Introduction to GMM with Weak Instruments

Eric Zivot

December 7, 2011

Monte Carlo Experiment to Illustrate Problem with 2SLS with Weak Instruments

Reference: Zivot, Startz and Nelson (1998). Valid Confidence Intervals and Inference in the Presence of Weak Instruments, *International Economic Review*.

Model: Instrumental variables regression with single endogenous variable and multiple instruments

$$\begin{aligned} y_i &= z_i\delta + \varepsilon_i,\ i = 1, \ldots, n \\ z_i &= \underset{(1\times k)}{\mathbf{x}_i'}\underset{(k\times 1)}{\boldsymbol{\pi}} + v_i \\ &\begin{pmatrix} \varepsilon_i \\ v_i \end{pmatrix} \sim iid\ N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\varepsilon & \sigma_{\varepsilon v} \\ \sigma_{\varepsilon v} & \sigma_v \end{pmatrix}\right) \\ &\mathbf{x}_i \sim iid\ N(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{xx}) \text{ independent of } v_i \end{aligned}$$

Structural parameter δ is estimated by 2SLS (efficient GMM under conditional homoskedasticity) using instruments $\mathbf{x}$

Remarks

1. $\rho = corr(\varepsilon_i, v_i) = \frac{\sigma_{\varepsilon v}}{\sigma_\varepsilon \sigma_v}$ captures the degree of endogeneity of z_i; $\rho \approx 0 \Rightarrow$ low endogeneity; $\rho \approx 1 \Rightarrow$ high endogeneity

2. The reduced form (1st stage in 2SLS) $z_i = \mathbf{x}_i'\boldsymbol{\pi} + v_i$ links the endogenous variable to the instruments

3. $\boldsymbol{\pi}$ captures the strength of the instruments; $\boldsymbol{\pi} \approx \mathbf{0} \Rightarrow$ weak instruments. That is,

$$\begin{aligned} E[\mathbf{x}_i z_i] &= E[\mathbf{x}_i(\mathbf{x}_i'\boldsymbol{\pi} + v_i)] = E[\mathbf{x}_i\mathbf{x}_i']\boldsymbol{\pi} + E[\mathbf{x}_i v_i] \\ &= \boldsymbol{\Sigma}_{xx}\boldsymbol{\pi} \approx \mathbf{0} \end{aligned}$$

so the rank condition essentially fails.

Review of 2SLS estimation of δ

Errors are conditionally homoskedastic so efficient GMM is 2SLS

$$\begin{aligned}
\hat{\delta}_{2SLS} &= (\mathbf{z}'\mathbf{P}_X\mathbf{z})^{-1}\mathbf{z}'\mathbf{P}_X\mathbf{y} \\
&= \delta + (\mathbf{z}'\mathbf{P}_X\mathbf{z})^{-1}\mathbf{z}'\mathbf{P}_X\boldsymbol{\varepsilon} \\
\mathbf{P}_X &= \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'
\end{aligned}$$

Now,

$$\begin{aligned}
\mathbf{z}'\mathbf{P}_X\mathbf{z} &= \mathbf{z}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{z} \\
\mathbf{z}'\mathbf{P}_X\boldsymbol{\varepsilon} &= \mathbf{z}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon} \\
\mathbf{z} &= \mathbf{X}\boldsymbol{\pi} + \mathbf{v}
\end{aligned}$$

Under usual assumptions

$$\frac{\mathbf{X}'\mathbf{X}}{n} \xrightarrow{p} \mathbf{\Sigma}_{xx},$$
$$\frac{\mathbf{X}'\mathbf{v}}{n} \xrightarrow{p} \mathbf{0},$$
$$\frac{\mathbf{X}'\boldsymbol{\varepsilon}}{n} \xrightarrow{p} \mathbf{0},$$
$$\frac{\mathbf{X}'\mathbf{z}}{n} = \frac{1}{n}\mathbf{X}'(\mathbf{X}\boldsymbol{\pi} + \mathbf{v}) = \frac{1}{n}\mathbf{X}'\mathbf{X}\boldsymbol{\pi} + \frac{1}{n}\mathbf{X}'\mathbf{v} \xrightarrow{p} \mathbf{\Sigma}_{xx}\boldsymbol{\pi},$$
$$\frac{\mathbf{X}'\boldsymbol{\varepsilon}}{\sqrt{n}} = \frac{1}{\sqrt{n}}\mathbf{X}'\boldsymbol{\varepsilon} \xrightarrow{d} N(\mathbf{0}, \sigma_{\varepsilon}\mathbf{\Sigma}_{xx})$$

Then, for fixed $\boldsymbol{\pi} \neq \mathbf{0}$

$$
\begin{aligned}
&\hat{\delta}_{2SLS} - \delta \xrightarrow{p} 0, \\
&\sqrt{n}\left(\hat{\delta}_{2SLS} - \delta\right) \xrightarrow{d} N(0, \sigma_\varepsilon^2(\boldsymbol{\pi}'\boldsymbol{\Sigma}_{xx}\boldsymbol{\pi})^{-1}), \\
&\hat{\delta}_{2SLS} \overset{A}{\sim} N(\delta, n^{-1}\sigma_\varepsilon^2(\boldsymbol{\pi}'\boldsymbol{\Sigma}_{xx}\boldsymbol{\pi})^{-1})
\end{aligned}
$$

