NBER Summer Institute 2018

Methods Lectures

Weak Instruments and What To Do About Them

Isaiah Andrews, Harvard University

James H. Stock, Harvard University

July 22, 2018

(updated July 25, 2018)

3-4:20pm	1. Weak instruments in the wild 2. Detecting weak instruments	Stock Stock
4:20-4:40pm	*Break*	
4:40-6pm	3. Inference with weak instruments 4. Open issues and recent research	Andrews Andrews

Overview and Summary

Topic: IV regression with a single included endogenous regressor, control variables, and non-homoskedastic errors.

- This covers heteroskedasticity, HAC, cluster, etc.
- We assume that consistent robust SEs exist for the reduced form & first stage regressions.
- Early literature (through ~2006): homoskedastic case
- **This mini-course focuses on weak instruments in the non-homoskedastic case** (i.e., the relevant case).

Outline

1) So what?
2) Detecting weak instruments
3) Estimation (brief)
4) Weak instrument-robust inference about parameter of interest (β)
5) Extensions

So what? (1) Theory

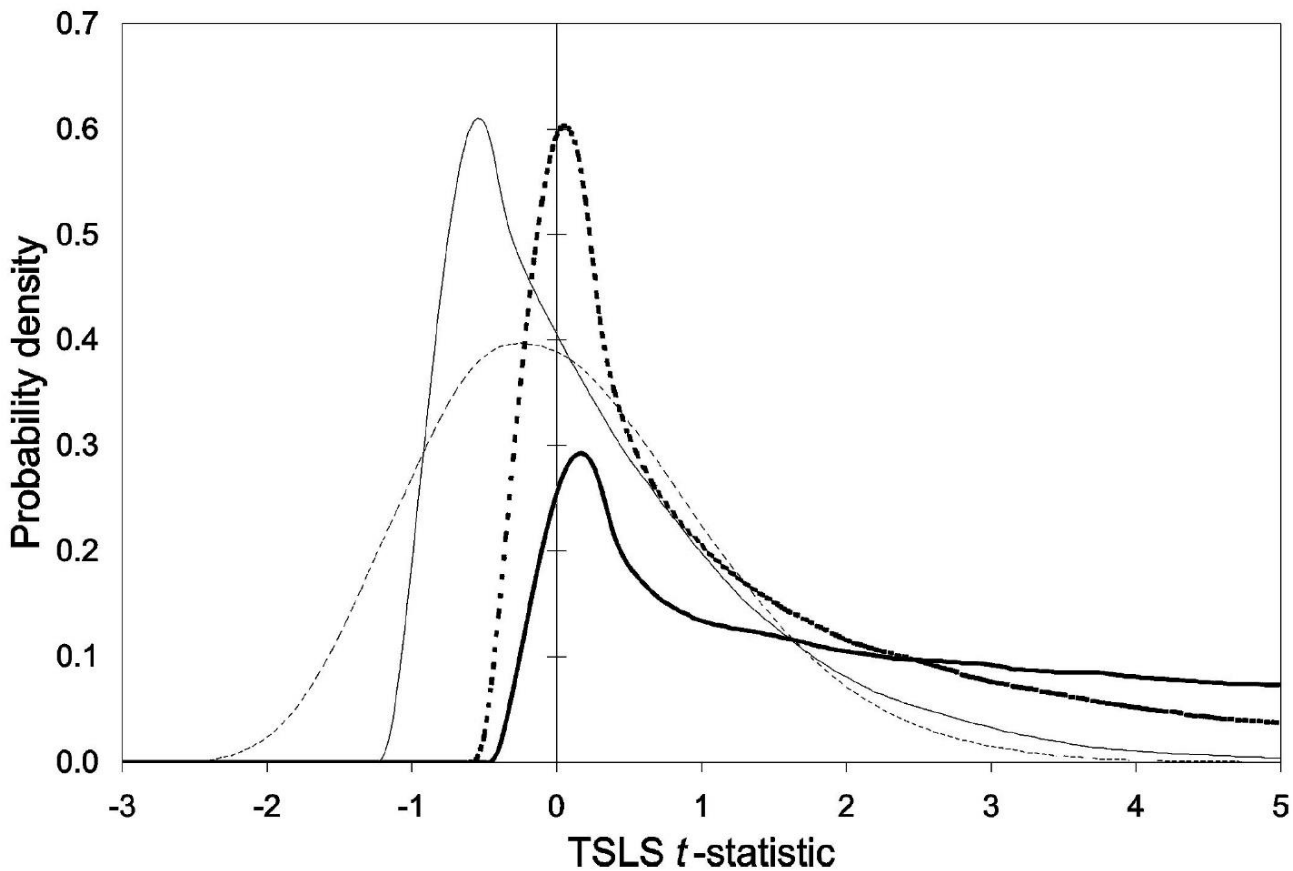
An instrumental variable is weak if its correlation with the included endogenous regressor is small.

1. "small" depends on the inference problem at hand, and on the sample size

With weak instruments, TSLS is biased towards OLS, and TSLS tests have the wrong size.

Distribution of the TSLS *t*-statistic (Nelson-Startz (1990a,b))

- Dark line = irrelevant instruments
- dashed light line = strong instruments
- intermediate cases = weak instruments

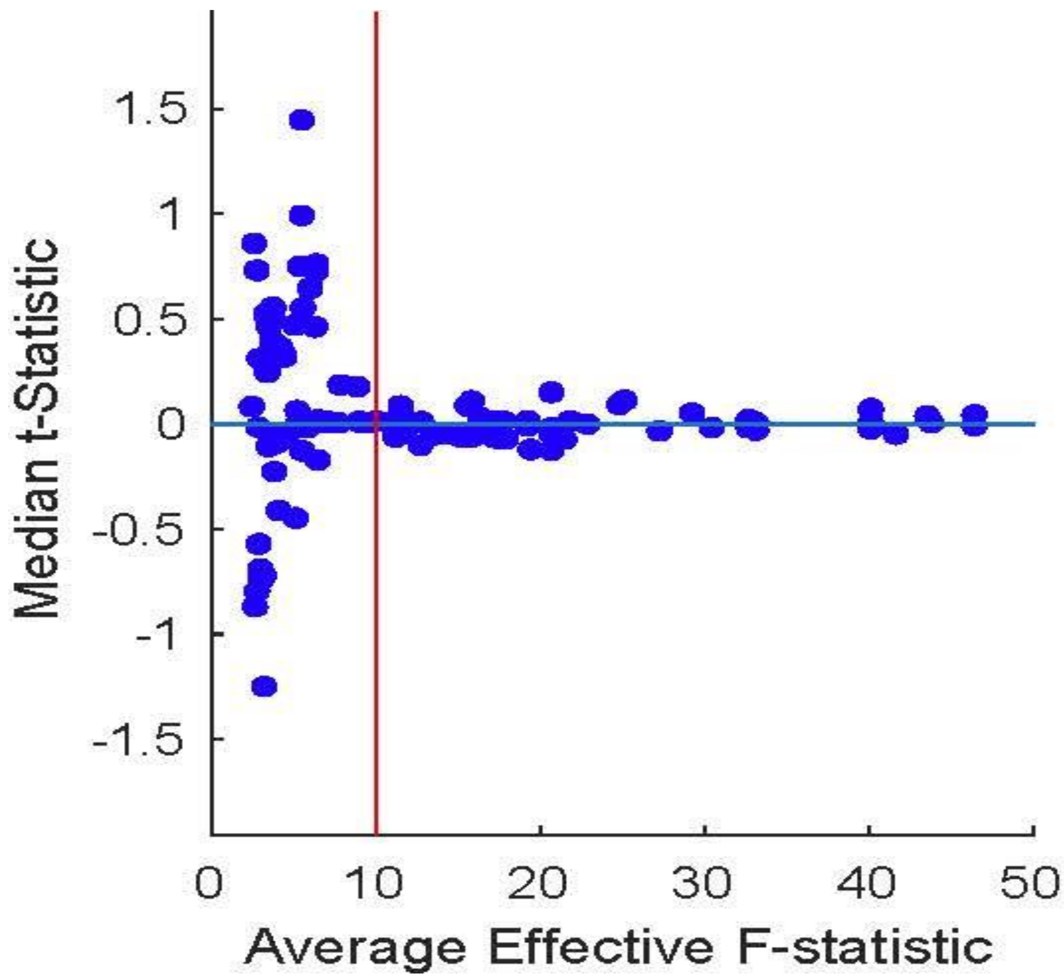


So what? (2) Simulation

DGP: 8 AER papers 2014-2018

(Sample: 17 that use IV; 16 with a single *X*; 8 in simulation sample)

