w_{2,j}$ for $K_{2,j}$ such that it forms a null space of $M_{2,j} = [q_{1,1} : \cdots : q_{1,j-1}]'$ for $1 \leq j \leq k$.

   (e) If the sign restrictions are satisfied, proceed. Otherwise, repeat step 2.

3. Set $\tilde{B}^\star = PQ$ and $\beta^\star = S_b \text{vec}(\tilde{B}^\star)$.

Note that by construction $\Sigma = \tilde{B}\tilde{B}'$ and furthermore, $\tilde{B}$ will satisfy the desired zero block restrictions on the upper right part as well as on Φ if the exogeneity restrictions of equation (B.2) are imposed additionally.

In the following, we give the density implied by this proposal distribution. Denote the mapping $[w', \text{vec}(\Sigma)']' \xrightarrow{f} \beta^\star$ and its inverse by $\beta^\star \xrightarrow{f^{-1}} [w', \text{vec}(\Sigma)']'$, where $w = [w'_{1,1}, \ldots, w'_{1,n}, w'_{2,1}, \ldots, w'_{2,k}]'$. Then, a draw from $\beta^\star \sim p^\star(\beta; v, S)$ has density value:

$$p^\star(\beta; v, S) \propto \det(\tilde{B}^\star \tilde{B}^{\star\prime})^{-\frac{v+\tilde{n}+1}{2}} \exp\left(-\frac{1}{2}\text{tr}(S(\tilde{B}^\star \tilde{B}^{\star\prime})^{-1})\right) v_{f^{-1}}(\tilde{B}^\star), \tag{B.3}$$

where the first part comes from the inverse Wishart density of Σ, and $v_{f^{-1}}(\tilde{B})$ is the "volume element" as denoted in Arias et al. (2018), which accounts for the change in variables when transforming draws from Σ, Q to $\tilde{B}$. In our case, we have that following Theorem 2 of Arias et al. (2018):

$$v_{f^{-1}}(\tilde{B}) = |\det(J_{f^{-1}}(\tilde{B}) J_{f^{-1}}(\tilde{B})')|^{\frac{1}{2}}, \tag{B.4}$$

where $J_{f^{-1}}(\tilde{B})$ is the Jacobian of f^{-1} evaluated at $\tilde{B}$. Note that this holds only if S_b in $\beta^\star = S_b \text{vec}(\tilde{B})$ is specified as to include all zero constraints, that is those on the upper right block of $\tilde{B}$, as well as those on Φ if exogeneity restrictions are specified as in equation (B.2). Note that otherwise, Theorem 3 of Arias et al. (2018) would apply.

To ensure that the mappings f and f^{-1} are differentiable and one to one, we follow Appendix A.3 of Arias et al. (2018) to compute $K_{1,j}$ and $K_{2,j}$ by the QR decomposition using the Gram Schmidt process. In order to evaluate the Jacobian, we use numerical derivatives of f^{-1}. Given that the dimension of $\beta^\star$ is usually relatively small, the computational costs are not very high. This is particularly an advantage over Arias et al. (2021), given that in their approach, the mapping underlying the Jacobian is of various magnitudes larger since they include the whole SVAR parameters, that is also the autoregressive parameters.

B.2 Convergence properties of the algorithm

When applying our methods, an important question is how many draws are needed to get reliable answers from the MCMC algorithm. To this end, in Figure 1 we have plotted a random snapshot (of length 5000) of the posterior simulation output for the elements of B in the empirical oil market model of Section 3.1. Visual inspection suggests stationary output with no visible autocorrelation patterns. This is confirmed by Gewekes Relative Numerical Efficiency (RNE) statistic printed into each subplots title. As described in Geweke (1992), the RNE carries the interpretation of the ratio of number of replications required to achieve the same efficiency than drawing iid from the posterior. The RNE values documented for each element of B suggest very high numerical efficiency of the algorithm. The analysis suggests that with several thousand draws one might obtain fairly reliable posterior inference for most of the structural quantities one would typically consider. However, for the computation of Bayes factors we recommend a much higher number of draws, as it requires

to explore sufficiently well the entire set of models. For our applications, we used 100'000 draws when computing Bayes factors, and 10'000 draws for impulse response functions.

Note that Figure 1 is fairly representative of the output obtained in the other empirical applications and identification schemes. Hence, we do not repeat this analysis for each of the empirical applications and make them available upon request.

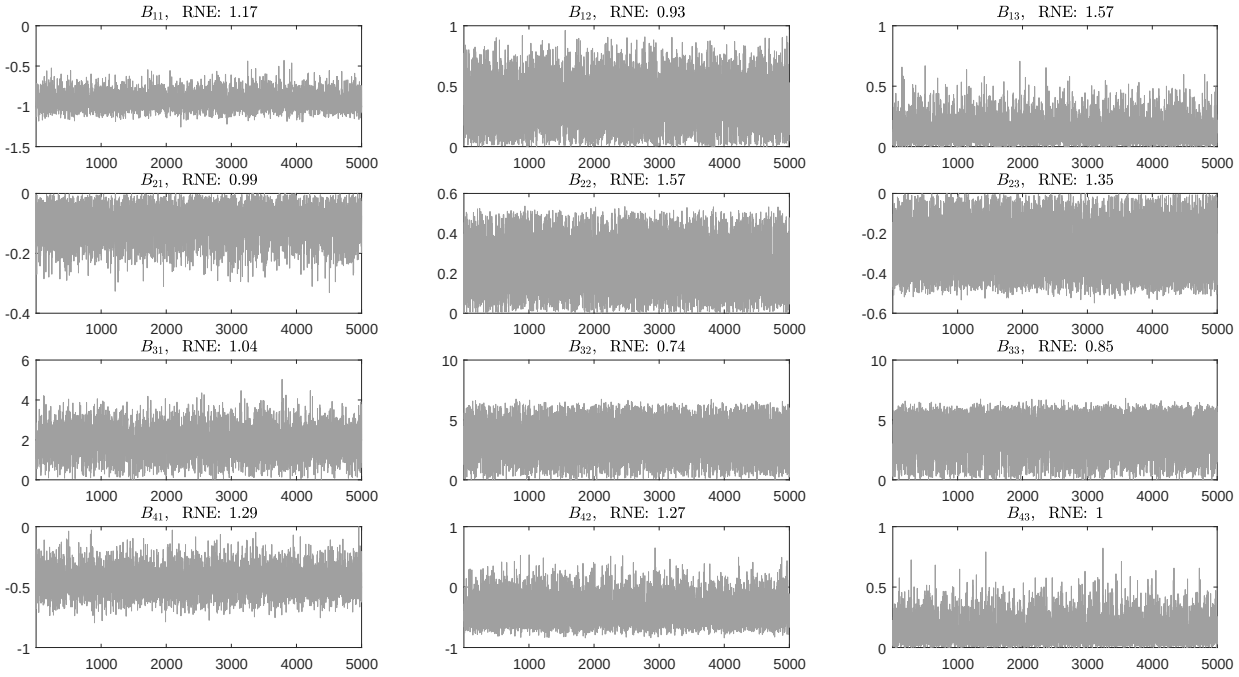

Figure 1: Exemplary Markov Chain Monte Carlo output of each element in B for the oil market model identified by R1 and R2 (see Section 3.1). The title of each subplot also includes the relative numerical efficiency suggested in Geweke (1992).