Here,

$$
\text{avar}(\hat{\delta}_{2SLS}) = \sigma_\varepsilon^2(\boldsymbol{\pi}'\boldsymbol{\Sigma}_{xx}\boldsymbol{\pi})^{-1}
$$

Clearly, if $\boldsymbol{\pi} = 0$ then $\boldsymbol{\pi}'\boldsymbol{\Sigma}_{xx}\boldsymbol{\pi} = \mathbf{0}$ and avar$(\hat{\delta}_{2SLS})$ is not defined.

Intuition: if $\boldsymbol{\pi} \approx \mathbf{0}$ then avar$(\hat{\delta}_{2SLS})$ should be very big indicating much uncertainty about δ.

Unfortunately, this intuition is not always right!!!!

Monte Carlo Design Parameters

$$
\begin{aligned}
\delta &= 1,\ \sigma_\varepsilon = \sigma_v = 1 \\
\rho &= 0.99 \text{ (very high endogeneity)} \\
k &= 1, 4;\ n = 100 \\
\boldsymbol{\mu}_x &= 0,\ \boldsymbol{\Sigma}_{xx} = \mathbf{I}_k \\
\text{avar}(\hat{\delta}_{2SLS}) &= (\boldsymbol{\pi}'\boldsymbol{\pi})^{-1}
\end{aligned}
$$

Instrument cases

1. Irrelevant instruments:

$$
\begin{aligned}
k &= 1 : \pi = 0 \\
k &= 4 : \boldsymbol{\pi} = (0, 0, 0, 0)'
\end{aligned}
$$

2. Weak instruments:

$$
\begin{aligned}
k &= 1 : \pi = 0.1 \\
k &= 4 : \boldsymbol{\pi} = (0.1, 0, 0, 0)'
\end{aligned}
$$

3. Strong instruments:

$$
\begin{aligned}
k &= 1 : \pi = 1 \\
k &= 4 : \boldsymbol{\pi} = (1, 0, 0, 0)'
\end{aligned}
$$

Monte Carlo Experiment

Goal: evaluate inference on δ using 2SLS

1. Simulate data from model 10,000 times

2. Compute $\hat{\delta}_{2SLS}$ and t-stat for testing $\delta = 1$

$$t = \frac{\hat{\delta}_{2SLS} - 1}{\widehat{\mathsf{SE}}(\hat{\delta}_{2SLS})}$$

3. Compute asymptotic 95% confidence interval for δ

$$\hat{\delta}_{2SLS} \pm 1.96 \cdot \widehat{\mathsf{SE}}(\hat{\delta}_{2SLS})$$

4. Determine if $|t| > 1.96$ (equivalently, $\delta \in$ 95% CI)

Empirical size of 5% t-test

$$\# \text{ times } |t| > 1.96$$

Monte Carlo Results

$$H_0 : \delta = 1 \text{ vs. } H_1 : \delta \neq 1$$

Empirical Size of 5% t-test		
Instrument Quality	k=1	k=4
Strong	.055	.084
Weak	.193	.855
Irrelevant	.632	.987

Conclusion: Traditional asymptotic theory is not appropriate for 2SLS in the presence of weak instruments, particularly if endogeneity is high and there are many weak instruments.

Staiger-Stock Weak Instrument Asymptotics

References: Staiger and Stock (1997), "Instrumental Variables Regression with Weak Instruments", *Econometrica*.

Zivot and Wang (1998), "Inference on Structural Parameters in Instrumental Variables Regression with Weak Instruments", *Econometrica*.

$$\begin{aligned} y_i &= z_i\delta + \varepsilon_i,\ i = 1, \ldots, n \\ z_i &= \underset{(1\times k)}{\mathbf{x}_i'} \underset{(k\times 1)}{\boldsymbol{\pi}} + v_i \end{aligned}$$

$$\begin{pmatrix} \varepsilon_i \\ v_i \end{pmatrix} \sim iid\ N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\varepsilon^2 & \sigma_{\varepsilon v} \\ \sigma_{\varepsilon v} & \sigma_v^2 \end{pmatrix}\right)$$

$$\rho = \frac{\sigma_{\varepsilon v}}{\sigma_\varepsilon \sigma_v}$$

Assumptions

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i' \xrightarrow{p} E[\mathbf{x}_i\mathbf{x}_i'] &= \mathbf{\Sigma}_{xx} > 0, \\
\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\varepsilon_i \xrightarrow{p} E[\mathbf{x}_i\varepsilon_i] &= \mathbf{0}, \\
\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i v_i \xrightarrow{p} E[\mathbf{x}_i v_i] &= \mathbf{0}, \\
\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{x}_i\varepsilon_i \xrightarrow{d} N(0, \sigma_\varepsilon^2\mathbf{\Sigma}_{xx}) &= \boldsymbol{\psi}_{z\varepsilon}, \\
\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{x}_i v_i \xrightarrow{d} N(0, \sigma_v^2\mathbf{\Sigma}_{xx}) &= \boldsymbol{\psi}_{zv},
\end{aligned}$$

