INFERENCE IN SVARS IDENTIFIED WITH AN EXTERNAL INSTRUMENT.[1]

José L. Montiel Olea, James H. Stock and Mark W. Watson

APPENDIX A: MAIN RESULTS

A.1. *Weak-Instrument Asymptotic Distributions for Plug-in SVAR Estimators*

This section presents the weak-instrument distributions for the plug-in estimators of the target structural shock, the historical decompositions, and the forecast-error variance decompositions.

Set-up: The distribution of an SVAR-IV data set of size T, denoted P_T, is indexed by (A, Θ_0, F); where A is the matrix of VAR slope coefficients, Θ_0 is the matrix of contemporaneous responses, and F is the joint distribution of $\{\varepsilon_t, z_t\}_{t=1}^{\infty}$.

To allow for models in which the correlation between the external instrument and the target structural shock can be arbitrarily close to zero, consider a sequence $\{P_T\}_{T=1}^{\infty}$ such that Assumption 1 holds. This means that $\mathbb{E}_{P_T}[z_t \varepsilon_{1,t}] = \alpha_T$, $\mathbb{E}_{P_T}[z_t \varepsilon_{j,t}] = 0$ for $j \neq 1$, and $\alpha_T \to 0$.

A.1.1. *Weak-instrument distribution of impulse response coefficients*

Result 1 *Let $\{P_T\}_{T=1}^{\infty}$ be a sequence along which Assumptions 1 and 2 are satisfied. Suppose in addition that the covariance between the external instrument and the target shock is local-to-zero as in Staiger and Stock (1997); i.e.,*

$$\alpha_T = a/\sqrt{T}.$$

If AsyVar $(e_1'\sqrt{T}(\widehat{\Gamma}_T - \Gamma_T)) \neq 0$. Then:

$$\lambda_{k,i}(\widehat{A}_T, \widehat{\Gamma}_T) \xrightarrow{d} \lambda_{k,i}(A, \Theta_0) + \frac{\delta_{k,i}(A, \Theta_0)'\xi}{e_1'\xi + a\Theta_{0,11}},$$

where:

$$\delta_{k,i}(A, \Theta_0) \equiv (e_i' C_k(A) - \lambda_{k,i}(A, \Theta_0) e_1')' \in \mathbb{R}^n$$

and ξ is the limiting distribution of $\sqrt{T}(\widehat{\Gamma}_T - \Gamma_T)$.

Proof: Define the auxiliary statistics

$$\widehat{\Delta}_{N,T} \equiv (e_i' C_k(\widehat{A}_T)\widehat{\Gamma}_T - \lambda_{k,i}(A, \Gamma_T) e_1' \widehat{\Gamma}_T), \quad \widehat{\Delta}_{D,T} \equiv e_1' \widehat{\Gamma}_T,$$

[1]This version: September 21st, 2018.

and the difference between the plug-in IRF and the true IRF:

$$\widehat{\Delta}_T \equiv \lambda_{k.i}(\widehat{A}_T, \widehat{\Gamma}_T) - \lambda_{k,i}(A, \Gamma_T).$$

Algebra shows that $\widehat{\Delta}_T = \widehat{\Delta}_{N,T}/\widehat{\Delta}_{D,T}$. Moreover, the numerator $\widehat{\Delta}_{N,T}$ can be written as:

$$\begin{aligned} \widehat{\Delta}_{N,T} &= e_i'[C_k(\widehat{A}_T) - C_k(A)]\sqrt{T}(\widehat{\Gamma}_T - \Gamma_T) \\ &+ e_i'[C_k(\widehat{A}_T) - C_k(A)]a\Theta_{0,1} \\ &+ (e_i'C_k(A) - \lambda_{k,i}(A, \Gamma_T)e_1')\sqrt{T}(\widehat{\Gamma}_T - \Gamma_T) \\ &+ \sqrt{T}(e_i'C_k(A)\Gamma_T - \lambda_{k,i}(A, \Gamma_T)e_1'\Gamma_T). \end{aligned}$$

Assumption 2 and the continuity of $C_k(\cdot)$ imply that both of the first two terms in the last equation above, which are given by

$$e_i'[C_k(\widehat{A}_T) - C_k(A)]\sqrt{T}(\widehat{\Gamma}_T - \Gamma_T) \text{ and } e_i'[C_k(\widehat{A}_T) - C_k(A)]a\Theta_{0,11},$$

converge in probability to zero. In addition:

$$e_i'C_k(A)\Gamma_T - \lambda_{k,i}(A, \Gamma_T)e_1'\Gamma_T = 0,$$

as $\lambda_{k,i}(A, \Gamma_T) \equiv e_i'C_k(A)\Gamma_T/e_1'\Gamma_T$. Consequently, under our assumptions

$$\widehat{\Delta}_T = (e_i'C_k(A) - \lambda_{k,i}(A, \Gamma_T)e_1')\sqrt{T}(\widehat{\Gamma}_T - \Gamma_T) \;/\; (e_1'\sqrt{T}(\widehat{\Gamma}_T - \Gamma_T) + a\Theta_{0,11}) + o_p(1).$$

Implying:

$$\lambda_{k,i}(\widehat{A}_T, \widehat{\Gamma}_T) \xrightarrow{d} \lambda_{k,i}(A, \Theta_0) + \frac{\delta_{k,i}(A, \Theta_0)'\xi}{e_1'\xi + a\Theta_{0,11}},$$

where

$$\delta_{k,i}(A, \Theta_0) \equiv (e_i'C_k(A) - \lambda_{k,i}(A, \Theta_0)e_1')' \in \mathbb{R}^n$$

and ξ is such that $\sqrt{T}(\widehat{\Gamma}_T - \Gamma_T) \xrightarrow{d} \xi$. *Q.E.D.*

A.1.2. ***Weak-instrument distributions of the Target Structural Shock, Historical Decompositions, and Forecast-error Variance Decompositions***

TARGET STRUCTURAL SHOCK: Let $\tilde{\varepsilon}_{1,t} \equiv (\varepsilon_1/\sigma_1)$. We have shown that

$$\text{sign}(\alpha)\tilde{\varepsilon}_{1,t} = \Gamma'\Sigma^{-1}\eta_t/\sqrt{\Gamma'\Sigma^{-1}\Gamma}.$$

A plug-in estimator for the target structural shock (valid up to sign) is:

$$\widehat{\tilde{\varepsilon}}_{1,t} = \widehat{\Gamma}_T'\widehat{\Sigma}^{-1}\widehat{\eta}_t \Big/ (\widehat{\Gamma}_T'\widehat{\Sigma}^{-1}\widehat{\Gamma}_T)^{1/2}, \tag{A.1}$$

with $\widehat{\Sigma}$ a consistent estimator for Σ and $\widehat{\eta}_t$ are the estimated VAR reduced-form residuals. Assumption 2 and $\alpha_T = a/\sqrt{T}$ imply

$$\sqrt{T}\widehat{\Gamma}_T = \sqrt{T}(\widehat{\Gamma}_T - \Gamma_T) + \sqrt{T}\Gamma_T \xrightarrow{d} \Gamma^* \equiv \xi + a\Theta_{0,1}.$$

The Continuous Mapping Theorem gives:

$$\begin{aligned}
\widehat{\widetilde{\varepsilon}}_{1,t} &= [\sqrt{T}\widehat{\Gamma}_T]'\widehat{\Sigma}^{-1}\widehat{\eta}_t \Big/ (\sqrt{T}\widehat{\Gamma}_T'\widehat{\Sigma}^{-1}\sqrt{T}\widehat{\Gamma}_T)^{1/2}, \\
&\xrightarrow{d} \Gamma^{*\prime}\Sigma^{-1}\eta_t \Big/ (\Gamma^{*\prime}\Sigma^{-1}\Gamma^*)^{1/2} \\
&= (\xi + a\Theta_{0,1})'\Sigma^{-1}\eta_t \Big/ ((\xi + a\Theta_{0,1})'\Sigma^{-1}(\xi + a\Theta_{0,1}))^{1/2}.
\end{aligned}$$

We note that only as $a \to \infty$ the limiting distribution concentrates around the object of interest: $(\varepsilon_{1,t}/\sigma_1)$.

Historical Decompositions: The plug-in estimator for the contribution of $\varepsilon_{1,t}$ to η_t is:

$$\text{(A.2)} \qquad \widehat{\Theta_{0,1}\varepsilon_{1,t}} \equiv (\widehat{\Gamma}_T)\widehat{\Gamma}_T'\widehat{\Sigma}^{-1}\widehat{\eta}_t \Big/ (\widehat{\Gamma}_T'\widehat{\Sigma}^{-1}\widehat{\Gamma}_T).$$

Assumption 2 and $\alpha_T = a/\sqrt{T}$ imply

$$\begin{aligned}
\widehat{\Theta_{0,1}\varepsilon_{1,t}} &\xrightarrow{d} (\xi + a\Theta_{0,1})((\xi + a\Theta_{0,1})'\Sigma^{-1}\eta_t) \,/\, ((\xi + a\Theta_{0,1})'\Sigma^{-1}(\xi + a\Theta_{0,1})). \\
&= (\Gamma^*)(\Gamma^{*\prime}\Sigma^{-1}\eta_t) \,/\, (\Gamma^{*\prime}\Sigma^{-1}\Gamma^*).
\end{aligned}$$

The limiting distribution converges to $\Theta_{0,1}\varepsilon_{1,t}$ only as $a \to \infty$.

Forecast-error Variance Decompositions: Finally, the plug-in estimator for the forecast-error variance decompositions is:

$$\text{(A.3)} \quad \widehat{\text{FEVD}}_{k,i} \equiv \widehat{\Gamma}_T' \left(\sum_{s=0}^{k} C_s(\widehat{A}_T)'e_i e_i' C_s(\widehat{A}_T) \right) \widehat{\Gamma}_T \Big/ (\widehat{\Gamma}_T'\Sigma^{-1}\widehat{\Gamma}_T) \sum_{s=0}^{k} e_i' C_s(\widehat{A}_T)\widehat{\Sigma} C_s(\widehat{A}_T)'e_i.$$

Under the local to zero assumption:

$$\widehat{\text{FEVD}}_{k,i} \xrightarrow{d} \Gamma^{*\prime} \left(\sum_{s=0}^{k} C_s(A)'e_i e_i' C_s(A) \right) \Gamma^* \Big/ (\Gamma^{*\prime}\Sigma^{-1}\Gamma^*) \sum_{s=0}^{k} e_i' C_s(A)\Sigma C_s(A)'e_i.$$

A.2. ***Proofs of Proposition 1 and 2***

This section presents the proofs of the main propositions in the paper. Proposition 1 states that our proposed Anderson-Rubin confidence set is valid under weak and strong instruments. Proposition 2 states that the Hausdorff distance between the Anderson-Rubin confidence set and the standard delta-method confidence interval converges in probability to zero under strong instruments.

A.2.1. ***Proposition 1***

PROOF: Let $\lambda_{k,i}$ denote the true impulse response coefficient and consider the test statistic

$$X_T \equiv [\sqrt{T}(e_i' C_k(\widehat{A}_T) - \lambda_{k,i} e_1')\widehat{\Gamma}_T].$$

By definition of the Anderson-Rubin confidence interval:

$$P_T\left(\lambda_{k,i} \in \mathrm{CS}_T^{\mathrm{AR}}(1-\alpha)\right) = P_T\left(X_T^2 \leq z_{1-\alpha/2}^2\, \widehat{\sigma}_T^2(\lambda_{k,i})\right),$$

where $\widehat{\sigma}_T(\lambda_{k,i})$ is the estimator of the asymptotic variance of X_T.

The matrix Ω defined in Proposition 1 is positive definite by assumption and therefore $\sigma^2(\lambda_{k,i}) \neq 0$. Consequently,

$$X_T^2/\widehat{\sigma}_T^2(\lambda_{k,i}) \xrightarrow{d} \chi_1^2$$

follows from Assumption 1 and 2 and the differentiability of $C_k(\cdot)$ with respect to A, regardless of the instrument strength. Therefore

$$\lim_{T\to\infty} P_T(\lambda_{k,i} \in \mathrm{CS}_T^{\mathrm{AR}}(1-\alpha)) = 1-\alpha.$$

Q.E.D.