C Testing for instrument exogeneity

C.1 Densities underlying Bayes factors

From Section 2.5, recall that a simulation consistent estimator of the Bayes factor is given by:

$$\widehat{BF}_{10} = \frac{J_1^{-1} \sum_{i=1}^{J_1} p(\phi_r = \phi_{r,0} | \theta_{-\phi_r}^{(i)})}{J_2^{-1} \sum_{j=1}^{J_2} p(\phi_r = \phi_{r,0} | \tilde{Y}, \theta_{-\phi_r}^{(j)})},$$

where $\theta_{-\phi_r}^{(i)}$ and $\theta_{-\phi_r}^{(j)}$ are prior and posterior draws of the unrestricted model respectively.

Obtaining $p(\phi_r = \phi_{r,0} | \theta_{-\phi_r})$ is straightforward given conditional normality of Φ, that is $p(\Phi | \theta_{-\Phi}) \sim \mathcal{MN}(S_{21} S_{11}^{-1} B, \Sigma_\eta, B' S_{11}^{-1} B)$. The only missing ingredient is to further include ϕ_f into the conditioning set. Let $\phi = \text{vec}(\Phi)$ as well as $\mu_\phi = \text{vec}(S_{21} S_{11}^{-1} B)$ and $V_\phi = (B' S_{11}^{-1} B \otimes \Sigma_\eta)$ the moments of $p(\phi | \theta_{-\phi}) \sim \mathcal{N}(\mu_\phi, V_\phi)$. Exploiting standard results

on joint normality between ϕ_f and ϕ_r we obtain $p(\phi_r = \phi_{r,0}|\theta_{-\phi_r}) \sim \mathcal{N}(\mu_{\phi_r}, V_{\phi_r})$ where

$$\mu_{\phi_r} = S_r\mu_\phi + \left(S_rV_\phi S_f'\right)\left(S_fV_\phi S_f'\right)^{-1}(\phi_f - S_f\mu_\phi),$$
$$V_{\phi_r} = S_rV_\phi S_r' - \left(S_rV_\phi S_f'\right)\left(S_fV_\phi S_f'\right)^{-1}(S_fV_\phi S_r').$$

The posterior ordinate $p(\phi_r = \phi_{r,0}|\theta_{-\phi_r}, \tilde{Y})$ can be obtained following similar steps but departing from the conditional posterior of Φ: $p(\Phi|\theta_{-\Phi}, \tilde{Y}) \sim \mathcal{MN}(\bar{S}_{21}\bar{S}_{11}^{-1}B, \Sigma_\eta, B'\bar{S}_{11}^{-1}B)$. Note that in this case $\bar{S} = S_0 + \left(\tilde{Y} - X\tilde{A}\right)\left(\tilde{Y} - X\tilde{A}\right)'$, which follows from the conjugacy of the prior. Alternatively, and maybe more intuitively, one can write the posterior as the result of a standard regression formulation. Specifically, define $M = [m_1 : \ldots : m_T]$, $E = [\varepsilon_1 : \ldots : \varepsilon_T]$, $H = [\eta_1 : \ldots : \eta_T]$, $\mu_\Phi = S_{21}S_{11}^{-1}B$ and $V_\Phi = B'S_{11}^{-1}B$. Then, our framework implies the regression model $M = \Phi E + \Sigma_\eta^{1/2}H$. One can show that the posterior moments of $p(\Phi|\theta_{-\Phi}, \tilde{Y})$ can be expressed as $\bar{S}_{21}\bar{S}_{11}^{-1}B = \left(ME' + \mu_\Phi V_\Phi^{-1}\right)\left(EE' + V_\Phi^{-1}\right)^{-1}$ and $B'\bar{S}_{11}^{-1}B = \left(EE' + V_\Phi^{-1}\right)^{-1}$, which are standard posteriors for multivariate regression under a conjugate prior. This representation helps to understand the mechanics of the Bayes factors when testing IV restrictions. Assume the posterior of the sign-identified model implies structural shocks for which not only the shock(s) of interest ε_{2t} are able to predict m_t. Then, the posterior ordinate will get smaller and ultimately, the Bayes factor larger pointing towards evidence against instrument exogeneity. A similar line of argument also holds for testing instrument relevance.

C.2 Rotational invariance with respect to unidentified shocks

In this part of the appendix, we show that the Bayes factor developed to test the IV exclusion restrictions is not sensitive to how shocks unrelated to the instrument are identified. To set notation for this analysis, let us assume that the k instruments are related to the first k shocks. This leaves the $k - n$ remaining shocks to be assumed unrelated to the instrument under the null hypothesis. Such an ordering implies the following partition of Φ:

$$\mathrm{E}(m_t\varepsilon_t') = \Phi = [\underbrace{\Phi_f}_{k\times k}, \underbrace{\Phi_r}_{k\times(n-k)}].$$

In the following, we study how an orthogonal rotation of the $n-k$ remaining shocks affects the Bayes factor for $\Phi_{r,0} = 0$. Such a rotation can be obtained by post multiplying the impact matrix B (and hence also Φ_r) with $Q = \mathrm{diag}(I_k, Q_r)$ where $Q_r'Q_r = I_{n-k}$.

In the following, we show that starting from a model $\{B, \Phi\}$, the density $p(\Phi_r = 0|\tilde{Y}, \theta_{-\Phi_r})$ is the same than the density $p(\Phi_r^\star = 0|\tilde{Y}, \theta_{-\Phi_r^\star})$ for a rotated model $\{B^\star = BQ, \Phi^\star = \Phi Q\}$, as long as the rotation only affects the shocks unrelated to the instrument. As we shall see, the reason is that for the special case that $\Phi_{r,0} = \Phi_{r,0}^\star = 0$, the null hypothesis is invariant to those rotations.

In the following, define $\Phi^\star = \Phi Q = [\Phi_f, \Phi_r Q_r]$ and $B^\star = BQ = [B_f, B_rQ_r] = [B_f, B_r^\star]$.

