Identification of Structural Vector Autoregressions by Stochastic Volatility
Appendix

Dominik Bertsche

Department of Economics, University of Konstanz

and

Robin Braun

Bank of England

Appendix A Derivations and proofs

To ensure identification of impact matrix B in model (2.1)-(2.4) we show that under sufficient heterogeneity in the second moments of the structural shocks, i.e. $r \geq K-1$, there is no $\tilde{B}$ different from B except for column permutations and sign changes which yields an observationally equivalent model with the same time-varying second moment properties in reduced form errors u_t for all $t = 1, \ldots, T$. Furthermore, for $r < K-1$, we show which parameters in impact matrix B are identified and which are not. This also yields an identification scheme for this scenario. We start with the derivation of the autocovariance function of the second moments of reduced form residuals u_t.

A.1 Autocovariance function of the second moments

The autocovariance function of the second moments of the structural shocks for $\tau > 0$ is:

$$\mathrm{Cov}\left(\mathrm{vec}\left(\varepsilon_t \varepsilon_t'\right), \mathrm{vec}\left(\varepsilon_{t+\tau}\varepsilon_{t+\tau}'\right)\right) = \left[E\left(\varepsilon_{it}\varepsilon_{jt}\varepsilon_{k,t+\tau}\varepsilon_{l,t+\tau}\right) - E\left(\varepsilon_{it}\varepsilon_{jt}\right)E\left(\varepsilon_{k,t+\tau}\varepsilon_{l,t+\tau}\right)\right]_{ijkl}.$$

Since the structural shocks are uncorrelated and have independent variance processes, the law of iterated expectations yields that the entries of this expression are only non-zero if both $i = j = k = l$ and $i \leq r$ hold for $i, j, k, l \in \{1, \ldots, K\}$. Thus, it is:

$$\mathrm{Cov}\left(\mathrm{vec}\left(\varepsilon_t \varepsilon_t'\right), \mathrm{vec}\left(\varepsilon_{t+\tau}\varepsilon_{t+\tau}'\right)\right) = G_K M_\tau G_K',$$

with G_K being a selection matrix such that $\mathrm{vec}(D) = G_K d$ for a diagonal matrix $D = \mathrm{diag}(d)$ and $M_\tau = \mathrm{diag}(\gamma_1(\tau), \ldots, \gamma_r(\tau), 0_{K-r})$ with $\gamma_i(\tau) = \exp(\sigma_{h_i}^2)(\exp(\sigma_{h_i}^2\phi_i^\tau) - 1)$ and $\sigma_{h_i}^2 = s_i/(1-\phi_i^2)$. Briefly recall that we define $\xi_t = \mathrm{vech}\left(u_t u_t'\right) = L_K \mathrm{vec}\left(u_t u_t'\right)$ (Lewis

2019). Consequently, the autocovariance function in ξ_t reads:

$$\begin{aligned}\mathrm{Cov}\left(\xi_t, \xi_{t+\tau}\right) &= L_K \mathrm{Cov}\left(\mathrm{vec}\left(u_t u_t'\right), \mathrm{vec}\left(u_{t+\tau} u_{t+\tau}'\right)\right) L_K' \\ &= L_K \left(B \otimes B\right) \mathrm{Cov}\left(\mathrm{vec}\left(\varepsilon_t \varepsilon_t'\right), \mathrm{vec}\left(\varepsilon_{t+\tau} \varepsilon_{t+\tau}'\right)\right) \left(B \otimes B\right)' L_K' \\ &= L_K \left(B \otimes B\right) G_K M_\tau G_K' \left(B \otimes B\right)' L_K'.\end{aligned}$$

A.2 Proof of Proposition 1

Proof. Suppose $\tilde{B} = BQ$ and $\tilde{\varepsilon}_t = Q^{-1}\varepsilon_t$ with $Q = \begin{pmatrix} Q_1 & Q_3 \\ Q_2 & Q_4 \end{pmatrix}$, where $Q_1 \in \mathbb{R}^{r\times r}$, $Q_2, Q_3' \in \mathbb{R}^{(K-r)\times r}$ and $Q_4 \in \mathbb{R}^{(K-r)\times(K-r)}$ define an observationally equivalent model. Hence, the log-variances $\tilde{h}_i$ of $\tilde{\varepsilon}_i$ for $i = 1, \ldots, r$ are modeled by AR(1) processes (2.4) with parameters $|\tilde{\phi}_i| < 1$, $\tilde{\phi}_i \neq 0$ and $0 < \tilde{s}_i < \infty$. Consequently, restriction (2.5) implies:

$$\mathrm{E}(u_t u_t') = BQ\tilde{V}Q'B' = BVB', \tag{A.1}$$

where $V = \mathrm{E}(V_t) = \mathrm{diag}(V_1, I_{K-r})$, $V_1 = \mathrm{diag}(\exp(\sigma^2_{h_1}/2), \ldots, \exp(\sigma^2_{h_r}/2))$, $\sigma^2_{h_i} = s_i/(1-\phi_i^2)$ and $\tilde{V}$ analogue. Since $s_i, \tilde{s}_i > 0$ for $i = 1, \ldots, r$, the diagonal elements of $V_1, \tilde{V}_1$ are nonzero why they are of full rank. The diagonality of $Q\tilde{V}Q' = V$ due to (A.1) implies:

$$Q_2\tilde{V}_1Q_1' + Q_4Q_3' = 0, \tag{A.2}$$

$$Q_2\tilde{V}_1Q_2' + Q_4Q_4' = I_{K-r}. \tag{A.3}$$

Furthermore, the autocovariance function with lag $\tau > 0$ in the second moment of the reduced form errors $\xi_t = \mathrm{vech}(u_t u_t')$ defined in (2.8) imposes:

$$\begin{aligned}\mathrm{Cov}(\xi_t, \xi_{t+\tau}) &= L_K(B \otimes B)G_K M_\tau G_K'(B \otimes B)'L_K' \\ &= L_K(\tilde{B} \otimes \tilde{B})G_K \tilde{M}_\tau G_K'(\tilde{B} \otimes \tilde{B})'L_K' \\ &= L_K(B \otimes B)(Q \otimes Q)G_K \tilde{M}_\tau G_K'(Q \otimes Q)'(B \otimes B)'L_K',\end{aligned} \tag{A.4}$$

where $M_\tau = \mathrm{diag}(\gamma_1(\tau), \ldots, \gamma_r(\tau), 0_{K-r})$ with elements $\gamma_i(\tau) = \exp(\sigma^2_{h_i})(\exp(\sigma^2_{h_i}\phi_i^\tau) - 1)$ and $\tilde{M}_\tau$ analogue for the autocovariance in $\mathrm{vec}(\tilde{\varepsilon}_t\tilde{\varepsilon}_t')$. As $s_i, \tilde{s}_i > 0$ and $\phi_i, \tilde{\phi}_i \neq 0$, it is $\gamma_i(\tau), \tilde{\gamma}_i(\tau) \neq 0$ for $i = 1, \ldots, r$. Furthermore, (A.4) implies $(Q \otimes Q)G_K\tilde{M}_\tau G_K'(Q \otimes Q)' =$

$G_K M_\tau G_K'$ what yields the following conditions:

$$\forall i = 1, \ldots, r : \sum_{l=1}^{r} q_{il}^4 \tilde{\gamma}_l(\tau) = \gamma_i(\tau) \neq 0, \tag{A.5}$$

$$\forall a_j \in \{0,1,2,3\} : \sum_{j=1}^{K} a_j = 4 : \sum_{l=1}^{r} \left(\prod_{j=1}^{K} q_{jl}^{a_j} \right) \tilde{\gamma}_l(\tau) = 0. \tag{A.6}$$

Because of (A.6), it is $\sum_{l=1}^{r} q_{\bullet l} \underbrace{q_{il}^2 q_{jl} \tilde{\gamma}_l(\tau)}_{=:\lambda_{ijl}} = 0$ for all $i, j \in \{1, \ldots, K\}$ with $i \neq j$. As Q is a full rank matrix, its column vectors $q_{\bullet l}$ are linearly independent such that $\lambda_{ijl} = 0$ for all $l \in \{1, \ldots, r\}, i, j \in \{1, \ldots, K\} : i \neq j$. As in addition $\tilde{\gamma}_l(\tau) \neq 0$, considering the first r columns of Q, i.e. matrix $(Q_1', Q_2')'$, only one element per column can be different from zero.