A.2.2. ***Proposition 2***

PROOF: The Anderson-Rubin confidence set solves a quadratic inequality:

$$\mathrm{CS}_T^{\mathrm{AR}}(1-\alpha) \equiv \left\{\lambda \in \mathbb{R} \,:\, \lambda^2 \widehat{a}_{1-\alpha} + \lambda \widehat{b}_{1-\alpha} + \widehat{c}_{1-\alpha} \leq 0\right\},$$

where the coefficients $\widehat{a}_{1-\alpha}, \widehat{b}_{1-\alpha}, \widehat{c}_{1-\alpha}$ depend on the data and the confidence level. The results in Fieller (1954) and footnote 12 imply

$$\mathrm{CS}_T^{\mathrm{AR}}(1-\alpha, \lambda_{k,i}) = \Big[\frac{-\widehat{b}_{1-\alpha} - \sqrt{\Delta_{1-\alpha}}}{2\widehat{a}_{1-\alpha}}\,,\ \frac{-\widehat{b}_{1-\alpha} + \sqrt{\Delta_{1-\alpha}}}{2\widehat{a}_{1-\alpha}}\Big],$$

whenever:

$$\widehat{a}_{1-\alpha} \equiv T(e_1'\widehat{\Gamma}_T)^2 - z_{1-\alpha/2}^2\, \widehat{\omega}_{T,22} > 0,$$

where $\widehat{\omega}_{T,22}$ is the asymptotic variance of $\sqrt{T}(e_1'\widehat{\Gamma}_T - e_1'\Gamma)$. Therefore, under strong instruments, $P(\widehat{a}_{1-\alpha} > 0)$ goes to 1 as $T \to \infty$.[1] It is thus sufficient to focus on the Hausdorff distance between the Anderson-Rubin confidence interval

$$\left[\frac{-\widehat{b}_{1-\alpha} - \sqrt{\Delta_{1-\alpha}}}{2\widehat{a}_{1-\alpha}}\, , \, \frac{-\widehat{b}_{1-\alpha} + \sqrt{\Delta_{1-\alpha}}}{2\widehat{a}_{1-\alpha}}\right], \quad \Delta_{1-\alpha} \equiv \widehat{b}^2_{1-\alpha} - 4\widehat{a}_{1-\alpha}\widehat{c}_{1-\alpha},$$

and the delta-method/plug-in confidence interval

$$\left[\widehat{\lambda}_{k,i} - \sqrt{\frac{z^2_{1-\alpha/2}}{T}}\frac{\widehat{\sigma}_T(\widehat{\lambda}_{k,i})}{|e_i'\widehat{\Gamma}_1|}\, , \, \widehat{\lambda}_{k,i} + \sqrt{\frac{z^2_{1-\alpha/2}}{T}}\frac{\widehat{\sigma}_T(\widehat{\lambda}_{k,i})}{|e_1'\widehat{\Gamma}_1|}\right].$$

Direct computation shows that the Hausdorff distance between two intervals $[a, b]$ and $[c, d]$ is given by:

$$\max\{|c - a|\, , \, |d - b|\}.$$

We complete the proof establishing two results.

Step 1: We show first that:

$$-\frac{\widehat{b}_{1-\alpha}}{2\widehat{a}_{1-\alpha}} = \widehat{\lambda}_{k,i} + O_p(1/T).$$

Algebra shows that

$$\widehat{b}_{1-\alpha} = 2z_{1-\alpha/2}\widehat{\omega}_{T,12} - 2T(e_i'C_k(\widehat{A}_T)\widehat{\Gamma}_T)(e_1'\widehat{\Gamma}_T)$$

Therefore

$$-\frac{\widehat{b}_{1-\alpha}}{2\widehat{a}_{1-\alpha}} = \frac{2T(e_i'C_k(\widehat{A}_T)\widehat{\Gamma}_T)(e_1'\widehat{\Gamma}_T) - 2z_{1-\alpha/2}\widehat{\omega}_{T,12}}{2T(e_1'\widehat{\Gamma}_T)^2 - 2z^2_{1-\alpha/2}e_1'\widehat{\omega}_{T,22}} = \widehat{\lambda}_{k,i} + O_p(1/T),$$

provided the probability limit of $e_1'\widehat{\Gamma}_T$ is different from zero.

Step 2: We now show that under strong instruments:

$$\frac{\sqrt{\Delta_{1-\alpha}}}{2\widehat{a}_{1-\alpha}} = \left(\sqrt{z^2_{1-\alpha/2}}\widehat{\sigma}_T(\widehat{\lambda}_{k,i})/(|\sqrt{T}e_1'\widehat{\Gamma}_T|)\sqrt{1 + o_p(1)}\right) + O_p(1/T),$$

[1] This happens because of two reasons. First $\widehat{\omega}_{T,22}$ converges in probability to $\omega_{2,2}$, by assumption. Second, Assumption 1 implies $e_1'\Gamma = \alpha e_1'\Theta_{0,1} = \alpha \neq 0$ and Assumption 2 implies $e_1'\widehat{\Gamma}_T \xrightarrow{p} e_1'\Gamma$. Consequently, $T(e_1'\widehat{\Gamma}_T)^2$ diverges to infinity.

where

$$\widehat{\sigma}^2_T(\widehat{\lambda}_{k,i}) = \widehat{\omega}_{1,T} - 2\widehat{\lambda}_{k,i}\widehat{\omega}_{T,12} + \widehat{\lambda}^2_{k,i}\widehat{\omega}_{T,22}.$$

Consider the square of the desired expression:

$$\frac{\Delta_{1-\alpha}}{4\widehat{a}^2_{1-\alpha}} = \frac{\Delta_{1-\alpha}}{4T^2(e_1'\widehat{\Gamma}_T)^4}\frac{4T^2(e_1'\widehat{\Gamma}_T)^4}{4\widehat{a}^2_{1-\alpha}} = \frac{(\widehat{b}^2_{1-\alpha} - 4\widehat{a}_{1-\alpha}\widehat{c}_{1-\alpha})}{4T^2(e_1'\widehat{\Gamma}_T)^4}\frac{4T^2(e_1'\widehat{\Gamma}_T)^4}{4\widehat{a}^2_{1-\alpha}}.$$

First, we study the term:

$$\begin{aligned}
\frac{\widehat{b}^2_{1-\alpha}}{4T^2(e_1'\widehat{\Gamma}_T)^4} &= \left(-\frac{\widehat{b}_{1-\alpha}}{2T(e_1'\widehat{\Gamma})^2}\right)^2, \\
&= \left(\frac{2T(e_i'C_k(\widehat{A}_T)\widehat{\Gamma}_T)(e_1'\widehat{\Gamma}_T) - 2z_{1-\alpha/2}\widehat{\omega}_{T,12}}{2T(e_1'\widehat{\Gamma}_T)^2}\right)^2, \\
&= (\widehat{\lambda}_{k,i} - ((z_{1-\alpha/2})/T)\widehat{\upsilon}_0)^2, \quad \widehat{\upsilon}_0 \equiv \widehat{\omega}_{T,12}/(e_1\widehat{\Gamma}_T)^2, \\
&= \widehat{\lambda}^2_{k,i} - ((z_{1-\alpha/2})/T)2\widehat{\lambda}_{k,i}\widehat{\upsilon}_0 + O_p(1/T^2).
\end{aligned}$$

Second, we look at

$$\frac{4\widehat{a}_{1-\alpha}\widehat{c}_{1-\alpha}}{4T^2(e_1'\widehat{\Gamma}_T)^4} = \frac{\widehat{a}_{1-\alpha}}{T(e_1'\widehat{\Gamma}_T)^2}\frac{\widehat{c}_{1-\alpha}}{T(e_1'\widehat{\Gamma}_T)^2}.$$

Algebra shows that

$$\widehat{c}_{1-\alpha} = T(e_i'C_k(\widehat{A}_T)\widehat{\Gamma}_T)^2 - z_{1-\alpha/2}\widehat{\omega}_{T,11}.$$

Consequently:

$$\begin{aligned}
\frac{4\widehat{a}_{1-\alpha}\widehat{c}_{1-\alpha}}{4T^2(e_1'\widehat{\Gamma}_T)^4} &= \frac{\widehat{a}_{1-\alpha}}{T(e_1'\widehat{\Gamma}_T)^2}\frac{\widehat{c}_{1-\alpha}}{T(e_1'\widehat{\Gamma}_T)^2}, \\
&= \left(1 - (z_{1-\alpha/2}/T)\widehat{\upsilon}_1\right)\left(\widehat{\lambda}^2_{k,i} - (z_{1-\alpha/2}/T)\widehat{\upsilon}_2\right), \\
&\quad \left(\widehat{\upsilon}_1 \equiv \widehat{\omega}_{T,22}/(e_1'\widehat{\Gamma}_T)^2 \text{ and } \widehat{\upsilon}_2 \equiv \widehat{\omega}_{T,11}/\left(e_1'\widehat{\Gamma}_T\right)^2\right) \\
&= \widehat{\lambda}^2_{k,i} - (z^2_{1-\alpha/2}/T)\widehat{\upsilon}_1\widehat{\lambda}^2_{k,i} - (z^2_{1-\alpha/2}/T)\widehat{\upsilon}_2 - O_p(1/T^2).
\end{aligned}$$

Therefore

$$\frac{(\widehat{b}^2_{1-\alpha} - 4\widehat{a}_{1-\alpha}\widehat{c}_{1-\alpha})}{4T^2(e_1'\widehat{\Gamma}_T)^4} = \frac{(z^2_{1-\alpha/2}/T)}{(e_1'\widehat{\Gamma}_T)^2}\widehat{\sigma}^2_T(\widehat{\lambda}_{k,i}) + O_p(1/T^2),$$

implying

$$\begin{aligned}
\frac{\Delta_{1-\alpha}}{4\widehat{a}^2_{1-\alpha}} &= \left(\frac{(z^2_{1-\alpha/2}/T)}{(e_1'\widehat{\Gamma}_T)^2}\widehat{\sigma}^2_T(\widehat{\lambda}_{k,i}) + O_p(1/T^2)\right)\left(1 + o_p(1)\right) \\
&= (z^2_{1-\alpha}/T)(\widehat{\sigma}^2_T(\widehat{\lambda}_{k,i})/(e_1'\widehat{\Gamma}_T)^2) + o_p(1/T).
\end{aligned}$$

Combining Step 1 and Step 2 we can conclude that the bounds of our confidence interval are approximately equal to:

$$\widehat{\lambda}_{k,i} \pm \sqrt{\frac{z_{1-\alpha}^2}{T}\frac{\widehat{\sigma}_T^2(\widehat{\lambda}_{k,i})}{(e_1'\widehat{\Gamma}_T)^2} + o_p\left(\frac{1}{T}\right)} + O_p\left(\frac{1}{T}\right),$$

which can be written as:

$$\widehat{\lambda}_{k,i} \pm \frac{\sqrt{z_{1-\alpha}^2}}{\sqrt{T}}\frac{\widehat{\sigma}_T(\widehat{\lambda}_{k,i})}{|e_1'\widehat{\Gamma}_T|}\sqrt{1+o_p\left(1\right)} + O_p\left(\frac{1}{T}\right).$$

The probability limit of $\widehat{\sigma}_T^2(\widehat{\lambda}_{k,i})$ is not zero by assumption. Therefore, for large enough T

$$\sqrt{T}d_H\left(\mathrm{CS}_T^{\mathrm{AR}}(1-\alpha,\lambda_{k,i})\,,\,\left[\widehat{\lambda}_{k,i} - \sqrt{\frac{z_{1-\alpha/2}^2}{T}}\frac{\widehat{\sigma}_T(\widehat{\lambda}_{k,i})}{|e_1'\widehat{\Gamma}_T|}\,,\,\widehat{\lambda}_{k,i} + \sqrt{\frac{z_{1-\alpha/2}^2}{T}}\frac{\widehat{\sigma}_T(\widehat{\lambda}_{k,i})}{|e_1'\widehat{\Gamma}_T|}\right]\right)$$

equals the absolute value of

$$\sqrt{z_{1-\alpha}^2}\frac{\widehat{\sigma}_T(\widehat{\lambda}_{k,i})}{|e_1'\widehat{\Gamma}_T|}\left(\sqrt{1+o_p\left(1\right)}-1\right) + O_p\left(\frac{1}{\sqrt{T}}\right),$$

which is the difference between the bounds of our confidence set and the plug-in confidence interval. Since under strong instruments the probability limit of $e_1'\widehat{\Gamma}_T$ is different from zero, the desired result follows.

Q.E.D.

A.2.3. *Proposition 2 and local power comparison*

We now show that Proposition 2 implies that the tests for the null hypothesis $\lambda_{k,i} = \lambda_0$ corresponding to $\mathrm{CS}^{\text{Plug-in}}\,(1-\alpha)$ and $\mathrm{CS}^{\mathrm{AR}}\,(1-\alpha)$ have the same local power.

The delta-method confidence interval is a set of the form $[\widehat{a},\widehat{b}]$. Proposition 2 has shown that for T large enough the $1-\alpha$ Anderson-Rubin confidence set is an interval of the form $[\widehat{c},\widehat{d}]$. Consequently, for large T the Hausdorff distance between the delta-method confidence interval and the Anderson-Rubin confidence set is $\widehat{d}_H \equiv \max\{|\widehat{c}-\widehat{a}|, |\widehat{d}-\widehat{c}|\}$.

