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

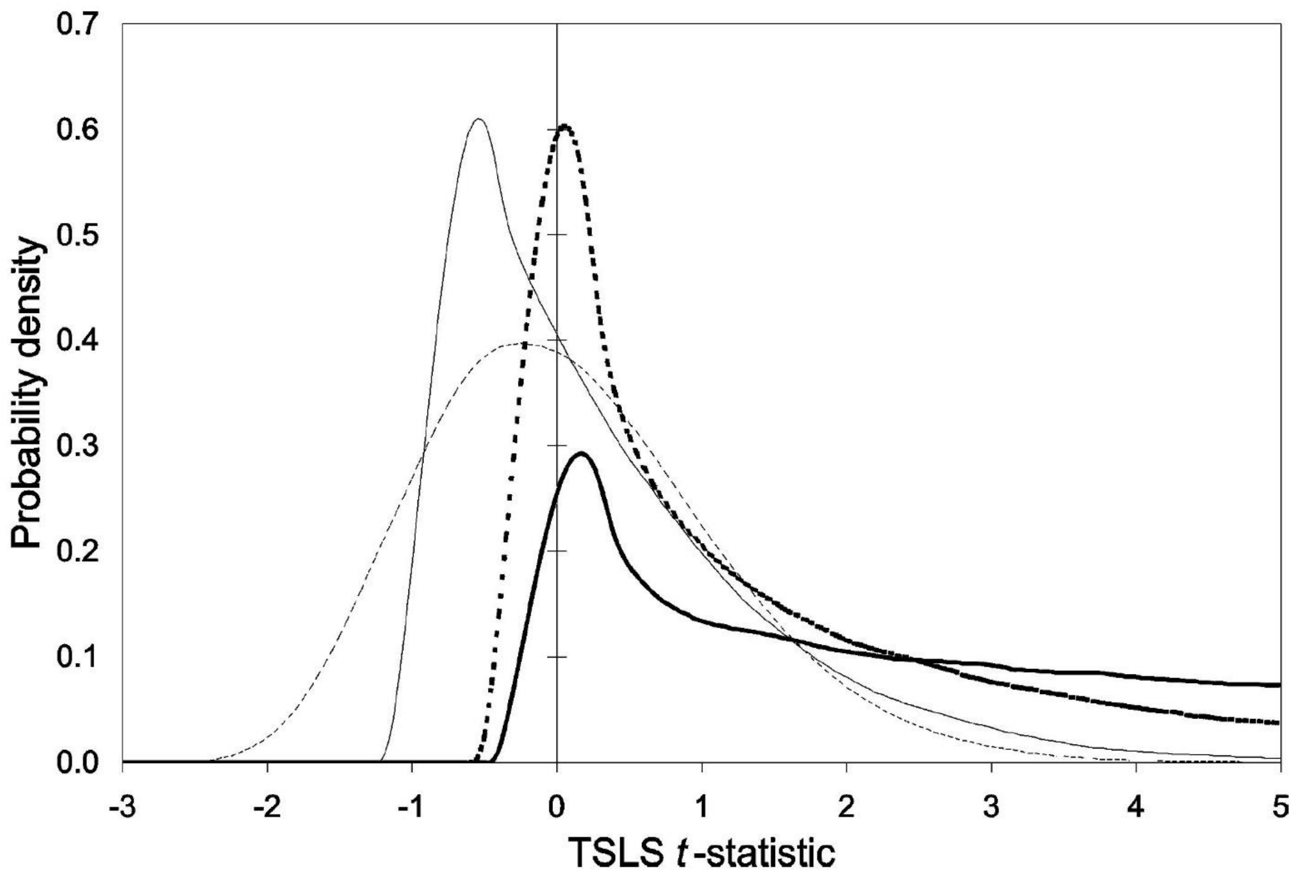
An instrumental variable is weak if its correlation with the included endogenous regressor is small.