Then, we need to show that:

$$p(\Phi_r = 0|\tilde{Y}, \theta_{-\Phi_r}) = p(\Phi_r^\star = 0|\tilde{Y}, \theta^\star_{-\Phi_r^\star}),$$

where $\theta_{-\Phi_r} = \{\Phi_f, B, \Sigma_\eta, A\}$ and $\theta^\star_{-\Phi_r^\star} = \{\Phi_f, B^\star, \Sigma_\eta, A\}$. Using Bayes Theorem, this is equivalent to show that denominator and numerator are equal for both sides of the equation:

$$\frac{p(\Phi_f, \Phi_{r,0} = 0|\tilde{Y}, \theta_{-\Phi})}{p(\Phi_f|\tilde{Y}, \theta_{-\Phi_f})} = \frac{p(\Phi_f, \Phi^\star_{r,0} = 0|\tilde{Y}, \theta^\star_{-\Phi^\star})}{p(\Phi_f|\tilde{Y}, \theta^\star_{-\Phi_f})}. \tag{C.1}$$

Based on our results of Section 2 and defining $\hat{S}_1 = \overline{S}_{21}\overline{S}_{11}^{-1}$ and $\hat{S}_2 = \overline{S}_{11}^{-1}$, the density of $\Phi^\star$ in the rotated model is given by:

$$\begin{aligned}
p(\Phi^\star|\tilde{Y}, \theta_{-\Phi^\star}) &\sim \mathcal{MN}(\hat{S}_1 B^\star, \Sigma_\eta, (B^\star)'\hat{S}_2 B^\star) \\
&= c^{-1}|\Sigma_\eta|^{-n/2}|(B^\star)'S_{11}^{-1}B^\star|^{k/2}\exp\left(-\frac{1}{2}\text{tr}\left(\left((B^\star)'S_{11}^{-1}B^\star\right)^{-1}(\Phi^\star - \hat{S}B^\star)'\Sigma_\eta^{-1}(\Phi^\star - \hat{S}_1 B^\star)\right)\right) \\
&= c^{-1}|\Sigma_\eta|^{-n/2}|Q'B'\hat{S}_2 BQ|^{k/2} \\
&\times \exp\left(-\frac{1}{2}\text{tr}\left(Q'\left(B'\hat{S}_2 B\right)^{-1}Q(\Phi^\star - \hat{S}_1 BQ)'\Sigma_\eta^{-1}((\Phi^\star - \hat{S}_1 BQ)\right)\right)
\end{aligned}$$

First, let's look at the numerator of equation (C.1). Note that at $\Phi_{r,0} = \Phi^\star_{r,0} = 0$ it holds that $\Phi^\star_{r,0} = \Phi_{r,0}Q_r$ implying that $[\Phi_f, \Phi^\star_{r,0}] = [\Phi_f, \Phi_{r,0}Q_r] = [\Phi_f, \Phi_{r,0}]Q$. Hence, the numerator is equal to:

$$\begin{aligned}
p(\Phi_f, \Phi^\star_{r,0} = 0|\tilde{Y}, \theta^\star_{-\Phi^\star}) &= c^{-1}|\Sigma_\eta|^{-n/2}|B'\hat{S}_2 B|^{k/2}|Q|^k \\
&\times \exp\left(-\frac{1}{2}\text{tr}\left(Q'\left(B'\hat{S}_2 B\right)^{-1}Q([\Phi_f, \Phi^\star_{r,0}] - \hat{S}_1 BQ)'\Sigma_\eta^{-1}(([\Phi_f, \Phi^\star_{r,0}] - \hat{S}_1 BQ)\right)\right) \\
&= c^{-1}|\Sigma_\eta|^{-n/2}|B'\hat{S}_2 B|^{k/2} \\
&\times \exp\left(-\frac{1}{2}\text{tr}\left(Q'\left(B'\hat{S}_2 B\right)^{-1}Q([\Phi_f, \Phi_{r,0}]Q - \hat{S}_1 BQ)'\Sigma_\eta^{-1}(([\Phi_f, \Phi_{r,0}]Q - \hat{S}_1 BQ)\right)\right) \\
&= c^{-1}|\Sigma_\eta|^{-n/2}|B'\hat{S}_2 B|^{k/2} \\
&\times \exp\left(-\frac{1}{2}\text{tr}\left(\left(B'\hat{S}_2 B\right)^{-1}([\Phi_f, \Phi_{r,0}] - \hat{S}_1 B)'\Sigma_\eta^{-1}(([\Phi_f, \Phi_{r,0}] - \hat{S}_1 B)\right)\right) \\
&= p(\Phi_f, \Phi_{r,0} = 0|\tilde{Y}, \theta_{-\Phi})
\end{aligned}$$

Also the denominator stays the same, since the marginal of Φ_f is not affected by the rotation. Specifically, we have that $p(\Phi_f|\tilde{Y}, \theta^\star_{-\Phi_f}) \sim \mathcal{MN}(\hat{S}_1 B_f, \Sigma_\eta, B_f'\hat{S}_2 B_f)$ and hence $p(\Phi_f|\tilde{Y}, \theta^\star_{-\Phi_f}) = p(\Phi_f|\tilde{Y}, \theta_{-\Phi_f})$. This shows that the posterior ordinate in the Savage Dickey Density Ratio will be unaffected by the way we identify the last $n - k$ shocks, as long as it is based on the prior we study in this paper. Note that a similar derivation can be done to show that the prior ordinate is equal, that is $p(\Phi_r = 0|\theta_{-\Phi_r}) = p(\Phi_r^\star = 0|\theta^\star_{-\Phi_r^\star})$.

D Simulation Evidence Bayes Factor

D.1 Testing sign restrictions as overidentifying

To demonstrate the performance of the Bayes factor as a tool to test sign restrictions, we simulate data from the following *static* structural model of supply and demand:

$$\text{supply: } q_t = \alpha p_t + \varepsilon_t^s \tag{D.1}$$

$$\text{demand: } q_t = \beta p_t + \varepsilon_t^d \tag{D.2}$$

$$\text{instrument: } m_t = \phi_1 \varepsilon_t^s + \eta_t \tag{D.3}$$

Here, we can think of q_t and p_t as quantity and price, and therefore $\alpha > 0$ and $\beta < 0$ as the supply and demand (price) elasticity respectively. The last equation relates an instrument m_t to the supply shock via the measurement error equation discussed in Section 2. We calibrate the underlying parameters $\{\alpha, \beta, \phi_1, \sigma_s^2, \sigma_d^2, \sigma_\eta^2\}$ based on solving an empirical covariance matrix of $\tilde{u}_t = [u_t^{prod}, u_t^{rpo}, m_t^{k08}]$, where $[u_t^{prod}, u_t^{rpo}]$ are VAR(24) forecast errors obtained from a bivariate VAR of the (log) real oil price and (log) global oil production, and m_t is the K08 shock series also considered in Section 3.1. Solving for the structural parameters we obtain $\{\alpha = 0.07, \beta = -0.5, \phi_1 = 0.16, \sigma_s = -1.46, \sigma_d = 3.72, \sigma_\eta = 0.71\}$. However, we alter the value to $\phi_1 = 0.5$ such that the structural shock explains about 30% of the instruments variance. This guarantees that the model is well identified for the small sample sizes we consider in our simulation design ($T = 50, 100, 150$) compared to the empirical application ($T = 543$). Written as an augmented B-model, the impact matrix (normalized to increase the oil price) is then given by:

$$\tilde{B} = \begin{pmatrix} -1.28 & 0.46 & 0 \\ 2.59 & 6.60 & 0 \\ -0.5 & 0 & 0.71 \end{pmatrix}.$$