Note: $\boldsymbol{\psi}_{z\varepsilon}$ and $\boldsymbol{\psi}_{zv}$ are correlated random vectors

Weak Instrument Assumption

$$\begin{aligned}\underset{k\times 1}{\boldsymbol{\pi}} &= \boldsymbol{\pi}_n = \frac{1}{\sqrt{n}}\cdot\mathbf{g}, \\ \underset{k\times 1}{\mathbf{g}} &= \text{fixed vector of constants}\end{aligned}$$

Intuition:

$\boldsymbol{\pi}$ measures the quality of the instruments.

If $\boldsymbol{\pi} \neq \mathbf{0}$ is fixed, then as $n \to \infty$ it follows that $F_{\pi=0} \xrightarrow{p} \infty$ and we conclude that instruments are strong (eventually). That is, we reject $H_0 : \boldsymbol{\pi} = \mathbf{0}$ with large enough n regardless how close $\boldsymbol{\pi}$ is to zero.

By setting $\boldsymbol{\pi} = \mathbf{g}/\sqrt{n}$, as $n \to \infty$ then $\boldsymbol{\pi} \to \mathbf{0}$ at rate $\sqrt{n}$ so that $F_{\pi=0} \xrightarrow{d}$ bounded random variable. Hence, we do not reject $H_0 : \boldsymbol{\pi} = \mathbf{0}$ for large n with probability 1.

Implication: δ is only weakly identified as $n \to \infty$

Remark: Setting $\boldsymbol{\pi} = \mathbf{g}/\sqrt{n}$ is exactly the trick that is used when studying local power of tests (recall, this is the Pitman drift model). Here, we concentrate on the special case in which $\boldsymbol{\pi}$ is local to zero as $n \to \infty$.

Weak Instrument Asymptotics

$$\begin{aligned}
\boldsymbol{\pi} &= \boldsymbol{\pi}_n = \frac{1}{\sqrt{n}} \cdot \mathbf{g} \\
z_i &= \mathbf{x}_i' \boldsymbol{\pi}_n + v_i
\end{aligned}$$

The usual manipulations give

$$\begin{aligned}
\hat{\delta}_{2SLS} - \delta &= (n^{-1} \mathbf{z}' \mathbf{P}_X \mathbf{z})^{-1} n^{-1} \mathbf{z}' \mathbf{P}_X \boldsymbol{\varepsilon} \\
&= \left(\frac{\mathbf{z}'\mathbf{X}}{n} \left(\frac{\mathbf{X}'\mathbf{X}}{n} \right)^{-1} \frac{\mathbf{X}'\mathbf{z}}{n} \right)^{-1} \frac{\mathbf{z}'\mathbf{X}}{n} \left(\frac{\mathbf{X}'\mathbf{X}}{n} \right)^{-1} \frac{\mathbf{X}'\boldsymbol{\varepsilon}}{n}
\end{aligned}$$

Now, by assumption

$$\frac{\mathbf{X}'\mathbf{X}}{n} \xrightarrow{p} E[\mathbf{x}_i \mathbf{x}_i'] = \boldsymbol{\Sigma}_{xx}, \ \frac{\mathbf{X}'\boldsymbol{\varepsilon}}{n} \xrightarrow{p} 0$$

But

$$
\begin{aligned}
\frac{\mathbf{X}'\mathbf{z}}{n} &= \frac{1}{n}\mathbf{X}'(\mathbf{X}\boldsymbol{\pi}_n + \mathbf{v}) \\
&= \frac{1}{n}\mathbf{X}'\left(\frac{\mathbf{X}\mathbf{g}}{\sqrt{n}} + \mathbf{v}\right) \\
&= \frac{1}{n^{3/2}}\mathbf{X}'\mathbf{X}\mathbf{g} + \frac{1}{n}\mathbf{X}'\mathbf{v} \xrightarrow{p} \mathbf{0}
\end{aligned}
$$

Therefore,

$$
\frac{\mathbf{z}'\mathbf{X}}{n}\left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)^{-1}\frac{\mathbf{X}'\mathbf{z}}{n} \xrightarrow{p} 0
$$

and so

$$
\hat{\delta}_{2SLS} - \delta \xrightarrow{p} (0)^{-1}0
$$

which is not well defined.