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))

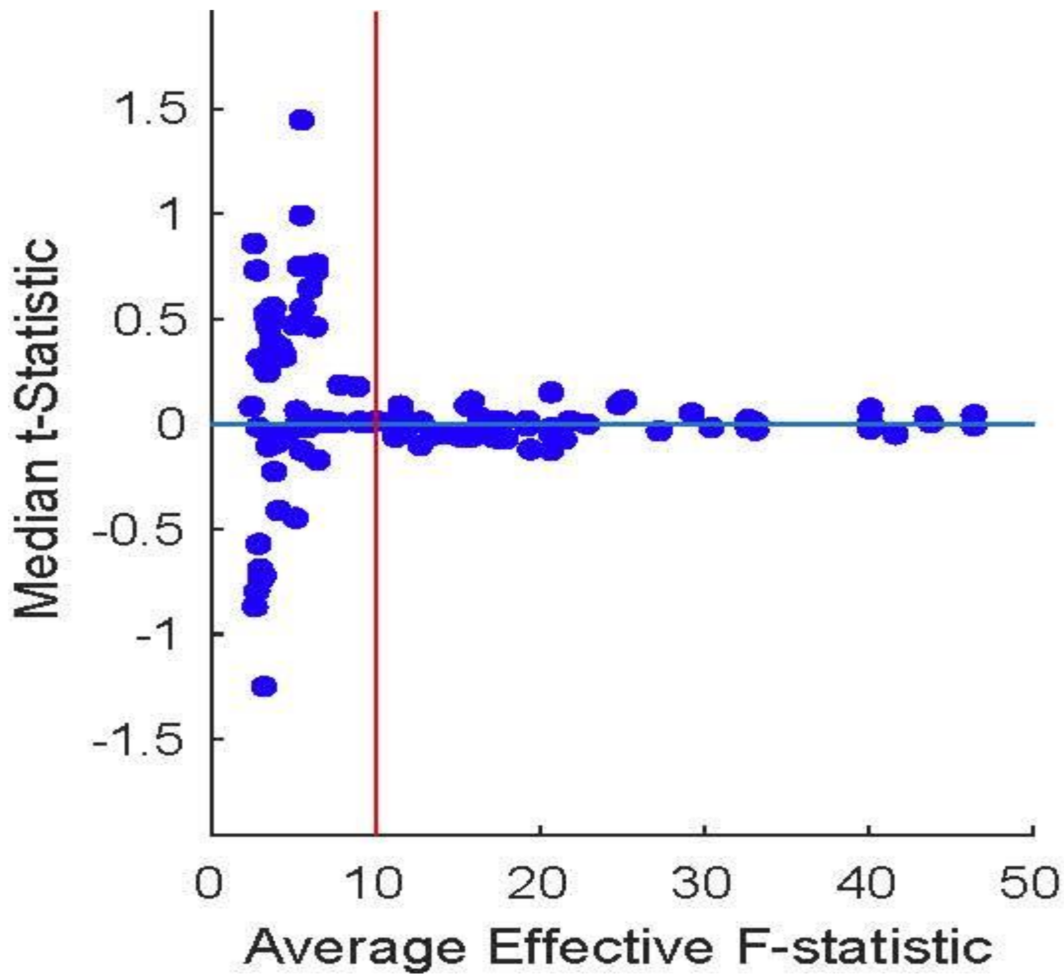
- Dark line = irrelevant instruments
- dashed light line = strong instruments
- intermediate cases = weak instruments



So what? (2) Simulation

DGP: 8 AER papers 2014-2018
(Sample: 17 that use IV; 16 with a single X; 8 in simulation sample)