After simulating data from a multivariate normal distribution $\tilde{u}_t \sim \mathcal{N}(0, BB')$, we are going to use the methodology developed in Section 2.5 to test the simple sign restriction that b_{11} and b_{22} have the same sign. Hence, our model M_1 is a model where we identify the first shock via IV restrictions. The model under the null hypothesis imposes the additional restriction that both elements of the first column are of the same sign. Given the true impact matrix, one would expect the Bayes factors to point towards evidence against M_0 with increasing sample size. For our simulation purpose, we generate 500 datasets generated from the model of length $T = 150$, and use subsamples at both $T = 50$ and $T = 100$ to assess the effects of an increasing sample size. With respect to the prior, we use the first 15 observations as a training sample. Specifically, let $[\tilde{u}_1^{(s)}, \dots, \tilde{u}_T^{(s)}]$ be the dataset from the sth simulation. Then, we simply set $v_0 = 10$ and S_0 a diagonal matrix with entries $S_{0,ii} = \sum_{t=1}^{v_0} \left(\tilde{u}_{it}^{(s)}\right)^2$.

The resulting distribution of the Bayes factors are given in table 1. As expected, with increasing sample sizes, the Bayes factors increasingly point towards strong evidence against model M_0.

	5%	16%	50%	84%	95%
$T = 50$	-0.30	1.05	3.50	8.17	11.37
$T = 100$	0.62	2.74	6.41	12.88	Inf
$T = 150$	2.53	4.34	9.56	15.65	Inf

Table 1: Distribution of Bayes factors to test equal sign of the first column of B, using 500 simulated datasets. Note: a value of infinity (Inf) may appear in our simulations if none of the 10000 posterior draws of the unrestricted model M_1 satisfies the additional sign restrictions of model M_0.

D.2 Testing instrumental variables restrictions as overidentifying

We proceed with a small simulation exercise for the second case. Here, we are confident in a set of sign restrictions and want to test IV restrictions as overidentifying. To demonstrate the performance of the Bayes factor under a default prior, we iteratively simulate datasets from the same structural model as in D.1. However, we alternate the instrument equation to reflect varying degrees of endogeneity. One would expect the Bayes factor to detect such endogeneity and reject instrument exogeneity as soon as the models identified by sign restrictions are at odds with instrument exogeneity.

To keep the setting as close as possible to D.1, we change the instrument equation without affecting neither the variance of the instrument nor the covariance of the forecast errors. To achieve this, we rotate the last column by a orthogonal (Givens) matrix $G(\theta) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix}$ for $\theta = [\frac{1}{10}\pi, \frac{2}{10}\pi, \frac{3}{10}\pi, \frac{4}{10}\pi]$. This yields models to simulate from with an increasing degree of endogeneity. The corresponding impact matrices $\tilde{B}_1$ to $\tilde{B}_4$ are

printed at the left of the following panel:

$$\tilde{B}_1 = \begin{pmatrix} -1.28 & 0.46 & 0 \\ 2.59 & 6.60 & 0 \\ -0.48 & 0.15 & 0.70 \end{pmatrix} \qquad \tilde{B}_{1,IV} = \begin{pmatrix} -1.36 & 0.04 & 0 \\ 0.42 & 7.07 & 0 \\ -0.50 & 0 & 0.70 \end{pmatrix} \qquad \alpha_{1,IV} = 0.01, \beta_{1,IV} = -3.22 \tag{D.4}$$

$$\tilde{B}_2 = \begin{pmatrix} -1.28 & 0.46 & 0 \\ 2.59 & 6.60 & 0 \\ -0.40 & 0.29 & 0.70 \end{pmatrix} \qquad \tilde{B}_{2,IV} = \begin{pmatrix} 1.31 & -0.39 & 0 \\ 1.78 & 6.86 & 0 \\ 0.50 & 0 & 0.70 \end{pmatrix} \qquad \alpha_{2,IV} = -0.05, \beta_{2,IV} = 0.73 \tag{D.5}$$

$$\tilde{B}_3 = \begin{pmatrix} -1.28 & 0.46 & 0 \\ 2.59 & 6.60 & 0 \\ -0.29 & 0.40 & 0.70 \end{pmatrix} \qquad \tilde{B}_{3,IV} = \begin{pmatrix} 1.12 & -0.77 & 0 \\ 3.82 & 5.97 & 0 \\ 0.50 & 0 & 0.70 \end{pmatrix} \qquad \alpha_{3,IV} = -0.13, \beta_{3,IV} = 0.29 \tag{D.6}$$

$$\tilde{B}_4 = \begin{pmatrix} -1.28 & 0.46 & 0 \\ 2.59 & 6.60 & 0 \\ -0.15 & 0.48 & 0.70 \end{pmatrix} \qquad \tilde{B}_{4,IV} = \begin{pmatrix} 0.83 & -1.08 & 0 \\ 5.47 & 4.50 & 0 \\ 0.50 & 0 & 0.70 \end{pmatrix} \qquad \alpha_{4,IV} = -0.24, \beta_{4,IV} = 0.15 \tag{D.7}$$

To understand our simulation design better, we have also printed the values of $\tilde{B}_{i,IV}, \alpha_{i,IV}$ and $\beta_{i,IV}$ (for $i = 1, \ldots, 4$) which correspond to population values under the assumption of instrument exogeneity. These matrices show that as the instrument increasingly reflects variation of the second shock, wrongly assuming IV validity leads to the first column of B reflecting more and more the impact coefficient of the second shock. Hence, we expect the Bayes factor to increasingly point towards statistical evidence against model M_0.

For each of the four impact matrices, we simulate $N = 500$ datasets, each of length $T = 150$ and compute Bayes factors to compare the two models. The first (M_0) is identified via the sign restrictions that $0 < \alpha < 0.1$ and $-0.8 < \beta < 0$, while the second assumes instrument exogeneity in addition. Similar to our first simulation exercise, we use the first 15 observations as a training sample to train (a diagonal) S_0 and set $v_0 = 15$.

The resulting distribution of the Bayes factors are given in Table 1. As expected, as the instrument starts to increasingly reflect variation from the second shock, the Bayes factors point towards strong evidence against the null hypothesis of instrument exogeneity.