For T large enough the Anderson-Rubin test rejects the null λ_0 whenever $\lambda_0 \notin [\widehat{c},\widehat{d}]$. This means that the power of the Anderson-Rubin test under an alternative of the form $\lambda_T = \lambda_0 + l/\sqrt{T}$ is

$$
\begin{aligned}
P_{\lambda_T}(\lambda_0 \notin [\widehat{c}, \widehat{d}]) &= 1 - P_{\lambda_T}(\widehat{c} \leq \lambda_0 \leq \widehat{d}) \\
&= 1 - P_{\lambda_T}((\widehat{c} - \widehat{a}) + \widehat{a} \leq \lambda_0 \leq (\widehat{d} - \widehat{b}) + \widehat{b}) \\
&= 1 - P_{\lambda_T}(-\sqrt{T}(\widehat{d} - \widehat{b}) - z_{1-\alpha/2}\widehat{s} \leq \sqrt{T}(\widehat{\lambda}_{k,i} - \lambda_0) \leq -\sqrt{T}(\widehat{c} - \widehat{a}) + z_{1-\alpha/2}\widehat{s}),
\end{aligned}
$$

where the last line has used the fact that $\widehat{a}$ and $\widehat{b}$ are the endpoints of the delta-method confidence interval and $\widehat{s} = \widehat{\sigma}_T(\widehat{\lambda}_{k,i})/|e_1'\widehat{\Gamma}_T|$.

Proposition 2 implies that $\sqrt{T}(\widehat{d} - \widehat{b})$ and $\sqrt{T}(\widehat{c} - \widehat{a})$ converge in probability to zero. Therefore

$$
P_{\lambda_T}(\lambda_0 \notin [\widehat{c}, \widehat{d}]) \to 1 - P(z_{1-\alpha/2} \leq N(l/s, 1) \leq z_{1-\alpha/2}),
$$

where s is the probability limit of $\widehat{s}$. The right hand side above is the local power curve of the delta-method confidence interval.

A.3. ***Inference for the over-identified case***

We discuss the extensions of our main results to models with more than one external instrument for a single structural shock. This situation could arise, for example, in a monetary SVAR where several popular proxy variables for monetary shocks are available: the series of shocks in Romer and Romer (2004), the shock to the monetary policy reaction function in Smets and Wouters (2007), and the series of variance shocks in Sims and Zha (2006).

The extension of the Anderson-Rubin confidence set to the 'over-identified' case is conceptually straightforward. We note, however, that contrary to the 'just-identified' case there is no guarantee that the Anderson-Rubin confidence set performs as well as that based on an 'efficient' estimator for the parameter $\lambda_{k,i}$ when the external instrument is strong. This limitation is well-understood in the context of linear IV regression. Examples of weak-instrument robust procedures with better (local) power properties under strong instruments are the Lagrange Multiplier of Kleibergen (2002) and the Conditional Likelihood Ratio test of Moreira (2003).

In this section we also show that appropriate versions of the Lagrange Multiplier and Conditional Likelihood Ratio test of Moreira (2003) can be constructed for the SVAR-IV model.

A.3.1. ***Anderson-Rubin test for over-identified SVAR-IV models***

Suppose there are $M > 1$ external instruments, $z_{m,t}$, for a target shock $\varepsilon_{1,t}$. Let:

$$
\widehat{\Gamma}_{m,T} \equiv (1/T) \sum_{t=1}^{T} z_{m,t}\widehat{\eta}_t.
$$

Note that for each of the estimators $\widehat{\Gamma}_{m,T}$, one could construct the statistic:

$$s_{m,T}(\lambda) \equiv \sqrt{T}(e_i' C_k(\widehat{A}_T) - \lambda e_1')\widehat{\Gamma}_{m,T}.$$

This means that if the vector:

$$\left(\sqrt{T}(\text{vec}(\widehat{A}_T) - \text{vec}(A))', \quad \sqrt{T}(\widehat{\Gamma}_{1,T} - \mathbb{E}_{P_T}[z_{1,t}\eta_t])', \quad \ldots, \quad \sqrt{T}(\widehat{\Gamma}_{M,T} - \mathbb{E}_{P_T}[z_{M,t}\eta_t])'\right)'$$

is asymptotically multivariate nomal (which extends our Assumption 2), then the vector $s_T(\lambda) \equiv (s_{1,T}(\lambda), \ldots, s_{M,T}(\lambda))'$, will be asymptotically normal as well; provided λ is the true impulse response coefficient. If $\widehat{W}_T(\lambda)$ is a consistent estimator for the covariance of such vector, then the analogous of our Anderson-Rubin type confidence interval would collect the values of λ such that:

$$s_T'(\lambda)\widehat{W}_T(\lambda)^{-1}s_T(\lambda) \leq \chi^2_{M,1-\alpha}.$$

This extends our Anderson-Rubin procedure to over-identified models.

A.3.2. ***Quasi-Conditional Likelihood Ratio Test for over-identified models***

A natural question to ask is whether there exists a confidence interval that is robust to the presence of weak external IVs but, at the same time, is as accurate as the best confidence interval that would be used if instruments were known to be strong. Using a Linear IV interpretation of external instruments—in which $e_i' C_k(\widehat{A}_T)\widehat{\eta}_t$ is the outcome variable, $\widehat{\eta}_{1,t}$ is the endogenous regressor, and $Z_t = (z_{1,t}, \ldots, z_{M,t})'$ the vector of instrumental variables—it is possible to derive Lagrange multiplier and Quasi-Conditional Likelihood Ratio tests as those discussed in Kleibergen (2007) to conduct inference that is as 'efficient' as when the instruments are known to be strong.

To formalize this argument, we start by defining what we mean by efficient inference when the external instruments are strong. We use a typical minimum-distance framework. For each of the M external instruments, let $\widehat{\lambda}^m_{k,i}$ denote the plug-in estimator for $\lambda_{k,i}$. Consider the class of minimum-distance estimators—indexed by the weighting matrix S—given by:

$$\widehat{\lambda}_{k,i}(S) \equiv \underset{\lambda \in \mathbb{R}}{\arg\min} \left(\widehat{\lambda}^1_{k,i} - \lambda, \quad \ldots, \quad \widehat{\lambda}^M_{k,i} - \lambda\right) S \left(\widehat{\lambda}^1_{k,i} - \lambda, \quad \ldots, \quad \widehat{\lambda}^M_{k,i} - \lambda\right)'.$$

The standard theory of minimum-distance estimation (e.g., Newey and McFadden (1994) or Hayashi (2000)) implies that the minimum-distance estimator with the smallest asymptotic variance corresponds to the weighting matrix:

$$S^* \equiv \text{AsyVar}\left(\sqrt{T}\left(\widehat{\lambda}^1_{k,i} - \lambda, \quad \ldots, \quad \widehat{\lambda}^M_{k,i} - \lambda\right)'\right)^{-1}.$$

Thus, one way to define efficiency is to use the local power curve of a test of hypothesis for $\lambda_{k,i}$ based on the efficient estimator $\widehat{\lambda}_{k,i}(S^*)$. Direct calculation shows that such power curve is given by the tail of a non-central chi-squared distribution with one degree of freedom and centrality parameter $(1_M' S^* 1_M)l$, where 1_M denotes the vector of ones in $\mathbb{R}^M$ and $l \in \mathbb{R}$ is the local alternative. We show that the Lagrange multiplier and Quasi-Conditional Likelihood Ratio tests are indeed weak-instrument robust procedures that achieve such local power curve. Details are provided below.

Overview: As suggested in Müller (2011), Moreira and Moreira (2015), Andrews (2016) the weak-IV robust procedures for linear IV regression (with a single right-hand endogenous regressor) can be described using the following statistical model for the OLS reduced-form estimators:

$$\begin{pmatrix} \widehat{\gamma}_1 \\ \widehat{\gamma}_2 \end{pmatrix} \sim \mathcal{N}_{2M}\left(\begin{pmatrix} \beta \\ 1 \end{pmatrix} \otimes \pi \, , \, \Sigma/T \right), \tag{A.4}$$

where $\beta \in \mathbb{R}$ is the coefficient of the right-hand endogenous regressor, π is the vector of first-stage coefficients, and Σ is the asymptotic variance of the reduced-form estimators. Consider the following SVAR-IV statistics:

$$\begin{pmatrix} (1/T)\sum_{t=1}^{T}(e_i' C_k(\widehat{A}_T)\widehat{\eta}_t) Z_t \\ \\ (1/T)\sum_{t=1}^{T}(e_1'\widehat{\eta}_t) Z_t \end{pmatrix}, \tag{A.5}$$

which correspond to the covariances between the external instruments and linear combinations of the reduced-form residuals. Let $\lambda_{k,i}$ denote the true (k,i)-th IRF coefficient and let α_m denote the covariance between instrument $z_{m,t}$ and $\varepsilon_{1,t}$. If we were to ignore—just to simplify exposition—the sampling variability in the statistics above coming from the estimation of $\widehat{A}_T$ and $\widehat{\eta}_t$, the vector (A.5) would be centered at:

$$\begin{pmatrix} \mathbb{E}[e_i' C_k(A)\eta_t z_{1,t}] \\ \vdots \\ \mathbb{E}[e_i' C_k(A)\eta_t z_{M,t}] \\ \mathbb{E}[e_1'\eta_t z_{1,t}] \\ \vdots \\ \mathbb{E}[e_1'\eta_t z_{M,t}] \end{pmatrix} = \begin{pmatrix} \lambda_{k,i}\alpha_1\Theta_{0,11} \\ \vdots \\ \lambda_{k,i}\alpha_M\Theta_{0,11} \\ \alpha_1\Theta_{0,11} \\ \vdots \\ \alpha_M\Theta_{0,11} \end{pmatrix} = \begin{pmatrix} \lambda_{k,i} \\ 1 \end{pmatrix} \otimes \begin{pmatrix} \alpha_1\Theta_{0,11} \\ \vdots \\ \alpha_M\Theta_{0,11} \end{pmatrix}$$

This observation, combined with a Central Limit Theorem and the normalization $\Theta_{0,11} =$

1 would imply that:

$$\widehat{\gamma}_{\text{SVAR-IV}} \equiv \begin{pmatrix} (1/T)^{-1}\sum_{t=1}^{T}(e_i' C_k(\widehat{A}_T)\widehat{\eta}_t)Z_t \\ \\ (1/T)^{-1}\sum_{t=1}^{T}(e_1'\widehat{\eta}_t)Z_t \end{pmatrix} \overset{\text{approx}}{\sim} \mathcal{N}_{2M}\left(\begin{pmatrix} \lambda_{k,i} \\ 1 \end{pmatrix} \otimes \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_M \end{pmatrix}, \; \Sigma/T \right),$$

which is the same statistical model as in (A.4). Below we show that tests that are analogous to the Lagrange Multiplier test and the Quasi-Conditional Likelihood Ratio test of Kleibergen (2007) based on the statistics (A.4) achieve the same local power curve as the Wald test based on minimum-variance minimum-distance estimator.

Lagrange Multiplier and Conditional Likelihood Ratio Test: Consider thus the model:

$$\widehat{\gamma}_{\text{SVAR-IV}} \equiv \begin{pmatrix} (1/T)\sum_{t=1}^{T}(e_i' C_k(\widehat{A}_T)\widehat{\eta}_t)Z_t \\ \\ (1/T)\sum_{t=1}^{T}(e_1'\widehat{\eta}_t)Z_t \end{pmatrix} \sim \mathcal{N}_{2M}\left(\begin{pmatrix} \lambda_{k,i} \\ 1 \end{pmatrix} \otimes \begin{pmatrix} \alpha_1\Theta_{0,11} \\ \vdots \\ \alpha_M\Theta_{0,11} \end{pmatrix}, \ \Sigma/T \right),$$

and treat Σ as known. The statistic $\widehat{\gamma}_{\text{SVAR-IV}}$ is the SVAR version of the OLS estimators of the reduced-form coefficients in a linear IV model. Let λ_0 denote the hypothesized value of the (k,i)-th IRF coefficient. Define:

$$b_0 \equiv (1, -\lambda_0)' \quad a_0 \equiv (\lambda_0, 1)',$$

and consider the following rotation of the statistic $\widehat{\gamma}_{\text{SVAR-IV}}$:

$$\begin{pmatrix} S_T \\ T_T \end{pmatrix} \equiv \begin{pmatrix} B_0^{-1/2} & 0_{M\times M} \\ 0_{M\times M} & A_0^{-1/2} \end{pmatrix} \begin{pmatrix} b_0' \otimes \mathbb{I}_M \\ (a_0' \otimes \mathbb{I}_M)\,\Sigma^{-1} \end{pmatrix} \sqrt{T}\widehat{\gamma}_{\text{SVAR-IV}},$$

where $B_0 \equiv (b_0' \otimes \mathbb{I}_M)\,\Sigma\,(b_0 \otimes \mathbb{I}_M)$ and $A_0 \equiv (a_0' \otimes \mathbb{I}_M)\,\Sigma^{-1}\,(a_0 \otimes \mathbb{I}_M)$. The rotation defining the statistics S_n and T_n is common in the linear IV literature, and it is often used to standardize and orthogonalize the OLS estimators of the reduced-form parameters. The Lagrange Multiplier statistic is usually defined as:

$$(T_T' S_T)^2 / T_T' T_T,$$

see for example p. 722 in Andrews, Moreira, and Stock (2006). Define the following SVAR-IV version of the LM statistic:

$$\text{LM}_{\text{SVAR-IV}} \equiv (T_T' A_0^{-1/2} B_0^{-1/2} S_T)^2 / T_T' A_0^{-1/2} B_0^{-1} A_0^{-1/2} T_T. \tag{A.6}$$

Under weak instrument asymptotics and our regularity assumptions:

$$\begin{pmatrix} S_T \\ T_T \end{pmatrix} \xrightarrow{d} \begin{pmatrix} S \\ T \end{pmatrix},$$

where $(S', T')'$ are independent multivariate normal random vectors, and S is mean zero. This implies that under weak external instruments $\text{LM}_{\text{SVAR-IV}} \xrightarrow{d} \chi^2_1$.