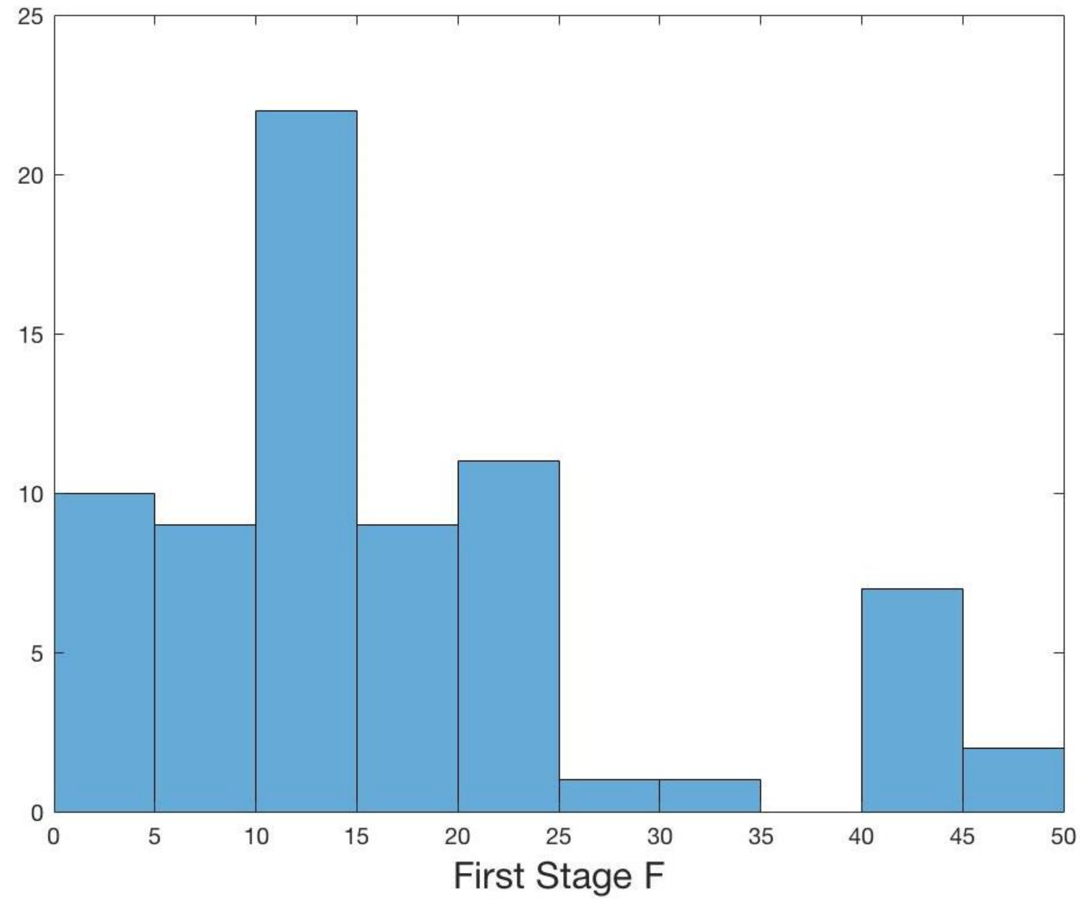
Median of TSLS *t*-statistic under the null



So what? (3) Practice (the "in the wild" bit)

Histogram of first-stage *F*s in AER papers (108 specifications), 2014-2018

- The first-stage *F* tests the hypothesis that the first-stage coefficients are zero.
- Of the 17 papers, all but 1 report first-stage *F*s for at least one specification; the histogram is of the 108 specifications that report a first-stage *F* (72 of which are <50 and are in the plot).



- ***Great that authors/editors/referees are aware of the potential importance of weak instruments, as evidence by nearly all papers reporting first stages Fs.***

- The spike at $F = 10$ is "interesting"

Detecting Weak Instruments

It is convenient to have a way to decide if instruments are strong (TSLS "works") or weak (use weak-instrument robust methods).

The standard method is "the" first-stage F. Candidates:

F^N – nonrobust

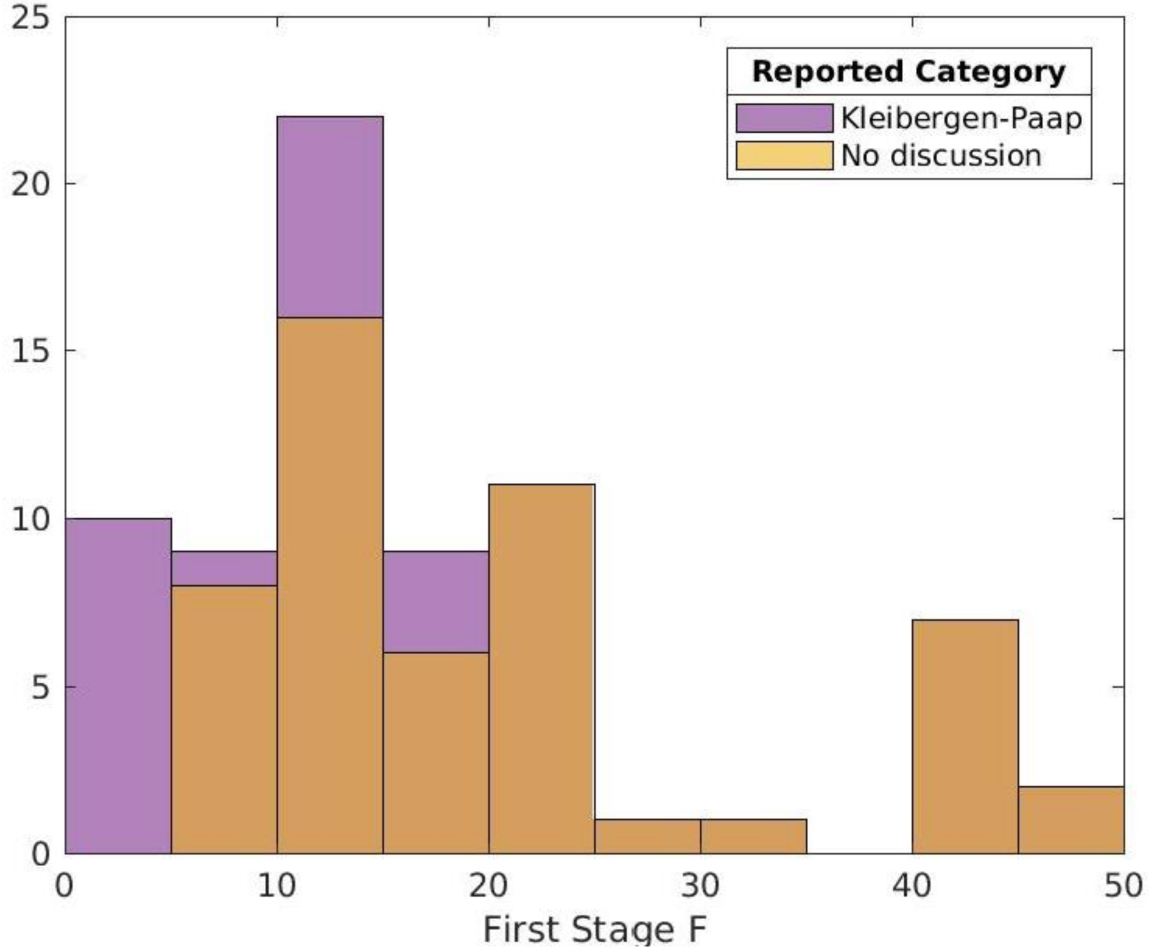
F^R – robust (HR, HAC, cluster), also called Kleibergen-Paap (2006)