	5%	16%	50%	84%	95%
$\tilde{B}_1$	−2.65	−1.52	0.29	1.96	3.51
$\tilde{B}_2$	0.55	2.25	5.53	10.31	13.23
$\tilde{B}_3$	9.33	12.44	18.00	23.35	27.13
$\tilde{B}_4$	11.26	14.61	20.44	25.58	28.88

Table 2: Distribution of Bayes factors to test equal sign of the first column of B, using 500 simulated datasets of length $T = 150$.

E Interpretation of Bayes Factors

A widely acknowledged reference point for interpretation of Bayes factor magnitudes is the paper of Kass & Raftery (1995). In the following, we tabulate the main categories given therein. These are explicitly expressed as twice the natural logarithm of the Bayes factor B_{10}, as to coincide with the scale of the more familiar likelihood ratio test statistic. Note that if the log Bayes factor is negative, the statistical evidence is in favour of the null hypothesis (and against M_1). A simple redefinition as $2\ln(BF_{01}) = -2\ln(BF_{10})$ allows to stick to the reference values below.

Table 3: Categories of interpretation according to Kass & Raftery (1995)

$2\ln(BF_{10})$	B_{10}	Evidence against M_0
0 to 2	1 to 3	Not worth more than a bare mention
2 to 6	3 to 20	Positive
6 to 10	20 to 150	Strong
> 10	> 150	Very Strong

At this point, we note that a variety of other papers have used Bayes factors in combination with the reference values of Kass & Raftery (1995) to test identifying restrictions in SVARs, including Woźniak & Droumaguet (2015), Lütkepohl & Woźniak (2020), Lanne & Luoto (2020) and Nguyen (2019).

F Reconstructing and extending Kilian's oil supply shock

In Section 3.1, we have used a monthly series of oil supply shocks as in Kilian (2008). Only a quarterly time series for the 'exogenous' oil price shock from 1973Q2-2004Q3 is available on Lutz Kilian's homepage. A corresponding time series on the monthly frequency is not readily available and we would also like to use a more recent sample period. Therefore, we have reconstructed the monthly series shock series using updated oil production data from the US Energy Information Administration (see Monthly Energy Review, Table 11.1a and Table 11.b, https://www.eia.gov/totalenergy/data/monthly/index.php). As described in Kilian (2008) the construction is based on computing oil supply shortfalls based on counterfactual oil growth rates for countries that have been exposed to exogenous oil supply disruption caused e.g. by geopolitical turmoils and wars (see the Kilian paper for a precise description of the shock construction methodology). Reconstructing the series allows us to extend the shock measure to the sample 1973M02-2017M12 used in our paper. For this period, we have added two more exogenous events that affected oil production in Libya. The first event is related to the Libyan war in 2011, which led to a sharp drop of oil production. We start the counterfactual in March 2011 and it ends in April 2012. Since no other OPEC country was affected by the civil war, the benchmark of all OPEC countries'

production minus Libya. The second event was triggered in May 2013 by a series of militia attacks that started the civil unrest. Consequently, we start a second counterfactual for Libya starting in that period. Using the information from the oil market reports, it is clear that Libya never managed to resolve the civil unrest with two rival governments in the country. For this reason the counterfactual continues until the end of our sample in 2017M12. For this second event, we have removed Iran from the benchmark group in the period May 2013 to December 2015, as Iran faced international sanctions that led to problems for the oil industry. For the time between May 2016 until the end of our sample, sanctions on Iran were less stringent due to a political deal and consequently, we have included Iran in the benchmark during this period. Starting in January 2016, we have also removed Venezuela from the benchmark as this country faced its own problems related to a political and economic crisis.

The resulting shock series is shown in Figure 2. Note that transforming our shock series to quarterly frequency and comparing it with the original Kilian quarterly shock series shows a correlation of about 0.995.

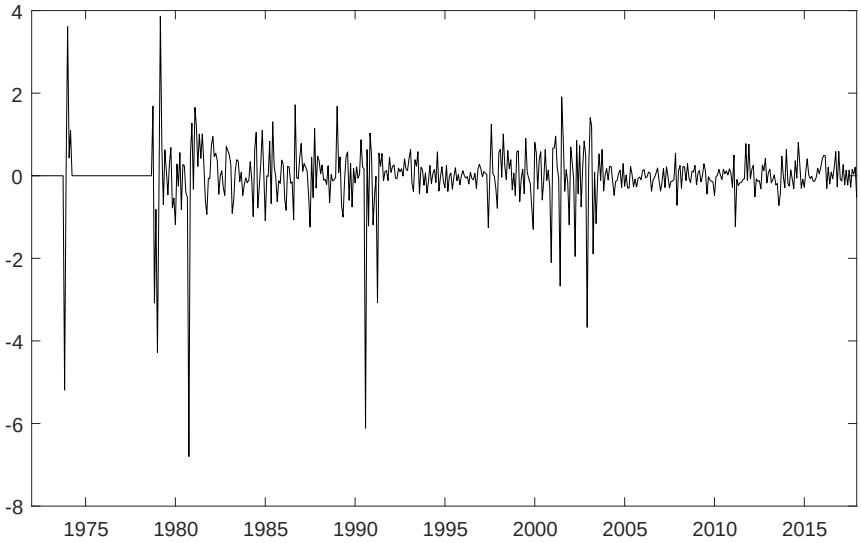

Figure 2: Exogenous oil production shortfall series as in Kilian (2008) (extended). Sample period: 1973M01 - 2017M12.

G Revisiting the oil market model of Baumeister and Hamilton (2019)

As noted in Section 3.1, Baumeister & Hamilton (2019) (BH19) express their prior beliefs on the oil price elasticities in a different way than Kilian & Murphy (2014). To assess the robustness of our empirical results to how we define indicators of the elasticity, we use this part of the appendix to repeat our oil market exercise within the model of BH19. As opposed to the B-type of model we considered in the main part of our paper, this involves a mixture of sign- and exclusion restrictions on $\mathsf{A} = B^{-1}$ plus formulation of prior densities

for the underlying parameters that can be obtained by further standardizing each row of A. We start our analysis by broadly replicating the results of BH19 using the methodology proposed in this paper. We then proceed by documenting how the results change once we use Kilian's oil production shortfall series as an IV for the SVAR supply shock.