Because of (A.5), in each row of $r \times r$ matrix Q_1 at least one element has to be non-zero. Following the previous argument, these r non-zero entries correspond to the r non-zero entries in $(Q_1', Q_2')'$. This directly implies that Q_2 is a zero matrix and Q_1 has exactly one element different from zero per row and column.

The fact that $Q_2 = 0$ and (A.3) directly imply $Q_4 Q_4' = I_{K-r}$, i.e. Q_4 is an orthogonal matrix. Then, (A.2) yields that $Q_3 = 0$.

Since Q_1 has exactly one non-zero entry per row and column, it can be decomposed into $Q_1 = D_1 P_1 S_1$ where D_1 is diagonal with ± 1 entries, P_1 is a permutation matrix and S_1 is any diagonal matrix that rescales the columns of B. Regardless of sign switches and permutations, think of rescaled structural shocks $\tilde{\varepsilon}_{jt} = c_j \varepsilon_{jt}$. For the reduced form errors this means:

$$\begin{aligned} u_{it} &= \sum_{j=1}^{K} \tilde{b}_{ij} \tilde{\varepsilon}_{jt} = \sum_{j=1}^{K} \tilde{b}_{ij} c_j \varepsilon_{jt} = \sum_{j=1}^{K} \tilde{b}_{ij} c_j \exp\left(\frac{h_{jt}}{2}\right) \eta_{jt} \\ &= \sum_{j=1}^{K} \tilde{b}_{ij} \exp\left(\frac{h_{jt} + 2\log(c_j)}{2}\right) \eta_{jt} = \sum_{j=1}^{K} \tilde{b}_{ij} \exp\left(\frac{\tilde{h}_{jt}}{2}\right) \eta_{jt}. \end{aligned}$$

Since the log-variance process $\tilde{h}_{jt}$ is restricted to have zero mean, it is:

$$\mathrm{E}(\tilde{h}_{jt}) = \underbrace{\mathrm{E}(h_{jt})}_{=0} + 2\log(c_j) = 0 \quad \Leftrightarrow \quad c_j = 1.$$

Hence, the restriction of the log-variance process to mean zero fixes the scaling of B (Kastner et al. 2017), i.e. $S_1 = I_r$. Therefore, it is $Q_1 = D_1P_1$ and thus an orthogonal matrix why also full matrix Q is orthogonal. Moreover, it is shown that block B_1 is identified up to permutation and sign switches.

□

A.3 Proof of Corollary 1

Using Proposition 1 shows that an observationally equivalent model with the same autocovariance function in the second moment of the reduced form errors can be obtained by $\tilde{B} = BQ$ if and only if Q has the structure $\begin{pmatrix} Q_1 & 0 \\ 0 & Q_4 \end{pmatrix}$, $Q_1 = D_1P_1$ with D_1 a diagonal matrix with ± 1 entries on the diagonal, P_1 a permutation matrix and $Q_4 \in \mathbb{R}^{(K-r)\times(K-r)}$ any orthogonal matrix. Thus, the decomposition $B = (B_1, B_2)$ with $B_1 \in \mathbb{R}^{K\times r}$ and $B_2 \in \mathbb{R}^{K\times(K-r)}$ yields uniqueness of B_1 apart from multiplication of its columns by -1 and permutation. Moreover, in case that $r = K - 1$, column vector B_2 is also unique up to multiplication with -1:

Proof. For $r = K - 1$, matrix Q_4 is a scalar with $Q_4^2 = 1 \Rightarrow Q_4 = \pm 1$. So, full matrix Q can be decomposed in a diagonal matrix with ± 1 entries and a permutation matrix that has an entry of one in the very right bottom corner. This proves the uniqueness of the full matrix B apart from sign reversal of its columns and permutation of its first $r = K - 1$ columns. □

A.4 Proof of Corollary 2

Proof. Let $Q = \begin{pmatrix} Q_1 & 0 \\ 0 & Q_4 \end{pmatrix}$ be a $K \times K$ matrix such that $BQ = \begin{pmatrix} B_{11}Q_1 & B_{12}Q_4 \\ B_{21}Q_1 & B_{22}Q_4 \end{pmatrix}$ has the same structure as B, i.e. $B_{22}Q_4$ is still a lower triangular matrix. Thereby, it directly follows that Q_4 is a lower triangular matrix itself. Moreover, because Q_4 is orthogonal, it is also normal and therefore diagonal. Any diagonal and orthogonal matrix has ± 1 entries on the diagonal. So, full matrix Q can be decomposed in a diagonal matrix D having ± 1 entries and a permutation matrix P having an identity block in the lower right $(K - r) \times (K - r)$ block. Thus, matrix B is unique up to multiplication of its columns with -1 and permutation of its first r columns. □

A.5 Partial identification of A-model

Let B be unrestricted and partitioned as in Corollary 2. If BQ should imply an observationally equivalent model with the same autocovariance function in the second moment of the reduced form errors, it is $Q = \text{diag}(Q_1, Q_4)$ with $Q_1 = D_1 P_1$ with D_1 a diagonal matrix with ± 1 entries on the diagonal and P_1 a permutation matrix. For the corresponding A-model this implies:

$$A = (BQ)^{-1} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},$$

with $A_{11} = Q_1' B_{11}^{-1} + Q_1' B_{11}^{-1} B_{12}(B_{22} - B_{21} B_{11}^{-1} B_{12})^{-1} B_{21} B_{11}^{-1}$ and $A_{12} = -Q_1' B_{11}^{-1} B_{12}(B_{22} - B_{21} B_{11}^{-1} B_{12})^{-1}$ (Magnus & Neudecker 2019). Hence, the first r rows of A do not depend on Q_4 but only on $Q_1 = D_1 P_1$. Consequently, they are identified up to permutation and sign switches. The structural model in A-model form reads:

$$\mathrm{A} y_t = \tilde{\nu} + \sum_{j=1}^{p} \tilde{A}_j y_{t-j} + \varepsilon_t,$$

with $\tilde{\nu} = \mathrm{A}\nu$ and $\tilde{A}_j = \mathrm{A} A_j$ $(j = 1, \ldots, p)$. Consequently, the first r equations of the A-model are identified up to permutation and sign switches.