F^E – Effective first-stage F statistic of Montiel Olea and Plueger (2013)

Actually there are other candidates too, not used and not to be discussed here including Hahn-Hausman (2002), Shea's (1997) partial R^2

Detecting weak instruments in practice

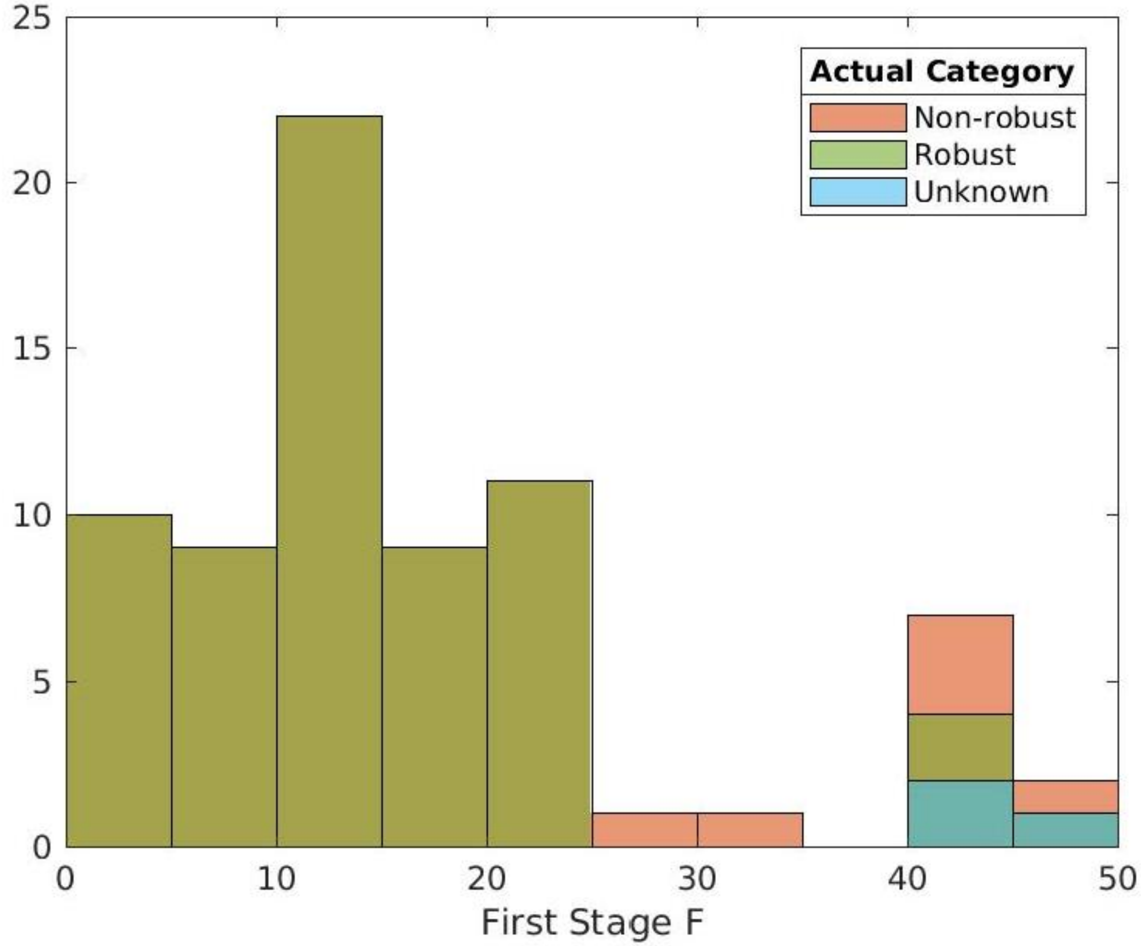
Reported first-stage *F*'s: what authors say they use



Candidates: F^N – nonrobust
F^R – robust (HR, HAC, cluster), also called Kleibergen-Paap (2006)
F^E – Effective first-stage F statistic of Montiel Olea and Plueger (2013)

Detecting weak instruments in practice, ctd

Actual first-stage F's: what authors actually use



Candidates: F^N – nonrobust

F^R – robust (HR, HAC, cluster), also called Kleibergen-Paap (2006)

F^E – Effective first-stage F statistic of Montiel Olea and Plueger (2013)

Our recommendations (1 included endogenous regressor)

- Do:
  - Use the Montiel Olea-Pflueger (2013) effective first-stage F statistic

    $F^{Eff} = F^N \times$ correction factor for non-homoskedasticity
  - Report F^{Eff}
  - Compare F^{Eff} to MOP critical values (`weakivtest.ado`), or to 10.
  - If $F^{Eff} \geq$ MOP critical value, or $\geq$ 10 for rule-of-thumb method, use TSLS inference; else use weak-instrument robust inference.

- Don't
  - use/report p-values of test of $\pi = 0$ (null of irrelevant instruments)
  - use/report nonrobust first stage F (F^N)
  - use/report usual robust first-stage F (except OK for $k = 1$ where $F^R = F^{Eff}$)
  - use/report Kleibergen-Paap (2006) statistic (same thing).
  - compare HR/HAC/Kleibergen-Paap to Stock-Yogo critical values
  - reject a paper because $F^{Eff} < 10$!

    Instead, tell the authors to use weak-IV robust inference.

Notation and Review of IV Regression

IV regression model with a single endogenous regressor and *k* instruments

$$Y_i = X_i\beta + W_i'\gamma_1 + \varepsilon_i \qquad \text{(Structural equation)} \qquad (1)$$

$$X_i = Z_i'\pi + W_i'\gamma_2 + V_i \qquad \text{(First stage)} \qquad (2)$$

where *W* includes the constant. Substitute (2) into (1):

$$Y_i = Z_i'\delta + W_i'\gamma_3 + U_i \qquad \text{(Reduced form)} \qquad (3)$$

where $\delta = \pi\beta$ and $\varepsilon_i = U_i - \beta V_i$.

- OLS is in general inconsistent: $\hat{\beta}^{OLS} \xrightarrow{p} \beta + \dfrac{\sigma_{X\varepsilon}}{\sigma_X^2}$.
- β can be estimated by IV using the *k* instruments *Z*.
- By Frisch-Waugh, you can eliminate *W* by regressing *Y*, *X*, *Z* against *W* and using the residuals. This applies to everything we cover in the linear model so we drop *W* henceforth.

Setup: $Y_i = X_i\beta + \varepsilon_i$ (Structural equation) (1)

$X_i = Z_i'\pi + V_i$ (First stage) (2)

$Y_i = Z_i'\delta + U_i, \quad \delta = \pi\beta, \varepsilon = U - \beta V.$ (Reduced form) (3)

The two conditions for instrument validity

(i) Relevance: cov(Z,X) $\neq 0$ or $\pi \neq 0$ (general k)

(ii) Exogeneity: cov(Z,ε) = 0

The IV estimator when k = 1 (Wright 1926)

$$\text{cov}(Z,Y) = \text{cov}(Z, X\beta + \varepsilon) = \text{cov}(Z,X)\beta + \text{cov}(Z,\varepsilon)$$

$$= \text{cov}(Z,X)\beta \qquad \text{by (i)}$$

so

$$\beta = \frac{\text{cov}(Z,Y)}{\text{cov}(Z,X)} \qquad \text{by (ii)}$$

IV estimator:

$$\hat{\beta}^{IV} = \frac{n^{-1}\sum_{i=1}^{n} Z_i Y_i}{n^{-1}\sum_{i=1}^{n} Z_i X_i} = \frac{\hat{\delta}}{\hat{\pi}}$$