Trick: To get a meaningful asymptotic result, we need to use a different normalization:

$$\hat{\delta}_{2SLS} - \delta = \left(\frac{\mathbf{z}'\mathbf{X}}{\sqrt{n}} \left(\frac{\mathbf{X}'\mathbf{X}}{n} \right)^{-1} \frac{\mathbf{X}'\mathbf{z}}{\sqrt{n}} \right)^{-1} \frac{\mathbf{z}'\mathbf{X}}{\sqrt{n}} \left(\frac{\mathbf{X}'\mathbf{X}}{n} \right)^{-1} \frac{\mathbf{X}'\boldsymbol{\varepsilon}}{\sqrt{n}}$$

Now,

$$\begin{aligned} \frac{\mathbf{X}'\mathbf{z}}{\sqrt{n}} &= \frac{\mathbf{1}}{\sqrt{n}}\mathbf{X}'(\mathbf{X}\boldsymbol{\pi}_n + \mathbf{v}) = \frac{\mathbf{1}}{\sqrt{n}}\mathbf{X}' \left(\frac{\mathbf{X}\mathbf{g}}{\sqrt{n}} + \mathbf{v} \right) \\ &= \frac{\mathbf{1}}{n}\mathbf{X}'\mathbf{X}\mathbf{g} + \frac{\mathbf{1}}{\sqrt{n}}\mathbf{X}'\mathbf{v} \xrightarrow{d} \boldsymbol{\Sigma}_{xx}\mathbf{g} + N(\mathbf{0}, \sigma_v^2 \boldsymbol{\Sigma}_{xx}) \\ &= \boldsymbol{\Sigma}_{xx}\mathbf{g} + \boldsymbol{\psi}_{xv} \equiv N(\boldsymbol{\Sigma}_{xx}\mathbf{g}, \sigma_v^2 \boldsymbol{\Sigma}_{xx}) \end{aligned}$$

where we used

$$\frac{\mathbf{X}'\boldsymbol{\varepsilon}}{\sqrt{n}} \xrightarrow{d} \boldsymbol{\psi}_{x\varepsilon}$$

Therefore,

$$\begin{aligned}\hat{\delta}_{2SLS} - \delta \xrightarrow{d} & \left([\mathbf{\Sigma}_{xx}\mathbf{g} + \boldsymbol{\psi}_{xv}]' \mathbf{\Sigma}_{xx}^{-1} [\mathbf{\Sigma}_{xx}\mathbf{g} + \boldsymbol{\psi}_{xv}] \right)^{-1} \\ & \times \ [\mathbf{\Sigma}_{xx}\mathbf{g} + \boldsymbol{\psi}_{xv}]' \mathbf{\Sigma}_{xx}^{-1} \boldsymbol{\psi}_{x\varepsilon} \\ & = \text{ random variable!}\end{aligned}$$

Results:

1. Under the Staiger-Stock weak instrument asymptotics, $\hat{\delta}_{2SLS}$ is not consistent!

2. Rearrangement shows that

$$\hat{\delta}_{2SLS} \xrightarrow{d} \delta + \text{ ratio of quadratic form in correlated normal vectors}$$

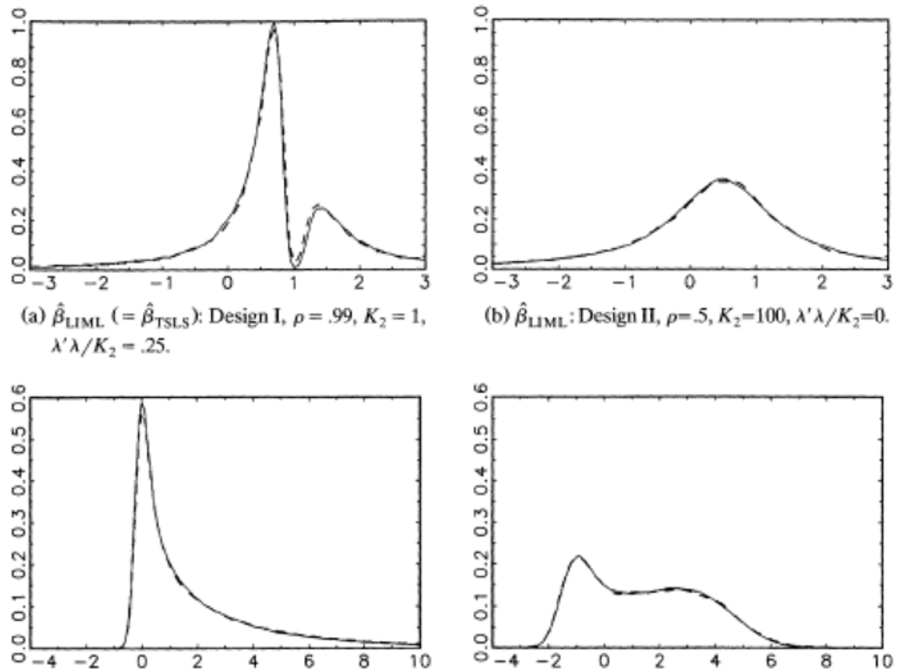
3. Staiger and Stock show that weak instrument asymptotic result is a much better approximation to the finite sample distribution of $\hat{\delta}_{2SLS}$ in the presence of weak instruments than the usual asymptotic normal distribution based on strong instruments.