Appendix B Estimation

B.1 Importance density

To derive the Gaussian approximation of the (unrestricted) IS density $\pi_G(h_{1:r}|\theta, \varepsilon_{1:r})$, we closely follow the exposition of Chan & Grant (2016). We start with an application of Bayes' theorem which gives the zero variance importance density:

$$\log p(h_{1:r}|\theta, \varepsilon_i) \propto \log p(\varepsilon_{1:r}|\theta, h_{1:r}) + \log p(h_{1:r}). \tag{B.1}$$

The assumption of normality in both the transition and measurement equation gives:

$$\log p(h_{1:r}) \propto -\frac{1}{2} h_{1:r}' Q h_{1:r}, \tag{B.2}$$

$$\log p(\varepsilon_{1:r}|\theta, h_{1:r}) \propto \sum_{i=1}^{r} \sum_{t=1}^{T} -\frac{1}{2}\left(h_{it} + \varepsilon_{it}^2 e^{-h_{it}}\right). \tag{B.3}$$

Since the measurement equation is nonlinear in h_i, the normalizing constant of the smoothing distribution in equation (B.1) is not known. An approximate distribution, however, can be obtained by a second order Taylor approximation of the measurement equation (B.3). The corresponding partial derivatives are given as:

$$\begin{aligned} \frac{\partial \log p(\varepsilon_{it}|\theta, h_{it})}{\partial h_{it}} &= -\frac{1}{2} + \frac{1}{2}\varepsilon_{it}^2 e^{-h_{it}} =: f_{it} \quad \Rightarrow \quad f = (f_{1:r,1}, \ldots, f_{1:r,T})', \\ -\frac{\partial^2 \log p(\varepsilon_{it}|\theta, h_{it})}{\partial h_{it}^2} &= \frac{1}{2}\varepsilon_{it}^2 e^{-h_{it}} =: c_{it} \quad \Rightarrow \quad C = \operatorname{diag}(c_{1:r,1}, \ldots, c_{1:r,T}), \end{aligned}$$

with $f_{1:r,t} = [f_{1t}, \ldots, f_{rt}]$ and $c_{1:r,t} = [c_{1t}, \ldots, c_{rt}]$. A second order Taylor approximation around $\tilde{h}_{1:r}^{(0)}$ then yields:

$$\begin{aligned} \log p(\varepsilon_{1:r}|\theta, h_{1:r}) &\approx \log p(\varepsilon_{1:r}|\theta, \tilde{h}_{1:r}^{(0)}) + (h_{1:r} - \tilde{h}_{1:r}^{(0)})' f - \frac{1}{2}(h_{1:r} - \tilde{h}_{1:r}^{(0)})' C (h_{1:r} - \tilde{h}_{1:r}^{(0)}) \\ &\propto -\frac{1}{2}[h_{1:r}' C h_{1:r} - 2h_{1:r}' \underbrace{(f + C\tilde{h}_{1:r}^{(0)})}_{=:b}]. \end{aligned} \tag{B.4}$$

Combining (B.1), (B.2) and (B.4) provides an approximation of the smoothing distribution which takes the form of a normal kernel:

$$\log p(h_i|\theta, \varepsilon_i) \mathrel{\underset{\sim}{\propto}} -\frac{1}{2}[h_{1:r}' \underbrace{(C + Q)}_{=:\bar{Q}} h_{1:r} - 2h_{1:r}' b].$$

Consequently, the approximate smoothing density is:

$$\pi_G(h_{1:r}|\theta, \varepsilon_{1:r}) \sim \mathcal{N}\left(\bar{\delta}, \bar{Q}^{-1}\right), \qquad \text{with} \quad \bar{\delta} = \bar{Q}^{-1} b.$$

The restricted density $\pi_G^c(h_{1:r}|\theta, \varepsilon_{1:r})$ is constructed as outlined in Section 3. Note that $\pi_G^c(h_{1:r}|\theta, \varepsilon_{1:r})$ yields a good approximation only if $\tilde{h}_{1:r}^{(0)}$ is chosen appropriately. In the following, we sketch how the Newton-Raphson method is used to evaluate the IS density at the mode of the smoothing distribution (B.1).

B.2 Newton-Raphson method

The Newton-Raphson method is implemented as follows: $h_{1:r}$ is initialized by some vector $h_{1:r}^{(0)}$ satisfying the linear constraint, i.e. $A_h h_{1:r}^{(0)} = 0_{r\times 1}$. Then, $h_{1:r}^{(l)}$ is used to evaluate $\bar{Q}$, $\bar{\delta}$

and to iterate:

$$\tilde{h}_{1:r}^{(l+1)} = h_{1:r}^{(l)} + \bar{Q}^{-1}\left(-\bar{Q}h_{1:r}^{(l)} + b\right) = \bar{Q}^{-1}b = \bar{\delta},$$
$$h_{1:r}^{(l+1)} = \tilde{h}_{1:r}^{(l+1)} - \bar{Q}^{-1}A_h'\left(A_h\bar{Q}^{-1}A_h'\right)^{-1}A_h\tilde{h}_{1:r}^{(l+1)},$$

for $l \geq 0$ until convergence, i.e. until $\left\|h_{1:r}^{(l+1)} - h_{1:r}^{(l)}\right\| < \varepsilon$ holds for a specified tolerance level ε.

B.3 EM algorithms

To fix notation, define the following quantities:

$$\begin{aligned}
Y^0 &:= (y_1, \ldots, y_T) && K \times T,\\
A &:= (\nu, A_1, \ldots, A_p) && K \times (Kp+1),\\
Y_t^0 &:= \left(y_{t-1}', \ldots, y_{t-p}'\right)' && Kp \times 1,\\
x_t &:= \left(1, Y_t^{0'}\right)' && (Kp+1) \times 1,\\
X &:= (x_1, \ldots, x_T) && (Kp+1) \times T,\\
y^0 &:= \mathrm{vec}(Y^0) && KT \times 1,\\
\alpha &:= \mathrm{vec}(A) && [K(Kp+1)] \times 1,\\
U &:= (u_1, \ldots, u_T) && K \times T,\\
u &:= \mathrm{vec}(U) && KT \times 1,\\
V_{(-1)} &:= (\exp(-h_1), \ldots, \exp(-h_T)) && K \times T.
\end{aligned}$$

Using this, VAR equation (2.1) can be compactly written as:

$$y^0 = Z\alpha + u,$$

with $Z = (X' \otimes I_K)$.

This yields the following compact representation of the complete data log-likelihood:

$$
\begin{aligned}
\mathcal{L}_c(\theta) \propto & - T \ln |B| - \frac{1}{2} \left(y^0 - Z\alpha\right)' \left(I_T \otimes B^{-1}\right)' \Sigma_e^{-1} \left(I_T \otimes B^{-1}\right) \left(y^0 - Z\alpha\right) \\
& + \sum_{i=1}^{r} \left\{ -\frac{T}{2} \ln(s_i) + \frac{1}{2} \ln\left(1-\phi_i^2\right) - \frac{1}{2s_i} \left(\left[1-\phi_i^2\right] h_{i1}^2 + \sum_{t=2}^{T} \left(h_{it} - \phi_i h_{i,t-1}\right)^2 \right) \right. \\
& \left. + \frac{1}{2} \ln \left(\frac{s_i}{\left(1-\phi_i^2\right)} \frac{T\left(1-\phi_i^2\right) - 2\phi_i \left(1-\phi_i^T\right)}{T^2 \left(1-\phi_i\right)^2} \right) \right\},
\end{aligned} \tag{B.5}
$$

where $\Sigma_e^{-1} = \operatorname{diag}(\operatorname{vec}(V_{(-1)}))$ and the last term origins from the constraint imposed on the prior (see equation (3.2)). In particular, it is obtained when multiplying out $\frac{1}{2}\ln(A_h Q^{-1} A_h')$, the normalizing constant of $\pi_2(A_h h_{1:r}|\theta)$.