If the external instruments are strong, the SVAR-IV version of the LM test achieves the same local power as the Wald test based on the minimum-variance minimum-distance estimator for $\lambda_{k,i}$. To see this, let $\lambda_T = \lambda_0 + l/\sqrt{T}$ and $(\alpha_1, \alpha_2, \dots, \alpha_M)' \neq 0_{M\times 1}$ then:

$$\begin{pmatrix} S_T \\ T_T/\sqrt{T} \end{pmatrix} \xrightarrow{d}_{\lambda_T} \mathcal{N}_{2M}\left(\begin{pmatrix} B_0^{-1/2} l\pi \\ A_0^{1/2}\pi \end{pmatrix}, \begin{pmatrix} \mathbb{I}_M & 0_{M\times M} \\ 0_{M\times M} & 0_{M\times M} \end{pmatrix} \right), \text{ where } \pi \equiv \begin{pmatrix} \alpha_1\Theta_{0,11} \\ \vdots \\ \alpha_M\Theta_{0,11} \end{pmatrix}.$$

This implies that under strong instrument asymptotics and local alternatives:

$$\mathrm{LM}_{\text{SVAR-IV}} \xrightarrow{d} \mathcal{N}((\pi' B_0^{-1}\pi)^{1/2} l, 1)^2.$$

Consequently, the local power of an α-level test for the hypothesis $\lambda_{k,i} = \lambda_0$ based on $\mathrm{LM}_{\text{SVAR-IV}}$ has a local power curve given by:

$$\mathbb{P}\left(\chi_1^2(\pi' B_0^{-1}\pi l^2) > \chi^2_{1,1-\alpha}\right).$$

All we need to show to establish the desired equivalence between local power curves is that $\pi' B_0^{-1}\pi$ equals $1_M' S^* 1_M$. To establish this result, note that B_0 is the asymptotic variance of the vector:

$$(1/\sqrt{T})\sum_{t=1}^{T}\left[(e_i' C_k(\widehat{A}_T)\widehat{\eta}_t)Z_t - \lambda_0(e_1'\widehat{\eta}_t)Z_t\right].$$

Such vector can be expanded as:

$$\begin{pmatrix} (1/\sqrt{T})\sum_{t=1}^{T} e_i' C_k(\widehat{A}_T)\widehat{\eta}_t z_{1,t} - \lambda_0 e_1'\widehat{\eta}_{1,t} z_{1,t} \\ \vdots \\ (1/\sqrt{T})\sum_{t=1}^{T} e_i' C_k(\widehat{A}_T)\widehat{\eta}_t z_{M,t} - \lambda_0 e_1'\widehat{\eta}_{1,t} z_{M,t} \end{pmatrix},$$

which is equal to:

$$\begin{pmatrix} e_1'\widehat{\Gamma}_{1,T}\sqrt{T}[\widehat{\lambda}^1_{k,i} - \lambda_0] \\ \vdots \\ e_1'\widehat{\Gamma}_{M,T}\sqrt{T}[\widehat{\lambda}^M_{k,i} - \lambda_0] \end{pmatrix},$$

where $\widehat{\Gamma}_{m,T} \equiv (1/T)\sum_{t=1}^{T} e_1'\widehat{\eta}_t z_{m,t}$. This simple algebra shows that

$$B_0 = \begin{pmatrix} e_1'\Gamma_1, & 0, & \dots, & 0 \\ 0, & e_1'\Gamma_2, & \dots, & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0, & 0, & \dots, & e_1'\Gamma_M \end{pmatrix} (S^*)^{-1} \begin{pmatrix} e_1'\Gamma_1, & 0, & \dots, & 0 \\ 0, & e_1'\Gamma_2, & \dots, & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0, & 0, & \dots, & e_1'\Gamma_M \end{pmatrix},$$

where Γ_m is the probability limit of $\widehat{\Gamma}_{m,T}$. Since Γ_m equals $\alpha_m\Theta_{0,1}$ under the relevance and exogeneity assumption and the fact that $\pi \equiv (\alpha_1\Theta_{0,11}, \dots, \alpha_M\Theta_{0,11})'$, it follows that:

$\pi' B_0^{-1} \pi$ is the same as $1_M' S^* 1_M$ whenever $\alpha_m \neq 0$ for all $m = 1, \ldots M$.

Once we have found an analogous version of the LM statistic for SVAR-IVs we can define the SVAR-IV version of the Quasi-Conditional Likelihood Ratio as:

$$S_T' S_T - r(T_T) + \left((S_T' S_T - r(T_T))^2 + 4r(T_T)\mathrm{LM}_{\text{SVAR-IV}} \right)^{1/2},$$

where $r(T_T) \equiv T_T' A_0^{-1/2} B_0^{-1} A_0^{-1/2} T_T$, and the critical values are computed conditional on T_T.

A.4. ***Bootstrap Implementation of the Anderson-Rubin Confidence Set***

To implement the confidence interval for $\lambda_{k,i}$, we relied on typical delta-method approximations. It is well understood that the nonlinearity of the impulse-response functions can compromise the quality of the delta-method approximation; see for example Kilian (1998), Sims and Zha (1999), and Benkwitz, Neumann, and Lütekpohl (2000). With this observation in mind, we now discuss a bootstrap-type approach to implement our confidence interval.

Our suggestion is to use draws from the asymptotic distribution of $(\text{vec}(\widehat{A}_T)', \widehat{\Gamma}_T')'$ to compute the quantile of a test statistic over a grid of values for $\lambda_{k,i}$. The bootstrap-type implementation eliminates the need of closed-form formulae for the Anderson-Rubin confidence set, but it is computationally more expensive (because it requires re-sampling from the reduced-form parameters and constructing quantiles of a test statistic over a grid of possible values for the impulse response coefficients).

Description: We have explained that the intuition behind our inference approach is that the square of the statistic:

$$\sqrt{T}(e_i' C_k(\widehat{A}_T) - \lambda e_1')\widehat{\Gamma}_T,$$

should be small if λ were the true impulse response coefficient. Since we have assumed that the distribution of $(\text{vec}(\widehat{A}_T)', \widehat{\Gamma}_T')'$ can be approximated by a normal random vector centered at $(\text{vec}(A)', \Gamma')'$ with covariance matrix W/T, we suggest the following procedure:

1. Let $\widehat{A}_T$ and $\widehat{\Gamma}_T$ denote the estimators of A and Γ.

2. Generate M i.i.d. draws $\{\text{vec}(A)_m, \Gamma_m\}_{m=i}^{M}$ from the model:

$$(\text{vec}(A)_m', \Gamma_m')' \sim \mathcal{N}_{n^2p+n}\Big((\text{vec}(\widehat{A}_T)', \widehat{\Gamma}_T')' \,,\, \widehat{W}_T/T\Big),$$

   where $\widehat{W}_T$ is a consistent estimator for W.

3. For each value λ_g in a grid $\Lambda(G) \equiv \{\lambda_1, \lambda_2, \ldots, \lambda_G\}$ and conditioning on the data compute:

$$\Big\{(\sqrt{T}(e_i' C_k(A_m) - \lambda_g e_1')\Gamma_m - \sqrt{T}(e_i' C_k(\widehat{A}_T) - \lambda_g e_1')\widehat{\Gamma}_T)\Big\}_{m=1}^{M},$$

   and let $\widehat{q}_{\alpha/2}$ and $\widehat{q}_{g,1-\alpha/2}$ denote its $\alpha/2$ and $1-\alpha/2$ quantiles. Standard arguments based on the differentiability of the function

$$g_\lambda(A, \Gamma) = e_i' C_k(A)\Gamma - \lambda e_1'\Gamma,$$

with respect to (A,Γ) imply that the quantiles of

$$\sqrt{T}\Big(g_\lambda(A_m,\Gamma_m)-g_\lambda(\widehat{A}_T,\widehat{\Gamma}_T)\Big)\;\Big|\;(\widehat{A}_T,\widehat{\Gamma}_T,\widehat{W}_T),$$

approximate the quantiles of

$$\sqrt{T}\Big(g_\lambda(\widehat{A}_T,\widehat{\Gamma}_T)-g_\lambda(A,\Gamma)\Big)=\sqrt{T}(e_i'C_k(\widehat{A}_T)-\lambda_g e_i')\widehat{\Gamma}_T.$$

4. The bootstrap-type confidence interval is then given by:

$$\Big\{\lambda_g\in\Lambda(G)\mid \widehat{q}_{g,\alpha/2}\le(\sqrt{T}(e_i'C_k(\widehat{A}_T)-\lambda_g e_i')\widehat{\Gamma}_T)\le\widehat{q}_{g,1-\alpha/2}\Big\}.$$

Figure 1 reports 68% and 95% bootstrap Anderson-Rubin confidence intervals for IRFs and compares them with the delta-method implementation.

Two comments:

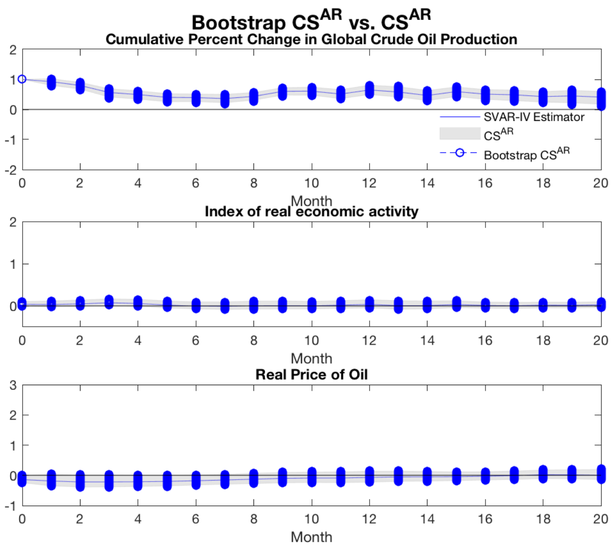
i) Step 2 above, which re-samples the values of the SVAR-IV reduced-form parameters, could be replaced by any other bootstrap procedure, such as the block bootstrap for proxy SVARs recently suggested by Jentsch and Lunsford (2019). One could use their block bootstrap procedure to re-sample the data first, and then obtain the reduced-form parameters for each data realization; instead of implementing step 2.

ii) Step 3 above is the crucial step of our bootstrap-type implementation. The 'standard' bootstrap algorithm computes

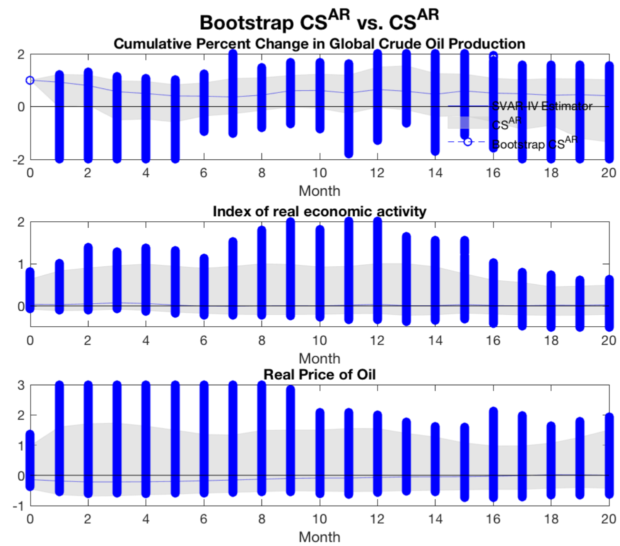
$$\lambda_m=e_i'C_k(A_m)\Big/e_i'\Gamma_m$$

for each re-sampled value of (A_m,Γ_m). A valid confidence interval under strong instruments can be obtained from the quantiles of $\{\lambda_m\}_{m=1}^M$. We also report such bootstrap-like confidence interval in our Matlab suite for comparison to standard delta-method inference. We remark, however, that such procedure is not valid under weak instrument asymptotics.

Figure 1



(a) 68% Bootstrap CS^{AR}



(b) 95% Bootstrap CS^{AR}

A.5. *Details of the Monte Carlo exercise*

We conduct a simple Monte Carlo exercise to analyze the finite-sample coverage of our confidence set. The data generating process for the SVAR-IV model is parameterized by the matrix of autoregressive coefficients, the matrix of contemporaneous impulse-response coefficients, the variance of the structural innovations, and the joint distribution of the external instrument and target shock.