Following BH19, we use $y_t = [100 \times \Delta q_t, 100 \times \Delta y_t, 100 \times \Delta p_t, \Delta i_t]'$, where q_t is the log of global crude oil production (in million barrels per day), y_t the log of world industrial production index, p_t is the log of the real oil price and Δi_t the proxy for OECD oil inventories expressed as a fraction of previous month's global crude oil production. As in their paper, we set $p = 12$ lags in the VAR and use a slightly updated dataset covering 1974m2 to 2019m4. Their structural oil market model (abstracting from lags and difference notation) is given by the following simultaneous equation system:

$$\begin{aligned}
\textit{Supply} \qquad & q_t = \alpha_{qp} p_t + \varepsilon_t^s, && \text{(G.1)}\\
\textit{Economic activity} \qquad & y_t = \alpha_{yp} p_t + \varepsilon_t^{ad}, && \text{(G.2)}\\
\textit{Consumption demand} \qquad & q_t - i_t^\star = \beta_{qy} y_t + \beta_{qp} p_t + \varepsilon_t^{cd}, && \text{(G.3)}\\
\textit{Inventory demand} \qquad & i_t^\star = \psi_1 q_t + \psi_3 p_t + \varepsilon_t^{id}, && \text{(G.4)}\\
\textit{Measurement error} \qquad & i_t = \chi i_t^\star + \varepsilon_t^{me}. && \text{(G.5)}
\end{aligned}$$

In this model, $\alpha_{qp} > 0$ is the (unique) oil supply elasticity, $\alpha_{yp} < 0$ is the systematic reaction of global production to oil price changes, $\beta_{qy} > 0$ the income elasticity of oil demand, $\beta_{qp} < 0$ the oil demand elasticity, and $0 < \chi < 1$ carries the interpretation of a fraction of latent oil inventories ($i_t^\star$) observed under a measurement error specification. Furthermore, the structural shocks are assumed to be mutually orthogonal with each variance $\sigma_i^2, i = 1, \dots, 4$. Written in terms of observable VAR forecast errors, the model is given by:

$$\underbrace{\begin{pmatrix} 1 & 0 & -\alpha_{qp} & 0 \\ 0 & 1 & -\alpha_{yp} & 0 \\ 1 & -\beta_{qy} & -\beta_{qp} & -\chi^{-1} \\ -\psi_1 & 0 & -\psi_3 & 1 \end{pmatrix}}_{\mathsf{A}^\star} \underbrace{\begin{pmatrix} u_t^q \\ u_t^y \\ u_t^p \\ u_t^i \end{pmatrix}}_{u_t} = \underbrace{\begin{pmatrix} \varepsilon_t^s \\ \varepsilon_t^{ea} \\ \varepsilon_t^{cd} - \chi^{-1}\varepsilon_t^{me} \\ \chi^{-1}\varepsilon_t^{id} + \varepsilon_t^{me} \end{pmatrix}}_{\tilde{\varepsilon}_t}. \qquad \text{(G.6)}$$

To further orthogonalize the latter last two shocks, BH19 premultiply the system by a matrix

$$\Gamma = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \rho & 1 \end{pmatrix},$$

where $\rho = \frac{\chi^{-1}\sigma_{me}^2}{\sigma_{cd}^2 + \chi^{-2}\sigma_{me}^2}$ which yields mutually orthogonal shocks $\varepsilon_t^{cd\star}$ and $\varepsilon_t^{id\star}$ and further transforms the last row of $\mathsf{A}^\star$. Augmented by an equation for our IV, the model is then

given by:

$$\underbrace{\begin{pmatrix} 1 & 0 & -\alpha_{qp} & 0 \\ 0 & 1 & -\alpha_{yp} & 0 \\ 1 & -\beta_{qy} & -\beta_{qp} & -\chi^{-1} \\ \psi_1^\star & \psi_2^\star & \psi_3^\star & \psi_4^\star \end{pmatrix}}_{\mathsf{A}^\star} \underbrace{\begin{pmatrix} u_t^q \\ u_t^y \\ u_t^p \\ u_t^i \end{pmatrix}}_{u_t} = \underbrace{\begin{pmatrix} \varepsilon_t^s \\ \varepsilon_t^{ea} \\ \varepsilon_t^{cd\star} \\ \varepsilon_t^{id\star} \end{pmatrix}}_{\varepsilon_t^\star}. \tag{G.7}$$

$$m_t = \phi_1 \varepsilon_t^s + \phi_2 \varepsilon_t^{ea} + \phi_3 \varepsilon_t^{cd\star} + \phi_4 \varepsilon_t^{id\star} + \eta_t \tag{G.8}$$

Here, $\psi_1^\star = \rho - \psi_1$, $\psi_2^\star = -\rho\beta_{qy}$, $\psi_3^\star = -\rho\beta_{qp} - \psi_1$ and $\psi_4^\star = -\rho\chi^{-1} + 1$. The last equation allows us to further exploit the information of the IV if further constraints are imposed on ϕ.

We compare results obtained under the following two identification schemes. In model R1 we closely follow BH19 and combine the exclusion restrictions on B^{-1} expressed in equation (G.7) with a series of prior distributions that put larger weight on *a priori* plausible structural parameters of the simultaneous equation formulation given by equation (G.7).

$$\begin{aligned} p_{R1}(\tilde{B}) \propto |\tilde{B}|^{-(v_0+n+k)} \exp\left(-\frac{1}{2}\mathrm{tr}\left(S_0\left(\tilde{B}\tilde{B}'\right)^{-1}\right)\right) \\ \times p\left(\alpha_{qp}(B)\right) p\left(\alpha_{yp}(B)\right) p(\beta_{qy}\left(B\right)) p\left(\beta_{qp}(B)\right) p\left(\chi(B)\right). \end{aligned} \tag{G.9}$$

For the exact density specifications of each parameter we refer to the paper of BH19. Note that within our methodology, α_{qp}, α_{yp}, β_{qy}, β_{qp} and χ can be obtained as a function of B, by standardizing the first three rows of B^{-1} such that unit elements are given exactly as in equation (G.7). More specifically, denote by $\mathsf{A}^\star_{1:3,\bullet}$ the first three rows of $\mathsf{A}^\star$ which contains the structural parameters used in equation (G.9). Furthermore, denote as $\mathsf{A}_{1:3,\bullet}$ the first three rows of $\mathsf{A} = B^{-1}$. Then, it holds that $\mathsf{A}^\star_{1:3,\bullet} = D^{-1}\mathsf{A}_{1:3,\bullet}$ where $D = \mathrm{diag}(\mathsf{a}_{11}, \mathsf{a}_{22}, \mathsf{a}_{31})$ and a_{ij} is the row i column j element of A.

We note that in contrast to the prior considered in Section 2, the density in (G.9) is informative about certain rotations that imply a priori reasonable structural parameters. Also, note that BH19 also specify additional priors on ρ and $\psi_{1/2}$ and determinants of $\mathsf{A}^\star$ which we do not further consider in our paper as they are not necessary to replicate the results of BH19.