The EM algorithm requires starting values, which we simply set:

$$
\begin{aligned}
\hat{\alpha}^{(0)} &= \left(\left[(XX')^{-1}X\right] \otimes I_K\right) y^0, \\
\hat{\beta}^{(0)} &= S_B \operatorname{vec}\left(\hat{B}^{(0)}\right), \quad \text{with } \hat{B}^{(0)} = (T^{-1}\hat{U}\hat{U}')^{\frac{1}{2}} Q, \text{ and } \hat{U} = Y^0 - \hat{A}X,
\end{aligned}
$$

where Q is a $K \times K$ orthogonal matrix uniformly drawn from the space of K-dimensional orthogonal matrices. In case that $r < K-1$, we postmultiply Q with a fixed orthogonal matrix Q_2 that rotates $\hat{B}^{(0)}$ such that the lower right $(K-r)\times(K-r)$ block of $\hat{B}^{(0)}Q_2$ is lower triangular. Furthermore, we set the $r \times 1$ vectors:

$$
\begin{aligned}
\hat{\phi}^{(0)} &= [0.95, \ldots, 0.95]', \\
\hat{s}^{(0)} &= [0.02, \ldots, 0.02]',
\end{aligned}
$$

which correspond to persistent heteroskedasticity with initial kurtosis of about 3.7 for the estimated structural shocks $\hat{\varepsilon}_i, i = 1, \ldots, r$.

Based on starting values $\theta^{(0)} = \left[\hat{\alpha}^{(0)\prime}, \hat{\beta}^{(0)\prime}, \hat{\phi}^{(0)\prime}, \hat{s}^{(0)\prime}\right]'$, the EM algorithm iteratively cycles through the following steps for $l \geq 1$:

E-Steps

Recall that the E-step is computing the expected value of the complete data log-likelihood:

$$
Q(\theta; \theta^{(l-1)}) = \mathrm{E}_{\theta^{(l-1)}}\left[\mathcal{L}_c(\theta)\right],
$$

where the expectations are built with respect to the smoothing distribution $p(h_{1:r}|\theta^{(l-1)}, y)$. The expected complete data log-likelihood is given by:

$$\begin{aligned}
E_{\theta^{(l-1)}}[\mathcal{L}_c(\theta)] \propto & -T\ln|B| - \frac{1}{2}\left(y^0 - Z\alpha\right)'\left(I_T \otimes B^{-1}\right)' \mathrm{E}_{\theta^{(l-1)}}[\Sigma_e^{-1}]\left(I_T \otimes B^{-1}\right)\left(y^0 - Z\alpha\right) \\
& + \sum_{i=1}^{r}\left\{-\frac{T}{2}\ln(s_i) + \frac{1}{2}\ln\left(1-\phi_i^2\right) + \frac{1}{2}\ln\left(\frac{T\left(1-\phi_i^2\right) - 2\phi_i\left(1-\phi_i^T\right)}{T^2\left(1-\phi_i\right)^2}\right.\right. \\
& \left. \times\frac{s_i}{(1-\phi_i^2)}\right) - \frac{1}{2s_i}\left(\left[1-\phi_i^2\right]\mathrm{E}_{\theta^{(l-1)}}[h_{i1}^2] + \sum_{t=2}^{T}\left(\mathrm{E}_{\theta^{(l-1)}}[h_{it}^2]\right.\right. \\
& \left.\left.\left. -2\phi_i \mathrm{E}_{\theta^{(l-1)}}[h_{it}h_{i,t-1}] + \phi_i^2 \mathrm{E}_{\theta^{(l-1)}}[h_{i,t-1}^2]\right)\right)\right\}.
\end{aligned}$$

Therefore, we require computing the expectations of $\mathrm{E}_{\theta^{(l-1)}}[\Sigma_e^{-1}]$, with elements $\mathrm{E}_{\theta^{(l-1)}}[\exp(-h_{it})]$, $\mathrm{E}_{\theta^{(l-1)}}[h_{it}^2] = \mathrm{Var}_{\theta^{(l-1)}}[h_{it}] + (\mathrm{E}_{\theta^{(l-1)}}[h_{it}])^2$, and $\mathrm{E}_{\theta^{(l-1)}}[h_{it}h_{i,t-1}] = \mathrm{Cov}_{\theta^{(l-1)}}[h_{it}, h_{i,t-1}] + \mathrm{E}_{\theta^{(l-1)}}[h_{it}]\mathrm{E}_{\theta^{(l-1)}}[h_{i,t-1}]$.

1. EM-1: Here, we compute the moments based on the Gaussian approximation of the smoothing density $\pi_G^c\left(h_{1:r}|\theta^{(l-1)}, \varepsilon_{1:r}^{(l-1)}\right)$. The first two moments are given by:

$$\mathrm{E}\left(h_{1:r}|\theta^{(l-1)}, \varepsilon_{1:r}^{(l-1)}, A_h h_{1:r}{=}0\right) = \bar{\delta} - \bar{Q}^{-1}A_h'(A_h\bar{Q}^{-1}A_h')^{-1}A_h\bar{\delta}, \tag{B.6}$$

$$\mathrm{Cov}\left(h_{1:r}|\theta^{(l-1)}, \varepsilon_{1:r}^{(l-1)}, A_h h_{1:r}{=}0\right) = \bar{Q}^{-1} - \bar{Q}^{-1}A_h'(A_h\bar{Q}^{-1}A_h')^{-1}A_h\bar{Q}^{-1}. \tag{B.7}$$

   Computation of both variances $\mathrm{Var}_{\theta^{(l-1)}}[h_{it}]$ and first order autocovariances $\mathrm{Cov}_{\theta^{(l-1)}}[h_{it}, h_{i,t+1}]$ can be obtained without computing the whole inverse of $\bar{Q}$ using sparse matrix routines (Rue et al. 2009). An efficient implementation in Matlab is available at the MathWorks File Exchange (see *sparseinv* by Tim Davis). Finally, we compute:

$$\mathrm{E}_{\theta^{(l-1)}}[\exp(-h_{it})] = \exp\left(-\mathrm{E}_{\theta^{(l-1)}}[h_{it}] + 0.5\mathrm{Var}_{\theta^{(l-1)}}[h_{it}]\right), \tag{B.8}$$

   which follows from (approximate) normality of $h_{1:r}$.