The population parameters in the Monte Carlo (henceforth, MC) design depend on the estimators obtained from the oil SVAR in Kilian (2009). We compute the MC coverage of our confidence interval and also the MC coverage of the standard delta-method confidence set. The details are as follows:

1. Specification of $(A_1, A_2, \ldots, A_p)$: We use Kilian (2009)'s data to estimate the constant term and slope coefficients of the model:

$$Y_t = \mu + A_1 Y_{t-1} + A_2 Y_{t-1} + \ldots + A_{24} Y_{t-1} + \eta_t,$$

with a sample size of $T = 356$. We let $\widehat{\mu}_T$ and $\widehat{A}_T$ denote the least-squares estimators of the parameters μ and A, and we let $\widehat{\Sigma}$ denote the estimated covariance matrix of the reduced-form residuals; which is given by:

$$\widehat{\Sigma} \equiv \frac{1}{T} \sum_{t=1}^{T} \widehat{\eta}_t \widehat{\eta}_t'; \quad \widehat{\eta}_t = Y_t - \widehat{\mu}_T - \widehat{A}_1 Y_{t-1} - \ldots - \widehat{A}_P Y_{t-1}.$$

2. Specification of the first column of the matrix $\Theta_0 = [\Theta_{0,1}, \Theta_{0,2}, \Theta_{0,3}]$: We specify the matrix Θ_0 in three steps. First, we set $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 1$. Second, we specify the first column, denoted $\Theta_{0,1}$. Third, we specify the elements $[\Theta_{0,2}, \Theta_{0,3}]$.

   (a) We propose a DGP in which $\Theta_{0,1}$ is proportional to $e = [1, 1, -1]'$. The signs of this vector are in line with the typical interpretation of an expansionary supply shock. To guarantee that $\Theta_{0,1}$ is still the first column of a square root of $\widehat{\Sigma}$ we set:

$$\widehat{\Theta}_{0,1} = e / \sqrt{e' \widehat{\Sigma}^{-1} e}.$$

   This yields the vector $[2.8276, 2.8276, -2.8276]'$.

   (b) Specification of the second and third column of the matrix $\Theta_0 = [\Theta_{0,1}, \Theta_{0,2}, \Theta_{0,3}]$: To specify the remaining columns of the matrix Θ, we exploit the following observation. Let $D = \text{diag}(\sigma_1^2, \sigma_2^2, \sigma_3^2)$. It is straightforward to show that $BDB' = \Sigma$ holds if and only if:

$$\Theta_{0,l}' \Sigma^{-1} \Theta_{0,m} = 0, \text{ and } \Theta_{0,m}' \Sigma^{-1} \Theta_{0,m} = 1/\sigma_m^2, \text{ for } \quad l, m = 1, 2, 3.$$

Thus, we can compute the orthogonal complement of $\widehat{\Sigma}^{-1/2}\widehat{\Theta}_{0,1}$ and find an orthornormal basis $[\widehat{\gamma}_2, \widehat{\gamma}_3] \in \mathbb{R}^{3\times 2}$ for such space.[2] We can then define the vectors:

$$\widehat{\Theta}_{0,2} \equiv \widehat{\Sigma}^{1/2}\widehat{\gamma}_2, \quad \widehat{\Theta}_{0,3} \equiv \widehat{\Sigma}^{1/2}\widehat{\gamma}_3.$$

Note that since the columns of $[\widehat{\gamma}_2, \widehat{\gamma}_3]$ have unit norm it follows that:

$$\widehat{\Theta}'_{0,j}\widehat{\Sigma}^{-1}\widehat{\Theta}_{0,j} = (\widehat{\Sigma}^{1/2}\widehat{\gamma}_j)'\widehat{\Sigma}^{-1}(\widehat{\Sigma}^{1/2}\widehat{\gamma}_j) = 1, \quad j = 2, 3.$$

Moreover, because the elements $[\widehat{\gamma}_2, \widehat{\gamma}_3]$ are orthogonal then:

$$\widehat{\Theta}'_{0,2}\widehat{\Sigma}^{-1}\widehat{\Theta}_{0,3} = (\widehat{\Sigma}^{1/2}\widehat{\gamma}_2)'\widehat{\Sigma}^{-1}(\widehat{\Sigma}^{1/2}\widehat{\gamma}_3) = 0.$$

Since $[\widehat{\gamma}_2, \widehat{\gamma}_3]$ are both in the orthogonal complement of $\widehat{\Sigma}^{-1/2}\widehat{\Theta}_{0,1}$ it follows that:

$$\widehat{\Theta}'_{0,j}\widehat{\Sigma}^{-1}\widehat{\Theta}_{0,1} = (\widehat{\Sigma}^{1/2}\widehat{\gamma}_j)'\widehat{\Sigma}^{-1}\widehat{\Theta}_{0,1} = 0, \quad j = 2, 3.$$

This means that we can set Θ_0 as:

$$\widehat{\Theta} = [\widehat{\Theta}_{0,1}, \widehat{\Theta}_{0,2}, \widehat{\Theta}_{0,3}] \in \mathbb{R}^{3\times 3},$$

and, by construction, $\widehat{\Theta}$ is guaranteed to be a square-root of $\widehat{\Sigma}$. This gives the matrix:

$$\begin{pmatrix} 2.8276 & -14.1971 & 9.7074 \\ 2.8276 & 1.6411 & 1.7045 \\ -2.8276 & 2.5595 & 3.5324 \end{pmatrix}$$

3. Finally, we propose a joint distribution for the structural innovations and the external instrument. Under the unit variance assumption for the structural shock $\Gamma_1'\Sigma^{-1}\Gamma = \alpha^2$. Thus, we set

$$\widehat{\alpha} \equiv \sqrt{\widehat{\Gamma}'\Sigma^{-1}\widehat{\Gamma}}, \quad \widehat{\Gamma} = \frac{1}{T}\sum_{t=1}^{T} z_t\widehat{\eta}_t.$$

   We introduce an auxiliary variable `auxparam`, define $\tilde{\alpha} \equiv$ `auxparam` $\cdot\, \widehat{\alpha}$ and assume that the data is generated according to:

[2] In Matlab, we find the orthogonal complement of $\widehat{\Sigma}^{-1/2}\widehat{B}_1$ using `null(`$\widehat{B}_1'\widehat{\Sigma}^{-1/2}$`)`

$$(A.7) \qquad Y_t = \widehat{\mu}_T + \widehat{A}_1 Y_{t-1} + \ldots + \widehat{A}_p Y_{t-p} + \widehat{B}\varepsilon_t,$$

$$(A.8) \qquad z_t = \widehat{\mu}_z + \tilde{\alpha}\, \varepsilon_{1,t} + \sqrt{\widehat{\text{Var}}(z_t) - \tilde{\alpha}^2} v_t,$$

$$\begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \\ v_t \end{pmatrix} \sim \mathcal{N}_4(\mathbf{0}, \mathbb{I}_4), \quad \text{i.i.d.}$$

with a vector of p initial conditions equal to the first p observations of Y_t in the data. Note that the specification for z_t in (A.8) corresponds to a simple, linear measurement error model for the external instrument z_t.[3] The parameters of the model for z_t are chosen so that

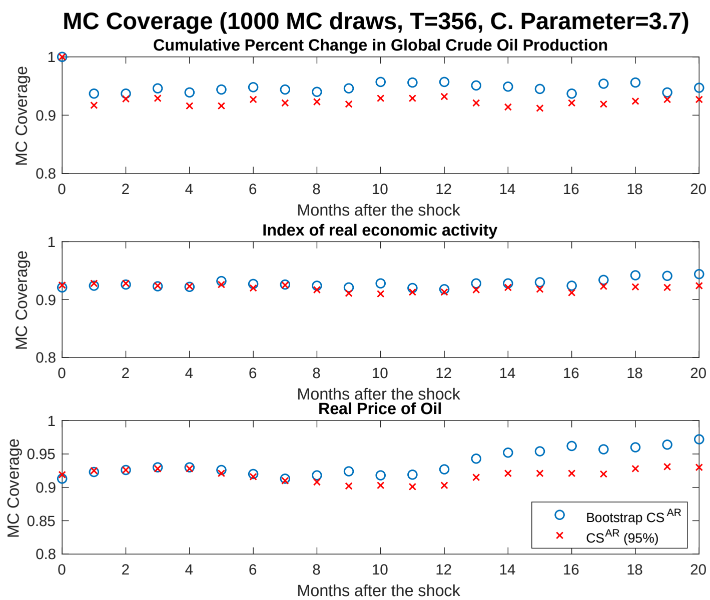
$$\mathbb{E}[z_t] = \widehat{\mu}_z = -0.0182, \quad \text{Var}(z_t) = \widehat{\text{Var}}(z_t) = 0.7436, \quad \text{Cov}(z_t, \epsilon_{1,t}) = \tilde{\alpha}.$$

Under this design, `auxparam` controls the correlation between the external instrument and the target shock. We consider two different values for `auxparam`: 2.3452 and 4.4441. Each of these values correspond to a concentration parameter of 3.7 and 10.09, respectively.

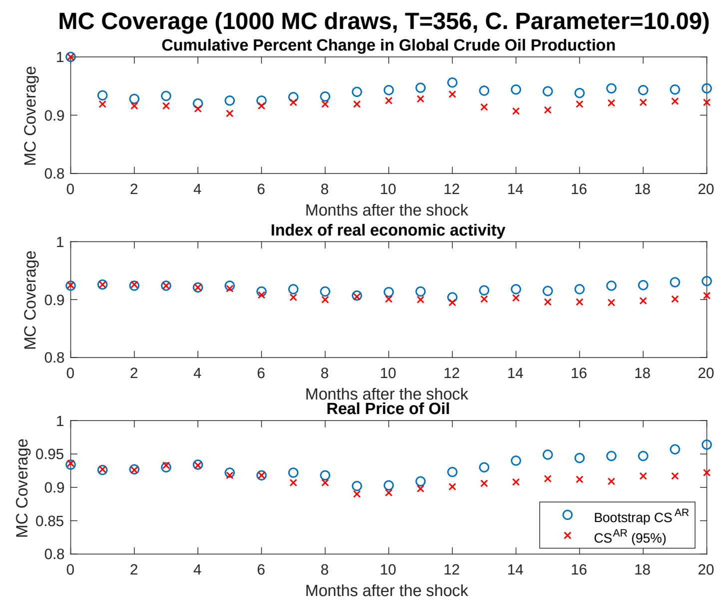
Figure 2 presents the results of the MC coverage for a sample size of $T = 356$ and two different values of the concentration parameter. The comparison is between the CS^{AR} and its bootstrap version, which complements the results reported in the main body of the paper. Figure 3 reports the MC coverage for the standard CS^{AR} and $\text{CS}^{\text{Plug-in}}$ sample size of $T = 1500$.

[3]As we mentioned in the main body of the paper, the validity of our theoretical results do not require a linear measurement error model for z_t. The only restriction we place on the joint distribution of $\{z_t, \varepsilon_t\}_{t=1}^{T}$ are Assumptions 1 and 2. The process is constructed to guarantee that z_t has the same variance as the one estimated from the data.

Figure 2: Coverage of the nominal 95% $\mathrm{CS}^{\mathrm{AR}}$ and bootstrap $\mathrm{CS}^{\mathrm{AR}}$, $T = 356$

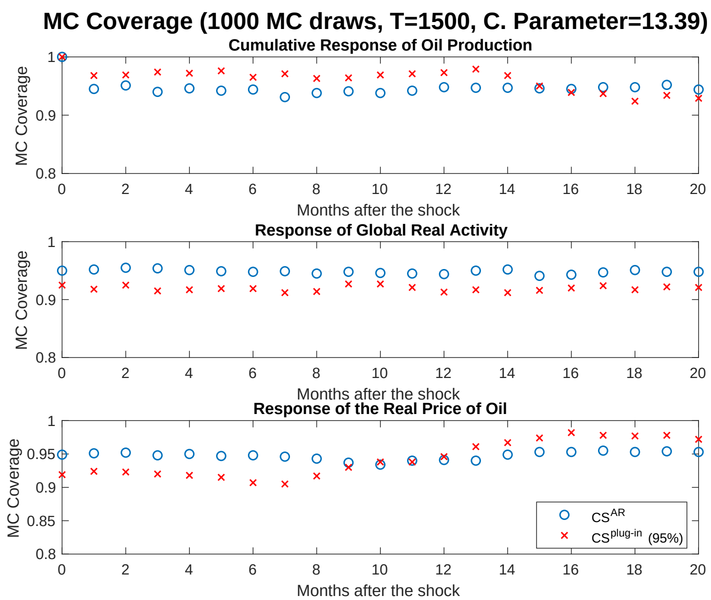


(a) Concentration Parameter = 3.7 (`auxparam`= 2.3452)