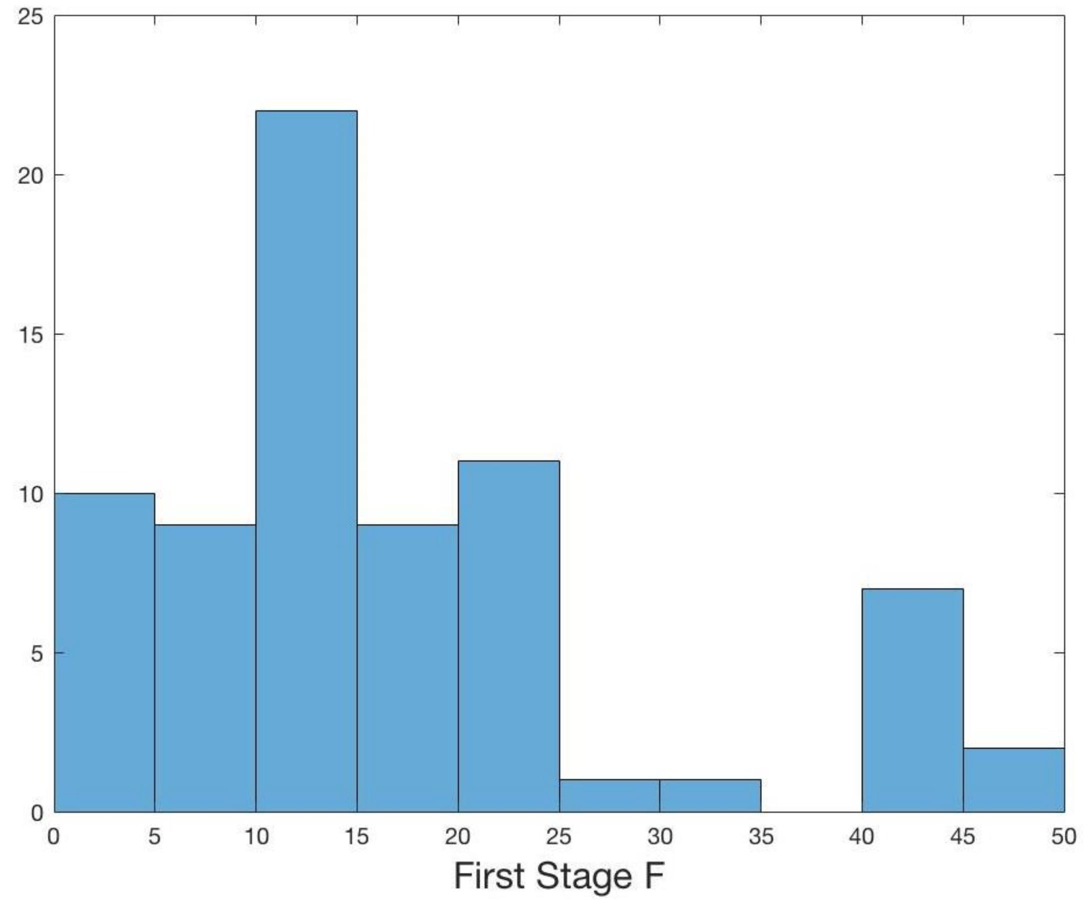
Median of TSLS *t*-statistic under the null



