

ONLINE APPENDIX:

Generalised External Instrument SVAR Analysis

Mario Forni¹, Luca Gambetti², and Giovanni Ricco³

¹*Università di Modena e Reggio Emilia, CEPR and RECent*

²*Universitat Autònoma de Barcelona, BSE, Università di Torino, CCA*

³*École Polytechnique, University of Warwick, OFCE-SciencesPo, and CEPR*

17th April 2026

Abstract

This Online Appendix provides additional results, proofs, and materials for the paper ‘Generalised External Instrument SVAR Analysis’. Section [A](#) presents a number of additional results. Section [B](#) contains all the proofs of the results reported in the main text. Sections [C](#), [D](#), and [E](#) present additional examples and simulations, respectively. Section [F](#) contains robustness exercises for the empirical section.

JEL classification: C32, E32.

Keywords: Proxy-SVAR, SVAR-IV, Impulse response functions, Variance Decomposition, Historical Decomposition, Monetary Policy Shock.

Contents

A	Additional Results	2
A.1	Bounds for absolute IRFs	2
A.2	Bounds for variance decomposition	2
A.3	A few remarks	3
B	Proofs	6
B.1	Proof of Proposition 1	6
B.2	Proof of Proposition 2	6
B.3	Proof of Proposition 3	7
B.4	Proof of Proposition 4	7
B.5	Proof of Proposition 5	7
B.6	Proof of Proposition 6	8
B.7	Proof of Proposition 7	8
B.8	Proof of Proposition 1	8
C	Examples	8
C.1	A square system	8
C.2	Example 2: A ‘short’ system	10
C.3	Example 3: FVR and FVD	11
D	Additional Simulations	13
D.1	Simulation 3: Size and power for the invertibility test	13
D.2	Simulation 4: Some guidance for lag selection	14
E	Simulations based on the fiscal foresight model	16
E.1	Simulation E1: a recoverable shock	16
E.2	Simulation E2: a comparison with the Internal-Instrument SVAR	18
E.3	Simulation E3: a nonrecoverable shock	19
E.4	Simulation E4: invertibility test	20
E.5	Simulation E5: recoverability test	22
E.6	Simulation E6: additional comparison with the Internal Instrument method	23
F	Empirics: robustness checks	25
F.1	Changing p , r and the instrument	25
F.2	Using levels and changing the VAR specification and the time span	26

A Additional Results

A.1 Bounds for absolute IRFs

Proposition 1. (Upper and lower bounds) *Letting $\bar{\alpha}^2$ and $\underline{\alpha}^2$ be the upper and the lower bound of α^2 , respectively, we have*

$$\begin{aligned}\alpha^2 &\leq \sigma_z^2 = \bar{\alpha}^2 \\ \alpha^2 &\geq \alpha^2 \sup_{\theta \in (0, \pi]} R_r^2(\theta) = \sigma_z^4 \sup_{\theta \in (0, \pi]} \psi'(e^{j\theta}) \Sigma_\varepsilon^{-1} \psi(e^{-j\theta}).\end{aligned}\tag{1}$$

The proof is reported in Section B below. In practice, the upper bound can be easily estimated as $\hat{\alpha} = \hat{\sigma}_z$; the lower bound can be estimated by taking the maximum of $\hat{\psi}'(e^{j\theta}) \hat{\Sigma}_\varepsilon^{-1} \hat{\psi}(e^{-j\theta})$ over the Fourier frequencies $\theta_f = (2\pi/T)f$, for $f = 1, \dots, T$, T being the number of observations:

$$\hat{\alpha} = \hat{\sigma}_z^2 \max_{0 < f \leq T} \sqrt{\hat{\psi}'(e^{j\theta_f}) \hat{\Sigma}_\varepsilon^{-1} \hat{\psi}(e^{-j\theta_f})}.$$

Having an estimate of the two bounds, the lower bound for the absolute impulse response functions can be estimated as

$$\bar{b}_i(L) = \hat{\gamma}(L) \hat{\sigma}_z^2 / \hat{\alpha} = \hat{\gamma}(L) \hat{\sigma}_z = \hat{C}(L) \hat{\psi}(L) \hat{\sigma}_z.\tag{2}$$

Similarly, the upper bound for the impulse response functions can be estimated as

$$\bar{b}_i(L) = \frac{\hat{\gamma}(L) \hat{\sigma}_z^2}{\hat{\alpha}} = \frac{\hat{C}(L) \hat{\psi}(L)}{\max_{0 < f \leq T} \sqrt{\hat{\psi}'(e^{j\theta_f}) \hat{\Sigma}_\varepsilon^{-1} \hat{\psi}(e^{-j\theta_f})}}.\tag{3}$$

A.2 Bounds for variance decomposition

sec:app-bounds

The upper and lower bounds for α can also be used to get upper and lower bounds for the fraction of variance accounted for by the shock of interest (see subsection 3.3).

Combining equations (26) and (19) we get

$$\alpha^2 c_h(\theta_1, \theta_2) = \frac{\sigma_z^4 \int_{\theta_1}^{\theta_2} \gamma_h(e^{-j\theta}) \gamma_h(e^{j\theta}) d\theta}{\int_{\theta_1}^{\theta_2} S_h(\theta) d\theta}.\tag{4}$$

Since α cannot be estimated, we cannot perform the variance decomposition. However, the

upper and lower bounds for α^2 provided in subsection 3.2.1 provide respectively lower and upper bounds for $c_h(\theta_1, \theta_2)$. Precisely, we can estimate the lower bound as

$$\hat{c}_h(\theta_1, \theta_2) = \hat{\sigma}_z^2 \frac{\int_{\theta_1}^{\theta_2} \hat{\gamma}_h(e^{-j\theta}) \hat{\gamma}_h(e^{j\theta}) d\theta}{\int_{\theta_1}^{\theta_2} \hat{S}_h(\theta) d\theta} \quad (5)$$

and the upper bound as

$$\hat{c}_h(\theta_1, \theta_2) = \hat{\sigma}_z^4 \frac{\int_{\theta_1}^{\theta_2} \hat{\gamma}_h(e^{-j\theta}) \hat{\gamma}_h(e^{j\theta}) d\theta}{\underline{\hat{\alpha}}^2 \int_{\theta_1}^{\theta_2} \hat{S}_h(\theta) d\theta}, \quad (6)$$

where $\theta_f = (2\pi/T)f$, $f = 1, \dots, T$.

We can also evaluate the fraction of total variance explained by the shock of interest at each frequency, i.e.

$$\hat{c}_h(\theta) = \frac{\hat{b}_{ih}(e^{-j\theta}) \hat{b}_{ih}(e^{j\theta})}{\hat{S}_h(\theta)}. \quad (7)$$

The lower and upper bounds for this frequency-by-frequency decomposition can be computed as

$$\hat{c}_h(\theta) = \hat{\sigma}_z^2 \frac{\hat{\gamma}_h(e^{-j\theta}) \hat{\gamma}_h(e^{j\theta})}{\hat{S}_h(\theta)} \quad (8)$$

$$\hat{c}_h(\theta) = \frac{\hat{\sigma}_z^4 \hat{\gamma}_h(e^{-j\theta}) \hat{\gamma}_h(e^{j\theta})}{\underline{\hat{\alpha}}^2 \hat{S}_h(\theta)}. \quad (9)$$

A.3 A few remarks

Remark 1. (Global recoverability and square systems) An immediate consequence of Proposition 2 is that, if u_t is globally recoverable, then $s_t = 0$ and $u_t = D'(F)\varepsilon_t$. Hence from (5) we see that $D'(F)$ is a left-inverse of $Q(L)$. This can be true in our setting, where $n \leq q$, only if $q = n$ (so that $D'(F)$ is the inverse of $Q(L)$). The converse is also true, i.e. if $q = n$, then u_t is recoverable. For, from (5) we get $\Sigma_\varepsilon = Q(z)Q'(z^{-1})$. Hence, if $\det Q(z)$ vanishes for $z = z^*$, then $\det Q(z^{-1})$ must have a pole in z^* , which is impossible if $|z^*| = 1$, since ε_t is stationary. Hence $Q(L)$ is invertible (possibly toward the future) and $u_{it} \in \mathcal{H}$ for any i . In conclusion, u_t is globally recoverable if and only if $q = n$.¹

Remark 2. (Global recoverability and Blaschke matrices) Lippi and Reichlin (1994), in the context of a square model, show that the structural IRFs and shocks are related to the

¹See also Chahrour and Jurado (2021).

Cholesky IRFs and shocks by a Blaschke matrix, i.e. a matrix $M(z)$ of rational functions in the complex variable z such that $M(z)M'(1/z) = I$. This can be easily seen using the above results. From (3) we see that the Cholesky representation of y_t is $y_t = [C(L)H]\eta_t$, where H is the Cholesky factor of Σ_ε , such that $HH' = \Sigma_\varepsilon$, and $\eta_t = H^{-1}\varepsilon_t$. We have seen in Remark 1 that, under global recoverability, $u_t = D'(F)\varepsilon_t$. Hence $u_t = M'(F)\eta_t$, where $M'(F) = D'(F)H$. Moreover, $M(z)$, z being a complex variable, is a Blaschke matrix. For, we have seen in Remark 1 that, under global recoverability, $D'(F)Q(L) = I_n$. Since $Q(L) = \Sigma_\varepsilon D(L)$ by (8), we have $M'(F)M(L) = M(L)M'(F) = I_n$. Hence $\eta_t = M(L)u_t$ and the structural impulse response functions are related to the Cholesky impulse-response functions by the equation $B(L) = [C(L)H]M(L)$.