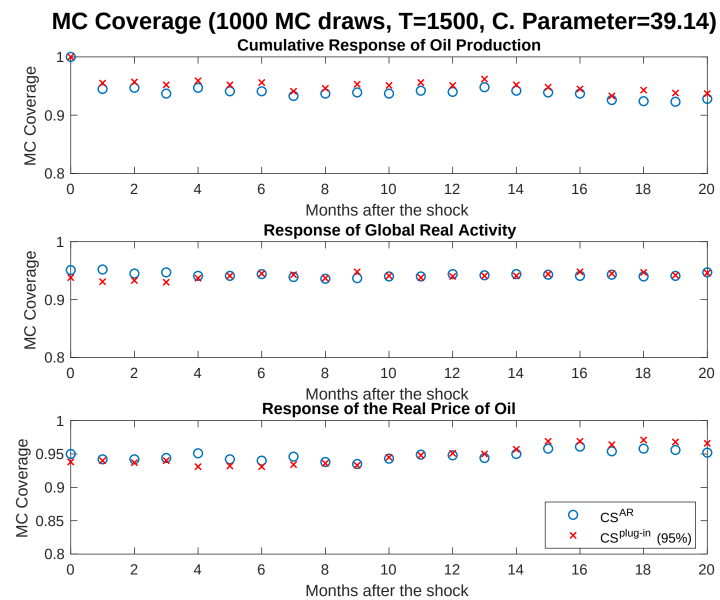


(b) Concentration Parameter = 10.09 (`auxparam`= 4.4441)

Figure 3: Coverage of the nominal 95% standard CS^{AR} and $CS^{\text{Plug-in}}$, $T = 1500$



(a) Concentration Parameter = 13.39 (`auxparam`= 2.3452)



(b) Concentration Parameter = 39.14 (`auxparam`= 4.4441)

A.6. ***An overview of*** r ***external instruments and*** r ***target shocks***

A.6.1. *Notation and Identification*

Let ε_t denote the vector containing the n structural shocks of the model. As in the main text, let

$$\eta_t \equiv \Theta_0 \varepsilon_t = \sum_{j=1}^{n} \Theta_{0,j} \varepsilon_{j,t},$$

and let $D \equiv \mathbb{E}(\varepsilon_t \varepsilon_t')$ be a diagonal matrix. We are interested in the impulse responses of Y_{t+k} to the first r elements of ε_t, denoted by $\varepsilon_{1:r,t} = (\varepsilon_{1,t}, \varepsilon_{2,t}, ..., \varepsilon_{r,t})'$. As shown in equation (1.6) in the main text, these impulse responses are a linear function of the first r columns of Θ_0, which we denote by $\Theta_{0,1:r}$. In this subsection we discuss identification of $\Theta_{0,1:r}$ and $D_{1:r,1:r} = E(\varepsilon_{1:r,t} \varepsilon_{1:r,t}')$ using a $r \times 1$ vector of external instruments z_t. The $\mathbb{R}^r$-valued random vector z_t satisfies the natural extension of Assumption 1:

Assumption 1*:

1. $\mathbb{E}\left[\varepsilon_{1:r,t} z_t'\right] = \Phi$, where $\Phi \in \mathbb{R}^{r \times r}$ has full rank (*relevance*)
2. $\mathbb{E}\left[\varepsilon_{j,t} z_t'\right] = \mathbf{0}$ for $j > r$ (*exogeneity*).

Under these assumptions:

$$\Gamma \equiv \mathbb{E}(\eta_t z_t') = \Theta_{0,1:r} \Phi. \tag{A.9}$$

The identification problem involves determining the value of $(\Theta_{0,1:r}, D_{1:r,1:r})$ from the second moments of (η_t, z_t).

As a prelude to solving this problem, suppose for a moment that the random vector

$$a_t = \Theta_{0,1:r} \varepsilon_{1:r,t}$$

was observed. In this case, the covariance matrix of a_t *partially identifies* the model's parameters by the equation

$$\Sigma_{aa} \equiv \Theta_{0,1:r} D_{1:r,1:r} \Theta_{0,1:r}. \tag{A.10}$$

This means $(\Theta_{0,1:r}, D_{1:r,1:r})$ could be determined from Σ_{aa} after imposing $r(r+1)/2$ additional *a priori* restrictions on $(\Theta_{0,1+r}, D_{1:r,1:r})$. This is the usual identification problem in structural vector autoregressions, but now involving only r shocks instead of the original n shocks. As an example, when $r = 1$, only one additional restriction is required, and in the main text this was the unit-effect normalization $\Theta_{0,11} = 1$. When $r > 1$, the standard 'Wold' (or Cholesky) restriction is that $\Theta_{0,1:r}$ is lower triangular, the unit-effect normalization could be used to impose unit coefficients on the diagonal, and together these yield the required $r(r+1)/2$ restrictions. Other restrictions could be used instead; for example, a long-run Wold causal ordering, or, as we discuss in detail in an example below, the assumptions used in Mertens and Ravn (2013).

With this prelude, the identification problem for $(\Theta_{0,1:r}, D_{1:r,1:r})$ becomes the problem of determining Σ_{aa} from the second moments of (η_t, z_t). A direct calculation yields[4]

$$\Sigma_{aa} = \Gamma(\Gamma' \Sigma_{\eta\eta}^{-1} \Gamma)^{-1} \Gamma'. \tag{A.11}$$

An intuitive derivation of this result can be given using linear projection arguments. Let $Proj(Y|X) = \Sigma_{YX}\Sigma_{XX}^{-1}X$ denote the projection of the random vector Y onto the random vector X, then

$$Proj(z_t|\eta_t) = P(z_t|\varepsilon_t) = P(z_t|\varepsilon_{1:r,t}) = \Phi' D_{1:r,1:r}^{-1} \varepsilon_{1:r,t}$$

where the first equality follows from $\varepsilon_t = \Theta_0^{-1}\eta_t$ and the final two equalities follow from Assumption 1*. This means that, by our assumptions, the best linear predictor of the instrument in terms of the structural shocks of interest is exactly the linear combination that entangles the impulse response coefficients. Thus the linear combination a_t can be re-expressed as the best linear predictor of the residuals η_t in terms of $Proj(z_t|\eta_t)$:

$$\begin{aligned} a_t &= \Theta_{0,1:r}\varepsilon_{1:r,t} \\ &= Proj(\eta_t|\varepsilon_{1:r,t}) \\ &= Proj(\eta_t|\Phi' D_{1:r,1:r}^{-1}\varepsilon_{1:r,t}) \\ &= Proj(\eta_t|Proj(z_t|\eta_t)) \end{aligned}$$

where the third equality follows from the non-singularity of Φ (Assumption 1*). A direct

[4]Since $\eta \equiv \Theta_0\varepsilon_t$, then $\Sigma_{\eta\eta}^{-1} = (\Theta_0')^{-1} D^{-1} (\Theta_0)^{-1}$. Let s_r denote the matrix that collects the first r columns of the identity matrix of dimension $r \times r$. Since $\Gamma = \Theta_{0,1:r}\Phi = \Theta_0 s_r \Phi$, then

$$\Gamma' \Sigma_{\eta\eta}^{-1} \Gamma = (\Phi' s_r' \Theta_0) \left((\Theta_0')^{-1} D^{-1} (\Theta_0)^{-1} \right) \Theta_0 s_r \Phi = \Phi' D_{1:r,1:r}^{-1} \Phi.$$

Under the assumption that Φ has full rank, then

$$\Gamma(\Gamma' \Sigma_{\eta\eta}^{-1} \Gamma)^{-1} \Gamma' = (\Theta_{0,1:r}\Phi)\,(\Phi)^{-1} D_{1:r,1:r} (\Phi')^{-1} (\Phi' \Theta_{0,1:r}') = \Sigma_{aa}$$

calculation yields the desired result

$$\Sigma_{aa} = \Gamma(\Gamma'\Sigma_{\eta\eta}^{-1}\Gamma)^{-1}\Gamma'$$

so that Σ_{aa} is identified. Thus, the identification of $(\Theta_{0,1:r}, D_{1:r,1:r})$ follows from the factorization of $\Gamma(\Gamma'\Sigma_{\eta\eta}^{-1}\Gamma)^{-1}\Gamma' = \Theta_{0,1:r}D_{1:r,1:r}\Theta_{0,1:r}$ after imposing $r(r+1)/2$ additional *a priori* restrictions so the factorization provides a unique solution.

This calculation highlights the role of z_t in the analysis: the covariance of z_t and η_t isolates the linear combinations of η_t associated with $\varepsilon_{1:r,t}$. That is, the instruments reduce the standard SVAR identification problem from n dimensions (requiring $n(n+1)/2$ identifying restrictions) to r dimensions (requiring only $r(r+1)/2$ restrictions).

A.6.2. *Weak instrument robust inference for impulse responses*

Let ϑ denote the $j \times 1$ vector of (possibly dynamic) impulse responses with respect some of to the structural shocks in $\varepsilon_{1:r,t}$. From equation (1.6) in the main text, we can write $\vartheta = C(A)vec(\Theta_{0,1:r})$, where the matrix $C(A)$ of dimension $j \times nr$ depends on the relevant horizons and variables of interest. In this subsection we outline weak-instrument robust methods for inference about ϑ.

It is useful to review the weak instrument inference strategy discussed in the main text for $r = 1$. There, equation (A.9) together with the unit-effect normalization $\Theta_{0,11} = 1$, allowed us to solve for the value Φ as a linear function of Γ; in particular, $\Phi = \Gamma_{11}$, so that (A.9) becomes $\Gamma - \Theta_{0,1}\Gamma_{11} = 0$. Using this, linear restrictions on $\Theta_{0,1}$, say $C\Theta_{0,1} = c$, imply linear restrictions on Γ. And, because $\hat{\Gamma} \overset{a}{\sim} N(\Gamma, V)$ holds regardless of the value of Φ (that is, regardless of the strength of the instruments), standard Wald tests of these linear restrictions on Γ provide weak-instrument robust tests of the null hypothesis that $C\Theta_{0,1} = c$.

This insight generalizes directly to the model with $r > 1$, with one important caveat: solving Φ as a *linear function* of Γ using the equation $\Gamma = \Theta_{0,1:r}\Phi$, requires r^2 *a priori* restrictions on $(\Theta_{0,1:r}, \Phi)$. When $r = 1$, the unit-effect normalization suffices, but when $r > 1$ more restrictions are needed.[5]

The next section discusses an example of these r^2 restrictions motivated by the empirical model studied in Mertens and Ravn (2013). In any event, with these restrictions in hand, solving for Φ can be shown to yield a linear estimator for the nuisance parameter $vec(\Phi)$; that is, we can show that the additional restrictions imply: $vec(\Phi) = Bvec(\Gamma)$, where B is a $r^2 \times nr$ matrix that depends on the r^2 *a priori* restrictions imposed on $(\Theta_{0,1:r}, \Phi)$. This,

[5]For example, in some applications these restrictions might impose that the first r rows of $\Theta_{0,1:r}$ are the identity matrix: this yields $\Phi = \Gamma_{1:r,1:r}$. Or, in another application, Φ might be restricted to be diagonal (so that each instrument is correlated with a unique element of ε), and the r diagonal elements of $\Theta_{0,1:r}$ restricted to be unity (the unit-effect normalization); in this case the diagonal elements of Φ are given by the diagonal elements of Γ.

together with $\Gamma = \Theta_{0,1:r}\Phi$, implies $vec(\Gamma) - (\mathbb{I}_r \otimes \Theta_{0,1:r})(Bvec(\Gamma)) = 0$. It is then possible to construct a weak-instrument robust confidence interval for $\Theta_{0,1:r}$ (inverting a Wald test) and project the resulting confidence region to conduct inference about ϑ. As in the model with $r = 1$, the resulting tests are robust to weak instruments.

As discussed in the last subsection, identification of $\Theta_{0,1:r}$ required Assumption 1* together with $r(r+1)/2$ *a priori* restrictions. In constrast, the weak-instrument robust inference outlined in the preceding paragraph requires more restrictions. In particular, we show that r^2 *a priori* restrictions suffice. When $r > 1$, this implies that we will have $r(r-1)/2$ over-identifying restrictions. These over-identifying restrictions are easily tested under strong-instruments, but weak-instrument robust tests for over-identifying restrictions are more difficult to construct, and we left this question out for future research.

A.7. *An example: two external instruments for two target shocks*

This subsection works through a specific example inspired by the empirical model in Mertens and Ravn (2013). In this example, $r = 2$, and the goal is to identify $[\Theta_{0,1}\Theta_{0,2}]$. We will start by showing how to build a weak-instrument robust confidence set for the full vector of contemporaneous impulse responses, $\Theta_{0,1:2}$. The test will be based on the S-test of Stock and Wright (2000). We then focus on the problem of inference about a vector of dynamic impulse responses, where we focus on the case in which the object of interest is the response of variable i to the two structural shocks of the model $(\varepsilon_{1,t}, \varepsilon_{2,t})$, k periods ahead in the future.

A.7.1. *Identifying restrictions.*

As in the discussion above, 2 instruments and $r(r+1)/2 = 3$ identifying restrictions suffice for identification when Φ has full rank. The first identifying restriction we will impose is

$$c'\Theta_{0,1} = 0, \tag{A.12}$$

where c is an $\mathbb{R}^{n\times 1}$ vector that is allowed to depend on (A, Σ). This type of restriction is used in Mertens and Ravn (2013).