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

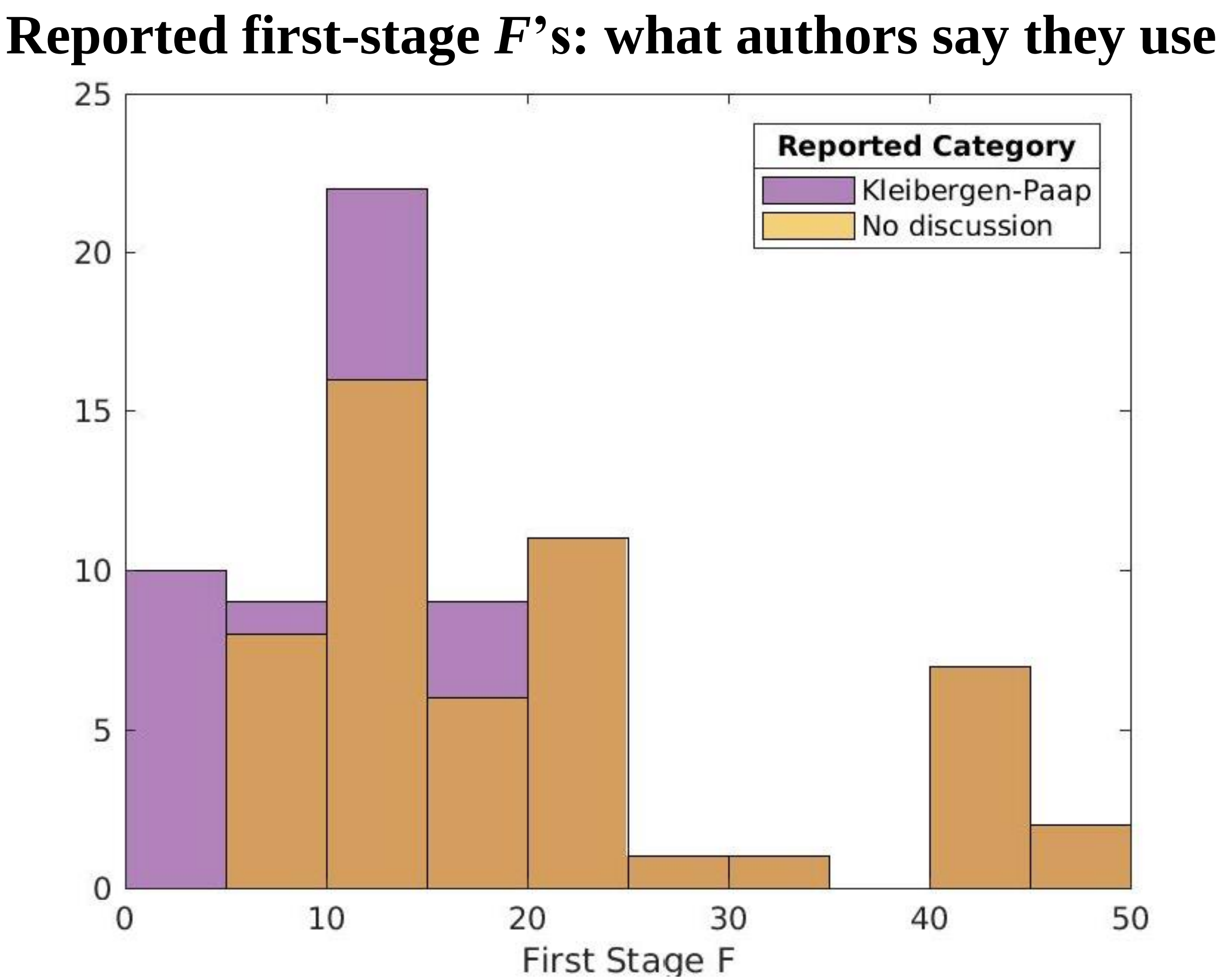
It is convenient to have a way to decide if instruments are strong (TSLS "works") or weak (use weak-instrument robust methods).