4. The weak instrument asymptotic distribution of $\hat{\delta}_{2SLS}$ depends on the nuisance parameters ρ and $\mathbf{g}$ that cannot be consistently estimated from the data. Hence, the weak instrument asymptotic distribution is not practically useful.

$$
\begin{aligned}
\rho &= corr(\varepsilon_i, v_i),\ \hat{\rho} = \frac{\hat{\sigma}_{\varepsilon v}}{\hat{\sigma}_{\varepsilon}\hat{\sigma}_{v}} = \frac{\hat{\boldsymbol{\varepsilon}}'\hat{\mathbf{v}}}{\sqrt{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}\hat{\mathbf{v}}'\hat{\mathbf{v}}}} \overset{p}{\nrightarrow} \rho \\
\hat{\boldsymbol{\varepsilon}} &= \mathbf{y} - \mathbf{z}\hat{\delta}_{2SLS},\ \hat{\mathbf{v}} = \mathbf{z} - \mathbf{X}\hat{\boldsymbol{\pi}}
\end{aligned}
$$

Here

$$\hat{\boldsymbol{\varepsilon}} \neq \boldsymbol{\varepsilon} + o_p(1) \text{ b/c } \hat{\delta}_{2SLS} - \delta \xrightarrow{d} \text{ random variable}$$



(a) $\hat{\beta}_{\text{LIML}}$ ($= \hat{\beta}_{\text{TSLS}}$): Design I, $\rho = .99$, $K_2 = 1$, $\lambda'\lambda/K_2 = .25$.

(b) $\hat{\beta}_{\text{LIML}}$: Design II, $\rho = .5$, $K_2 = 100$, $\lambda'\lambda/K_2 = 0$.

(c) t_{LIML} ($= t_{\text{TSLS}}$): Design I, $\rho = .99$, $K_2 = 1$, $\lambda'\lambda/K_2 = .25$.

(d) t_{LIML}: Design II, $\rho = .5$, $K_2 = 100$, $\lambda'\lambda/K_2 = 0$.

FIGURE 1—Asymptotic (solid line) and finite-sample (dashed line) pdf's of the LIML estimator and t statistic. True $\beta_0 = 0$; plim($\hat{\beta}_{\text{OLS}}$) $= \rho$.

Staiger-Stock Weak Instrument Asymptotic Approximation

Remarks

1. Weak instrument asymptotic distribution of $\hat{\delta}_{2SLS}$ can be asymmetric and bi-modal (Nelson and Startz 1991, Ecta result)

2. If $\boldsymbol{\pi} = \mathbf{0}$ so that δ is unidentified it can be shown that

$$\hat{\delta}_{2SLS} - \delta \xrightarrow{d} \rho\frac{\sigma_\varepsilon}{\sigma_v} + (1-\rho^2)^{1/2}\frac{\sigma_\varepsilon}{\sigma_v} \cdot \frac{t_k}{\sqrt{k}}$$

$$\begin{aligned} t_k &= \text{ Student's t with } k \text{ d.f.} \\ \rho\frac{\sigma_\varepsilon}{\sigma_v} &= p\lim_{n\to\infty} \hat{\delta}_{OLS} \\ k &= \text{number of instruments} \end{aligned}$$

3. If $\mathbf{g}$ is sufficiently large so that $\boldsymbol{\pi}_n = \mathbf{g}/\sqrt{n}$ is not too small then

$$\hat{\delta}_{2SLS} \sim N(\delta, n^{-1}\sigma_\varepsilon(\boldsymbol{\pi}_n'\boldsymbol{\Sigma}_{xx}\boldsymbol{\pi}_n)^{-1})$$

which is the usual strong instrument asymptotic result.

Inference in the Presence of Weak Instruments

$$H_0 : \delta = \delta_0 \text{ vs. } H_1 : \delta \neq \delta_0$$

Result: The usual 2SLS t-statistic is not normally distributed

$$t_{\delta=\delta_0} = \frac{\hat{\delta}_{2SLS} - \delta_0}{\widehat{\mathsf{SE}}(\hat{\delta}_{2SLS})} = \frac{\hat{\delta}_{2SLS} - \delta_0}{\left(\hat{\sigma}_{\varepsilon}^2(\mathbf{z}'\mathbf{P}_X\mathbf{z})^{-1}\right)^{1/2}} \overset{A}{\nsim} N(0,1)$$

$$\hat{\delta}_{2SLS} - \delta_0 \xrightarrow{d} \text{non normal random variable}$$

$$\hat{\sigma}_{\varepsilon}^2 \xrightarrow{d} \text{positive random variable}$$

$$\mathbf{z}'\mathbf{P}_X\mathbf{z} \xrightarrow{d} \text{positive random variable}$$

Conclusion: Don't use 2SLS t-statistics or Wald statistics to perform inference in the presence of weak instruments! In the presence of weak instruments, $t_{\delta=\delta_0}$ is not asymptotically pivotal: its distribution depends on nuisance parameters, some of which cannot be consistently estimated.

Test Statistics Robust to Weak Instruments

$$\begin{aligned} y_i &= z_i\delta + \varepsilon_i, \ i = 1, \ldots, n \\ z_i &= \underset{(1\times k)}{\mathbf{x}_i'} \underset{(k\times 1)}{\boldsymbol{\pi}} + v_i \end{aligned}$$

Hypotheses to be tested

$$\begin{aligned} H_0 &: \delta = \delta_0 \\ H_1 &: \delta \neq \delta_0 \end{aligned}$$

Goal: Find an asymptotically pivotal test statistic (i.e., a test statistic whose asymptotic distribution does not depend on nuisance parameters) regardless of the quality of the instruments.