The other two restrictions are "unit effect normalization" restrictions. In particular, we require that each of the two shocks have a unit impact effect of $Y_{1,t}$:

$$e_1'[\Theta_{0,1}, \Theta_{0,2}] = [1, 1]. \tag{A.13}$$

To verify that these these assumptions identify $\Theta_{0,1:2}$, note that under Assumption 1*

$$[e_1, c]'\Gamma = \begin{pmatrix} 1 & 1 \\ 0 & c'\Theta_{0,2} \end{pmatrix} \Phi. \tag{A.14}$$

Φ has full rank. If we assume further that $c'\Theta_{0,2} \neq 0$ then the left hand side of equation (A.14) has also full rank.[6] Consequently,

$$[\Theta_{0,1}, \Theta_{0,2}] = \Gamma\Phi^{-1} = \Gamma([e_1, c]'\Gamma)^{-1} \begin{pmatrix} 1 & 1 \\ 0 & c'\Theta_{0,2}. \end{pmatrix}. \tag{A.16}$$

This implies that $\Theta_{0,1}$ is identified as

$$\Theta_{0,1} = \Gamma([e_1, c]'\Gamma)^{-1}(1, 0)'. \tag{A.17}$$

Next, since Θ_0 is also assumed to be invertible, then $\Theta_{0,2}$ satisfies the restriction $\Theta_{0,1}'\Sigma^{-1}\Theta_{0,2} = 0$. Using an analogous argument to the one used to get an expression for $\Theta_{0,1}$ we get

$$\Theta_{0,2} = \Gamma([e_1, \Sigma^{-1}\Theta_{0,1}]'\Gamma)^{-1}(1, 0)'. \tag{A.18}$$

These identification results have been used elsewhere in the literature; for example in the work of Mertens and Ravn (2013). The results we presented in this section simplify some of their algebra, and we present them here for the sake of exposition.

The identification results also imply that strong-identification inference (derived under the assumption that Φ has full rank and that $c'\Theta_{0,2} \neq 0$) is straightforward under the following extension of Assumption 2:

Assumption 2*:**

$$\sqrt{T} \begin{pmatrix} \text{vec}(\widehat{A}_T - A) \\ \text{vec}(\widehat{\Gamma}_T - \Gamma_T) \\ \text{vech}(\widehat{\Sigma}_T - \Sigma) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \zeta \\ \xi \\ \phi \end{pmatrix} \sim \mathcal{N}(0, W) \tag{A.19}$$

[6]Crucially, in order to distinguish the structural shocks we also require that

$$c'\Theta_{0,2} \neq 0. \tag{A.15}$$

This means that the zero restriction is binding for the response of the first structural shock, but not for the second. Intuitively, this is necessary to be able to distinguish the responses to the two structural shocks.

A.7.2. S-*test for contemporaneous impulse responses*

Throughout this section, and for simplicity in the exposition, we assume that c is a deterministic vector (this happens, for example, when the zero restrictions are imposed on the contemporaneous impulse responses).

Consider the null hypothesis

$$H_0 : [\Theta_{0,1}, \Theta_{0,2}] = [\Theta^*_{0,1}, \Theta^*_{0,2}]$$

The null hypothesis above restricts all of the contemporaneous impulse responses to the two structural shocks of interest. This also imposes $2n$ restrictions, which is larger than r^2 as long as $n > 2$. It is therefore possible to construct a linear estimator for the nuisance parameter Φ. Our off-the-shelf suggestion is based on the S-test of Stock and Wright (2000).

Under the null hypothesis, we can use equation (A.14) to estimate Φ using the linear estimator:

$$\widehat{\Phi}_0 \equiv \begin{pmatrix} 1 & 1 \\ 0 & c'\Theta^*_{0,2} \end{pmatrix}^{-1} [e_1, c]'\widehat{\Gamma},$$

This estimator belongs to the general class we discussed in Section A.6.2, since standard properties of the *vec* operator imply $vec(\widehat{\Phi}_0)$ is a linear combination of $vec(\widehat{\Gamma})$. Assumption 2^* readily implies that $\widehat{\Phi}_0$ is consistent and asymptotically normal regardless of the identification strength. The S-test uses the statistic

$$\Psi(\Theta^*_{0,1}, \Theta^*_{0,2}) \equiv \sqrt{T}\text{vec}\left(\widehat{\Gamma} - [\Theta^*_{0,1}, \Theta^*_{0,2}]\widehat{\Phi}_0\right). \tag{A.20}$$

Assumption 1^* and 2^* imply that under the null hypothesis

$$\sqrt{T}\text{vec}\left(\widehat{\Gamma} - [\Theta^*_{0,1}, \Theta^*_{0,2}]\widehat{\Phi}_0\right)$$

equals

$$\begin{aligned} & \sqrt{T}\text{vec}\left(\widehat{\Gamma} - [\Theta^*_{0,1}, \Theta^*_{0,2}]\begin{pmatrix} 1 & 1 \\ 0 & c'\Theta^*_{0,2} \end{pmatrix}^{-1} [e_1, c]'\widehat{\Gamma}\right), \\ = \ & \sqrt{T}\text{vec}\left((\widehat{\Gamma} - \Gamma) - [\Theta^*_{0,1}, \Theta^*_{0,2}]\begin{pmatrix} 1 & 1 \\ 0 & c'\Theta^*_{0,2} \end{pmatrix}^{-1} [e_1, c]'(\widehat{\Gamma} - \Gamma)\right), \end{aligned}$$

where the last line uses the fact that under the null

$$\Gamma = [\Theta^*_{0,1}, \Theta^*_{0,2}]\Phi_0, \quad \text{and} \quad \Phi_0 = \begin{pmatrix} 1 & 1 \\ 0 & c'\Theta^*_{0,2} \end{pmatrix}^{-1} [e_1, c]'\Gamma.$$

Therefore, under Assumptions 1*-2* and the null hypothesis:

$$\Psi(\Theta^*_{0,1}, \Theta^*_{0,2}) \xrightarrow{d} \mathcal{N}_{2n}(0, M_0), \tag{A.21}$$

where algebra provides a closed-form expression for M_0.[7]

The $(1-\alpha)100\%$ S-test for the null hypothesis $H_0 : [\Theta_{0,1}, \Theta_{0,2}] = [\Theta^*_{0,1}, \Theta^*_{0,2}]$ can then be defined as the test that rejects whenever

$$S(\Theta^*_{0,1}, \Theta^*_{0,2}) \equiv \Psi(\Theta^*_{0,1}, \Theta^*_{0,2})' \widehat{M}_0^{-1} \Psi(\Theta^*_{0,1}, \Theta^*_{0,2}) > \chi^2_{2n,1-\alpha}, \tag{A.22}$$

where $\chi^2_{2n,1-\alpha}$ denotes the $1-\alpha$ upper quantiles of a χ^2 random variable with $2n$ degrees of freedom (n is the dimension of the SVAR). The algebra above (in particular, equation A.21) shows that S-test is a valid test for $H_0 : [\Theta_{0,1}, \Theta_{0,2}] = [\Theta^*_{0,1}, \Theta^*_{0,2}]$, regardless of the rank of Φ.

A.7.3. *S-region for dynamic impulse responses*

The S-region, defined as the collection of all values of $[\Theta_{0,1}, \Theta_{0,2}]$ that cannot be rejected by the S-test, provides a weak-instrument robust confidence set for the full vector of *contemporaneous* impulse responses.

In this subsection, we how show to combine the Projection and Bonferroni methods to construct a confidence region for *dynamic* impulse responses.

The k-period ahead response of variable i to a shock in $\varepsilon_{1,t}$ is defined as

$$\lambda_{k,i,1} \equiv e_i' C_k(A) \Theta_{0,1}$$

This parameter depends not only on $\Theta_{0,1}$ but also on the autoregressive coefficient, A. If A were known, a standard application of the projection method would yield a valid (but conservative) confidence set for the impulse response coefficient. We can account for the uncertainty in A by further relying on Bonferroni's method, in addition to the Projection method.

Let $CS(\Theta_{0,1}, 1-\beta)$ be a $(1-\beta)100\%$ confidence region for $\Theta_{0,1}$ (for example, the projection of the S-region obtained by inverting the S-test in A.22) and let $CS(A, 1-\eta)$ be a $(1-\eta)100\%$ confidence region for the parameter A (obtained, for example, using a typical Wald ellipse).

[7] In particular, the expression is

$$\left(\mathbb{I}_2 \quad \otimes \quad \left(\mathbb{I}_n - [\Theta^*_{0,1}, \Theta^*_{0,1}] \begin{pmatrix} 1 & 1 \\ 0 & c'\Theta^*_{0,2} \end{pmatrix}^{-1} [e_1, c]' \right)\right) W \left(\mathbb{I}_n - [\Theta^*_{0,1}, \Theta^*_{0,1}] \begin{pmatrix} 1 & 1 \\ 0 & c'\Theta^*_{0,2} \end{pmatrix}^{-1} [e_1, c]' \right)'.$$

Consider the confidence region $CS(\lambda,\beta,\eta)$ for $\lambda_{k,i,1}$ defined by

$$\{\lambda \in \mathbb{R} \mid \lambda = e_i' C_k(A)\Theta_{0,1} \quad \text{for} \quad \Theta_{0,1} \in CS(\Theta_{0,1}, 1-\beta), A \in CS(A, 1-\eta)\}. \tag{A.23}$$

Let $\lambda_0 = e_i' C_k(A_0)\Theta_{0,1}$ denote the true value of $\lambda_{k,i,1}$. Then,

$$\begin{aligned}
\mathbb{P}(\lambda_0 \in CS(\lambda,\beta,\eta)) &\geq \mathbb{P}(A_0 \in CS(A, 1-\eta) \text{ and } \Theta_{0,1} \in CS(\Theta_{0,1}, 1-\beta)), \\
&= 1 - P(A_0 \notin CS(A, 1-\eta) \quad \text{or} \quad \Theta_{0,1} \notin CS(\Theta_{0,1}, 1-\beta)), \\
&\geq 1 - P(A_0 \notin CS(A, 1-\eta) - P(\Theta_{0,1} \notin CS(\Theta_{0,1}, 1-\beta)), \\
&\geq 1 - (\eta + \beta).
\end{aligned}$$

Thus, if $\eta + \beta \leq \alpha$ the confidence set in (A.23) provides a valid (but conservative) confidence interval for $\lambda_{k,i,1}$. This shows that a relative off-the-shelf application of the results in Stock and Wright (2000) allows us to conduct valid (but conservative) weak-identification robust inference for dynamic impulse responses, as long as we maintain the assumption that $c'\Theta_{0,2}$ is bounded away from zero.

A.7.4. *Confidence Sets for the dynamic response of one variable to both structural shocks*

We have shown that the S-test can be used to construct a confidence region for the full vector of contemporaneous impulse responses of interest ($\Theta_{0,1}$ and $\Theta_{0,2}$). We have also shown that we can combine Projection/Bonferroni adjustments to construct a valid (but conservative) confidence region for dynamic impulse responses.

In this section we argue that we can construct a confidence region for the dynamic responses of a particular variable to both shocks by inverting a test for the following null hypothesis

$$\mathbb{H}_0 : \lambda_{k,i,1} = a_0,\ \lambda_{k,i,2} = b_0,\ c'\Theta_{0,2} = c_0.$$

Note that instead of postulating a null value for the parameters $[\Theta_{0,1}, \Theta_{0,2}]$ we only postulate a value for the three scalars: the dynamic response of a particular variable at a particular horizon, and the value of $c'\Theta_{0.2}$). An important remark is that we use the hypothesized value c_0 in order to have an estimator of Φ that is robust to the identification strength. Namely:

$$\widehat{\Phi}_{c_0} \equiv \begin{pmatrix} 1 & 1 \\ 0 & c_0 \end{pmatrix}^{-1} [e_1, c]'\widehat{\Gamma}. \tag{A.24}$$

We can then suggest a test for the null hypothesis using the statistic

$$Q_T(a_0, b_0, c_0) \equiv \sqrt{T}\text{vec}\left(e_i'C_k(\widehat{A})\widehat{\Gamma} - [a_0, b_0]\widehat{\Phi}_{c_0}\right). \quad \text{(A.25)}$$

The expression above can be viewed as a natural extension of the statistic we used to construct the Anderson-Rubin test in the case with only one structural shock. The key difference is that in a model with two instruments and two shocks we also need to postulate a value for $c'\Theta_{0,2}$. As we mentioned before, this value is used to construct an estimator of the nuisance parameter Φ whose performance does not depend on the identification strength.

Assumptions 1^* and 2^* imply that under the null hypothesis

$$Q_T(a_0, b_0, c_0) \xrightarrow{d} \mathcal{N}_2(0, V_0), \quad \text{(A.26)}$$

where V_0 can be readily obtained by a straightforward application of the δ-method. Thus, a valid α-level test for the null hypothesis (regardless of the rank of Φ) rejects whenever

$$Q_T(a_0, b_0, c_0)'V_0^{-1}Q_T(a_0, b_0, c_0) > \chi^2_{2,1-\alpha}. \quad \text{(A.27)}$$

Interestingly, when Φ has full rank, this test has the same local power as the test for $H_0 : \lambda_{i,k,1} = a_0, \lambda_{i,k,2} = b_0$ based on the plug-in estimators of the dynamic impulse responses (assuming $c'\Theta_{0,2}$ is known).