Remark 3. (Recoverability measure) Let us consider the spectral density function of the projection of u_{it} onto $\mathcal{H}$, i.e. $d'_i(F)\varepsilon_t = q'_i(F)\Sigma_\varepsilon^{-1}\varepsilon_t$. We have

$$R_r^2(\theta) = d'_i(e^{j\theta})\Sigma_\varepsilon d_i(e^{-j\theta}) = q'_i(e^{j\theta})\Sigma_\varepsilon^{-1}q_i(e^{-j\theta}),$$

where j denotes the imaginary unit. This quantity represents the fraction of the total variance of u_{it} explained by the projection, decomposed by frequency, and is equal to 1 at all frequencies if u_{it} is recoverable. Correspondingly, the variance of the projection, i.e.

$$R_r^2 = \sum_{k=0}^{\infty} d'_{ik}\Sigma_\varepsilon d_{ik} = \sum_{k=0}^{\infty} q'_{ik}\Sigma_\varepsilon^{-1}q_{ik}$$

is the fraction of total variance explained by the ε 's and is equal to 1 when u_{it} is recoverable. Hence $R_r^2(\theta)$ and R_r^2 can be regarded as measures of recoverability.

Remark 4. (Global fundamentalness) An immediate consequence of Proposition 2 (ii) is that, if u_t is globally fundamental for y_t , then it is a contemporaneous linear combination of the ε 's, that is $u_t = D'\varepsilon_t$. Hence using Proposition 1 we get $Q(L) = \Sigma_\varepsilon D'$, so that $Q(L) = Q = B_0$ (since $A_0 = I_n$, see equation 5). Moreover, we see from equation (5) that $D'\varepsilon_t = u_t = D'Qu_t$, so that $D'Q = I_n$. Since we have global fundamentalness, B_0, Q and D must be square (see Remark 1), so that D' is the inverse of the matrix of the impact effects B_0 .

Remark 5. (Fundamentalness measure) In analogy with Remark 3, we can define a measure of fundamentalness as $R_f^2 = d'_i\Sigma_\varepsilon d_i$. This measure corresponds to the fundamentalness measure proposed by Forni et al. (2019).

Remark 6. (Consistency) The estimator (19) is an elementary functions of the estimators appearing on the right-hand side of the equation, so that consistency is ensured by consistency of such estimators. Notice that the population VAR is possibly of infinite order and projection (16) is necessarily of infinite order, so that, when setting p and r , we truncate such relations. The approximation works since the coefficients shrink the further you go into the past; however, consistency requires that p and r go to infinity with T at a suitable rate.

Remark 7. (An alternative estimator for the IRFs) An alternative strategy for the estimation of the relative impulse response functions is to perform the opposite projection, i.e. the OLS projection of z_t onto the present and future of ε_t , until a maximum lead r , to get an estimate of $\delta(F) = \alpha d_i(F)$ (see equation 7) and use the equality $b_i(L) = C(L)\Sigma_\varepsilon d_i(L)$ (see Formula 9). This strategy however implies estimation of a single equation with $(r+1)n$ regressors as against the $r+1$ regressors of each one of the n equations in (16). Our simulations, not reported here, confirm that the procedure proposed above performs better than this alternative. On the other hand, the alternative estimator is valid under somewhat milder conditions for the instrument; namely, we do not need z_t to be serially uncorrelated.

Remark 8. (An alternative estimator for the shock) An alternative strategy for the estimation of the shock is to use the equality $q_i(F) = \psi(F)\sigma_z/\alpha$ (see Equation 18), which, coupled with equation (10), gives $\tilde{u}_{it} = \hat{\psi}(F)\hat{\Sigma}_\varepsilon^{-1}\hat{\sigma}_z/\hat{\alpha}$. We do not follow this alternative route since our simulations, not reported here, show that the estimator $\hat{u}_{it}$ performs better than $\tilde{u}_{it}$.

Remark 9. (Measuring instrument validity) Having an estimate of the shock, we can measure the Instrument Relevance (IR), and test for it, by using the correlation coefficient of $\hat{u}_{it}$ and z_t :

$$\widehat{\text{IR}} = \text{corr}(\hat{u}_{it}, z_t).$$

Remark 10. (Estimating the fundamentalness measure) If fundamentalness does not hold, but recoverability does, we can estimate consistently the degree of fundamentalness R_f^2 (see Remark 4). From Proposition 6 we see that the fraction of the variance of u_{it} explained by the current VAR residuals is $R_f^2 = \delta'_0 \Sigma_\varepsilon \delta_0 / \sum_{k=0}^\infty \delta'_k \Sigma_\varepsilon \delta_k$. Hence we can estimate R_f^2 as

$$\hat{R}_f^2 = \hat{\delta}'_0 \hat{\Sigma}_\varepsilon \hat{\delta}_0 / \sum_{k=0}^r \hat{\delta}'_k \hat{\Sigma}_\varepsilon \hat{\delta}_k.$$

B Proofs

B.1 Proof of Proposition 1

Let $\mathcal{H}$ be the linear space spanned by the present, past and future of the shocks in ε_t : $\mathcal{H} = \overline{\text{span}}(\varepsilon_{j,t-k}, j = 1, \dots, n, k \in \mathbb{Z})$. Let $\mathcal{H}_t^+$ be the ‘half-space’ spanned by present and future ε ’s, i.e. $\mathcal{H}_t^+ = \overline{\text{span}}(\varepsilon_{j,t-k}, j = 1, \dots, n, k \leq 0)$, and $\mathcal{H}_t^-$ be the half-space spanned by present and past ε ’s, i.e. $\mathcal{H}_t^- = \overline{\text{span}}(\varepsilon_{j,t-k}, j = 1, \dots, n, k \geq 0)$. Then $\mathcal{H}$ is the direct sum of the orthogonal subspaces $\mathcal{H}_t^+$ and $\mathcal{H}_{t-1}^-$, so that, denoting with P the linear projection operator, $P(u_t|\mathcal{H}) = P(u_t|\mathcal{H}_t^+) + P(u_t|\mathcal{H}_{t-1}^-)$. But the latter projection is zero, since u_t is white noise, and therefore is orthogonal to $\varepsilon_{t-k} = Q(L)u_{t-k}$ for any $k > 0$. This establishes (i). To prove (ii) we use the cross-covariance generating function of ε_t and u_t , say $G_{\varepsilon u}(z)$, z being a complex variable.² From (5) we get $G_{\varepsilon u}(z) = Q(z)$. Now let us consider the opposite covariance, i.e. the cross-covariance generating function of u_t and ε_t : from (7) we get $G_{u\varepsilon}(z) = D'(z^{-1})\Sigma_\varepsilon$. But $G_{\varepsilon u}(z) = G'_{u\varepsilon}(z^{-1}) = \Sigma_\varepsilon D(z)$. The first equation in (8) follows. As for the second, we have just to show that Σ_ε is invertible. Now, the rank of the spectral density matrix of y_t , which is equal to $B(e^{-j\theta})B'(e^{j\theta})$, j being the imaginary unit, is equal to n everywhere in $[-\pi, \pi)$ by Assumption 1 (ii). But $B(e^{-j\theta})B'(e^{j\theta}) = C(e^{-j\theta})\Sigma_\varepsilon C'(e^{j\theta})$, so that Σ_ε has rank n . Finally, (iii) is obtained by pre-multiplying equation (5) by $C(L) = A(L)^{-1}$ and using (ii).

B.2 Proof of Proposition 2

(i) We see from (3) and (4) that $\varepsilon_{t-k} \in \mathcal{H}^y$ (see Definition 1) for any k and $y_{t-k} \in \mathcal{H}$ for any k . Hence $\mathcal{H}^y = \mathcal{H}$, so that u_{it} is recoverable if and only if it belongs to $\mathcal{H}$. It follows that, if u_{it} is recoverable, in equation (7) $s_{it} = 0$ and $u_{it} = d'_i(F)\varepsilon_t$. This proves (10). Finally, by taking the autocovariance generating function of all members of (10) we get $1 = d'_i(z^{-1})\Sigma_\varepsilon d_i(z) = q'_i(z^{-1})\Sigma_\varepsilon^{-1}q_i(z)$.

(ii) We see from (3) and (4) that ε_{t-k} , $k \geq 0$, span the same linear space as y_{t-k} , $k \geq 0$. Hence if u_{it} is fundamental it belongs to $\mathcal{H}_t^-$. On the other hand, if u_{it} is fundamental, it is also recoverable, so that, by Proposition 2 (i), we have $u_{it} = d_i(F)\varepsilon_t$. By projecting both sides of

²Let a_t and e_t be two zero-mean weakly stationary variables. Let $\Gamma_k = E(a_t b'_{t-k})$. The cross-covariance generating functions is defined as the z -transform of the cross-covariance function, i.e. $G_{ae}(z) = \sum_{k=-\infty}^{\infty} \Gamma_k z^k$. A well-known result in time series theory is the following. Let v_t be a vector white noise with covariance matrix Σ_v . Let $a_t = M(L)v_t + p_t$, where p_t is orthogonal to v_t at all leads and lags and $M(L) = \sum_{k=-\infty}^{\infty} M_k L^k$. Finally, let $e_t = N(L)v_t + s_t$, where p_t is orthogonal to v_t and p_t at all leads and lags and $N(L) = \sum_{k=-\infty}^{\infty} N_k L^k$. Then the cross-covariance generating function of a_t and e_t is given by $G_{ae}(z) = M(z)\Sigma_v N'(z^{-1})$.

this equation onto $\mathcal{H}_t^-$ we get $u_{it} = P(u_{it}|\mathcal{H}_t^-) = d_{i0}\varepsilon_t$, since $d_{ik}\varepsilon_{t+k}$ is orthogonal to $\mathcal{H}_t^-$ for $k > 0$. Finally, equation (12) is an immediate consequence of equation (9) and Proposition 2 (i).