Setup:

$$Y_i = X_i\beta + \varepsilon_i \qquad \text{(Structural equation)} \qquad (1)$$

$$X_i = Z_i'\pi + V_i \qquad \text{(First stage)} \qquad (2)$$

$$Y_i = Z_i'\delta + U_i, \quad \delta = \pi\beta,\ \varepsilon = U - \beta V. \qquad \text{(Reduced form)} \qquad (3)$$

---

k > 1: Two stage least squares (TSLS)

$$\hat{\beta}^{TSLS} = \frac{n^{-1}\sum_{i=1}^{n}\hat{X}_iY_i}{n^{-1}\sum_{i=1}^{n}\hat{X}_i^2}, \qquad \text{where } \hat{X}_i = \text{predicted value from first stage}$$

$$= \frac{\mathbf{X'Z(Z'Z)^{-1}Z'Y}}{\mathbf{X'Z(Z'Z)^{-1}Z'X}}$$

$$= \frac{\hat{\pi}'\hat{Q}_{ZZ}\hat{\delta}}{\hat{\pi}'\hat{Q}_{ZZ}\hat{\pi}}, \qquad \text{where } \hat{Q}_{ZZ} = n^{-1}\sum_{i=1}^{n}Z_iZ_i'$$

The weak instruments problem is a "divide by zero" problem

- $cov(Z,X)$ is nearly zero; or π is nearly zero; or
- $\hat{\pi}'\hat{Q}_{ZZ}\hat{\pi}$ is noisy
- Weak IV is a subset of weak identification (Stock-Wright 2000, Nelson-Starts 2006, Andrews-Cheng 2012)

Statistics for measuring instrument strength

Non-robust: $$F^N = n\frac{\hat{\pi}'\hat{Q}_{ZZ}\hat{\pi}}{k\hat{\sigma}_V^2}$$

Robust: $$F^R = \frac{\hat{\pi}'\hat{\Sigma}_{\pi\pi}^{-1}\hat{\pi}}{k}$$

MOP Effective *F*: $$F^{Eff} = \frac{\hat{\pi}'\hat{Q}_{ZZ}\hat{\pi}}{tr\left(\hat{\Sigma}_{\pi\pi}^{1/2}\hat{Q}_{ZZ}\hat{\Sigma}_{\pi\pi}^{1/2\prime}\right)} = \frac{k\hat{\sigma}_V^2}{tr\left(\hat{\Sigma}_{\pi\pi}^{1/2}\hat{Q}_{ZZ}\hat{\Sigma}_{\pi\pi}^{1/2\prime}\right)}F^N$$

compare to TSLS: $$\hat{\beta}^{TSLS} = \frac{\hat{\pi}'\hat{Q}_{ZZ}\hat{\delta}}{\hat{\pi}'\hat{Q}_{ZZ}\hat{\pi}}$$

Intuition

- F^N measures the right thing ($\pi'Q_{ZZ}\pi$), but gets the SEs wrong
- F^R measures the wrong thing ($\pi'\Sigma_{\pi\pi}^{-1}\pi$), but gets the SEs right
- F^{Eff} measures the right thing and gets SEs right "on average"

Distributional assumptions

Setup: $X_i = Z_i'\pi + V_i$ (First stage) (2)

$Y_i = Z_i'\delta + U_i, \quad \delta = \pi\beta, \varepsilon = U - \beta V.$ (Reduced form) (3)

CLT: $\begin{pmatrix} \sqrt{n}\left(\hat{\delta} - \delta\right) \\ \sqrt{n}\left(\hat{\pi} - \pi\right) \end{pmatrix} \xrightarrow{d} N\left(0, \Sigma^*\right)$, Σ^* is HR/HAC/Cluster (henceforth, "HR")

(i) CLT limit holds exactly: $\begin{pmatrix} \hat{\delta} \\ \hat{\pi} \end{pmatrix} \sim N\left(\begin{pmatrix} \delta \\ \pi \end{pmatrix}, \Sigma\right)$, where $\Sigma = \begin{pmatrix} \Sigma_{\delta\delta} & \Sigma_{\delta\pi} \\ \Sigma_{\pi\delta} & \Sigma_{\pi\pi} \end{pmatrix} = n^{-1}\Sigma^*$

(ii) Reduced form variance & moment matrices are all known: Σ, Q_{ZZ}

<u>A lot is going on here!</u>

- HR/HAC/cluster variance estimators are consistent
- 1950s-1970s finite-sample normal (fixed Z's) literature

A lot is going on here, ctd

From $$\begin{pmatrix} \sqrt{n}\left(\hat{\delta}-\delta\right) \\ \sqrt{n}\left(\hat{\pi}-\pi\right) \end{pmatrix} \xrightarrow{d} N\left(0,\Sigma^*\right)$$

to $$\begin{pmatrix} \hat{\delta} \\ \hat{\pi} \end{pmatrix} \sim N\left(\begin{pmatrix} \delta \\ \pi \end{pmatrix},\Sigma\right), \text{ where } \Sigma = \begin{pmatrix} \Sigma_{\delta\delta} & \Sigma_{\delta\pi} \\ \Sigma_{\pi\delta} & \Sigma_{\pi\pi} \end{pmatrix} = n^{-1}\Sigma^*$$

- Weak IV asymptotics (Staiger-Stock 1997): $\pi = C/\sqrt{n}$.

$$kF^R = \hat{\pi}'\hat{\Sigma}_{\pi\pi}^{-1}\hat{\pi} = \left(\sqrt{n}\hat{\pi}\right)'\left(\hat{\Sigma}_{\pi\pi}^{-1}/n\right)\left(\sqrt{n}\hat{\pi}\right)$$

$$= \left(\sqrt{n}\left(\hat{\pi}-\pi\right)+\sqrt{n}\pi\right)'\hat{\Sigma}_{\pi\pi}^{*-1}\left(\sqrt{n}\left(\hat{\pi}-\pi\right)+\sqrt{n}\pi\right)$$

$$= \left(\sqrt{n}\left(\hat{\pi}-\pi\right)+C\right)'\hat{\Sigma}_{\pi\pi}^{*-1}\left(\sqrt{n}\left(\hat{\pi}-\pi\right)+C\right) \xrightarrow{d} \chi^2_{k;C'\Sigma_{\pi\pi}^{*}C}$$

- Limit experiment interpretation (Hirano-Porter 2015)
- Uniformity (D. Andrews-Cheng 2012)

Homework problem

Let k = 2 and $\hat{Q}_{ZZ} = I_2$. Suppose $\Sigma = \begin{pmatrix} \sigma_U^2 & \sigma_{UV} \\ \sigma_{UV} & \sigma_V^2 \end{pmatrix} \otimes \begin{pmatrix} \omega^2 & 0 \\ 0 & \omega^{-2} \end{pmatrix} / n$.

1) Show that:

a) $tr(\Sigma_{\pi\pi}^{1/2} Q_{ZZ} \Sigma_{\pi\pi}^{1/2\prime}) = \left(\omega^2 + \omega^{-2}\right)\sigma_V^2 / n$.

b) $F^N \cong \frac{1}{2}\left[\left(\lambda_1 + z_{\pi,1}\right)^2 \omega^2 + \left(\lambda_2 + z_{\pi,2}\right)^2 \omega^{-2}\right]$

c) $F^R \cong \frac{1}{2}\left(\lambda + z_\pi\right)'\left(\lambda + z_\pi\right)$

d) $F^{Eff} \cong \dfrac{\left(\lambda_1 + z_{\pi,1}\right)^2 \omega^2 + \left(\lambda_2 + z_{\pi,2}\right)^2 \omega^{-2}}{\omega^2 + \omega^{-2}}$

2) Adopt the weak instrument nesting $\pi = n^{-1/2}C$, where C_1, $C_2 \neq 0$. Show that as $\omega^2 \rightarrow \infty$:

a) "bias" of $\hat{\beta}^{TSLS} - \beta \cong \sigma_{\varepsilon V} / \sigma_V^2 = \text{plim}\left(\hat{\beta}^{OLS} - \beta\right)$

b) $F^N \xrightarrow{p} \infty$

c) $F^R \xrightarrow{p} \infty$

d) $F^{Eff} \xrightarrow{d} \chi_1^2$

3) Discuss

Work out the details for *k* = 1 first.