2. EM-2: Here, we compute the corresponding moments by Importance Sampling. In particular, we approximate the moments by a Monte Carlo integral:

$$\mathrm{E}_{\theta^{(l-1)}}[g(h_{1:r})] = \int g(h_{1:r})p\left(h_{1:r}|\varepsilon_{1:r}^{(l-1)}, \theta^{(l-1)}\right)dh_{1:r} \approx R^{-1}\sum_{j=1}^{R} w_{(j)}g(h_{1:r}^{(j)}), \tag{B.9}$$

where $h_{1:r}^{(j)}$ is drawn from $\pi_G^c\left(h_{1:r}|\theta^{(l-1)}, \varepsilon_{1:r}^{(l-1)}\right)$ and $w_{(j)} \propto \frac{p\left(h_{1:r}^{(j)}|\theta^{(l-1)},\varepsilon_{1:r}^{(l-1)}\right)}{\pi_G^c\left(h_{1:r}^{(j)}|\theta^{(l-1)},\varepsilon_{1:r}^{(l-1)}\right)}$. Note that in practice, it is computationally more efficient to repeat the IS estimators of equation (B.9) for each $h_i, i = 1, \ldots, r$ separately. Furthermore, we recommend to start iterating with EM-2 only after EM-1 has converged. To facilitate convergence analysis, we compute the expectation always with the same underlying uniform random numbers until convergence. We find that for sample sizes typically used in macroeconomics ($T \approx 500$), one should choose $R >> 10\,000$ in order to guarantee a sufficient level of accuracy. In our application, we set $R = 50\,000$.

M-Steps

Conditional on the approximate smoothing density of log-variances h_i ($i = 1, \ldots, r$), we update parameters of both state and measurement equation of the SV-SVAR model.

1. Update ϕ_i and s_i for $i = 1, \ldots, r$:
   Conditional on the moments of the approximate smoothing density we maximize the expected value of the complete data log-likelihood (B.5) with respect to the state equation parameters. Therefore, define:

$$\nabla G(\hat{\phi}, \hat{s}) = \mathrm{E}\left[\frac{\partial \mathcal{L}_c}{\partial \phi'}, \frac{\partial \mathcal{L}_c}{\partial s'}\right]'_{\phi=\hat{\phi}, s=\hat{s}}, \qquad H(\hat{\phi}, \hat{s}) = \mathrm{E}\begin{pmatrix} \frac{\partial^2 \mathcal{L}_c}{\partial \phi \partial \phi'} & \frac{\partial^2 \mathcal{L}_c}{\partial \phi \partial s'} \\ \frac{\partial^2 \mathcal{L}_c}{\partial s \partial \phi'} & \frac{\partial^2 \mathcal{L}_c}{\partial s \partial s'} \end{pmatrix}_{\phi=\hat{\phi}, s=\hat{s}}.$$

   The detailed expressions for first and second derivatives of the expected complete data log-likelihood are printed in Section B.4. Then, set $\hat{\phi}_{(k)} = \hat{\phi}^{(l-1)}$ and $\hat{s}_{(k)} = \hat{s}^{(l-1)}$ and update parameters using Newton-Raphson, i.e. set:

$$\begin{pmatrix} \hat{\phi}_{(k+1)} \\ \hat{s}_{(k+1)} \end{pmatrix} = \begin{pmatrix} \hat{\phi}_{(k)} \\ \hat{s}_{(k)} \end{pmatrix} - \left(H\left(\hat{\phi}_{(k)}, \hat{s}_{(k)}\right)\right)^{-1} \nabla G\left(\hat{\phi}_{(k)}, \hat{s}_{(k)}\right),$$

   until $\left\| \begin{pmatrix} \hat{\phi}_{(k+1)} \\ \hat{s}_{(k+1)} \end{pmatrix} - \begin{pmatrix} \hat{\phi}_{(k)} \\ \hat{s}_{(k)} \end{pmatrix} \right\|$ is smaller than a specified threshold, e.g. 0.001. Then, set $\hat{\phi}^{(l)} = \hat{\phi}_{(k+1)}$ and $\hat{s}^{(l)} = \hat{s}_{(k+1)}$.

2. Update α. Let $Z = (X' \otimes I_K)$, then:

$$\hat{\alpha}^{(l)} = (Z'\tilde{\Sigma}_u^{-1}Z)^{-1}(Z'\tilde{\Sigma}_u^{-1}y^0),$$

   with $\tilde{\Sigma}_u^{-1} = \left(I_T \otimes \hat{B}^{(l-1)'}\right)^{-1} \hat{\Sigma}_e^{-1} \left(I_T \otimes \hat{B}^{(l-1)}\right)^{-1}$ and $\hat{\Sigma}_e^{-1} = \mathrm{E}_{\theta^{(l-1)}}[\Sigma_e^{-1}]$.

3. Update β. Recall $\beta = S_B \operatorname{vec}(B)$, define $\hat{U} = Y^0 - \hat{A}^{(l)} X$ and set:

$$\begin{aligned}\hat{\beta}^{(l)} =& \arg\max_{\beta} \mathrm{E}\left[\mathcal{L}_c(\beta)\middle|\hat{A}^{(l)}, \hat{\phi}^{(l)}, \hat{s}^{(l)}, y\right] \\ \propto& -T\ln|B| - \frac{1}{2}\operatorname{vec}(B^{-1}\hat{U})'\hat{\Sigma}_e^{-1}\operatorname{vec}(B^{-1}\hat{U}).\end{aligned}$$

Finally, set $\theta^{(l)} = \left[\hat{\alpha}^{(l)\prime}, \hat{\beta}^{(l)\prime}, \hat{\phi}^{(l)\prime}, \hat{s}^{(l)\prime}\right]'$, $l = l+1$ and return to the E-step. We iterate between E-step and M-steps until the relative change in the expected complete data log-likelihood becomes negligible. To be more precise, the algorithm is a Generalized EM algorithm since the M-step of impact matrix B depends on VAR coefficients α.

B.4 Derivatives expected complete data log-likelihood

The respective derivatives of the expected complete data log-likelihood (B.5) are given in the following. If a certain cross derivative is not stated explicitly, it is zero. Define $S_{xx,i} = \sum_{t=1}^{T-1}\mathrm{E}[h_{it}]^2 + \operatorname{Var}(h_{it})$, $S_{yy,i} = \sum_{t=2}^{T}\mathrm{E}[h_{it}]^2 + \operatorname{Var}(h_{it})$ and $S_{xy,i} = \sum_{t=2}^{T}\mathrm{E}[h_{it}]\mathrm{E}[h_{i,t-1}] + \operatorname{Cov}(h_{it}, h_{i,t-1})$, and:

$$c_i = \frac{s_i}{(1-\phi_i^2)}\frac{T\,(1-\phi_i^2) - 2\phi_i\left(1-\phi_i^T\right)}{T^2\,(1-\phi_i)^2} = \frac{s_i}{T(1-\phi_i^2)} + \frac{2s_i}{T^2(1-\phi_i^2)}\sum_{j=1}^{T-1}(T-j)\phi_i^j.$$