(iii) The result is obtained by using (10) and taking the cross-covariance generating function of u_{it} and u_{jt} .

B.3 Proof of Proposition 3

Without loss of generality, let us reorder u_t in the structural representation of equation (1) in such a way that $u_t = (u_t^{f'} \ u_t^{r'} \ u_t^{n'})'$. Correspondingly, we partition $B(L)$ as $B(L) = (B^f(L) \ B^r(L) \ B^n(L))$, $Q(L)$ as $Q(L) = (Q^f \ Q^r(L) \ Q^n(L))$, and $D(L)$ as $D(L) = (D^f \ D^r(L) \ D^n(L))$. The equations in (13) are obtained from (1), (5), (9) and (12). Properties (i) and (ii) follow from Proposition 2.

B.4 Proof of Proposition 4

Notice first that z_t is a white noise process, since z_{t-k} , $k > 0$, belongs to the linear space spanned by $\tilde{z}_{t-k}$ and x_{t-k} , $k > 0$, to which z_t is orthogonal by construction, being the residual of the projection equation (14). Hence ψ_k is equal to the coefficient of the projection of ε_t onto z_{t-k} , i.e. $\psi_k = \text{cov}(\varepsilon_t, z_{t-k})/\sigma_z^2$. But w_t is orthogonal to ε_{t+k} , $k \geq 0$, by Assumption 2. It follows that ε_t is orthogonal to w_{t-k} , $k > 0$. Hence $\text{cov}(\varepsilon_t, z_{t-k}) = \alpha \text{cov}(\varepsilon_t, u_{i,t-k}) = \alpha q_{ik}$, where the last equality is obtained from equation (5). Equation (17) follows. Equation (18) is an immediate consequence of (9).

B.5 Proof of Proposition 5

From (17) and Remark 2 we get $\alpha^2 \sum_{k=0}^{\infty} q'_{ik} \Sigma_{\varepsilon}^{-1} q_{ik} = \alpha^2 R_r^2 = (\sigma_z^2)^2 \sum_{k=0}^{\infty} \psi'_k \Sigma_{\varepsilon}^{-1} \psi_k$. If u_{it} is recoverable, we have $R_r^2 = 1$, so that

$$\alpha = \sigma_z^2 \sqrt{\sum_{k=0}^{\infty} \psi'_k \Sigma_{\varepsilon}^{-1} \psi_k}. \quad (10)$$

The statement follows from Proposition 4.

B.6 Proof of Proposition 6

Let $\mathcal{H}_t^+ = \overline{\text{span}}(\varepsilon_{j,t-k}, j = 1, \dots, n, k \leq 0)$. We have $P(z_t|\mathcal{H}_t^+) = \alpha P(u_{it}|\mathcal{H}_t^+) + P(w_t|\mathcal{H}_t^+)$. But the latter projection is zero, since w_t is orthogonal to $\mathcal{H}_t^+$ by Assumption 2. Hence from equation (7) we get

$$z_t = \delta'(F)\varepsilon_t + v_t = \alpha d'_i(F)\varepsilon_t + (\alpha s_{it} + w_t).$$

where $\alpha d'_i(F)\varepsilon_t$ is the projection and $v_t = \alpha s_{it} + w_t$ is the residual. If u_{it} is recoverable, $s_{it} = 0$, so that $v_t = w_t$ and $\delta'(F)\varepsilon_t = \alpha u_{it}$. Hence

$$\alpha = \sqrt{\text{Var}(\delta'(F)\varepsilon_t)} = \sqrt{\sum_{k=0}^{\infty} \delta'_k \Sigma_{\varepsilon} \delta_k}.$$

The statement follows.

B.7 Proof of Proposition 7

Let us first notice that, if u_{it} is fundamental, ε_t is orthogonal to z_{t-k} for $k > 0$, since $u_{i,t-k}$ is a contemporaneous linear combination of ε_{t-k} (see Proposition 2 (ii)) and ε_t is a vector white noise. It follows that in equation (16) the coefficients of lagged z_t 's are all zero and $\psi(L) = \psi_0 = \psi$. Result (i) then follows from Proposition 5. For the same reason, in equation (22) the coefficients of leaded ε_t 's are all zero and $\delta(F) = \delta_0 = \delta$. Result (ii) then follows from Proposition 6.

B.8 Proof of Proposition 1

The first inequality is easily obtained from (15). As for the second, from (17) we get $(\sigma_z^2)^2 \psi'(e^{j\theta}) \Sigma_{\varepsilon}^{-1} \psi(e^{-j\theta}) = R_r^2(\theta) \alpha^2$, where $R_r^2(\theta)$ is the measure of recoverability defined in Remark 3 above. Inequality (1) follows from $R_r^2(\theta) \leq 1$.

C Examples

C.1 A square system

Consider the structural model

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = B(L)u_t = \begin{pmatrix} L & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}.$$

with $u_t \sim \mathcal{WN}(0, I_2)$. Here the system is square, i.e. $n = q = 2$. The model is noninvertible (toward the past) since $\det B(z) = 0$ for $z = 0$. Noninvertibility arises from the fact that u_{1t} has a delayed effect on y_{1t} . However, by inverting the matrix on the right-hand side toward the future we get

$$\begin{pmatrix} L^{-1} & 0 \\ -L^{-1} & 1 \end{pmatrix} \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}.$$

Hence both shocks are recoverable with respect to y_t (see Remark 1). By projecting y_t onto its past values one can find the VAR(1) representation³

$$A(L) \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} 1 & -0.5L \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}.$$

By inverting the VAR we get the MA(1) Wold representation

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = C(L)\varepsilon_t = \begin{pmatrix} 1 & 0.5L \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix},$$

where $\Sigma_\varepsilon = \begin{pmatrix} 0.5 & 0 \\ 0 & 2 \end{pmatrix}$. Hence

$$\begin{aligned} u_{1t} &= y_{1,t+1} = \varepsilon_{1,t+1} + 0.5\varepsilon_{2t} \\ u_{2t} &= -y_{1,t+1} + y_{2t} = -\varepsilon_{1,t+1} + 0.5\varepsilon_{2t} \end{aligned} \tag{11}$$

so that both u_{1t} and u_{2t} are linear combinations of the present and future values of ε_{1t} and ε_{2t} as stated in Proposition 2.

Now let us assume that

$$z_t = u_{1t} + w_t,$$

where w_t is orthogonal to u_t at all leads and lags and $\sigma_w^2 = 1$. By projecting the proxy onto the present and the future of ε_t we get $\delta(L^{-1}) = (L^{-1} \ 0.5)'$. The implied standard deviation α is 1. Hence $\delta(L) = (L \ 0.5)'$. According to Proposition 6 we get $u_{1t} = \delta'(L^{-1})\varepsilon_t = (L^{-1} \ 0.5)\varepsilon_t$, which corresponds to equation (11).

³The projection of y_{2t} on y_{t-1} is zero, so $\varepsilon_{2t} = y_{2t}$. The projection of y_{1t} onto y_{t-1} is $\frac{Var(u_{1,t-1})}{Var(y_{2,t-1})}y_{2,t-1} = 0.5y_{2,t-1}$.

Turning to the response functions, from (11) and the definition of z_t we get

$$\begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} = \begin{pmatrix} 0.5L & -0.5L \\ 1 & 1 \end{pmatrix} \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} = \begin{pmatrix} 0.25L \\ 0.5 \end{pmatrix} z_t + \begin{pmatrix} 0.25z_{t-1} - 0.5w_{t-1} - 0.5u_{2t} \\ 0.5z_t - w_t + u_{2t} \end{pmatrix}. \quad (12)$$

It can be easily verified that the second term on the right side is orthogonal to z_t and its past history, so that (12) is the projection equation of ε_t onto z_t and its lags. Hence $\psi(L) = (0.25L \ 0.5)'$. Applying formula (18) we get the first column of the matrix $B(L)$:

$$b_1(L)\alpha = b_1(L) = C(L)\psi(L)\sigma_z^2 = \begin{pmatrix} 1 & 0.5L \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0.25L \\ 0.5 \end{pmatrix} 2 = \begin{pmatrix} L \\ 1 \end{pmatrix},$$

which provides the response functions of interest.

C.2 Example 2: A ‘short’ system

Now let us modify the example above by adding a third structural shock affecting only y_{2t} :

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = B(L)u_t = \begin{pmatrix} L & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{pmatrix}.$$

with $u_t \sim \mathcal{WN}(0, I_3)$. Now the system is ‘short’, since $n = 2$ and $q = 3$. By projecting y_t onto its past values one can find the Wold representation, which is now⁴

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = C(L)\varepsilon_t = \begin{pmatrix} 1 & L/3 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

where $\Sigma_\varepsilon = \begin{pmatrix} 2/3 & 0 \\ 0 & 3 \end{pmatrix}$. Since $q > n$, we cannot have global recoverability (see Remark 1).