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



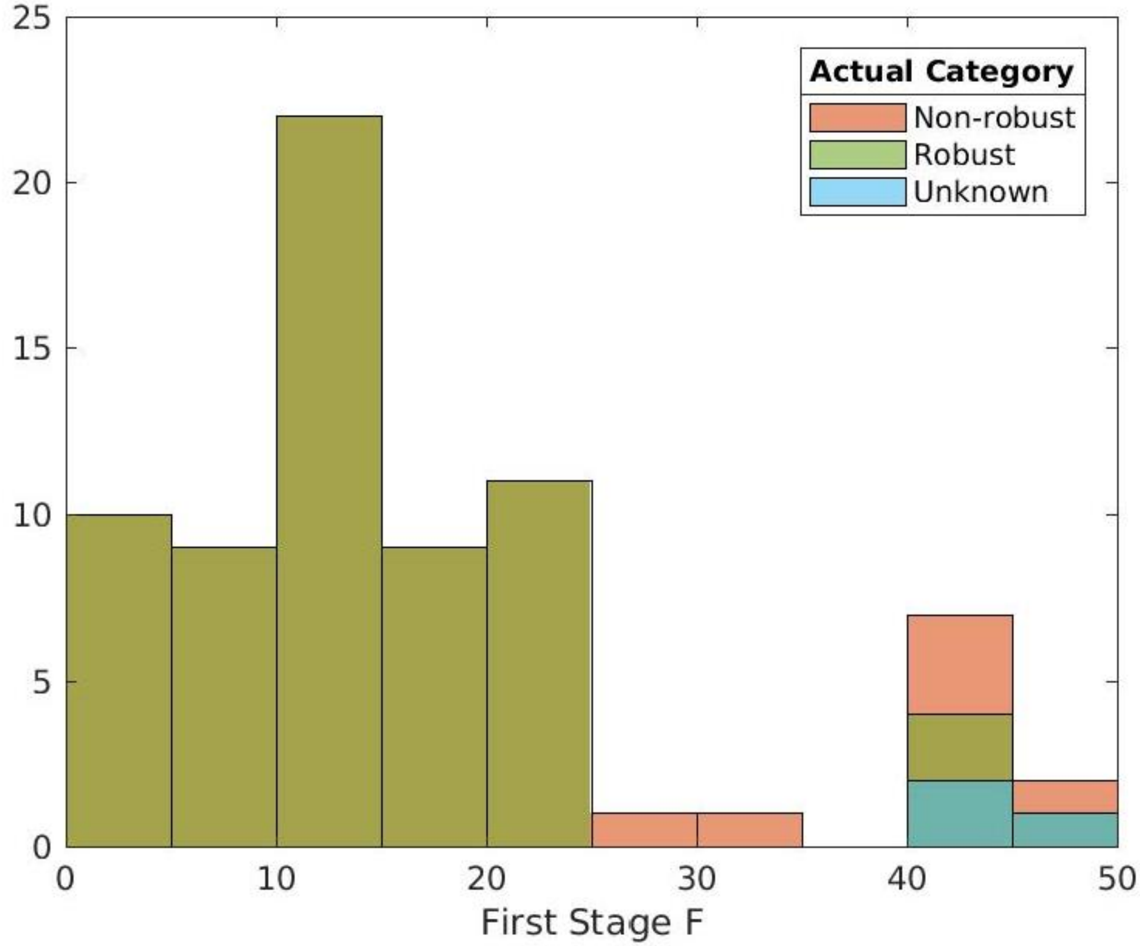
Candidates: F^N – nonrobust

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 ≥ 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:

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

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

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

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)

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

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

so

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

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)

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

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

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

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

<u>Intuition</u>

- 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 $\begin{pmatrix} \hat{\delta} \\ \hat{\pi} \end{pmatrix} \sim N\left(\begin{pmatrix} \delta \\ \pi \end{pmatrix},\Sigma\right)$, 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 $\omega^2 \to \infty$:

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:

$$\begin{aligned}\mathbf{(Z'Z)^{-1}Z'\varepsilon} &= \hat{\delta} - \hat{\pi}\beta \cong (\delta + \psi_{\delta}) - (\pi + \psi_{\pi})\beta = (\delta - \pi\beta) + (\psi_{\delta} - \psi_{\pi}\beta)\\ &= \psi_{\varepsilon}, \quad \text{where} \quad \psi_{\varepsilon} = \psi_{\delta} - \psi_{\pi}\beta\end{aligned}$$

(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)?

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

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

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

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

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

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 λ^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 *F*s, *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*),

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

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 $R = \Sigma_{\pi\pi}^{-1/2} \Sigma_{\varepsilon\pi} \Sigma_{\pi\pi}^{-1/2\prime} = \dfrac{\sigma_{\varepsilon V}}{\sigma_V^2} I_2$.

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?