Anderson-Rubin Statistic (1949, *Annals of Statistics*)

Step 1: subtract $z_i\delta_0$ from both sides of IV regression

$$y_i - z_i\delta_0 = z_i\delta - z_i\delta_0 + \varepsilon_i = z_i(\delta - \delta_0) + \varepsilon_i$$

Step 2: substitute $\mathbf{x}_i'\boldsymbol{\pi} + v_i$ for z_i

$$\begin{aligned} y_i - z_i\delta_0 &= \left(\mathbf{x}_i'\boldsymbol{\pi} + v_i\right)(\delta - \delta_0) + \varepsilon_i \\ &= \mathbf{x}_i'\boldsymbol{\pi}(\delta - \delta_0) + v_i(\delta - \delta_0) + \varepsilon_i \\ &= \mathbf{x}_i'\boldsymbol{\psi} + w_i \end{aligned}$$

where

$$\boldsymbol{\psi} = \boldsymbol{\pi}(\delta - \delta_0) \text{ and } w_i = v_i(\delta - \delta_0) + \varepsilon_i$$

Note: in the AR regression, $\mathbf{x}_i$ is uncorrelated with w_i so it can be estimated by OLS.

Testing $H_0 : \delta = \delta_0$ may be performed by testing

$$H_0 : \underset{k \times 1}{\boldsymbol{\psi}} = \mathbf{0}$$

That is,

$$H_0 : \delta = \delta_0 \Rightarrow \boldsymbol{\psi} = \boldsymbol{\pi}(\delta - \delta_0) = \mathbf{0}$$

The AR statistic is the usual F-statistic for testing $\boldsymbol{\psi} = \mathbf{0}$

$$\begin{aligned} \text{AR}(\delta_0) &= F_{\psi=0} = \frac{(RSS_R - RSS_{UR})/k}{RSS_{UR}/(n-k)} \\ &= \frac{(\mathbf{y} - \mathbf{z}\delta_0)'\mathbf{P}_X(\mathbf{y} - \mathbf{z}\delta_0)/k}{(\mathbf{y} - \mathbf{z}\delta_0)'\mathbf{Q}_X(\mathbf{y} - \mathbf{z}\delta_0)/(n-k)} \end{aligned}$$

where $\mathbf{P}_X = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ and $\mathbf{Q}_X = \mathbf{I}_n - \mathbf{P}_X$.

Distribution of AR Statistic under $H_0 : \delta = \delta_0$

1. In finite samples, with fixed regressors and normally distributed errors $\mathbf{y} - \mathbf{z}\delta_0 = \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma_\varepsilon \mathbf{I}_n)$ and

$$\begin{aligned} \text{AR}(\delta_0) &= F_{\psi=0} = \frac{\left(\frac{\varepsilon}{\sigma_\varepsilon}\right)' \mathbf{P}_X \left(\frac{\varepsilon}{\sigma_\varepsilon}\right)/k}{\left(\frac{\varepsilon}{\sigma_\varepsilon}\right)' \mathbf{Q}_X \left(\frac{\varepsilon}{\sigma_\varepsilon}\right)/(n-k)} \\ &= \frac{\chi^2(k)/k}{\chi^2(n-k)/(n-k)} \sim F_{k,n-k} \end{aligned}$$

regardless of the quality of the instruments!

2. Under the Staiger-Stock weak instrument asymptotics it can be shown that

$$k \cdot \text{AR}(\delta_0) = k \cdot F_{\psi=0} \xrightarrow{d} \chi^2(k)$$

regardless of the quality of the instruments.

Remarks:

1. In general, $k \cdot \text{AR}(\delta_0) \sim \chi^2(k)$ and not $\chi^2(1)$. As a result, the test may not have good power if the number of instruments k is large.

2. $\text{AR}(\delta_0)$ tests the joint hypothesis $\delta = \delta_0$ and $\mathbf{x}_i$ is uncorrelated with ε_i $(i = 1, \ldots, k)$. Hence, a large $\text{AR}(\delta_0)$ statistic can be due to $\delta \neq \delta_0$ or if some element of $\mathbf{x}_i$ is correlated with ε_i.

3. $k \cdot \text{AR}(\delta_0)$ is closely related to the CU-GMM J-statistics under conditional homoskedasticity. Recall, the CU-GMM objective function is

$$
\begin{aligned}
J(\delta, \hat{\mathbf{S}}^{-1}(\delta)) &= n\mathbf{g}_n(\delta)'\hat{\mathbf{S}}^{-1}(\delta)\mathbf{g}_n(\delta) \\
\mathbf{g}_n(\delta) &= \mathbf{S}_{xy} - \mathbf{S}_{xz}\delta = \frac{1}{n}\mathbf{X}'[\mathbf{y} - \mathbf{z}\delta] \\
\hat{\mathbf{S}}(\boldsymbol{\delta}) &= \hat{\sigma}^2(\boldsymbol{\delta})\mathbf{S}_{xx} = \frac{1}{n}(\mathbf{y} - \mathbf{z}\delta)'(\mathbf{y} - \mathbf{z}\delta) \times \mathbf{S}_{xx}
\end{aligned}
$$