Preliminaries:

(a) Use distributional assumption (i)

$$\begin{pmatrix}\hat{\delta}\\ \hat{\pi}\end{pmatrix} \sim N\left(\begin{pmatrix}\delta\\ \pi\end{pmatrix}, \Sigma\right), \quad \text{where } \Sigma = \begin{pmatrix}\Sigma_{\delta\delta} & \Sigma_{\delta\pi}\\ \Sigma_{\pi\delta} & \Sigma_{\pi\pi}\end{pmatrix} = n^{-1}\Sigma^{*}$$

to write,

$$\begin{matrix}\hat{\delta} \cong \delta + \psi_{\delta}\\ \hat{\pi} \cong \pi + \psi_{\pi}\end{matrix}, \text{ where } \begin{pmatrix}\psi_{\delta}\\ \psi_{\pi}\end{pmatrix} \sim N\left(0, \begin{pmatrix}\Sigma_{\delta\delta} & \Sigma_{\delta\pi}\\ \Sigma_{\pi\delta} & \Sigma_{\pi\pi}\end{pmatrix}\right)$$

(b) Connect to the structural regression:

$$\mathbf{(Z'Z)^{-1}Z'\varepsilon} = \hat{\delta} - \hat{\pi}\beta \cong (\delta + \psi_{\delta}) - (\pi + \psi_{\pi})\beta = (\delta - \pi\beta) + (\psi_{\delta} - \psi_{\pi}\beta)$$

$$= \psi_{\varepsilon}, \quad \text{where} \quad \psi_{\varepsilon} = \psi_{\delta} - \psi_{\pi}\beta$$

(c) Standardize:

$$\begin{matrix}\hat{\pi} \sim \pi + \psi_{\pi} = (\lambda + z_{\pi})\Sigma_{\pi\pi}^{1/2}\\ \psi_{\varepsilon} = z_{\varepsilon}\Sigma_{\varepsilon\varepsilon}^{1/2}\end{matrix}, \text{ where } \lambda = \Sigma_{\pi\pi}^{-1/2}\pi \text{ and } \begin{pmatrix}z_{\varepsilon}\\ z_{\pi}\end{pmatrix} \sim N\left(0, \begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}\right)$$

(d) Project & orthogonalize:

$$z_{\varepsilon} = \rho z_{\pi} + \eta \text{ , where } \eta \sim N(0, 1-\rho^{2}), \quad \eta \perp z_{v} \text{ , } \rho = \Sigma_{\varepsilon\pi} \Big/ \sqrt{\Sigma_{\varepsilon\varepsilon}\Sigma_{\pi\pi}}$$

What parameter governs departures from usual asymptotics (k = 1)?

$$\hat{\beta}^{IV} = \frac{\hat{\delta}}{\hat{\pi}}$$

$$= \frac{\hat{\pi}\beta + (\hat{\delta} - \hat{\pi}\beta)}{\hat{\pi}} \quad \text{add and subtract } \hat{\pi}\beta$$

$$\cong \beta + \frac{\psi_\varepsilon}{\pi + \psi_\pi} \quad \text{use representations in (a) and (b)}$$

$$= \beta + \frac{z_\varepsilon}{\lambda + z_\pi}\left(\frac{\Sigma_{\varepsilon\varepsilon}}{\Sigma_{\pi\pi}}\right)^{1/2} \text{standardize using representation in (c)}$$

$$= \beta + \frac{z_\pi}{\lambda + z_\pi}\left(\frac{\Sigma_{\varepsilon\pi}}{\Sigma_{\pi\pi}}\right) + \frac{\eta}{\lambda + z_\pi}\left(\frac{\Sigma_{\varepsilon\varepsilon}}{\Sigma_{\pi\pi}}\right)^{1/2} \quad \text{using projection (d)}$$

"bias" "noise"

Parameter measuring instrument strength (k = 1) is $\lambda^2 = \pi^2 / \Sigma_{\pi\pi}$

“Bias” part of IV representation

$$\hat{\beta}^{IV} - \beta \cong \frac{z_\pi}{\lambda + z_\pi}\left(\frac{\Sigma_{\varepsilon\pi}}{\Sigma_{\pi\pi}}\right), \text{ where } \lambda = \Sigma_{\pi\pi}^{-1/2}\pi$$

Instrument strength depends on λ^2

- Strong instruments: $\lambda^2 \to \infty$, usual asymptotic distribution
- Irrelevant instruments: $\pi = 0$ so $\lambda = 0$:

$$\hat{\beta}^{IV} - \beta \cong \frac{\Sigma_{\varepsilon\pi}}{\Sigma_{\pi\pi}} + \frac{\eta}{z_\pi}\left(\frac{\Sigma_{\varepsilon\varepsilon}^{1/2}}{\Sigma_{\pi\pi}^{1/2}}\right) \sim \text{Cauchy centered at } \frac{\Sigma_{\varepsilon\pi}}{\Sigma_{\pi\pi}}$$

  - In homoskedastic case, $\frac{\Sigma_{\varepsilon\pi}}{\Sigma_{\pi\pi}} = \frac{\sigma_{\varepsilon V}}{\sigma_V^2} = \text{plim}\left(\hat{\beta}^{OLS} - \beta\right)$

- In the homoskedastic case, λ^2 = the concentration parameter (old Edgeworth expansion/finite sample distribution literature)

Instrument strength, *k* = 1, ctd.

How big does λ need to be? A "bias" heuristic:

$$\frac{E\left(\hat{\beta}^{IV}-\beta\right)}{\Sigma_{\varepsilon\pi}/\Sigma_{\pi\pi}} = E\frac{z_\pi}{\lambda+z_\pi}$$

$$= E\frac{z_\pi/\lambda}{1+z_\pi/\lambda}$$

$$\approx E\left(\frac{z_\pi}{\lambda}\right)\left(1-\frac{z_\pi}{\lambda}+\ldots\right) = -E\left(\frac{z_\pi^2}{\lambda^2}\right) = -\frac{1}{\lambda^2}$$

- For bias, relative to unidentified case, to be <0.1, need $\lambda^2 > 10$.
- But we don't know λ! So, we need a statistic with a distribution that depends on λ, which we can use to back out an estimate/test/rule of thumb.
- This is the Nagar (1959) expansion for the bias
- *How do the three candidate first-stage Fs fare?*

Distributions of the three first-stage Fs, k = 1

First note that, when $k = 1$, $F^R = F^{Eff}$: $F^{Eff} = \dfrac{\hat{\pi}'\hat{Q}_{ZZ}\hat{\pi}}{tr\left(\hat{\Sigma}_{\pi\pi}^{1/2}\hat{Q}_{ZZ}\hat{\Sigma}_{\pi\pi}^{1/2\prime}\right)} = \dfrac{\hat{\pi}^2}{\hat{\Sigma}_{\pi\pi}} = F^R$

Distributions

$$F^{Eff}, F^R = \frac{\hat{\pi}^2}{\hat{\Sigma}_{\pi\pi}} \cong (\lambda + z_v)^2 \sim \chi^2_{1;\lambda^2}$$

$$F^N = n\frac{\hat{\pi}'\hat{Q}_{ZZ}\hat{\pi}}{\hat{\sigma}_V^2} = \frac{\hat{\pi}^2}{\hat{\sigma}_V^2 / n\hat{Q}_{ZZ}} \cong (\lambda + z_\pi)^2 \frac{\Sigma^*_{\pi\pi}}{\sigma_V^2 / Q_{ZZ}}$$

Implications

F^R, F^{Eff} can be used for inference about λ^2 when $k = 1$

- Estimation: $EF^{eff} = E\left(\lambda + z_V\right)^2 = \lambda^2 + 1$, so $\hat{\lambda}^2 = F^{Eff} - 1$
- Testing: H_0: "bias" ≤ 0.1. Reject H_0 if $F^{Eff} >$ critical value.
- Rule of thumb: $F^{eff} < 10$ will detect weak IVs with probability that increases as λ^2 gets smaller

Implications, ctd.

$$F^{N} \cong (\lambda + z_V)^2 \frac{\Sigma^*_{\pi\pi}}{\sigma_V^2 / Q_{ZZ}}$$

$$F^{Eff}, F^{R} \cong \left(\lambda + z_V\right)^2$$

F^N is misleading in the HR case.