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

~ weighted average of noncentral χ^2's – depends on full matrix H,
$0 \le$ eigenvalues (H) ≤ 1

Hierarchy of options

1. **Testing approach**: test null of $\lambda' H \lambda \ge$ 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),

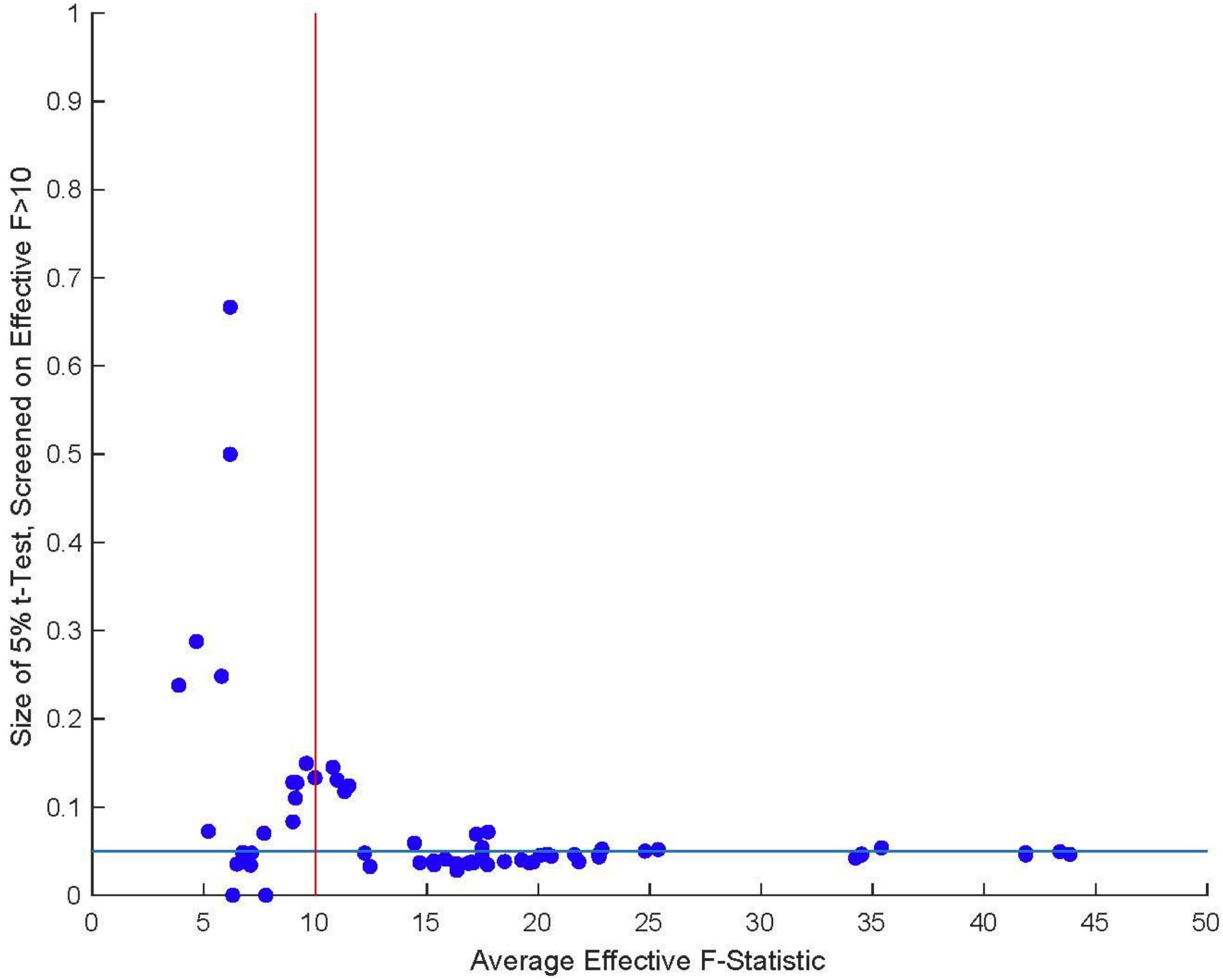
$$\hat{\beta}^{GMM} - \beta \Rightarrow \frac{\left(\lambda + z_\pi\right)' \Sigma_{\pi\pi}^{1/2}\Sigma_{\varepsilon\varepsilon}^{-1/2} z_\varepsilon}{\left(\lambda + z_\pi\right)' \Sigma_{\pi\pi}^{1/2}\Sigma_{\varepsilon\varepsilon}^{-1}\Sigma_{\pi\pi}^{1/2\prime}\left(\lambda + z_\pi\right)}$$