And after some algebra we have

$$J(\delta, \hat{\mathbf{S}}^{-1}(\delta)) = \frac{(\mathbf{y}-\mathbf{z}\delta)'\mathbf{P}_{\mathbf{X}}(\mathbf{y}-\mathbf{z}\delta)}{(\mathbf{y}-\mathbf{z}\delta)'(\mathbf{y}-\mathbf{z}\delta)/n}$$

Under $H_0 : \delta = \delta_0$ it is straightforward to show that

$$J(\delta_0, \hat{\mathbf{S}}^{-1}(\delta_0)) \xrightarrow{d} \chi^2(k)$$

Now, compare this to $k \cdot \mathsf{AR}(\delta_0)$

$$k \cdot \mathsf{AR}(\delta_0) = \frac{(\mathbf{y}-\mathbf{z}\delta_0)'\mathbf{P}_X(\mathbf{y}-\mathbf{z}\delta_0)}{(\mathbf{y}-\mathbf{z}\delta_0)'\mathbf{Q}_X(\mathbf{y}-\mathbf{z}\delta_0)/(n-k)}$$

Notice that

$$\begin{aligned}(\mathbf{y}-\mathbf{z}\delta_0)'\mathbf{Q}_X(\mathbf{y}-\mathbf{z}\delta_0) &= (\mathbf{y}-\mathbf{z}\delta_0)'(\mathbf{I}_n-\mathbf{P}_X)(\mathbf{y}-\mathbf{z}\delta_0)\\ &= (\mathbf{y}-\mathbf{z}\delta_0)'(\mathbf{y}-\mathbf{z}\delta_0)-(\mathbf{y}-\mathbf{z}\delta_0)'\mathbf{P}_X(\mathbf{y}-\mathbf{z}\delta_0)\end{aligned}$$

It follows that

$$\frac{(\mathbf{y}-\mathbf{z}\delta_0)'\mathbf{Q}_X(\mathbf{y}-\mathbf{z}\delta_0)}{n-k} = \frac{(\mathbf{y}-\mathbf{z}\delta_0)'(\mathbf{y}-\mathbf{z}\delta_0)}{n} + o_p(1)$$

Hence, we can re-write $k \cdot \text{AR}(\delta_0)$ as

$$\begin{aligned} k \cdot \text{AR}(\delta_0) &= \frac{(\mathbf{y}-\mathbf{z}\delta)'\mathbf{P}_\mathbf{X}(\mathbf{y}-\mathbf{z}\delta)}{(\mathbf{y}-\mathbf{z}\delta)'(\mathbf{y}-\mathbf{z}\delta)/n} + o_p(1) \\ &= J(\delta_0, \hat{\mathbf{S}}^{-1}(\delta_0)) + o_p(1) \end{aligned}$$

Result: $k \cdot \text{AR}(\delta_0)$ and $J(\delta_0, \hat{\mathbf{S}}^{-1}(\delta_0))$ are asymptotically equivalent and both statistics test the joint hypotheses

$$\begin{aligned} \delta &= \delta_0 \\ E[\mathbf{x}_i\varepsilon_i] &= 0 \end{aligned}$$

Remarks continued

4. You can invert the AR statistic to obtain a weak-instrument robust confidence interval for δ :

$$\mathsf{CI}_\alpha = \{\delta_0 : k \cdot \mathsf{AR}(\delta_0) \leq \chi^2_\alpha(k)\}$$

In words, the AR-based confidence interval contains all values of δ_0 such that you do not reject the joint hypotheses

$$\begin{aligned} \delta &= \delta_0 \\ E[\mathbf{x}_i \varepsilon_i] &= \mathbf{0} \end{aligned}$$

at the $\alpha \times 100\%$ level. We find CI_α by finding all δ_0 such that

$$\frac{(\mathbf{y} - \mathbf{z}\delta_0)'\mathbf{P}_X(\mathbf{y} - \mathbf{z}\delta_0)}{(\mathbf{y} - \mathbf{z}\delta_0)'\mathbf{Q}_X(\mathbf{y} - \mathbf{z}\delta_0)/(n-k)} < \chi^2_\alpha(k)$$

Confidence intervals found this way can have four distinct shapes: (1) closed interval $[\delta_{low}, \delta_{high}]$; (2) empty; (3); disconnected intervals $[-\infty, \delta_{low}] \cup [\delta_{high}, \infty]$; (4) unbounded $[-\infty, \infty]$

Kleibergen's CU-GMM Score Statistic (2002, *Econometrica*)