Then, the complete set of first and second derivatives are given by:

$$\begin{aligned}
\mathrm{E}\left[\frac{\partial\mathcal{L}_c(\theta)}{\partial s_i}\right] &= -\frac{T}{2s_i} + \frac{1-\phi_i^2}{2s_i^2}\mathrm{E}[h_{i1}^2] + \frac{1}{2s_i^2}(S_{yy,i} - 2\phi_i S_{xy,i} + \phi_i^2 S_{xx,i}) + \frac{\partial\frac{1}{2}\ln c_i}{\partial s_i},\\
\mathrm{E}\left[\frac{\partial\mathcal{L}_c(\theta)}{\partial\phi_i}\right] &= -\frac{\phi_i}{1-\phi_i^2} + \frac{\phi_i}{s_i}\mathrm{E}[h_{i1}^2] + \frac{1}{s_i}S_{xy,i} - \frac{1}{s_i}S_{xx,i}\phi_i + \frac{\partial\frac{1}{2}\ln c_i}{\partial\phi_i},\\
\mathrm{E}\left[\frac{\partial^2\mathcal{L}_c(\theta)}{\partial\phi_i\partial s_i}\right] &= -\frac{\phi_i}{s_i^2}\mathrm{E}[h_{i1}^2] - \frac{1}{s_i^2}S_{xy,i} + \frac{1}{s_i^2}S_{xx,i}\phi_i + \frac{\partial^2\frac{1}{2}\ln c_i}{\partial\phi_i\partial s_i},\\
\mathrm{E}\left[\frac{\partial^2\mathcal{L}_c(\theta)}{\partial s_i^2}\right] &= \frac{T}{2s_i^2} - \frac{1}{s_i^3}\left(S_{yy,i} - 2S_{xy,i}\phi_i + \phi_i^2 S_{xx,i} + (1-\phi_i^2)\mathrm{E}[h_{i1}^2]\right) + \frac{\partial^2\frac{1}{2}\ln c_i}{\partial s_i^2},\\
\mathrm{E}\left[\frac{\partial^2\mathcal{L}_c(\theta)}{\partial\phi_i^2}\right] &= -\frac{S_{xx,i}}{s_i} - \frac{1+\phi_i^2}{(1-\phi_i^2)^2} + \frac{\mathrm{E}[h_{i1}^2]}{s_i} + \frac{\partial^2\frac{1}{2}\ln c_i}{\partial\phi_i^2},
\end{aligned}$$

where the derivatives with respect to $\frac{1}{2}\ln c_i$ are given by:

$$
\begin{aligned}
\frac{\partial \frac{1}{2}\ln c_i}{\partial s_i} &= \frac{1}{2c_i}\frac{\partial c_i}{\partial s_i} = \frac{1}{2c_i}\left(\frac{1}{T(1-\phi_i^2)} + \frac{2}{T^2}\sum_{j=1}^{T-1}(T-j)\frac{\phi_i^j}{(1-\phi_i^2)}\right), \\
\frac{\partial \frac{1}{2}\ln c_i}{\partial \phi_i} &= \frac{1}{2c_i}\frac{\partial c_i}{\partial \phi_i} = \frac{1}{2c_i}\left(\frac{2\phi_i}{T(1-\phi_i^2)^2} + \frac{2}{T^2}\sum_{j=1}^{T-1}(T-j)\left(\frac{j\phi_i^{j-1}}{1-\phi_i^2} + 2\frac{\phi_i^{j+1}}{(1-\phi_i^2)^2}\right)\right)s_i, \\
\frac{\partial^2 \frac{1}{2}\ln c_i}{\partial \phi_i \partial s_i} &= \frac{1}{2c_i}\left(\frac{2\phi_i}{T(1-\phi_i^2)^2} + \frac{2}{T^2}\sum_{j=1}^{T-1}(T-j)\left(\frac{j\phi_i^{j-1}}{(1-\phi_i^2)} + \frac{2\phi_i^{j+1}}{(1-\phi_i^2)^2}\right)\right) - \frac{1}{2c_i^2}\frac{\partial c_i}{\partial s_i}\frac{\partial c_i}{\partial \phi_i}, \\
\frac{\partial^2 \frac{1}{2}\ln c_i}{\partial s_i^2} &= -\frac{1}{2c_i^2}\left(\frac{\partial c_i}{\partial s_i}\right)^2, \\
\frac{\partial^2 \frac{1}{2}\ln c_i}{\partial \phi_i^2} &= \frac{1}{2c_i}\frac{\partial^2 c_i}{\partial \phi_i^2} - \frac{1}{2c_i^2}\left(\frac{\partial c_i}{\partial \phi_i}\right)^2, \\
\frac{\partial^2 c_i}{\partial \phi_i^2} &= s_i\left(\frac{2(1+3\phi_i^2)}{T(1-\phi_i^2)^3}\right. \\
&\qquad \left. + \frac{2}{T^2}\sum_{j=1}^{T-1}(T-j)\left(\frac{(j-1)j\phi_i^{j-2}}{1-\phi_i^2} + \frac{2j\phi_i^j}{(1-\phi_i^2)^2} + \frac{2(j+1)\phi_i^j}{(1-\phi_i^2)^2} + \frac{8\phi_i^{j+2}}{(1-\phi_i^2)^3}\right)\right).
\end{aligned}
$$

Furthermore, let $\mathrm{E}[\Sigma_t^{-1}] = B^{-1\prime}\mathrm{E}[V_t^{-1}]B^{-1}$ whereas $\mathrm{E}[V_t^{-1}] = \mathrm{diag}\left(\mathrm{E}[V_{1t}^{-1}], I_{K-r}\right)$ and $\mathrm{E}[V_{1t}^{-1}] = \mathrm{diag}\left(\mathrm{E}[\exp(-h_{1t})], \ldots, \mathrm{E}[\exp(-h_{rt})]\right)$. Thereby, $\mathrm{E}[\exp(-h_{it})]$ is obtained using (B.8) for EM-1 and (B.9) for EM-2, respectively. Moreover, let $\beta = S_B \,\mathrm{vec}(B)$, $\alpha = \mathrm{vec}(A)$, $\tilde{X}_t = (x_t' \otimes I_K)$, such that $\mathrm{vec}(Ax_t) = \tilde{X}_t\alpha$ and $K^{(K,K)}$ be the $K^2 \times K^2$ commutation matrix. Then, the first and second derivatives of (B.5) with respect to α and β are given as:

$$
\begin{aligned}
\mathrm{E}\left[\frac{\partial \mathcal{L}_c(\theta)}{\partial \alpha'}\right] &= \left(\sum_{t=1}^{T} y_t' \mathrm{E}[\Sigma_t^{-1}]\tilde{X}_t\right) - \alpha' \left(\sum_{t=1}^{T} \tilde{X}_t' \mathrm{E}[\Sigma_t^{-1}]\tilde{X}_t\right),\\
\mathrm{E}\left[\frac{\partial \mathcal{L}_c(\theta)}{\partial \beta'}\right] &= \left[-T \operatorname{vec}\left(\left[B^{-1}\right]'\right)' + \operatorname{vec}\left(\sum_{t=1}^{T} \mathrm{E}[\Sigma_t^{-1}] u_t \varepsilon_t'\right)'\right] S_B',\\
\mathrm{E}\left[\frac{\partial^2 \mathcal{L}_c(\theta)}{\partial \alpha' \partial \beta}\right] &= -\sum_{t=1}^{T} \left[\left(\varepsilon_t' \otimes \tilde{X}_t' \mathrm{E}[\Sigma_t^{-1}]\right) + \left(\tilde{X}_t' \left[B^{-1}\right]' \otimes u_t' \mathrm{E}[\Sigma_t^{-1}]\right)\right] S_B',\\
\mathrm{E}\left[\frac{\partial^2 \mathcal{L}_c(\theta)}{\partial \alpha \partial \alpha'}\right] &= -\left(\sum_{t=1}^{T} \tilde{X}_t' \mathrm{E}[\Sigma_t^{-1}]\tilde{X}_t\right),\\
\mathrm{E}\left[\frac{\partial^2 \mathcal{L}_c(\theta)}{\partial \beta \partial \beta'}\right] &= S_B \left[T\left(B^{-1} \otimes \left[B^{-1}\right]'\right) K^{(K,K)} - \sum_{t=1}^{T} \left(\varepsilon_t \varepsilon_t' \otimes \mathrm{E}[\Sigma_t^{-1}]\right)\right.\\
&\quad \left. - \left(\sum_{t=1}^{T} \left(B^{-1} \otimes \mathrm{E}[\Sigma_t^{-1}] u_t \varepsilon_t'\right) + \left(\varepsilon_t u_t' \mathrm{E}[\Sigma_t^{-1}] \otimes \left[B^{-1}\right]'\right)\right) K^{(K,K)}\right] S_B'.
\end{aligned}
$$

B.5 Inference on impulse response functions and variance decompositions

In the following, we outline a few functions of interest in structural analysis, including different types of Impulse Response Functions (IRFs) and Forecast error variance decompositions (FEVD). We then describe a simple and generic algorithm to compute confidence intervals for any differentiable function of the SV-SVAR parameters.