- Suppose $\Sigma^*_{\pi\pi}$ is large (i.e., first stage HR SEs are a lot bigger than NR SEs)

$$F^{N} \cong (\lambda + z_V)^2 \frac{\Sigma^*_{\pi\pi}}{\sigma_V^2 / Q_{ZZ}} \sim \frac{\Sigma^*_{\pi\pi}}{\sigma_V^2 / Q_{ZZ}} \times \chi^2_{1;\lambda^2}$$

where $\lambda^2 = \pi^2/\Sigma_{\pi\pi}$. For $\Sigma^*_{\pi\pi}$ large, $\lambda^2 \approx 0$, and $F^N \sim \frac{\Sigma^*_{\pi\pi}}{\sigma_V^2 / Q_{ZZ}} \times \chi^2_1 \to \infty$

i.e., Instruments are in the limit irrelevant – but $F_N \to \infty$.

In the $k = 1$ case, $F^R = F^{Eff}$. These differ in the $k > 1$ case, where F^{Eff} is preferred.

Homework problem

Let $k = 2$ and $\hat{Q}_{ZZ} = I_2$. Suppose $\Sigma = \begin{pmatrix} \sigma_U^2 & \sigma_{UV} \\ \sigma_{UV} & \sigma_V^2 \end{pmatrix} \otimes \begin{pmatrix} \omega^2 & 0 \\ 0 & \omega^{-2} \end{pmatrix} / n$.

1) Show that:

a) $tr(\Sigma_{\pi\pi}^{1/2} Q_{ZZ} \Sigma_{\pi\pi}^{1/2\prime}) = \left(\omega^2 + \omega^{-2}\right)\sigma_V^2 / n$.

b) $F^N \cong \frac{1}{2}\left[\left(\lambda_1 + z_{\pi,1}\right)^2 \omega^2 + \left(\lambda_2 + z_{\pi,2}\right)^2 \omega^{-2}\right]$

c) $F^R \cong \frac{1}{2}\left(\lambda + z_\pi\right)'\left(\lambda + z_\pi\right)$

d) $F^{Eff} \cong \dfrac{\left(\lambda_1 + z_{\pi,1}\right)^2 \omega^2 + \left(\lambda_2 + z_{\pi,2}\right)^2 \omega^{-2}}{\omega^2 + \omega^{-2}}$

2) Adopt the weak instrument nesting $\pi = n^{-1/2}C$, where C_1, $C_2 \neq 0$. Show that as $\omega^2 \to \infty$:

a) "bias" of $\hat{\beta}^{TSLS} - \beta \sim \sigma_{\varepsilon V}/\sigma_V^2 = \text{plim}\left(\hat{\beta}^{OLS} - \beta\right)$

b) $F^N \xrightarrow{p} \infty$

c) $F^R \xrightarrow{p} \infty$

d) $F^{Eff} \xrightarrow{d} \chi_1^2$

3) Discuss

Homework problem solution

Let $k = 2$ and $\hat{Q}_{ZZ} = I_2$. Suppose $\Sigma = \begin{pmatrix} \sigma_U^2 & \sigma_{UV} \\ \sigma_{UV} & \sigma_V^2 \end{pmatrix} \otimes \begin{pmatrix} \omega^2 & 0 \\ 0 & \omega^{-2} \end{pmatrix} / n$ and $\pi_1, \pi_2 \neq 0$

1(a) Direct calculation: $tr(\Sigma_{\pi\pi}^{1/2} Q_{ZZ} \Sigma_{\pi\pi}^{1/2\prime}) = \left(\omega^2 + \omega^{-2}\right)\sigma_V^2 / n$

1(b)-(d): We have already done the work to get the expressions below following "~", and the final expressions come from substitution of Q_{ZZ} and Σ:

(b) $$F^N = \frac{n\hat{\pi}'\hat{Q}_{zz}\hat{\pi}}{k\sigma_V^2} \cong \frac{\left(\lambda + z_\pi\right)' n\Sigma_{\pi\pi}^{1/2}\hat{Q}_{zz}\Sigma_{\pi\pi}^{1/2\prime}\left(\lambda + z_\pi\right)}{k\sigma_V^2}$$

$$= \frac{1}{2}\left[\left(\lambda_1 + z_{\pi,1}\right)^2 \omega^2 + \left(\lambda_2 + z_{\pi,2}\right)^2 \omega^{-2}\right]$$

(c) $$F^R = \frac{\hat{\pi}'\hat{\Sigma}_{\pi\pi}^{-1}\hat{\pi}}{k} \cong \frac{\left(\lambda + z_V\right)'\left(\lambda + z_V\right)}{k} = \frac{1}{2}\left(\lambda + z_\pi\right)'\left(\lambda + z_\pi\right)$$

(d) $$F^{Eff} = \frac{\hat{\pi}'\hat{Q}_{ZZ}\hat{\pi}}{tr\left(\hat{\Sigma}_{\pi\pi}^{1/2}\hat{Q}_{ZZ}\hat{\Sigma}_{\pi\pi}^{1/2\prime}\right)} \cong \frac{\left(\lambda + z_\pi\right)'\Sigma_{\pi\pi}^{1/2}\hat{Q}_{zz}\Sigma_{\pi\pi}^{1/2\prime}\left(\lambda + z_\pi\right)}{tr\left(\Sigma_{\pi\pi}^{1/2}Q_{ZZ}\Sigma_{\pi\pi}^{1/2\prime}\right)}$$

$$= \frac{\left(\lambda_1 + z_{\pi,1}\right)^2 \omega^2 + \left(\lambda_2 + z_{\pi,2}\right)^2 \omega^{-2}}{\omega^2 + \omega^{-2}}$$

Homework problem solution, ctd.

2) Adopt the weak instrument nesting $\pi = n^{-1/2}C$, where C_1, $C_2 \neq 0$. Show that as $\omega^2 \to \infty$:

a) "bias" of $\hat{\beta}^{TSLS} - \beta \sim \sigma_{\varepsilon V} / \sigma_V^2 = \text{plim}\left(\hat{\beta}^{OLS} - \beta\right)$

Last part first: $\text{plim}\left(\hat{\beta}^{OLS} - \beta\right) = \sigma_{\varepsilon X} / \sigma_X^2 = \sigma_{\varepsilon V} / \sigma_V^2$ because $\pi = n^{1/2}C$.

Next obtain the expression (*several tedious steps*),

$$\text{“Bias” part } \hat{\beta}^{TSLS} - \beta \cong \frac{(\lambda + z_\pi)' H R z_\pi}{(\lambda + z_\pi)' H (\lambda + z_\pi)}$$

where $H = \Sigma_{\pi\pi}^{1/2} Q_{ZZ} \Sigma_{\pi\pi}^{1/2} = \begin{pmatrix} \omega^2 & 0 \\ 0 & \omega^{-2} \end{pmatrix} \sigma_V^2 / n$ and $R = \Sigma_{\pi\pi}^{-1/2} \Sigma_{\varepsilon\pi} \Sigma_{\pi\pi}^{-1/2\prime} = \frac{\sigma_{\varepsilon V}}{\sigma_V^2} I_2$.

For the weak instrument nesting,

$$\lambda = \Sigma_{\pi\pi}^{-1/2} \pi = \left[\sigma_V^2 \begin{pmatrix} \omega^2 & 0 \\ 0 & \omega^{-2} \end{pmatrix} / n \right]^{-/12} \pi$$

$$= \begin{pmatrix} \omega^{-1} & 0 \\ 0 & \omega \end{pmatrix} n^{1/2} \pi / \sigma_V = \begin{pmatrix} C_1 \omega^{-1} / \sigma_V \\ C_2 \omega / \sigma_V \end{pmatrix}$$

Homework problem solution, ctd.

