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:

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

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:

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

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:

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

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 $\sigma^2(\lambda_{k,i})$, which equals 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,

$$X_T^2/\widehat{\sigma}_T^2(\lambda_{k,i}) \xrightarrow{d} N(0,1)$$

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, $P\left(\widehat{a}_{1-\alpha} > 0\right)$ goes to 1 as $T \to \infty$. 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

$$\left[\widehat{\lambda}_{k,i} - \sqrt{\frac{z^2_{1-\alpha/2}}{T}\widehat{\sigma}^2_T(\widehat{\lambda}_{k,i})} \, , \, \widehat{\lambda}_{k,i} + \sqrt{\frac{z^2_{1-\alpha/2}}{T}\widehat{\sigma}^2_T(\widehat{\lambda}_{k,i})}\right].$$

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:

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

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:

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

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:

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

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:

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

which can be written as:

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

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

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

equals the absolute value of

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

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.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:

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

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:

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

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:

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

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 $\text{LM}_{\text{SVAR-IV}} \xrightarrow{d} \chi_1^2$.

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 $(\alpha_1, \alpha_2, \ldots, \alpha_M)' \neq 0_{M\times 1}$ 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

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

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 $\pi \equiv (\alpha_1\Theta_{0,11}, \ldots, \alpha_M\Theta_{0,11})'$, 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:

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

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

Two comments:

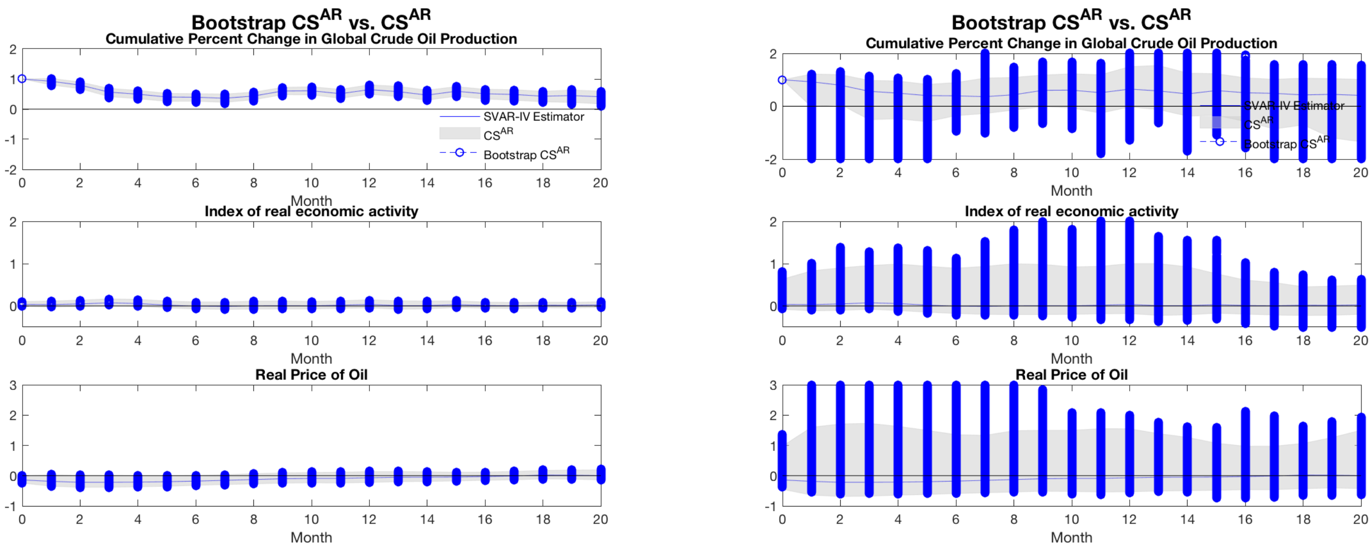
i) Step 2 above, which re-samples the values of the SVAR-IV reduced-form parameters, could be replaced by any other bootstrap procedure, such as the block bootstrap for proxy SVARs recently suggested by Jentsch and Lunsford (2016). One could use their block bootstrap procedure to re-sample the data first, 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 $\mathrm{CS}^{\mathrm{AR}}$ (b) 95% Bootstrap $\mathrm{CS}^{\mathrm{AR}}$

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.[1] 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:

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

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

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

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

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.[2] The parameters of the model for z_t are chosen so that

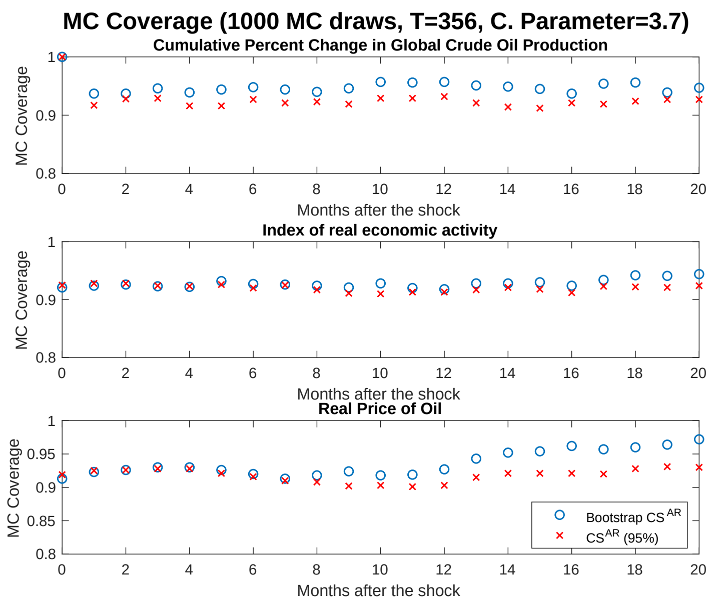
$$\mathbb{E}[z_t] = \widehat{\mu}_z = -0.0182, \quad \mathrm{Var}(z_t) = \widehat{\mathrm{Var}}(z_t) = 0.7436, \quad \mathrm{Cov}(z_t, \epsilon_{1,t}) = \tilde{\alpha}.$$

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.