Following Lütkepohl (2005), the IRFs are elements of the coefficient matrices $\Theta_i = \Phi_i B$ of the Vector Moving Average (VMA) representation of the model:

$$
y_t = \mu_y + \sum_{i=0}^{\infty} \Phi_i B \varepsilon_t, \qquad \varepsilon_t \sim (0, V_t^{\frac{1}{2}})
$$

where $\mu_y = (I_K - A_1 - \ldots - A_p)^{-1}\nu$ is the unconditional mean of y_t and $\Phi_i \in \mathbb{R}^{K \times K}$ $(i = 0, 1, \ldots)$ is a sequence of exponentially decaying matrices given as: $\Phi_i = J\mathbf{A}^i J'$ with $J = [I_K, 0, \ldots, 0]$ and:

$$
\mathbf{A} = \begin{pmatrix} A_1 & A_2 & \ldots & A_{p-1} & A_p \\ I_K & 0 & \ldots & 0 & 0 \\ 0 & I_K & & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & 0 \\ 0 & 0 & \ldots & I_K & 0 \end{pmatrix}.
$$

The elements of Θ_i, $\Theta_{jk,i}$'s are the impulse response functions in variable j to a structural innovation k after i periods. Since $\mathrm{E}[\varepsilon_t\varepsilon_t'] = V$, IRFs to shocks of size one standard deviation are given by $\tilde{\Theta}_i = \Phi_i B V^{\frac{1}{2}}$. In the empirical application, we further consider IRFs to unit shocks given by $\Theta_i^* = \Phi_i B \tilde{B}$ with $\tilde{B} = \mathrm{diag}(1_{K\times 1} \oslash b)$, $b = \mathrm{diag}(B)$ and $\oslash$ being the elementwise division, and its accumulated response $\Xi_n^* = \sum_{i=0}^{n} \Theta_i^*$. Finally, the proportion of the h-step ahead forecast error variance in variable k that is accounted for by innovations in variable j is given by $\zeta_{kj,h} = \sum_{i=0}^{h-1} (e_k' \tilde{\Theta}_i e_j)^2 / \mathrm{MSE}_k(h)$ where e_j is the j-th column of I_K and $\mathrm{MSE}_k(h) = \sum_{i=0}^{h-1} e_k' \Phi_i B V B' \Phi_i' e_k$.

In the working paper version of this article, confidence intervals (CIs) for IRFs and FEVDs were constructed via the Delta method. While computationally convenient, there is some evidence that the strong non-linearity of these functions might compromise the reliability of Delta method intervals in finite samples (Benkwitz et al. 2000). Therefore, in the following we outline an alternative procedure based on a bootstrap-type of algorithm (see also Montiel Olea et al. (forthcoming) and Mandel (2013)). This algorithm can be applied to any differentiable function of the SV-SVAR parameters without the need to work out complicated derivatives. In the following, denote by $g(\theta) : \mathbb{R}^{n_\theta} \to \mathbb{R}$ the function of interest (yielding e.g. $\Theta_{jk,i}$ or $\zeta_{kj,h}$), where n_θ is the dimension of the parameter vector. The algorithm proceeds as follows:

1. Compute the ML estimator $\hat{\theta}$ and an estimate of the asymptotic covariance matrix $\widehat{W} = \widehat{\mathcal{I}(\theta)}^{-1}$ (see section 3.3).

2. Generate M SV-SVAR models (i.i.d.) from the asymptotic distribution $\theta^{(m)} \sim \mathcal{N}(\hat{\theta}, \widehat{W}/T)$ for $m = 1, \ldots, M$.

3. Compute the α and $(1-\alpha)$ quantiles $\hat{q}_{\alpha/2}$ and $\hat{q}_{1-\alpha/2}$ of $\{g(\theta^{(m)})\}_{m=1}^{M}$.

4. Construct an $(1-\alpha)$ bootstrap-type CI as $\mathrm{CI}_{1-\alpha} = [\hat{q}_{\alpha/2}, \hat{q}_{1-\alpha/2}]$.

Appendix C Tests for identification

In this part, we quickly describe the tests for identification via heteroskedasticity proposed by Lanne & Saikkonen (2007) and Lütkepohl & Milunovich (2016). Recall the following sequence of tests:

$$H_0 : r = r_0 \qquad \text{vs} \qquad H_1 : r > r_0. \tag{C.1}$$

Suppose that r_0 is the true number of heteroskedastic errors, and separate the structural shocks $\varepsilon_t = B^{-1}u_t = (\varepsilon_{1t}', \varepsilon_{2t}')'$ into a heteroskedastic part $\varepsilon_{1t} \in \mathbb{R}^{r_0}$ and homoskedastic innovations $\varepsilon_{2t} \in \mathbb{R}^{K-r_0}$. Under the null ($r = r_0$), $\varepsilon_{2t} \sim (0, I_{K-r_0})$ is homoskedastic white noise. To test for remaining heteroskedasticity in ε_{2t}, Lanne & Saikkonen (2007) propose to use Portmanteau types of test statistics on the second moment of ε_{2t}. In particular, they construct the following time series:

$$\rho_t = \varepsilon_{2t}'\varepsilon_{2t} - T^{-1}\sum_{t=1}^{T}\varepsilon_{2t}'\varepsilon_{2t}, \tag{C.2}$$

$$\vartheta_t = \text{vech}(\varepsilon_{2t}\varepsilon_{2t}') - T^{-1}\sum_{t=1}^{T}\text{vech}(\varepsilon_{2t}\varepsilon_{2t}'), \tag{C.3}$$

with vech($\cdot$) being the half-vectorization operator as defined e.g. in Lütkepohl (2005). Based on these time series, autocovariances up to a prespecified horizon H are tested considering the following statistics:

$$Q_1(H) = T\sum_{h=1}^{H}\left(\frac{\tilde{\gamma}(h)}{\tilde{\gamma}(0)}\right)^2, \tag{C.4}$$

$$Q_2(H) = T\sum_{h=1}^{H}\text{tr}\left[\tilde{\Gamma}(h)'\tilde{\Gamma}(0)^{-1}\tilde{\Gamma}(h)\tilde{\Gamma}(0)^{-1}\right], \tag{C.5}$$