Recall, under conditional homoskedasticity, $\mathbf{S} = \sigma^2 \mathbf{\Sigma}_{xx}$, and the CU efficient GMM estimator becomes

$$\begin{aligned} \hat{\delta}(\hat{\mathbf{S}}_{\mathsf{CU}}^{-1}) &= \arg\min_{\boldsymbol{\delta}} \; n\mathbf{g}_n(\delta)'\hat{\mathbf{S}}^{-1}(\delta)\mathbf{g}_n(\delta) \\ \hat{\mathbf{S}}(\delta) &= \hat{\sigma}^2(\delta)\mathbf{S}_{xx} = \frac{1}{n}\sum_{i=1}^{n}(y_i - z_i\delta)^2 \times \mathbf{S}_{xx} \end{aligned}$$

A little algebra shows that

$$n\mathbf{g}_n(\delta)'\hat{\mathbf{S}}^{-1}(\delta)\mathbf{g}_n(\delta) = n\frac{(\mathbf{y} - \mathbf{z}\delta)'\mathbf{P}_{\mathbf{X}}(\mathbf{y} - \mathbf{z}\delta)}{(\mathbf{y} - \mathbf{z}\delta)'(\mathbf{y} - \mathbf{z}\delta)}$$

Result: $\hat{\delta}(\hat{\mathbf{S}}_{\mathsf{CU}}^{-1}) = \hat{\delta}_{\mathsf{LIML}}$ = *limited information maximum likelihood estimator* and minimizes the AR statistic

Kleibergen's test is the CU-GMM score statistic for testing

$$H_0 : \delta = \delta_0 \text{ vs. } H_1 : \delta \neq \delta_0$$

The CU-GMM score statistic is a quadratic form in the CU-GMM score

$$\begin{aligned}\frac{d}{d\delta} n\mathbf{g}_n(\delta)'\hat{\mathbf{S}}^{-1}(\delta)\mathbf{g}_n(\delta) &= \frac{d}{d\delta} n \frac{(\mathbf{y}-\mathbf{z}\delta)'\mathbf{P}_{\mathbf{X}}(\mathbf{y}-\mathbf{z}\delta)}{(\mathbf{y}-\mathbf{z}\delta)'(\mathbf{y}-\mathbf{z}\delta)} \\ &= \text{really ugly expression!}\end{aligned}$$

Result

$$\text{LM}_{CU-GMM}(\delta_0) \xrightarrow{d} \chi^2(1)$$

regardless of the quality of the instruments. It has a degrees-of-freedom advantage over the AR statistic and has better power for moderate-to-good instruments.

Weak Identification in Non-Linear GMM

Characterizing weak identification in the general nonlinear GMM estimation, when the K moment conditions $\mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta})$ are nonlinear functions of the p model parameters $\boldsymbol{\theta}$, is not straightforward (see Stock and Wright, 2000, Ecta). Recall, the identification conditions

Global identification of $\boldsymbol{\theta}_0$ requires that

$$\begin{aligned} E[\mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta}_0)] &= \mathbf{0} \\ E[\mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta})] &\neq \mathbf{0} \text{ for } \boldsymbol{\theta} \neq \boldsymbol{\theta}_0 \end{aligned}$$

Local Identification requires that the $K \times p$ matrix

$$\text{rank}(\mathbf{G}) = \text{rank}\left(E\left[\frac{\partial \mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} \right] \right) = p$$

has full column rank p.

Stock and Wright characterize weak identification when

$$E[\mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta})] \approx \mathbf{0} \text{ for } \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$$

and develop an asymptotic theory very similar to weak instrument approach of Staiger and Stock for the linear case.

For testing hypotheses of the form

$$H_0 : \underset{p \times 1}{\boldsymbol{\theta}} = \boldsymbol{\theta}_0 \text{ vs. } H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$$

Stock and Wright propose using a test statistic based on the CU-GMM objective function (generalization of the AR statistic to GMM)

$$S(\boldsymbol{\theta}_0) = J(\boldsymbol{\theta}_0, \hat{\mathbf{S}}^{-1}(\boldsymbol{\theta}_0)) = n\mathbf{g}_n(\boldsymbol{\theta}_0)'\hat{\mathbf{S}}^{-1}(\boldsymbol{\theta}_0)\mathbf{g}_n(\boldsymbol{\theta}_0)$$

Under $H_0 : \theta = \theta_0$ they show that

$$S(\boldsymbol{\theta}_0) \xrightarrow{d} \chi^2(K)$$

Kleibergen (2003) generalized his CU-GMM score statistic for linear GMM models to non-linear GMM. His test is the CU-GMM score statistic for testing

$$H_0 : \underset{p\times 1}{\boldsymbol{\theta}} = \boldsymbol{\theta}_0 \text{ vs. } H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$$

The CU-GMM score statistic is a quadratic form in the CU-GMM score

$$\frac{d}{d\delta} n\mathbf{g}_n(\boldsymbol{\theta_0})'\hat{\mathbf{S}}^{-1}(\boldsymbol{\theta}_0)\mathbf{g}_n(\boldsymbol{\theta_0}) = \text{really, really, ugly expression!}$$

such that

$$\mathrm{LM}_{CU-GMM}(\boldsymbol{\theta}_0) \xrightarrow{d} \chi^2(p)$$

regardless of the strength of indentification. It has a degrees-of-freedom advantage over the S statistic of Stock and Wright and has better power for moderate-to-good instruments.