Now substitute these expressions for λ, H, and R into the "bias" part:

$$\hat{\beta}^{TSLS} - \beta \cong \frac{(\lambda + z_\pi)'\begin{pmatrix} \omega^2 & 0 \\ 0 & \omega^{-2} \end{pmatrix} z_\pi}{(\lambda + z_\pi)'\begin{pmatrix} \omega^2 & 0 \\ 0 & \omega^{-2} \end{pmatrix}(\lambda + z_\pi)} \frac{\sigma_{\varepsilon V}}{\sigma_V^2}$$

$$= \frac{(C_1/\sigma_V + z_{\pi,1}\omega)z_{\pi,1}\omega + (C_2/\sigma_V + z_{\pi,2}\omega^{-1})z_{\pi,2}\omega^{-1}}{(C_1/\sigma_V + z_{\pi,1}\omega)^2 + (C_2/\sigma_V + z_{\pi,2}\omega^{-1})^2}\left(\frac{\sigma_{\varepsilon V}}{\sigma_V^2}\right)$$

$$= \left(1 + O_p(\omega^{-1})\right)\left(\frac{\sigma_{\varepsilon V}}{\sigma_V^2}\right)$$

Homework problem solution, ctd.

Remaining parts by substitution and taking limits:

(b)
$$F^N \cong \frac{1}{2}\left[\left(\lambda_1 + z_{\pi,1}\right)^2 \omega^2 + \left(\lambda_2 + z_{\pi,2}\right)^2 \omega^{-2}\right]$$
$$= \frac{1}{2}\left[\left(C_1/\sigma_V + z_{\pi,1}\omega\right)^2 + \left(C_2/\sigma_V + z_{\pi,2}\omega^{-1}\right)^2\right] \sim \frac{1}{2}\omega^2\chi_1^2 + O_p(\omega)$$

(c)
$$F^R \cong \frac{1}{2}\left(\lambda + z_\pi\right)'\left(\lambda + z_\pi\right)$$
$$= \frac{1}{2}\left[\left(C_1\omega^{-1}/\sigma_V + z_{\pi,1}\right)^2 + \left(C_2\omega/\sigma_V + z_{\pi,2}\right)^2\right]$$
$$= \frac{1}{2}\frac{C_2^2}{\sigma_V^2}\omega^2 + O_p(\omega) \to \infty$$

(d)
$$F^{Eff} = \frac{F^N}{\omega^2 + \omega^{-2}} \cong \frac{\omega^2 z_{\pi,1}^2 + O_p(\omega)}{\omega^2 + \omega^{-2}} = z_{\pi,1}^2 + O_p(\omega^{-1}) \sim \chi_1^2$$

3) Discuss

OK, F^{Eff} – but what cutoff?

$$F^{Eff} \cong (\lambda + z_\pi)' H (\lambda + z_\pi), \quad \text{where} \quad H = \frac{\Sigma_{\pi\pi}^{1/2} Q_{ZZ} \Sigma_{\pi\pi}^{1/2}}{tr(\Sigma_{\pi\pi}^{1/2} Q_{ZZ} \Sigma_{\pi\pi}^{1/2})}$$

~ weighted average of noncentral χ^2's – depends on full matrix H,

$0 \leq$ eigenvalues (H) ≤ 1

Hierarchy of options

1. **Testing approach**: test null of $\lambda' H \lambda \geq$ some threshold (e.g. 10% bias)
   a) (MOP Monte Carlo method) Given $\hat{H}$, compute cutoff $\lambda' \hat{H} \lambda$; critical value by simulation
   b) (MOP Paitnik-Nagar method) Approximate weighted average of noncentral χ^2's by noncentral χ^2; compute cutoff value of $\lambda' H \lambda$ using Nagar approximation to the bias, with some maximal allowable bias. Implemented in **`weakivtest.ado.`**
   c) (MOP simple method) Pick a maximal allowable bias (or size distortion) and use their "simple" critical values (based on noncentral χ^2 bounding distribution). *These are simple, but conservative.*
2. **Consistent sequence approach**: "Weak" if $F^{Eff} < \kappa_n$, $\kappa_n \to \infty$ (but what is κ_n?)
3. **Rule-of-thumb approach**: "Weak" if $F^{Eff} < 10$

k=1 case, additional comments about F^{Eff} and F^{R}

$$\hat{\beta}^{IV} - \beta \cong \frac{z_\pi}{\lambda + z_\pi}\left(\frac{\Sigma_{\varepsilon\pi}}{\Sigma_{\pi\pi}}\right), \text{ where } \lambda = \Sigma_{\pi\pi}^{-1/2}\pi$$

$$t^{IV} = \frac{\hat{\beta}^{IV} - \beta_0}{SE(\hat{\beta}^{IV})} \cong \frac{z_\varepsilon}{\left[1 - 2\left(\frac{z_\varepsilon}{\lambda + z_\pi}\right)\rho + \left(\frac{z_\varepsilon}{\lambda + z_\pi}\right)^2\right]^{1/2}}, \text{ where } \rho = \frac{\Sigma_{\pi\varepsilon}}{\left(\Sigma_{\pi\pi}\Sigma_{\varepsilon\varepsilon}\right)^{1/2}}$$

$$F^{R} = F^{Eff} \cong \left(\lambda + z_\pi\right)'\left(\lambda + z_\pi\right)$$

- By maximizing over ρ you can find worst case size distortion for usual IV t-stat testing β_0. This depends on λ, which can be estimated from $F^R = F^{Eff}$.
- These are the same expressions, with different definition of λ, as in homoskedastic case (special to $k = 1$)
- Critical values for $k = 1$ – two choices:
- Nagar bias ≤ 10%: 23 (5% critical value from $\chi^2_{1;\lambda^2=10}$) (MOP)
- Maximum t^{IV} size distortion of 0.10: 16.4; of 0.15: 9.0
- But with $k = 1$ there are fully robust methods that are easy and have very strong theoretical properties (AR) (Lecture 3).

Detecting weak instruments with multiple included endogenous regressors

Methods are based on multivariate *F*: Cragg-Donald statistic and robust variants

- Nonrobust:
  - Minimum eigenvalue of Cragg-Donald statistic, Stock-Yogo (2005) critical values
  - Sanderson-Windmeijer (2016)
- HR: Main method used is Kleibergen-Paap statistic, which is HR Cragg-Donald.
  - But recall that this doesn't work (theory) for 1 *X*, and having multiple *X*'s doesn't improve things.
- MOP Effective *F*: Hasn't been developed.

More work is needed….

What if you plan to use efficient 2-step GMM, not TSLS?

Everything above is tailored to TSLS!

- Suppose that, if you have strong instruments, you use efficient 2-step GMM:

$$\hat{\beta}^{GMM} = \frac{\hat{\pi}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\delta}}{\hat{\pi}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\pi}} \text{, where } \hat{\Sigma}_{\varepsilon\varepsilon} = \frac{1}{n}\sum_{i=1}^{n} Z_i Z_i' \left(\hat{\varepsilon}_i^{(1)}\right)^2$$

  where $\hat{\varepsilon}_i^{(1)}$ is the residual from a first-stage estimate of β, e.g. TSLS.