However, u_{1t} is still recoverable (whereas u_{2t} and u_{3t} are not), since

$$u_{1t} = y_{1,t+1} = \varepsilon_{1,t+1} + \varepsilon_{2t}/3. \quad (13)$$

Assuming again $z_t = u_{1t} + w_t$, when projecting z_t onto the present and future of ε_t we find $\delta(L^{-1}) = (L^{-1} \ 1/3)'$ and $\alpha = 1$. Hence $\delta(L) = (L \ 1/3)'$. According to (23) we get

⁴The projection of y_{2t} on y_{t-1} is zero, so $\varepsilon_{2t} = y_{2t}$. The projection of y_{1t} onto y_{t-1} is $\frac{\text{Var}(u_{1,t-1})}{\text{Var}(y_{2,t-1})}y_{2,t-1} = y_{2,t-1}/3 = \varepsilon_{2,t-1}/3$ and the innovation is $\varepsilon_{1t} = y_{1t} - \varepsilon_{2,t-1}/3$.

$u_{1t} = \delta'(L^{-1})\varepsilon_t = (L^{-1} \ 1/3)\varepsilon_t$, which corresponds to equation (13). The response functions can be found following the lines of Example 1.

Now let us assume that the shock of interest is u_{2t} , which is not recoverable, and we have the proxy $Z_t = u_{2t} + v_t$, where v_t is a unit-variance white noise process orthogonal to u_t at all leads and lags. Notice that $\sigma_Z^2 = 2$. Combining the Wold representation and the structural representation we find

$$\varepsilon_t = C(L)^{-1}B(L)u_t = \begin{pmatrix} 2L/3 & -L/3 & -L/3 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{pmatrix}.$$

Hence the projection of ε_t onto the present and past of u_{2t} is $(-L/3 \ 1)'u_{2t}$. By replacing the regressor u_{2t} with Z_t the variance of the regressor is doubled whereas the covariances are the same, so that we have an attenuation bias equal to one half and the projection is $\psi(L)Z_t = (-L/6 \ 1/2)'Z_t$. Applying formula (18) we get

$$b_2(L)\alpha = C(L)\psi(L)\sigma_z^2 = \begin{pmatrix} 1 & L/3 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} -L/6 \\ 1/2 \end{pmatrix} 2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

We can then normalise by dividing by the impact effect of the shock on the second variable, which is 1, to get the correct relative IRFs.

C.3 Example 3: FVR and FVD

As noted in the main text, the standard forecast error variance decomposition (FVD) can be computed only for globally invertible models. This is because the forecast error depends on all structural shocks and their corresponding impulse response functions. While an estimate of the IRFs of u_{it} allows us to compute the numerator of the ratio, the denominator cannot be estimated without estimating the entire structural model.

[Plagborg-Møller and Wolf \(2022\)](#) replace this denominator – i.e., the forecast error variance of the structural model – with the forecast error variance based on the present and past values of y_t , which can be estimated, and refer to this ratio as the FVR. If the model is globally invertible, the FVR and the FVD coincide. By contrast, if the shock of interest is invertible but the entire system of shocks is not, FVR and FVD differ. At horizon infinity, the two denominators are equal, since the forecast error coincides with the variable itself; however, in the short run, the past values of y_t are less informative than the structural shocks and

yield larger forecast errors. Consequently, the FVR is smaller than the FVD at short horizons (Forni et al., 2019, Section 3.4). Moreover, the FVR is not strictly a decomposition, since the sum of all contributions can be less than 100%.

The following example shows that the sum of the FVR over all shocks can be much smaller than 100% (indeed it can be any number between 0% and 100%), and, for the shock of interest, the FVR can be any number between 0% and the FVD.

Let us consider the following square structural model:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = B(L)u_t = \begin{pmatrix} 1 & 0 \\ -1 & 1 + 3L \end{pmatrix} \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix},$$

with $u_t \sim \mathcal{WN}(0, I_2)$. According to the structural model, for the information set given by the past of u_t , the one step ahead forecast error of y_{2t} is $-u_{1t} + u_{2t}$, whose variance is 2. The contribution of both shocks to the forecast error variance is $\text{FVD}(0) = 50\%$.

Since the model is square, all shocks are recoverable (see Remark 1). The shock u_{1t} is invertible, because it can be written as a linear combination of the present and past of the variables, being equal to y_{1t} . However, the model is not globally invertible, since the determinant $1 + 3L$ vanishes for $L = 1/3$, which is within the unit disk in the complex plane. The noninvertible shock is u_{2t} . The fundamental Cholesky representation is

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = B(L)u_t = \begin{pmatrix} 1 & 0 \\ -1 & 1 + L/3 \end{pmatrix} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix},$$

where

$$\varepsilon_{1t} = u_{1t} \quad \varepsilon_{2t} = \frac{1 + 3L}{1 + L/3} u_{2t} \quad \Sigma_\varepsilon = \begin{pmatrix} 1 & 0 \\ 0 & 9 \end{pmatrix}.$$

Hence, with respect to the information set given by the past of ε_t , the one step ahead forecast error is $-\varepsilon_{1t} + \varepsilon_{2t}$, whose variance is 10. The contribution to the forecast error variance, according to the ratio FVR, is $\text{FVR}(0) = 10\%$ for both shocks, since the numerator is 1 for both shocks, whereas the denominator is 10. The sum of the two contributions is just 20%, even though there are no other shocks in the system.

It is easily seen that, by replacing $1 + 3L$ with $1 + \alpha L$ in the structural MA matrix and choosing a suitable value for α we can get arbitrarily low values for $\text{FVR}(0)$, which therefore can be any number between 0% and 50%, the FVD. Indeed, the only one useful information provided by the FVR about the relative importance of the shock of interest is given by the fact that it is a lower bound for the FVD. Hence an high FVR is informative, but a low FVR

is not.

The relationship linking the VD, as suggested in the main text, to the FVD and the FVR is not straightforward. However, the VD over the entire interval $[-\pi, \pi)$ is the total variance of the variable accounted for by the shock of interest, divided by the total variance of the variable, and is therefore equivalent to both the FVD and the FVR at horizon infinity.

We recommend using the VD even when the shock is fundamental, since, as noted in the main text, the FVD requires global fundamentalness.

D Additional Simulations

D.1 Simulation 3: Size and power for the invertibility test

In this subsection, we provide results for the empirical size and power of the invertibility test. The Data Generating Process is the one reported in the main text. Results based on a different DGP are provided in Appendix E below. As for the simulations in the main text, we generate 2000 dataset.

	Empirical size (VAR I)					Empirical power (VAR III)				
	$r = 1$	$r = 2$	$r = 3$	$r = 4$	$r = 5$	$r = 1$	$r = 2$	$r = 3$	$r = 4$	$r = 5$
Sample size: $T = 240$										
$p = AIC$	6.8	8.0	8.0	6.9	6.0	86.8	88.0	83.5	75.8	71.0
$p = BIC$	5.8	6.6	5.9	5.7	5.4	82.5	83.0	77.2	69.6	63.6
Sample size: $T = 360$										
$p = AIC$	4.8	4.9	5.1	5.2	5.3	97.3	97.7	96.3	94.5	92.3
$p = BIC$	4.6	4.9	4.6	4.8	5.1	95.3	96.1	93.9	91.6	88.2
Sample size: $T = 480$										
$p = AIC$	5.9	6.0	5.3	5.4	5.3	99.7	99.8	99.5	98.9	98.4
$p = BIC$	5.4	5.0	5.1	4.9	5.4	99.1	99.3	98.5	98.0	96.5

TABLE 1: Results of Simulation 3. Generalised SVAR-EIV: Empirical size (VAR I) and power (VAR III) for the invertibility test. Percentage of rejections over 2000 replications. $T = 240, 360, 480$.

Table 1 shows the percentage of rejections for the invertibility test, applied to VAR I (left-hand side) and VAR III (right-hand side). Since the shock is invertible for VAR I, the numbers on the left side must be interpreted as the empirical size of the test. We see that the test is slightly oversized for the smaller sample size.

On the contrary, the shock is not invertible for VAR III, since such VAR is informationally deficient; hence, the numbers on the right-hand side should be interpreted as the empirical power. We see that, with this DGP, the test is more powerful for relatively small values of r ,

	SVAR-EIV, eq. (16)			SVAR-EIV, eq. (22)			SVAR-IIV
	$r = AIC$	$r = BIC$	$r = \text{LR-test}$	$r = AIC$	$r = BIC$	$r = \text{F-test}$	
proxy: $\tilde{z}_t = u_{1t}$							
$p = AIC$	81.8	80.7	82.4	67.7	79.3	72.6	59.6
$p = BIC$	80.9	63.0	83.0	63.6	62.4	66.5	69.9
proxy: $\tilde{z}_t = u_{1t} + 0.5w_t$							
$p = AIC$	102.9	88.4	107.9	85.3	87.3	92.2	68.9
$p = BIC$	99.9	69.4	106.4	76.5	69.2	81.1	75.2
proxy: $\tilde{z}_t = u_{1t} + 0.1y_{1t} + 0.5w_t$							
$p = AIC$	103.8	86.2	107.4	82.4	85.7	88.5	68.7
$p = BIC$	98.6	68.5	103.7	75.6	68.7	77.9	95.2

TABLE 2: Results of Simulation 4. Generalised SVAR-EIV: Mean Square Errors for different lag selection criteria for p and r over 2000 replications. $T = 240$. LR-test: recursive LR-test that the last lag in projection (16) is not significant at the 5% level. F-test: recursive F-test that the last lag in projection (22) is not significant at the 5% level. SVAR-IIV: Mean Square Errors for different lag selection criteria for p .

in particular $r = 2$, for all sample sizes considered. The test is very powerful for the sample sizes $T = 360, 480$.