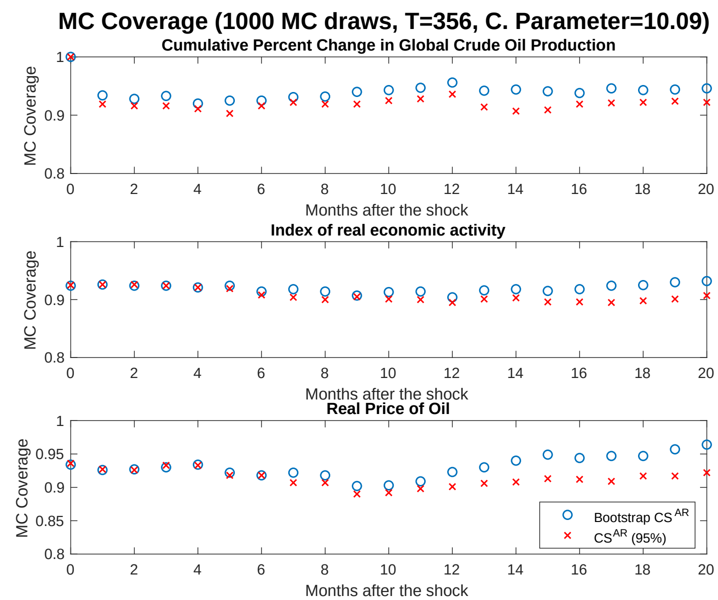
Figure 2 presents the results of the MC coverage for a sample size of $T = 356$ and two different values of the concentration parameter. The comparison is between the $\mathrm{CS}^{\mathrm{AR}}$ and its bootstrap version, which complements the results reported in the main body of the paper. Figure 3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

[2]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 $\mathrm{CS}^{\mathrm{AR}}$, $T = 356$

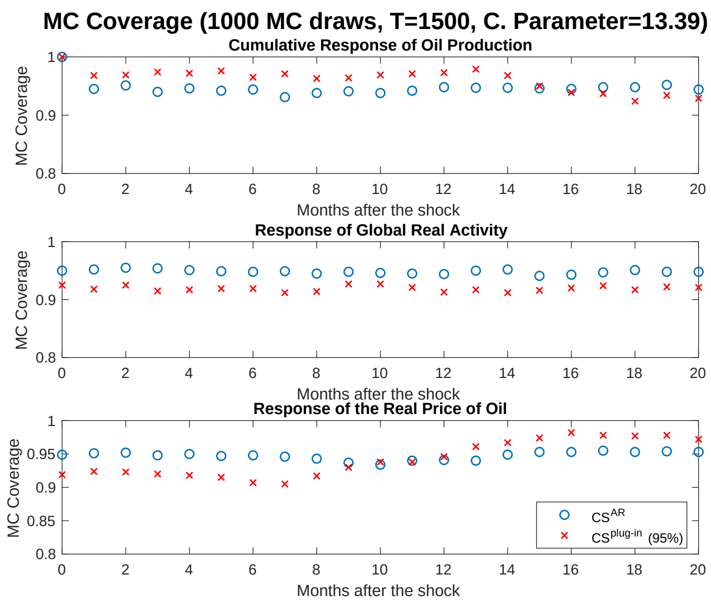


(a) Concentration Parameter = 3.7 (`auxparam`= 2.3452)

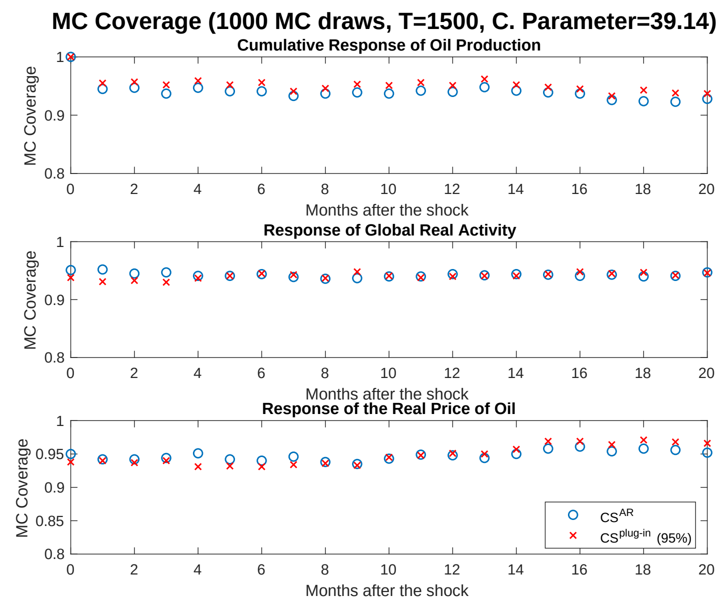


(b) Concentration Parameter = 10.09 (`auxparam`= 4.4441)

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



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



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

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.*

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.

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. Lunsford (2016): "Proxy SVARs: asymptotic theory, bootstrap inference, and the effects of income tax changes in the United States," *Working paper, Federal Reserve Bank of Cleveland.*

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.

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," *Econo-*

metrica, 79, 395–435.

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

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

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.