- Things get complicated because the first step (TSLS) isn't consistent with weak instruments.
  - $\hat{\Sigma}_{\varepsilon\varepsilon}$ converges in distribution to a random limit
  - If $\Sigma_{\varepsilon\varepsilon}$ were known (infeasible),

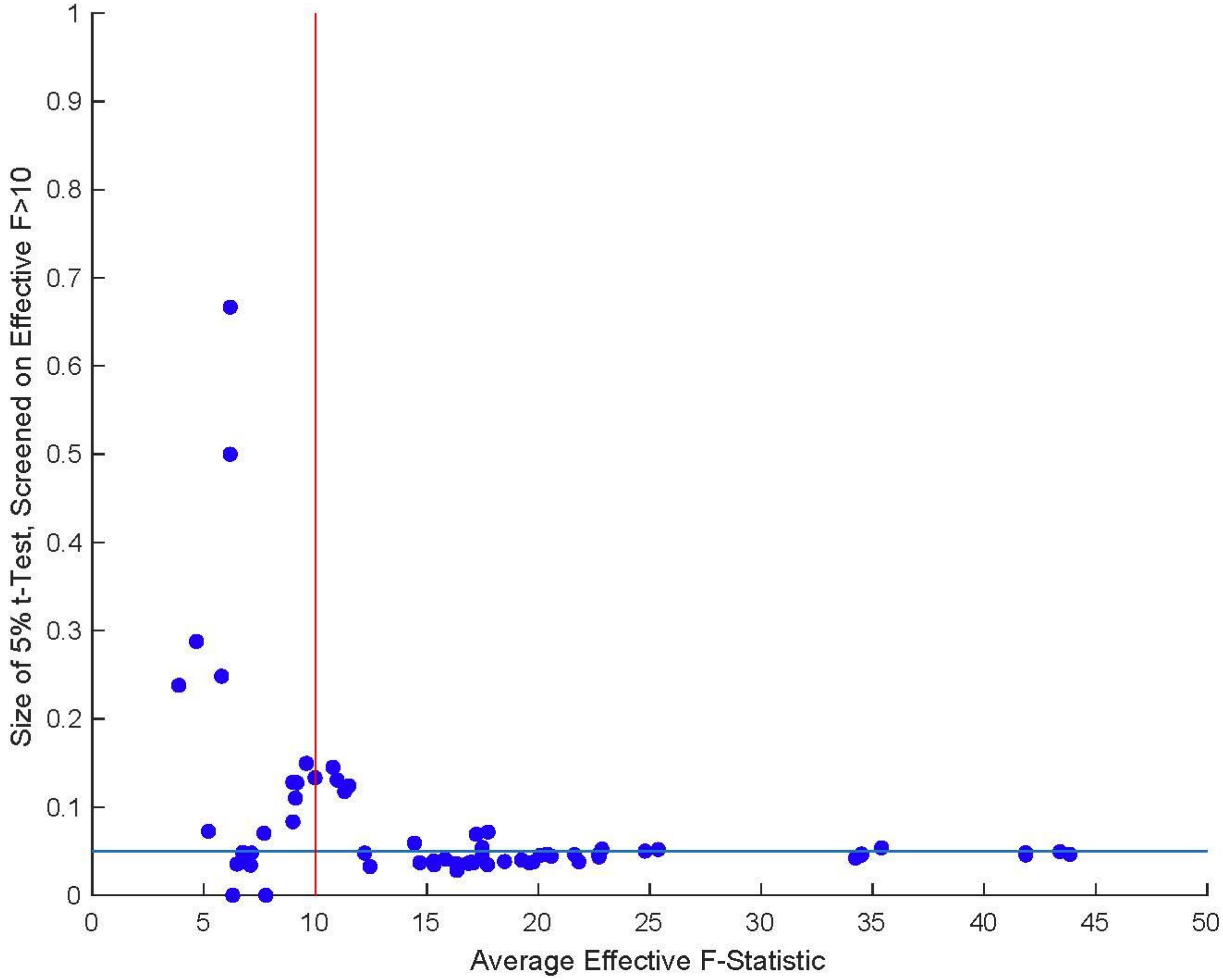
$$\hat{\beta}^{GMM} - \beta \Rightarrow \frac{\left(\lambda + z_\pi\right)' \Sigma_{\pi\pi}^{1/2}\Sigma_{\varepsilon\varepsilon}^{-1/2} z_\varepsilon}{\left(\lambda + z_\pi\right)' \Sigma_{\pi\pi}^{1/2}\Sigma_{\varepsilon\varepsilon}^{-1}\Sigma_{\pi\pi}^{1/2\prime}\left(\lambda + z_\pi\right)}$$

    In general none of the *F*'s discussed so far get at the right object, $\lambda'\Sigma_{\pi\pi}^{1/2}\Sigma_{\varepsilon\varepsilon}^{-1}\Sigma_{\pi\pi}^{1/2\prime}\lambda / tr(\Sigma_{\pi\pi}^{1/2}\Sigma_{\varepsilon\varepsilon}^{-1}\Sigma_{\pi\pi}^{1/2\prime})$. (And this is "right" only if $\Sigma_{\varepsilon\varepsilon}$ is known.)

OK – now what should you do if you have weak instruments?

Wrong answer: reject the paper.

Size distortion from screening based on first stage *F*



Isaiah's will discuss further…

Estimation – What have we learned/state of knowledge

k = 1

- There does not exist an unbiased or asymptotically-unbiased IV estimator (folk theorem; Hirano and Porter 2015).
- Only one moment condition, so weighting (HR) isn't an issue
- LIML=TSLS=IV doesn't have moments…
- Fuller seems to have advantage over IV in terms of "bias" (location) in simulations (e.g., Hahn, Hausman, Kuersteiner (2004), I. Andrews and Armstrong 2017) (so should *k*-class).
- If you know *a-priori* the sign of π, then unbiased, strong-instrument efficient estimation is possible (I. Andrews and Armstrong 2017)

Estimation – What have we learned/state of knowledge, ctd.

$k > 1$

- There does not exist an unbiased or asymptotically-unbiased IV estimator (folk theorem; Hirano and Porter 2015).
- The IV estimators that were developed in the 60s-90s (LIML, *k*-class, double *k*-class, JIVE, Fuller) are special to the homoskedastic case, and in general lose their good properties in the HR case
- Different IV estimators place different weights on the moments, and thus in general have different LATEs
- With heterogeneity, the LIML estimand (Fuller too?) can be outside the convex hull of the LATEs of the individual instruments (Kolesár 2013)
- For GMM applications estimating a structural parameter (e.g. New Keynesian Phillips Curve, etc.), the LATE concerns don't apply, however when the moment conditions are nonlinear in θ, things get difficult.
- If you know *a-priori* the sign of π, then unbiased estimation is possible (I. Andrews and Armstrong 2017)

References

Andrews, D.W.K. and X. Cheng (2012). "Estimation and Inference with Weak, Semi-Strong, and Strong Identification," *Econometrica* 80, 2153-2211.

Andrews, I. and T. Armstrong (2017). "Unbiased Instrumental Variables Estimation under Known First-Stage Sign," *Quantiative Economics* 8, 479-503.

Hahn, J., J. Hausman, and G. Kuersteiner (2004), "Estimation with weak instruments: Accuracy of higher order bias and MSE approximations," *Econometrics Journal*, 7, 272–306.

Kleibergen, F., and R. Paap (2006). "Generalized Reduced Rank Tests using the Singular Value Decomposition." *Journal of Econometrics* 133: 97–126.

Montiel Olea, J.L. and C.E. Pflueger (2013). "A Robust Test for Weak Instruments," *Journal of Business and Economic Statistics* 31, 358-369.

Nagar, A. L. (1959). "The bias and moment matrix of the general k-class estimators of the parameters in simultaneous equations," *Econometrica* 27: 575–595.

Nelson, C. R., and R. Startz (1990). "Some further results on the exact small sample properties of the instrumental variable estimator." *Econometrica* 58, 967–976.

Nelson, C. R., and R. Startz (2006). "The zero-information-limit condition and spurious inference in weakly identified models," *Journal of Econometrics* 138, 47-62.

Pflueger, C.E. and S. Wang (2015). "A Robust Test for Weak Instruments in Stata," *Stata Journal* 15, 216-225.

Sanderson, E. and F. Windmeijer (2016). "A Weak Instrument *F*-test in Linear IV Models with Multiple Endogenous Variables," *Journal of Econometrics* 190, 212-221.

Staiger, D., and J. H. Stock. 1997. Instrumental variables regression with weak instruments. *Econometrica* 65: 557–86.

Stock, J.H. and M. Yogo (2005). "Testing for Weak Instruments in Linear IV Regression," Ch. 5 in J.H. Stock and D.W.K. Andrews (eds), *Identification and Inference for Econometric Models: Essays in Honor of Thomas J. Rothenberg*, Cambridge University Press, 80-108.