Our approach will provide weak-instrument robust inference as long as $c'\Theta_{0,2}$ is bounded away from zero. Thus, our approach makes sense as long as the researcher's main concern is the strength of the correlation between the instruments and the target shocks, and not the validity of the additional zero restriction imposed.

A.8. ***Marginal Tax Rates and GDP***

In this section we use our inference tools to revisit the question of whether real economic activity in the United States (measured by the Gross Domestic Product, henceforth GDP) responds to cuts in marginal tax rates. This question received renewed attention in 2017-2018, during the discussion of the costs and benefits of the 2017 tax law (initially referred to as *Tax Cuts and Jobs Acts*). This tax reform implied, among its provisions, a 2.3 percentage point reduction in average marginal tax rates; see Barro and Furman (2018) p. 298.

We base our analysis on the recent work of Mertens and Montiel Olea (2018), which use the SVAR-IV framework outlined in this paper—along with the strong-instrument inferential methods herein suggested—to study the effect of exogenous changes in marginal tax rates over different macroeconomic variables. It is worth mentioning that even though the main focus of their work is the estimation of the *short-run elasticity of taxable income*, some of the references to their results—for example, the Economic Report of the President of the Council of Economic Advisers (2018)—emphasized their findings concerning GDP.

Consider the following VAR representation with two lags for the log of real GDP and average marginal tax rates

$$\begin{bmatrix} -\ln(1-AMTR_t) \\ \ln(GDP_t) \\ X_t \end{bmatrix} = \mu + \underbrace{A(L)}_{p=2} \begin{bmatrix} -\ln(1-AMTR_{t-1}) \\ \ln(income_{t-1}) \\ X_{t-1} \end{bmatrix} + \Theta_0 \begin{bmatrix} \varepsilon_t^{AMTR} \\ \varepsilon_t^{income} \\ \varepsilon_t^{x} \end{bmatrix},$$

where X_{t-1} is a vector of control variables of dimension $n-2$ (so that there are a total of n variables in the VAR) and the matrix Θ_0 need not be invertible (ε_t^x can have larger dimension than X_{t-1}).[8] The data are yearly, and available from 1948 to 2018, but we decide to focus on 1950-2008.[9] The construction of the external instrument is detailed in Section IV.C of Mertens and Montiel Olea (2018).[10] The first-stage statistic ξ_1 described in Section 4.2 of the main body of the paper (and using Newey and West (1987) standard errors with

[8]The controls we use are the same as in Mertens and Montiel Olea (2018), and include the unemployment rate, the log real stock market index, inflation and the Federal funds rate, log real government spending per tax unit (purchases and net transfers) and the change in log real federal government debt per tax unit.

[9]This is a common approach in practice, and avoids the inclusion of the major 1948 tax reform and the recession dummy for 1949.

[10]In a nutshell, the instrument is based on the Romer and Romer (2010) classification of postwar legislated tax changes but excludes changes that i) respond to current/planned changed in government spending (e.g., increase in payroll taxes due to Medicare 1965); ii) respond to current/expected economic conditions (e.g., tax cut in Tax Reduction act of 1975); iii) are legislated at least 1 year before becoming effective. There are fifteen tax reforms that satisfy such criteria, but only eight of them include direct changes to the income rate tax schedules (1948, 1964, 1978, 1981, 1986, 1990, 1993, and 2003). The external IV consists of the eight scored changes in the average marginal tax rate.

8 lags) is 7.0832. The usual HAC first-stage statistic is 34.0610.

We construct confidence intervals for the dynamic responses to innovations in ε_t^{AMTR} that decrease $-\ln(1 - AMTR)$ in 1% (this is the unit effect normalization).

Figure 4 reports 95% (strong-instrument) bootstrap and delta-method confidence intervals for the responses of income and GDP. The 1-period ahead response of GDP is estimated to be around .7956%. and is significant under both bootstrap and the delta-method inference. The 1-period ahead response of income (which is typically used as an estimator of the *short-run* elasticity of taxable income) is 1.3311. This elasticity parameter is significantly different from zero under both the delta method and the bootstrap.

Figure 5 reports the bootstrap and delta-method versions of the weak-instrument robust confidence interval suggested in this paper. The bootstrap confidence regions are represented by circles, and each of the circles represent one particular null hypothesis that we were not able to reject using the procedure described in A.4. The gray area represents the Anderson-Rubin confidence region computed using the delta method. This area is also obtained via test inversion, but the closed-form formulae discussed in Section 4.1 implies there is no need to conduct grid search.

The figures show that the bootstrap version of the Anderson-Rubin confidence set cannot reject the null hypothesis (at the 5% level) that the effect of cuts in marginal tax rates over GDP is zero. This stands in contrast to the results obtained using inference based on the assumption of strong instruments. The figure also shows that the short-run elasticity of taxable income remains statistically significant. The confidence interval for this parameter is wider than its strong-instrument counterpart, but most of the increase in uncertainty involves potentially higher values.

Figure 4

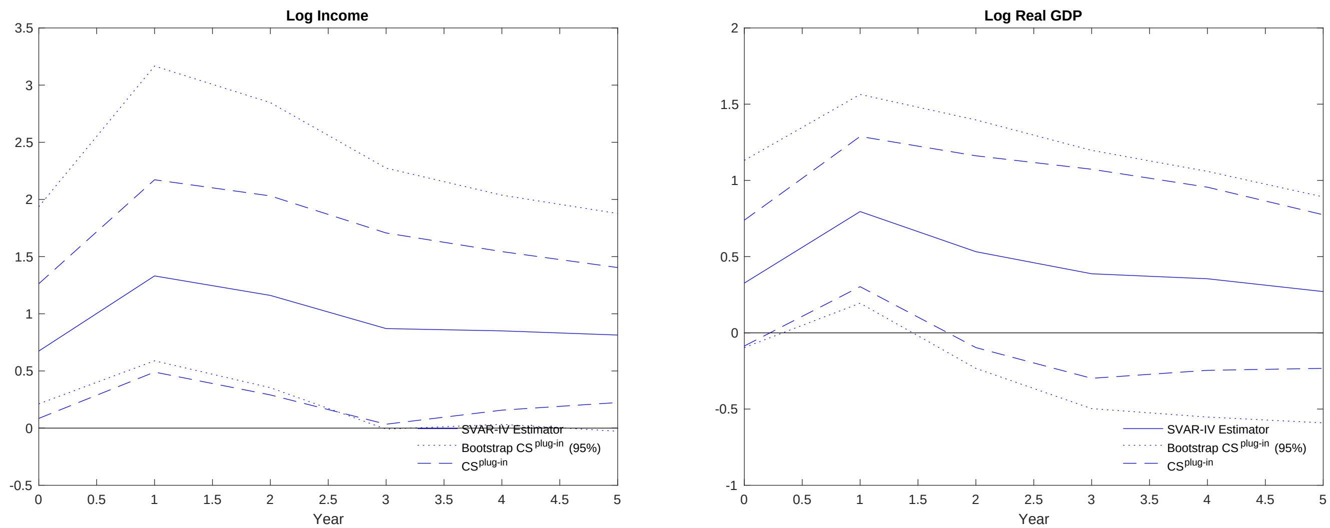
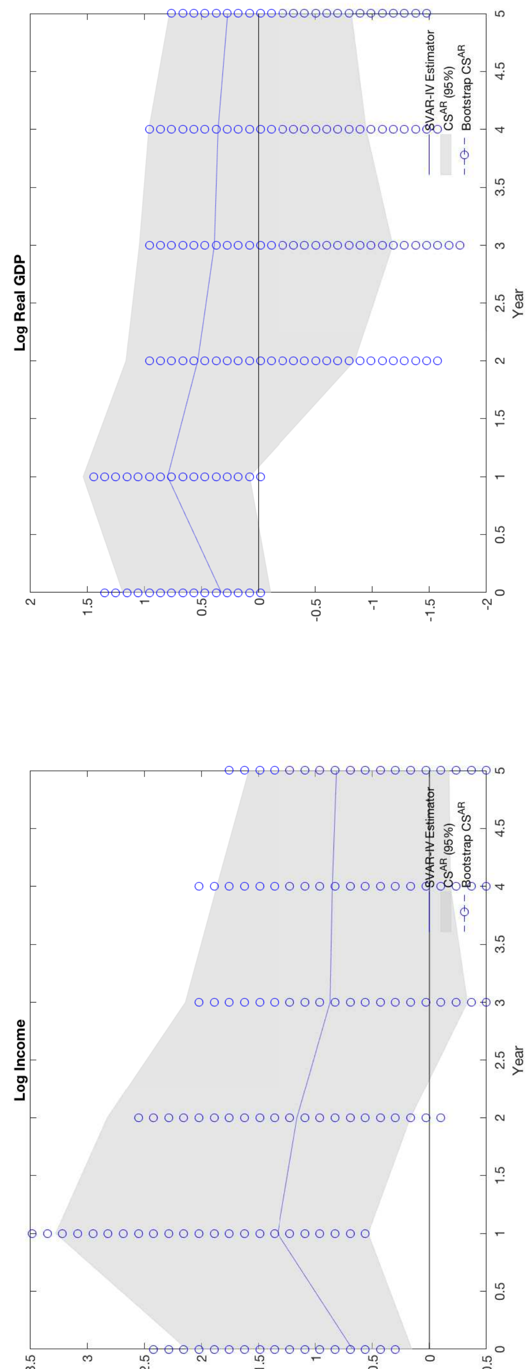

Log Income
Log Real GDP
Year
Year
SVAR-IV Estimator
Bootstrap CS$^{plug-in}$ (95%)
CS$^{plug-in}$

Figure 5


Log Income
Log Real GDP
SVAR-IV Estimator
CS^AR (95%)
Bootstrap CS^AR
Year

REFERENCES

Andrews, D., M. Moreira, and J. Stock (2006): "Optimal Two-sided Invariant Similar Tests for Instrumental Variables Regression," *Econometrica*, 74, 715–52.

Andrews, I. (2016): "Conditional Linear Combination Tests for Weakly Identified Models," *Forthcoming, Econometrica.*

Barro, R. J. and J. Furman (2018): "Macroeconomic effects of the 2017 tax reform," *Brookings papers on economic activity*, 2018, 257–345.

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

Council of Economic Advisers, . (2018): "Economic Report of the President, February 2018," .

Fieller, E. C. (1954): "Some Problems in Interval Estimation," *Journal of the Royal Statistical Society. Series B (Methodological)*, 175–185.

Hayashi, F. (2000): "Econometrics Princeton University Press," .

Jentsch, C. and K. G. Lunsford (2019): "Asymptotically Valid Bootstrap Inference for Proxy SVARs," .

Kilian, L. (1998): "Small-sample confidence intervals for impulse response functions," *Review of economics and statistics*, 80, 218–230.

——— (2009): "Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market," *The American Economic Review*, 1053-1069.

Kleibergen, F. (2002): "Pivotal statistics for testing structural parameters in instrumental variables regression," *Econometrica*, 70, 1781–1803.

——— (2007): "Generalizing weak instrument robust IV statistics towards multiple parameters, unrestricted covariance matrices and identification statistics," *Journal of Econometrics*, 139, 181–216.

Mertens, K. and J. L. Montiel Olea (2018): "Marginal tax rates and income: New time series evidence," *The Quarterly Journal of Economics*, 133, 1803–1884.

Mertens, K. and M. O. Ravn (2013): "The Dynamic Effects of Personal and Corporate

Income Tax Changes in the United States," *American Economic Review*, 103 (4), 1212–1247.

Moreira, M. (2003): "A conditional likelihood ratio test for structural models," *Econometrica*, 71, 1027–1048.

Moreira, M. and H. Moreira (2015): "Optimal Two-Sided Tests for Instrumental Variables Regression with Heteroskedastic and Autocorrelated Errors," Working Paper, EPFGV.

Müller, U. K. (2011): "Efficient Tests under a Weak Convergence Assumption," *Econometrica*, 79, 395–435.

Newey, W. and D. McFadden (1994): "Large sample estimation and hypothesis testing," *Handbook of econometrics*, 2111–2245.

Newey, W. K. and K. D. West (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Romer, C. D. and D. H. Romer (2004): "A new measure of monetary shocks: Derivation and implications," *American Economic Review*, 94.

——— (2010): "The macroeconomic effects of tax changes: Estimates based on a new measure of fiscal shocks," *The American economic review*, 100, 763–801.

Sims, C. A. and T. Zha (1999): "Error bands for impulse responses," *Econometrica*, 67, 1113–1155.

——— (2006): "Were there regime switches in US monetary policy?" *The American Economic Review*, 54–81.

Smets, F. and R. Wouters (2007): "Shocks and frictions in US business cycles: A Bayesian DSGE approach," *The American Economic Review*, 97, 586–606.

Staiger, D. and J. Stock (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557–586.

Stock, J. and J. Wright (2000): "GMM with Weak Identification," *Econometrica*, 68, 1055–1096.