D.2 Simulation 4: Some guidance for lag selection

This subsection provides some guidance for the lag selection. By equation (8) and Propositions 4 and 6, the two filters $\psi(L)$ and $\delta(F)$ have the same number of lags and leads, respectively. Hence we can apply information criteria to either projection (16) or (22). We consider here the sample size $T = 240$. As usual, we have 2000 replication.

Table 2 reports the results, in terms of Mean Square Error (MSE), obtained using the AIC and BIC criteria to select the VAR order p (rows) and the lag order r (columns). For the lag order r , we also report the results of the recursive tests that the coefficients of the largest lag are zero, a method sometimes used for lag selection. We employ the LR test for equation (16) and the F-test for equation (22). The instrumental variable is generated in three ways: $\tilde{z}_t = u_{1t}$, $\tilde{z}_t = u_{1t} + 0.5w_t$, and $\tilde{z}_t = u_{1t} + 0.1y_{1t-1} + 0.5w_t$, where w_t is i.i.d. Gaussian with unit variance and y_{1t} is GDP.

It is apparent from the table that using projection (22) yields better results than projection (16) in terms of mean square errors for all criteria considered. Furthermore, the best performance is obtained by using the BIC criterion for both the VAR and the projection of the VAR residuals onto the present and past values of the instrument. In contrast, the AIC criterion provides the best results for the SVAR-IIV method (last column). Accordingly, in Simulation 2, reported in the main text, we use the BIC criterion projection (22) to select p and r for our generalised SVAR-EIV and the AIC to select P for the SVAR-IIV method.

It should also be noted that the BIC is appropriate for our Simulation 2 in the main text, but not necessarily for real data. In our empirical application, for instance, the BIC suggests $r = 1$, which appears too parsimonious for monthly data. Regardless of the method used, we recommend performing a sensitivity analysis with different values of both p and r , as done in the robustness section of the main text.

E Simulations based on the fiscal foresight model

E.1 Simulation E1: a recoverable shock

In this Appendix we use a notable example of a non-invertible economy, the fiscal foresight model of [Leeper et al. \(2013\)](#) (LWY henceforth). The model is a simple Real Business Cycle model with log preferences, inelastic labor supply and two shocks: $u_{a,t}$, a technology shock, and $u_{\tau,t}$, a tax shock. A nonstandard feature of the model is the fact that the tax shocks are announced to the agents before being implemented, thus inducing fiscal foresight. The equilibrium capital accumulation is

$$k_t = \alpha k_{t-1} + a_t - \kappa \sum_{i=0}^{\infty} \theta^i E_t \tau_{t+i+1} \quad (14)$$

where $0 < \alpha < 1$, $0 < \theta < 1$, $\kappa = (1 - \theta)\tau/(1 - \tau)$, τ being the steady state tax rate, $0 \leq \tau < 1$, and a_t , k_t and τ_t are the log deviations from the steady state of technology, capital and the tax rate, respectively. Technology and taxes are given by

$$\begin{aligned} a_t &= u_{a,t} \\ \tau_t &= u_{\tau,t-2}, \end{aligned}$$

where $u_{\tau,t}$ and $u_{a,t}$ are i.i.d. shocks. Solving for k_t we obtain the following equilibrium MA representation for capital and taxes:

$$\begin{pmatrix} \tau_t \\ k_t \end{pmatrix} = \begin{pmatrix} L^2 & 0 \\ \frac{-\kappa(L + \theta)}{1 - \alpha L} & \frac{1}{1 - \alpha L} \end{pmatrix} \begin{pmatrix} u_{\tau,t} \\ u_{a,t} \end{pmatrix} = B(L)u_t. \quad (15)$$

The determinant of the above matrix vanishes for $L = 0$; hence the shocks are not fundamental. However, they are recoverable, since the system is square (see Remark 1). In particular, the tax shock is equal to taxes two periods ahead: $u_{\tau,t} = \tau_{t+2}$.

We generate 1000 different dataset with 240 time observations from model (15) using the parameterisation in [Leeper et al. \(2013\)](#): $\alpha = 0.36$, $\theta = 0.2673$ and $\tau = 0.25$ and $u_t \sim N(0, I)$. The proxy is generated according to the equation

$$\tilde{z}_t = u_{\tau,t} + 0.5z_{t-1} + 0.4k_{t-1} - 0.6\tau_{t-1} + v_t,$$

where $v_t \sim iid \mathcal{N}(0, 1)$; here the parameters are arbitrarily chosen. For each dataset, we

test for invertibility and recoverability, and estimate the tax shock along with its response functions as explained in the previous section, by setting $p = m = 2$, $r = 0$ and $p = m = 2$, $r = 4$. Invertibility is correctly rejected in all cases. Recoverability is (wrongly) rejected at the 5% level in 10% of the cases, showing that the test is somewhat oversized in small samples with this DGP. Simulations E4 and E5 below provide additional information about the empirical size and power of the invertibility and recoverability tests.

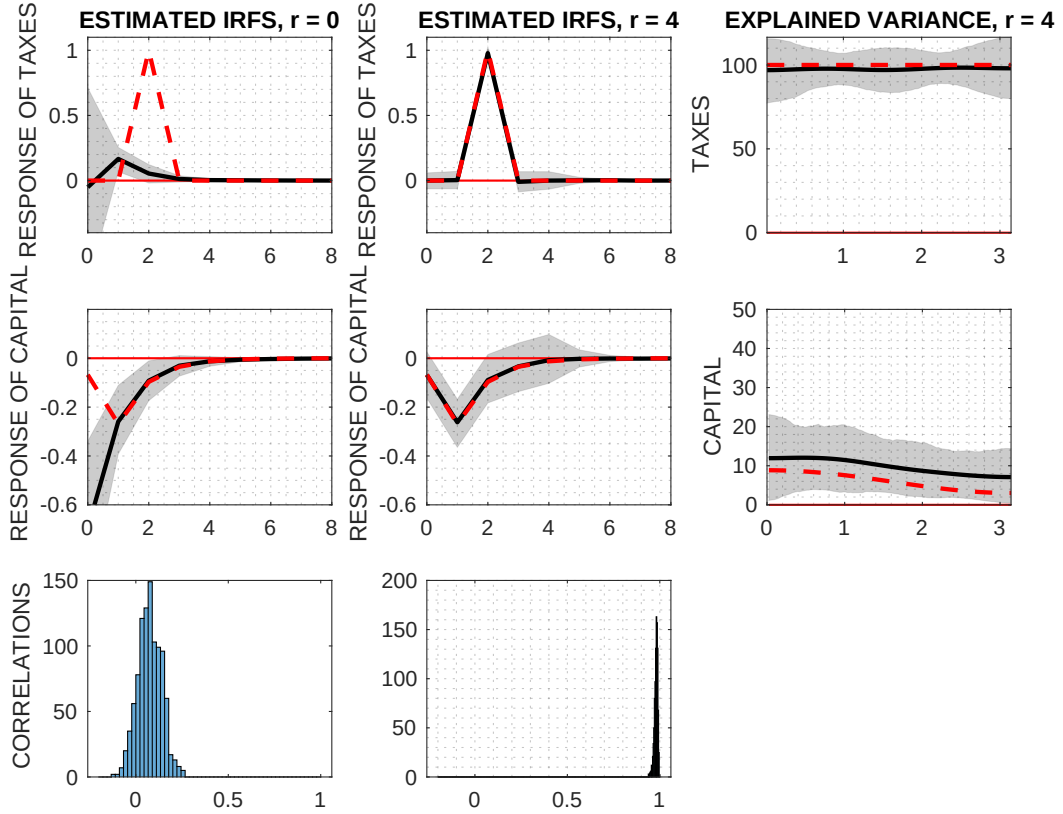


FIGURE 1: Simulation E1: LWY fiscal foresight model (recoverable shock). Left panels: IRF estimation with $r = 0$ (standard method). Middle panels: IRF estimation with $r = 4$. Right panels: estimation of the variance contribution according to equation (7), expressed in percentage terms. Upper and middle panels: red dashed line, true response functions and variance decomposition; black solid line, mean of the 1000 estimated response function and variance decompositions; grey area, 16th to 84th percentiles. Lower panels: frequency distribution of the correlation coefficients between the estimated shock and the true shock.

Estimation results are shown in Figure 1. The first column shows the estimates obtained with the standard method ($r = 0$). The upper and middle panels show the estimated response functions: the black solid line is the mean across the 1000 experiments; the grey area shows the estimates between the 16th and the 84th percentiles; the red dashed lines are the true IRFs. The bottom panel shows the frequency distribution of the correlation coefficients between the

estimated shock and the true shock. As expected, the estimates are dramatically wrong, since $r = 0$ requires invertibility, which does not hold in this case. The second column shows the estimates obtained with $r = 4$. Here the estimates are quite precise, because black and red lines are almost perfectly overlapping and the grey area is not large, particularly for taxes. The correlations coefficients between the estimated shock and the true shock are very close to 1. The third column shows the variance decomposition obtained according to equation (7), expressed in percentage terms. The estimates are fairly good, albeit somewhat less precise than those of the IRFs.