We compare results from model R1 to those of a second model. In this we relax the exclusion restrictions in the first equation and instead impose IV restrictions relating the K08 shortfall series to the SVAR supply shock. The model reads then:

$$\underbrace{\begin{pmatrix} 1 & -\alpha_{qy} & -\alpha_{qp} & -\alpha_{qi} \\ 0 & 1 & -\alpha_{yp} & 0 \\ 1 & -\beta_{qy} & -\beta_{qp} & -\chi^{-1} \\ \psi_1^\star & \psi_2^\star & \psi_3^\star & \psi_4^\star \end{pmatrix}}_{\mathsf{A}^\star} \underbrace{\begin{pmatrix} u_t^q \\ u_t^y \\ u_t^p \\ u_t^i \end{pmatrix}}_{u_t} = \underbrace{\begin{pmatrix} \varepsilon_t^s \\ \varepsilon_t^{ea} \\ \varepsilon_t^{cd\star} \\ \varepsilon_t^{id\star} \end{pmatrix}}_{\varepsilon_t^\star}. \tag{G.10}$$

$$m_t = \phi_1 \varepsilon_t^s + \eta_t. \tag{G.11}$$

As for the prior in this model R2, we use the exact same density used for R1 but disregard from the additional term on α_{qp} which we would like to test for. However, similar to the exercise conducted in the main part, we maintain the sign restriction that $\alpha_{qp} > 0$. Hence, the prior is given by:

$$p_{R2}(\tilde{B}) \propto |\tilde{B}|^{-(v_0+n+k)} \exp\left(-\frac{1}{2}\text{tr}\left(S_0\left(\tilde{B}\tilde{B}'\right)^{-1}\right)\right) \times p\left(\alpha_{yp}(B)\right) p(\beta_{qy}\left(B\right)) p\left(\beta_{qp}(B)\right) p\left(\chi(B)\right). \tag{G.12}$$

For both priors, we set S_0 and v_0 via a training sample based on the first five years.

Our empirical results are summarized in Table 4. First, in Panel A we provide the posterior credibility set of the supply elasticity α_{qp}. Model R1 is designed to replicate the results of BH19 and hence finds a very similar posterior of α_{qp}. The posterior distribution of the other structural parameters α_{yp}, β_{qy}, β_{qp} and χ also match those of BH19. The median estimate suggests a fairly large value of about 0.13 in comparison to the upper bound of HR20. In turn, once we replace the exclusion restrictions of the supply equation with the IV constraints (R2), we end up with considerably smaller values of α_{qp}. The posterior is remarkably narrow given that model R2 does not use explicit prior information on α_{qp}. We proceed by testing the competing priors used in the literature (BH19 and HR20) as overidentifying. The resulting Bayes factors are given in Panel B. Similar to the analysis in the main part of this paper, the likelihood of the densities increases from prior to posterior. Hence, there is positive support in favor of using either piece of information. The differences between HR20 and BH19 are not very large, however, closely resembling our findings in the main part of the text.

Finally, we compare the variance contribution of the supply shock to oil prices in the two models. In the model designed to replicate BH19 results (R1), we find that supply shocks are fairly important drivers of oil prices, with point estimates of about one third of the variance at both impact and 2 years horizon. Using model R2 plus the BH19 prior for α_{qp}, we arrive at much smaller estimates of between 7% and 11% depending on the horizon. If we use R2 plus the HR20 restriction, similar results are obtained in terms of magnitudes, although with considerable smaller confidence sets. Overall, the findings are similar to those of Section 3.1 despite relying on different identifying assumptions and indicators of the oil price elasticity.

Table 4: Posterior distribution of supply elasticities and Bayes factors for overidentifying restrictions using the model of Baumeister & Hamilton (2019).

Panel A: Posterior of supply elasticity α_{qp}				
Model	16%	50%	84%	
R1	0.096	0.134	0.181	
R2	0.013	0.039	0.081	
Panel B: Bayes factors testing restrictions on α_{qp}				
Restrictions	$\mathrm{E}_{\theta\|\tilde{Y}}[p_2(\theta)]$	$\mathrm{E}_{\theta}[p_2(\theta)]$	$2\ln\widehat{\mathrm{BF}}_{10}$	s.e.
BH19	3.61	0.98	-2.59	0.04
HR20	0.51	0.03	-5.51	0.17
Panel C: Contribution of ε_t^s to the FEVD of the real price of oil				
Model	$h=0$		$h=24$	
R1	0.33		0.32	
	(0.23, 0.45)		(0.22, 0.44)	
R2 + BH19	0.07		0.11	
	(0.03, 0.15)		(0.06, 0.19)	
R2 + HR20	0.04		0.07	
	(0.02, 0.06)		(0.04, 0.10)	

Bayes factors computed as described in Section 2.5. Here, the less restrictive model is identified using the IV restrictions combined with prior distributions on $\mathsf{A}^\star$ with the exception of α_{qp} (R2), while the more restrictive model additionally imposes prior information on the supply elasticity α_{qp}. For BH19, $p_2(\theta) : \alpha_{qp} \sim t(0.1, 0.2, 3)$ while for HR20 $p_2(\theta) : p(\alpha_{qp} \leq 0.04) = c$ and 0 else. The FEVD of the real oil price is computed at horizon h and values in brackets indicate pointwise 68% posterior credibility sets.

H The effects of monetary policy: supplementary results

H.1 Posterior distribution for parameters of the policy rule

Table 5: Posterior distribution for parameters of the policy rule

	R1			R2			R3		
quantile	ξ_y	ξ_π	ξ_{cp}	ξ_y	ξ_π	ξ_{cp}	ξ_y	ξ_π	ξ_{cp}
5%	−0.29	−1.30	−0.03	0.06	0.34	−0.24	0.03	0.09	−0.10
50%	−0.09	−0.74	0.00	0.63	2.26	−0.01	0.26	0.84	−0.00
95%	0.10	−0.19	0.03	2.27	3.81	0.22	0.65	1.85	0.08

Posterior quantiles of the parameters governing the monetary policy rule

In this part of the appendix, we report posterior quantiles for the parameters governing the monetary policy rule (Table 5) for the three identification schemes considered in Section 3.2. Estimates based on R1 suggest that m_t is highly informative about the monetary policy rule. However, using it as an instrument yields economically implausible parameters. For instance, the 90% posterior confidence sets of ξ_π suggest that the Fed systematically cuts the policy rate in response to higher inflation, contradicting standard macroeconomic thinking. As for the second identification strategy (R2), posterior probability intervals suggest that the data is not overly informative at all. For example, the 95% quantile of ξ_π implies that in reaction to a 1% increase in prices, the central bank systematic reaction is to increase the federal funds rate by almost 4 percentage points within the same month, which is very near to the upper bound of ACR. Adding the additional restriction on the relation between the policy shock and the R&R residual (R3) substantially narrows down these credibility sets. Values between 0.03 and 0.65 for ξ_y and between 0.09 and 1.85 for ξ_π seem reasonable and are more in line with conventional estimates of a Taylor Rule (Hamilton et al. 2011).