where $\tilde{\gamma}(h) = T^{-1}\sum_{t=h+1}^{T}\rho_t\rho_{t-h}$ and $\tilde{\Gamma}(h) = T^{-1}\sum_{t=h+1}^{H}\vartheta_t\vartheta_{t-h}'$. It is shown that under the null, $Q_1(H) \xrightarrow{d} \chi^2(H)$ and $Q_2(H) \xrightarrow{d} \chi^2\left(\frac{1}{4}H(K-r_0)^2(K-r_0+1)^2\right)$. Alternatively, Lütkepohl & Milunovich (2016) propose a test based on the auxiliary model:

$$\tilde{\vartheta}_t = \delta_0 + D_1\tilde{\vartheta}_{t-1} + \ldots + D_H\tilde{\vartheta}_{t-h} + \zeta_t, \tag{C.6}$$

where $\tilde{\vartheta}_t = \text{vech}(\varepsilon_{2t}\varepsilon_{2t}')$. Under the null hypothesis, $D_1 = \ldots = D_h = 0$, and a standard LM statistic is given by:

$$LM(H) = \frac{1}{2}T(K-r_0)(K-r_0+1) - T\,\text{tr}[\hat{\Sigma}_{\zeta}\tilde{\Gamma}(0)^{-1}],$$

where $\hat{\Sigma}_{\zeta}$ is the estimated residual covariance matrix from auxiliary model (C.6). Under the null, the test statistic converges to $LM(H) \xrightarrow{d} \chi^2\left(\frac{1}{4}H(K-r_0)^2(K-r_0+1)^2\right)$.

Appendix D Complementary results

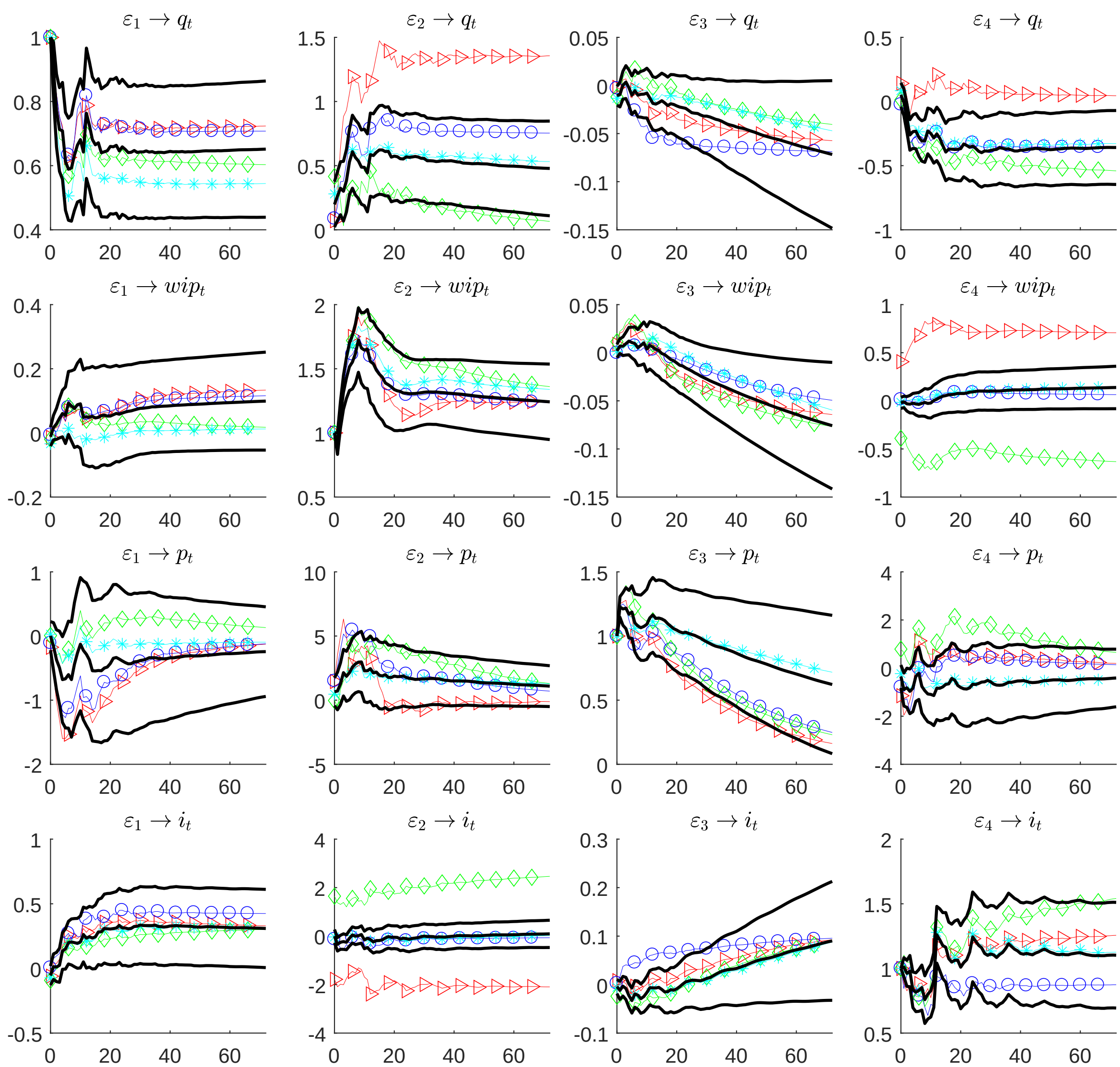

Figure 1: Complete set of standardized IRFs with 90% confidence intervals (solid lines). For comparison, we also provide estimates identified by alternative volatility estimators: MS(2) (triangle), MS(3) (circles), STVAR (diamonds) and GARCH (stars).

References

Benkwitz, A., Neumann, M. H. & Lütekpohl, H. (2000), 'Problems related to confidence intervals for impulse responses of autoregressive processes', *Econometric Reviews* **19**(1), 69–103.

Chan, J. C. C. & Grant, A. L. (2016), 'On the observed-data deviance information criterion for volatility modeling', *Journal of Financial Econometrics* **14**(4), 772–802.

Kastner, G., Frühwirth-Schnatter, S. & Lopes, H. F. (2017), 'Efficient bayesian inference for multivariate factor stochastic volatility models', *Journal of Computational and Graphical Statistics* **26**(4), 905–917.

Lanne, M. & Saikkonen, P. (2007), 'A multivariate generalized orthogonal factor GARCH model', *Journal of Business & Economic Statistics* **25**(1), 61–75.

Lewis, D. J. (2019), 'Identifying shocks via time-varying volatility', *working paper* .

Lütkepohl, H. (2005), *New Introduction to Multiple Time Series Analysis*, Springer Science & Business Media.

Lütkepohl, H. & Milunovich, G. (2016), 'Testing for identification in SVAR-GARCH models', *Journal of Economic Dynamics and Control* **73**, 241–258.

Magnus, J. R. & Neudecker, H. (2019), *Matrix Differential Calculus with Applications in Statistics and Econometrics*, John Wiley & Sons.

Mandel, M. (2013), 'Simulation-based confidence intervals for functions with complicated derivatives', *The American Statistician* **67**(2), 76–81.

Montiel Olea, J. L., Stock, J. H. & Watson, M. W. (forthcoming), 'Inference in svars identified with an external instruments', *Journal of Econometrics* .

Rue, H., Martino, S. & Chopin, N. (2009), 'Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations', *Journal of the Royal Statistical Society Series B-Statistical Methodology* **71**(2), 319–392.