E.2 Simulation E2: a comparison with the Internal-Instrument SVAR

In this second Monte Carlo exercise we compare the small-sample performance of our proposed procedure in estimating the relative IRFs with the one of the SVAR-IIV approach, using the LWY model of Simulation E1. In particular, we want to see whether the flexibility of the Generalised SVAR-EIV translates into smaller estimation errors. The instrument is generated as in the previous exercise. We focus on the relative IRFs of the tax shocks. Since $b_{i1}(0) = 0$, the normalisation in equation (20) cannot be used. Hence we normalise the IRFs by dividing by the effect on the first variable (taxes) at lag 2, i.e. we estimate

$$\mu(L) = \frac{\gamma(L)}{\gamma_{1,2}} = \begin{pmatrix} L^2 \\ -\kappa(L + \theta) \\ 1 - \alpha L \end{pmatrix}.$$

For our procedure, the instrument is preliminarily ‘cleaned’ by setting $x_t = y_t$ and the number of lags m is set according to the BIC. For the Internal-Instrument method, we estimate a VAR for the vector $(\tilde{z}_t \ y'_t)'$; the IRFs are identified as the ones corresponding to the first shock of a Cholesky scheme. As in the previous exercise, we generate 1000 different dataset with 240 time observations. The estimation error of the two competing methods is measured as the sum of the squared errors divided by the sum of the squared coefficients of the true IRFs:

$$100 \times \frac{\sum_{h=1}^n \sum_{k=0}^K (\hat{\mu}_{hk} - \mu_{hk})^2}{\sum_{h=1}^n \sum_{k=0}^K \mu_{hk}^2}. \quad (16)$$

We set $K = 10$ and $n = 2$. This ratio is equal to 100 for the flat estimate $\hat{\mu}(L) = 0$.

Results are shown in Table 3. The first column report results for the Internal-Instrument method. The second column reports results obtained with our procedure, with r selected according to the BIC, applied to regression (16). Columns 3-7 report results obtained by setting arbitrarily $r = 2, 3, 4, 5, 6$. Boldface numbers are the best estimates of each column.

VAR order	Internal IV	External IV					
		$r = BIC$	$r = 2$	$r = 3$	$r = 4$	$r = 5$	$r = 6$
$p = 1$	396.7	4.4	4.3	5.0	6.5	7.9	9.5
$p = 2$	36.0	5.9	5.2	5.9	6.5	8.0	9.5
$p = 3$	7.6	7.4	6.0	6.7	7.3	8.0	9.6
$p = 4$	9.5	9.1	6.9	7.7	8.3	9.0	9.7
$p = 5$	11.2	10.7	7.8	8.6	9.2	9.9	10.7
$p = 6$	12.9	12.3	8.7	9.5	10.1	10.9	11.6
$p = BIC$	7.6	4.4					

TABLE 3: Results of Simulation E2. Estimation errors for the relative IRFs, measured according to (16), for the internal instrument SVAR (first column) and our proposed procedure, with r determined by the BIC (second column) and $r = 2, 3, 4, 5, 6$ (columns 3-7). Boldface numbers are the best results obtained for each column.

Let us focus first on columns 3-7. Clearly, $r = 2$ is the best choice for our DGP. Column 2 shows that the BIC criterion is successful in detecting the correct value of r , since the errors are very close to those of the best column (indeed, using just one decimal figure, we cannot see any difference).

With this DGP, the Generalised SVAR-EIV performs better than the SVAR-IIV for all values of p (small values of p seem particularly effective with our method). When using the BIC for the VAR order p (last row), our method performs better than the internal proxy SVAR and the relative improvement is large (about 40%).

A similar Monte Carlo exercise with a different DGP, but similar results, is provided below (see Simulation E6).

E.3 Simulation E3: a nonrecoverable shock

Our third Monte Carlo exercise adopts the LWY model of Simulation D1, modified by adding a demand shock that affects taxes. Precisely, the data generating process is now

$$\begin{pmatrix} \tau_t \\ k_t \end{pmatrix} = \begin{pmatrix} L^2 & 0 & \frac{1}{1-\gamma L} \\ \frac{-\kappa(L+\theta)}{1-\alpha L} & \frac{1}{1-\alpha L} & 0 \end{pmatrix} \begin{pmatrix} u_{\tau,t} \\ u_{a,t} \\ u_{d,t} \end{pmatrix} = B(L)u_t.$$

We arbitrarily set $\gamma = 0.7$. Now the system is no longer square and the tax shock is no longer recoverable. As usual, we generate 1000 different dataset with 240 time observations. As in

Simulations 1 and 2, the proxy is generated according to the equation

$$\tilde{z}_t = u_{\tau,t} + 0.5z_{t-1} + 0.4k_{t-1} - 0.6\tau_{t-1} + v_t,$$

where $v_t \sim iid\mathcal{N}(0,1)$. For all estimates we set $p = m = 2$ and $r = 4$.

First, we test for recoverability and fundamentalness. Invertibility is correctly rejected in all cases; recoverability is correctly rejected in 85.5% of the cases. Then we estimate the relative IRFs of the tax shock. Since $b_{i1}(0) = 0$, the normalisation in equation (20) cannot be used, so that we normalise by dividing by $\hat{\gamma}_{1,2}$ i.e. by imposing that the effect on taxes at lag 2 is equal to 1. In addition, we estimate upper and lower bounds for the absolute IRFs, according to equations (3) and (2). Finally, we compute the lower and upper bounds for the variance decomposition according to equations (8) and (9).

In the left panels of Figure 2, one can observe that the estimates of the relative IRFs are consistently very close to the true relative IRFs.⁵ In the middle panes are reported the estimated upper and lower bounds for the IRFs. The black solid lines are the average upper bounds; the blue solid lines are the average lower bounds; as before, the red dashed lines are the true IRFs and the shaded areas are bounded by the 16th and the 84th percentile of the IRFs distribution. Similarly, in the right panels are reported the upper and lower bounds for the explained variance, decomposed by frequency, expressed in percentage terms (equations (8)-(9)). The basic message is that upper and lower bounds provide some valuable information about the importance of the shock for the two variables.

E.4 Simulation E4: invertibility test

The goal of our fourth experiment is to evaluate the small-sample performance of the fundamentalness test. In this simulation the DGP for y_t is the ARMA(1,1) model

$$(I - AL) \begin{pmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{pmatrix} = (I - BL) \begin{pmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{pmatrix},$$

where A and B are generated as follows. First, we produce a 3×3 matrix with random entries, uniformly distributed in the interval $(0 \ 1)$. Next we divide all entries by the maximum eigenvalue. Finally, to generate A we multiply by 0.5, whereas to generate B we multiply by β , with $\beta = 0.8, 1.1, 1.2, 1.3, 1.4, 1.5$. Hence the maximum eigenvalue of A is 0.5, whereas

⁵Here we do not report results for the shock, since the shock cannot be estimated consistently.

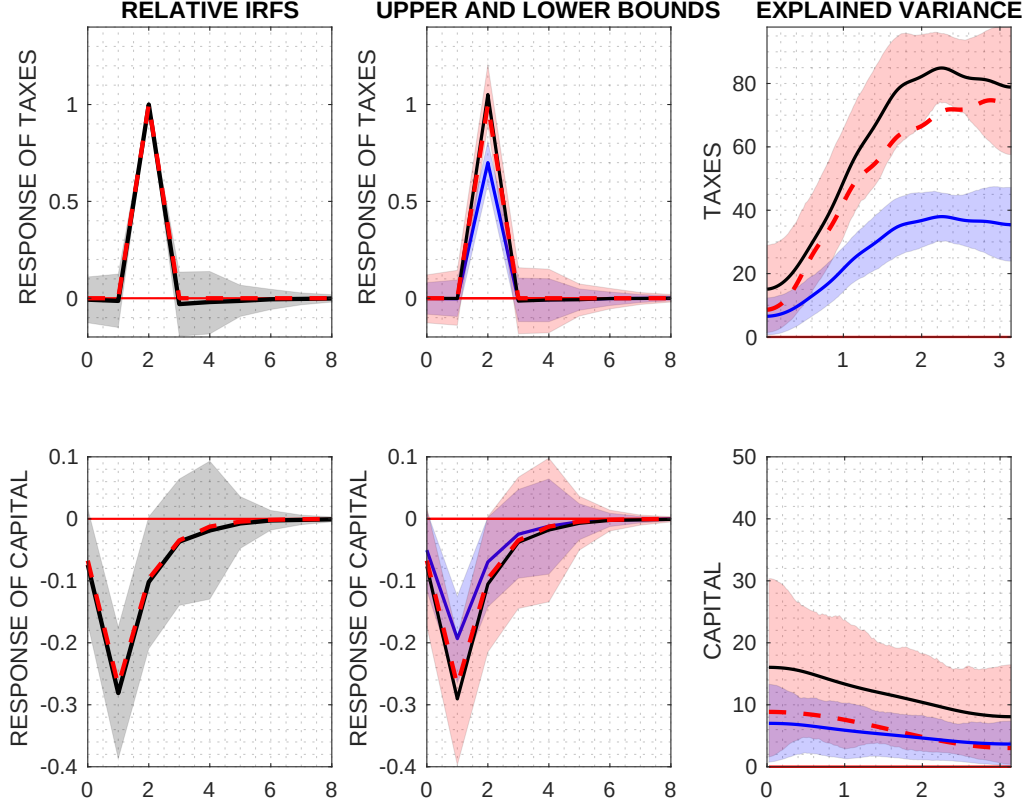


FIGURE 2: Simulation E3: Modified fiscal foresight model (nonrecoverable shock). Left panels: Estimated relative IRFs. Middle panels: estimated upper and lower bounds for the absolute IRFs. Right panels: estimated upper and lower bounds for the variance contribution according to equations (8)-(8), expressed in percentage terms. Red dashed lines: true response functions and variance decomposition; black solid lines, average of the estimated upper bounds; blue solid lines: average of the estimated lower bounds; shaded areas: 16th to 84th percentiles.