H.2 Impact of monetary policy on financial variables

In the following we demonstrate that a sharper identification is particularly useful if we are interested in estimating the effects of monetary policy on financial variables. To this end, we add one financial variable at a time to the baseline specification and recompute IRFs to a monetary policy shock. We consider real stock prices, measured as the log of consumer price deflated S&P500 index, the mortgage spread, defined as difference between 30-year fixed rate mortgage average and the 10-year treasury yield, the commercial paper spread, defined as 3-month AA financial commercial paper rate minus the 3-months T-bill rate, and the 'excess bond premium' measure of credit market tightness developed by Gilchrist & Zakrajšek (2012). Similar to the baseline model without financial variables, we document in Figure 3 that posterior credibility sets are much tighter if we exploit information from the R&R series in addition to the sign restrictions. For instance, in a model identified by R2 not much can be said on the response of stock prices and the excess bond premium since

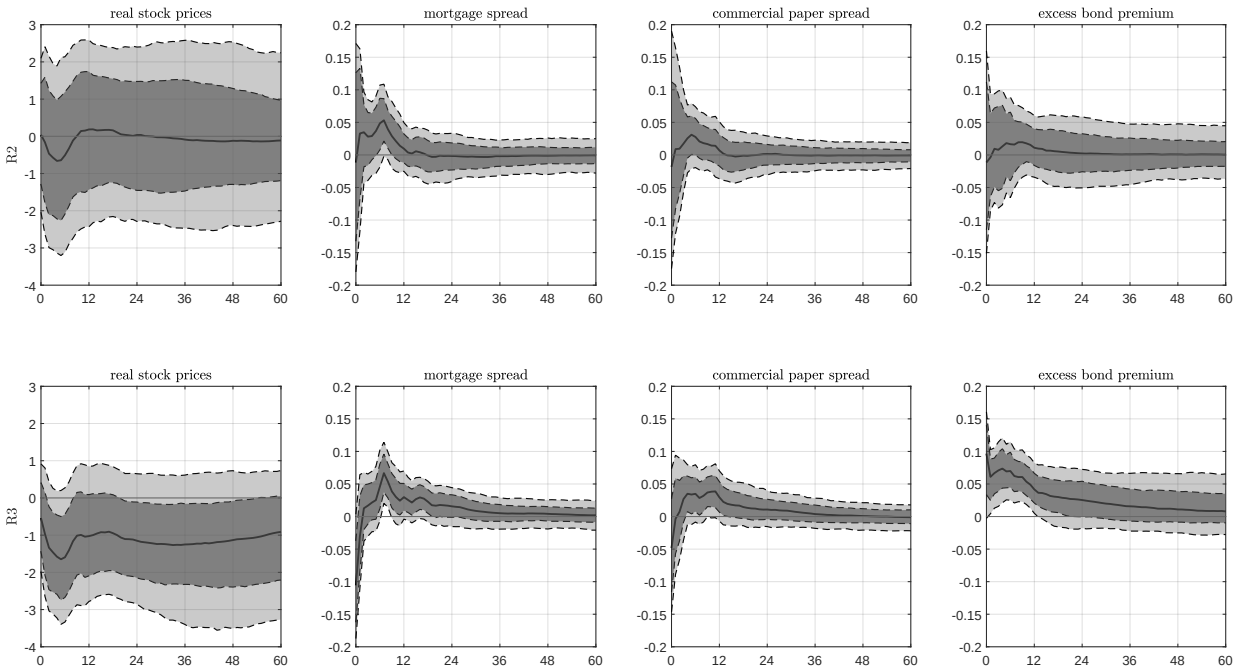

Figure 3: Impulse responses in the monetary policy SVAR augmented by one financial variable at a time. Posterior median (solid line), 68% and 90% posterior credibility sets (dotted lines). Sample periods: 1965M01-2007M12 (real stock prices, commercial paper spread), 1971M04-2007M12 (mortgage spreads), 1973M01-2007M12 (excess bond premium).

credibility sets are wide and include zero. In contrast, the picture is clearer when using R3. Here, real stock prices tend to fall and the excess bond premium responds positively. Furthermore, impulse responses are significantly different from zero, at least if judged by the 68% posterior credibility sets. A similar pattern arises for the responses of mortgage and commercial paper spreads.

References

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

Arias, J. E., Rubio-Ramírez, J. F. & Waggoner, D. F. (2021), 'Inference in Bayesian proxy-SVARs', *Journal of Econometrics* **225**(1), 88–106.

Baumeister, C. & Hamilton, J. D. (2019), 'Structural interpretation of vector autoregressions with incomplete identification: Revisiting the role of oil supply and demand shocks', *American Economic Review* **109**(5), 1873–1910.

Chib, S. & Greenberg, E. (1995), 'Understanding the Metropolis–Hastings algorithm', *The American Statistician* **49**(4), 327–335.

Geweke, J. (1992), 'Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments', *J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, Eds., Bayesian Statistics* **4**, 169–193.

Gilchrist, S. & Zakrajšek, E. (2012), 'Credit spreads and business cycle fluctuations', *American Economic Review* **102**(4), 1692–1720.

Hamilton, J. D., Pruitt, S. & Borger, S. (2011), 'Estimating the market-perceived monetary policy rule', *American Economic Journal: Macroeconomics* **3**(3), 1–28.

Kass, R. E. & Raftery, A. E. (1995), 'Bayes factors', *Journal of the American Statistical Association* **90**(430), 773–795.

Kilian, L. (2008), 'Exogenous oil supply shocks: How big are they and how much do they matter for the U.S. economy?', *The Review of Economics and Statistics* **90**(2), 216–240.

Kilian, L. & Murphy, D. P. (2014), 'The role of inventories and speculative trading in the global market for crude oil', *Journal of Applied Econometrics* **29**(3), 454–478.

Lanne, M. & Luoto, J. (2020), 'Identification of economic shocks by inequality constraints in Bayesian structural vector autoregression', *Oxford Bulletin of Economics and Statistics* **82**(2), 425–452.

Lütkepohl, H. & Woźniak, T. (2020), 'Bayesian inference for structural vector autoregressions identified by Markov-switching heteroskedasticity', *Journal of Economic Dynamics and Control* **113**, 103862.

Nguyen, L. (2019), Bayesian inference in structural vector autoregression with sign restrictions and external instruments, Technical report.

Tierney, L. (1994), 'Markov chains for exploring posterior distributions', *The Annals of Statistics* **22**(4), 1701–1728.

Woźniak, T. & Droumaguet, M. (2015), Assessing monetary policy models: Bayesian inference for heteroskedastic structural VARs, Working Paper, University of Melbourne.