the maximum eigenvalue of B is β , implying that $I - AL$ is invertible, whereas $I - BL$ is invertible only when $\beta = 0.8$; for the other values of β it is not, and the larger is β the larger is the departure from invertibility. The shocks are again $u_t \sim iid\mathcal{N}(0, I_3)$. The proxy is generated as in Simulations 1, 2 and 3.

Again, we generate 1000 dataset with $T = 240$ for each value of β , and estimate the shock u_{1t} as described in the main text, using $p = m = 6$ and $r = 1, 2, 3, 4, 5$. Then we test for fundamentalness as explained in the main text.

Table 4 shows the percentage of rejections obtained at the 5% level with each r, β configuration. In the first column we have $\beta = 0.8$, so that we have invertibility. Hence the percentage of rejections shows the empirical size of the test, which is reasonably close to the theoretical value (5%), even if there is some evidence that the test is oversized for this DGP.

r	Fundamentalness parameter					
	$\beta = 0.8$	$\beta = 1.1$	$\beta = 1.2$	$\beta = 1.3$	$\beta = 1.4$	$\beta = 1.5$
1	6.5	24.3	48.6	64.6	79.3	90.2
2	6.7	31.2	60.2	77.9	87.9	94.8
3	6.5	33.8	63.9	78.9	89.6	94.3
4	6.1	33.2	64.0	77.1	89.8	93.9
5	6.3	32.8	61.7	75.6	87.6	92.7

TABLE 4: Results of Simulation E4. Percentage of rejection of the fundamentalness test over 1000 replications for different values of the parameter governing invertibility.

In the other columns we do not have invertibility; hence the percentage of rejections shows the power of the test. With $\beta = 1.1, 1.2$ the model is close to invertibility, so that the power is relatively low; but with $\beta = 1.3$ the test exhibits a reasonably good power, which becomes excellent for $\beta = 1.4, 1.5$. Note that the power of the test improves considerably for $r > 1$.

E.5 Simulation E5: recoverability test

In this Monte Carlo exercise, we evaluate the performance of the recoverability test. The DGP is given by the ‘short’ model

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = B(L)u_t = \begin{pmatrix} \frac{\sum_{j=0}^4 \psi_j L^j}{1 - \psi_5 L} & 0 & \beta \\ \frac{\psi_6 L}{1 - \psi_7 L} & \frac{1 + \psi_8 L}{1 - \psi_7 L} & \beta \end{pmatrix} \begin{pmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{pmatrix}.$$

The shocks are $u_t \sim iid\mathcal{N}(0, I_2)$. We set $\psi_0 = 0.1$, $\psi_1 = 0.3$, $\psi_2 = 0.5$, $\psi_3 = 1$, $\psi_4 = -0.3$, $\psi_5 = 0.7$, $\psi_6 = 1$, $\psi_7 = 0.8$, $\psi_8 = 0.5$. These parameters are set arbitrarily. Moreover, we set $\beta = \sqrt{(\sigma_1^2 + \sigma_2^2)\theta/(1 - \theta)}$, where σ_1^2 and σ_2^2 are the variances of the components of y_{1t} and y_{2t} , respectively, driven by u_{1t} and u_{2t} . In this way, the parameter θ is the fraction of the total variance of the variables explained by the shock u_{3t} , so that the larger is θ the larger is the deviation of the model from recoverability. We set $\theta = 0, 0.2, 0.4, 0.6, 0.8$ and test for serial correlation of $\hat{u}_{1t}$ using the Liung-Box Q-test with maximum lag 6, 12, 18, and 24.

Table 6 shows the percentage of rejections over 1000 experiments. In the first column we have $\theta = \beta = 0$, so that the model is square and we have recoverability (see Remark 1, Appendix A.3). Hence the percentage of rejections shows the empirical size of the test, which is somewhat undersized for this DGP (the theoretical size is 5%). In the other columns we do not have recoverability; hence the percentage of rejections shows the power of the test. Of course, with small values of θ the power is relatively low; but with $\theta \geq 0.4$ we have a fairly

	Recoverability parameter				
	$\theta = 0$	$\theta = 0.2$	$\theta = 0.4$	$\theta = 0.6$	$\theta = 0.8$
max lag = 6	2.6	69.5	78.0	87.0	93.0
max lag = 12	2.1	52.1	68.3	78.6	89.3
max lag = 18	2.7	47.5	62.1	73.9	85.4
max lag = 24	2.8	46.3	57.7	71.1	83.2

TABLE 5: Results of Simulation E5. Percentage of rejection of the recoverability test over 1000 replications for different values of the parameter governing recoverability.

good power of the test, especially when the maximum lag is set to 6.

E.6 Simulation E6: additional comparison with the Internal Instrument method

In this exercise we compare the small-sample performance of our proposed procedure in estimating the relative IRFs with the one of the Internal Instrument SVAR approach, using the same DGP as in Simulation 5, with $\theta = \beta = 0$. The instrument is generated as in the previous simulations. We focus on the relative IRFs of the first shock. We normalise the IRFs by dividing by the effect on the first variable at lag 3, i.e. we estimate

$$\mu(L) = \frac{\gamma(L)}{\gamma_{1,3}} = \left(\frac{\sum_{j=0}^4 \psi_j L^j}{\frac{\psi_3(1 - \psi_5 L)}{\psi_6 L}} \right).$$

For our procedure, the instrument is preliminarily “cleaned” by setting $x_t = y_t$ and the number of lags m according to the BIC. For the Internal-Instrument method, we estimate a VAR for the vector $(\tilde{z}_t \ y'_t)'$; the IRFs are identified as the ones corresponding to the first shock of a Cholesky scheme. As in the previous exercises, we generate 1000 different dataset with 240 time observations. The estimation errors are measured according to equation (16) in the main text.

Results are reported in Table 6. The first column report results for the Internal Instrument method. The second column reports results obtained with External Instrument method, with r selected according to the BIC, applied to equation (16). Columns 3-7 report results obtained by setting arbitrarily $r = 3, 4, 5, 6, 7$. Boldface numbers are the best estimates of each column.

Let us focus first on columns 3-7. Clearly, $r = 4$ is the best choice for this DGP. Column 2 shows that the BIC criterion is fairly successful in detecting the correct value of r , since the

VAR order	Internal IV	External IV					
		$r = BIC$	$r = 3$	$r = 4$	$r = 5$	$r = 6$	$r = 7$
$p = 1$	255.2	12.2	21.6	11.9	12.0	12.6	13.1
$p = 2$	55.9	10.8	18.2	10.0	10.6	11.7	12.7
$p = 3$	42.6	12.8	15.7	12.0	12.3	12.7	13.4
$p = 4$	12.6	13.3	14.6	11.9	12.1	12.7	13.4
$p = 5$	14.2	14.5	14.9	12.5	12.9	13.4	13.9
$p = 6$	14.7	14.8	15.0	12.7	13.0	13.7	14.2
$p = BIC$	13.5	13.9					

TABLE 6: Results of Simulation E6. Estimation errors for the relative IRFs, measured according to (16), for the Internal-Instrument SVAR (first column) and our proposed procedure, with r determined by the BIC (second column) and $r = 3, 4, 5, 6, 7$ (columns 3-7). Boldface numbers are the best results obtained for each column.

errors are close to those of the best column.

The results of Simulation E2 are confirmed. The Generalised External Instrument SVAR performs similarly or better than the Internal Instrument SVAR for all values of p .

F Empirics: robustness checks

F.1 Changing p , r and the instrument

In this subsection, we explore the robustness of our baseline VAR to changes of the number

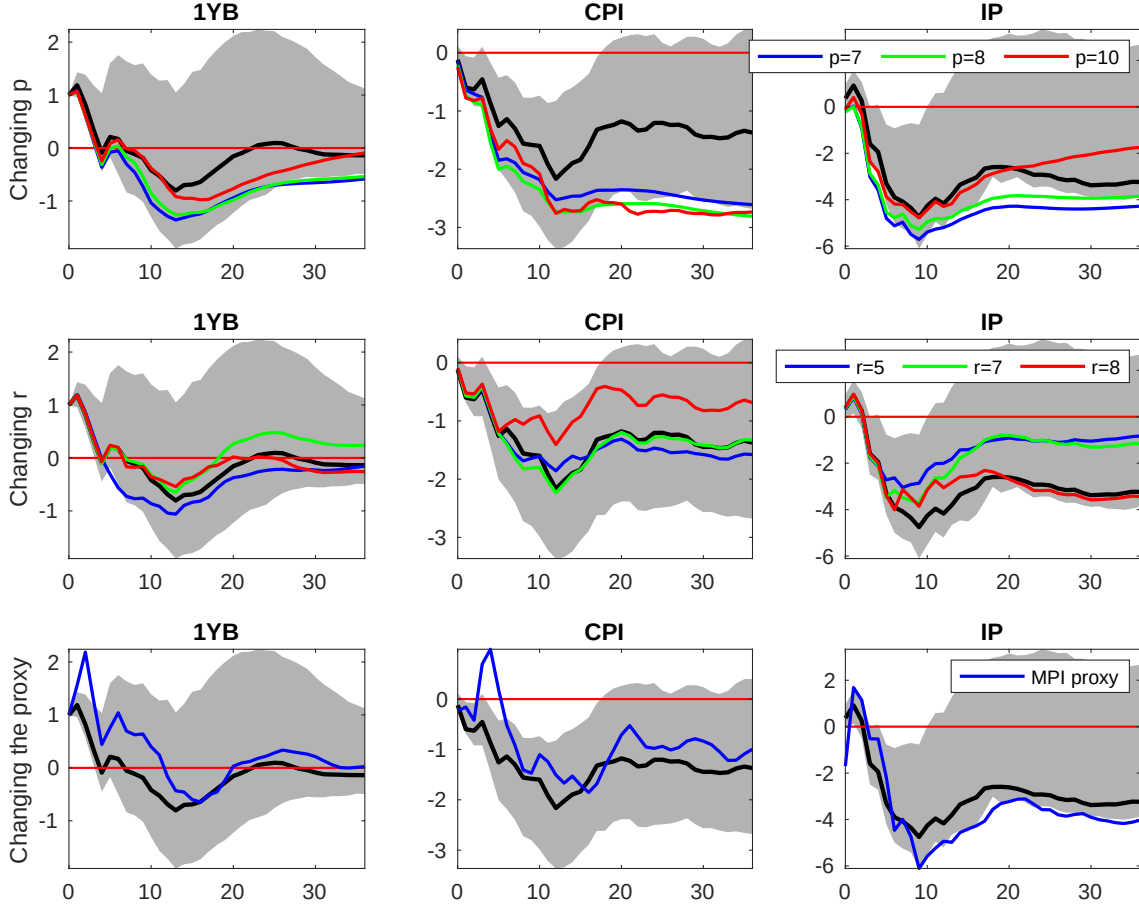


FIGURE 3: Top panels: changing p . Black line: $p = 12$; blue line: $p = 7$; green line: $p = 8$; red line: $p = 10$. Middle panels: changing r . Black line: $r = 6$; blue line: $r = 5$; green line: $r = 7$; red line: $r = 8$. Bottom panels: changing the instrument. Black line: GK instrument; blue line MPI instrument. Grey areas: 68% confidence bands for the baseline specification.

Figure 3, top panels, reports results for $p = 7$ (blue lines), $p = 8$ (red lines) and $p = 10$ (green lines), whereas $r = 6$ is kept fixed. The baseline IRFs $p = 12$ are the black solid lines; the shaded grey area is the 68% confidence region of the baseline. In the middle panels we show results obtained with $p = 12$ and different values of r , i.e. the benchmark specification ($r = 6$, black lines), $r = 5$ (blue lines), $r = 7$ (green lines), $r = 8$ (red lines). In the bottom panel we report the results obtained by using the MPI instrument (blue lines) in place of the GK instrument (black lines).

Our conclusion is that results are remarkably robust to these changes.

F.2 Using levels and changing the VAR specification and the time span

In this subsection, we explore the robustness of our baseline VAR to changes of the treatment of the variables, the VAR specification and the time span. The parameters are kept fixed to $p = 12$ and $r = 6$; the instrument is GK.

Results are reported in Figure 4. The top panels refer to the results obtained with CPI and IP in log-levels (blue lines). The effects on prices are similar to that of the benchmark, but the effect on IP is much smaller. The middle panels show the results obtained with different

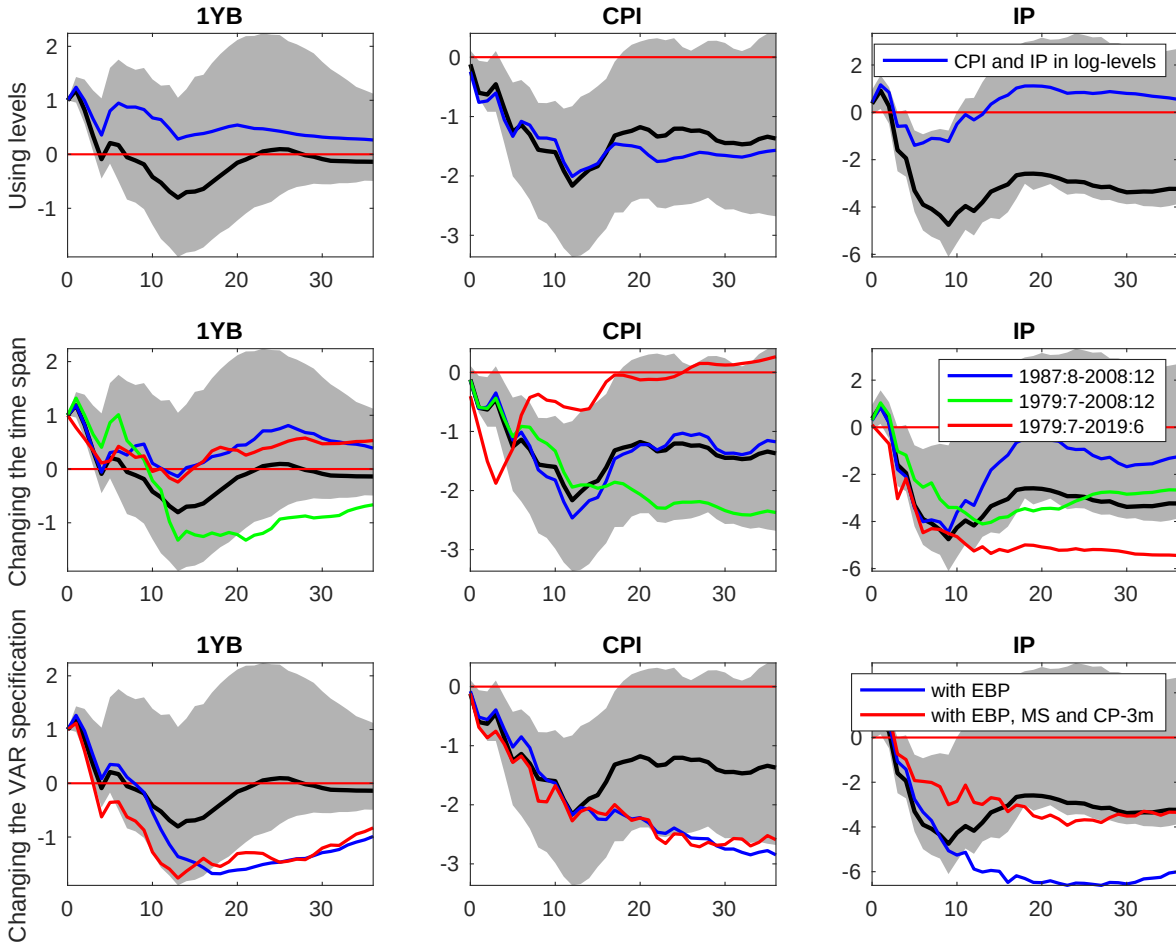


FIGURE 4: Top panels: changing the treatment of the variables. Black line: benchmark; blue line: CPI and IP are taken in log-levels. Middle panels: changing the time span. Black line: benchmark; blue line: 1987:8-2008:12; green line: 1979:7-2008:12; red line: 1979:7-2019:6. Bottom panels: changing the VAR specification. Black line: benchmark (Specification I); blue line: Specification II; red line: Specification III. Grey areas: 68% confidence bands for the baseline specification.

lines). Finally, the bottom panels show what happens when changing the VAR specification by adding the financial variables. The blue lines correspond to Specification II and the red lines to Specification III. Results are qualitatively similar to those of the benchmark.

Our overall conclusion is that results are reasonably robust to these important changes.

References

- Chahrour, Ryan and Kyle Jurado**, “Recoverability and Expectations-Driven Fluctuations,” *The Review of Economic Studies*, 03 2021, *89* (1), 214–239.
- Forni, Mario, Luca Gambetti, and Luca Sala**, “Structural VARs and noninvertible macroeconomic models,” *Journal of Applied Econometrics*, 2019, *34* (2), 221–246.
- Leeper, Eric M., Todd B. Walker, and Shu-Chun Susan Yang**, “Fiscal Foresight and Information Flows,” *Econometrica*, May 2013, *81* (3), 1115–1145.
- Lippi, Marco and Lucrezia Reichlin**, “VAR analysis, nonfundamental representations, Blaschke matrices,” *Journal of Econometrics*, July 1994, *63* (1), 307–325.
- Plagborg-Møller, Mikkel and Christian K. Wolf**, “Instrumental Variable Identification of Dynamic Variance Decompositions,” *Journal of Political Economy*, 2022, *130* (8), 2164–2202.