  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.

3: Robust Inference with Weak Instruments

July 22, 2018

The Story So Far...

- Conventional t-test-based confidence intervals can under-cover true parameter value when instruments are weak
- Effective First-stage F-statistic provides a guide to bias
  - But screening applications on F-statistics can induce size distortions
- This section: identification-robust confidence sets
  - Ensure correct coverage regardless of instrument strength
  - No need to screen on first stage
    - Avoids pretesting bias
    - Avoids throwing away applications with valid instruments just because weak
    - Confidence sets can be informative even with weak instruments

Reminder: Normal Model

- To discuss these issues, continue to consider the normal model

$$\begin{pmatrix} \hat{\delta} \\ \hat{\pi} \end{pmatrix} \sim N\left(\begin{pmatrix} \delta \\ \pi \end{pmatrix}, \Sigma \right)$$

  where
  - $\hat{\delta}$ is the reduced-form OLS coefficient
  - $\hat{\pi}$ is the first-stage OLS coefficient
  - Σ is known
- IV model implies $\delta = \pi\beta$

Negative Result

- Initial Question: can we obtain correct coverage by adjusting our standard errors?
  - Confidence interval $\left[\hat{\beta} \pm b\left(\hat{\delta}, \hat{\pi}\right)\right]$ for some $b(\cdot, \cdot)$
  - Answer: no (unless $b\left(\hat{\delta}, \hat{\pi}\right)$ can be infinite)
- Gleser and Hwang (1989) and Dufour (1997) show that for any robust confidence set CS with coverage $1-\alpha$,

$$Pr_{\beta,\pi}\left\{\beta \in CS\right\} \geq 1-\alpha \text{ for all } \beta,\ \pi,$$

  we must have

$$Pr_{\beta,\pi}\left\{CS \text{ has infinite length}\right\} > 0 \text{ for all } \beta,\ \pi$$

  - Inutition: in case with $\pi = 0$, must cover every value β with probability $1-\alpha$
- Adjusting our (finite) standard errors isn't enough: need alternative approach

Test Inversion

- Leading alternative: test inversion
- Idea: Define a family of tests $\phi(\cdot)$ where
  - $\phi(\beta_0)$ test for $H_0 : \beta = \beta_0$
  - $\phi(\beta_0) = 1$ if reject H_0, 0 otherwise
- Suppose $\phi(\beta_0)$ has size α for all β_0, i.e.

$$E_{\beta_0,\pi}\left[\phi(\beta_0)\right] \leq \alpha \text{ for all } \beta_0,\ \pi$$

- If we form CS by collecting the non-rejected values

$$CS = \{\beta : \phi(\beta) = 0\}$$

  then CS has coverage $1 - \alpha$
  - Called test inversion
- Hence, to form an identification-robust confidence set, we only need to form identification-robust tests of $H_0 : \beta = \beta_0$

Restriction Implied By IV Model

- To implement test inversion, need to find a test
- To construct robust test, use restrictions that hold regardless of instrument strength
  - IV model implies that $\delta - \pi\beta = 0$
- Under $H_0 : \beta = \beta_0$,

$$\hat{\delta} - \hat{\pi}\beta_0 \sim N\left(0, \Omega\left(\beta_0\right)\right)$$

for

$$\Omega(\beta_0) = \Sigma_{\delta\delta} - \beta_0(\Sigma_{\delta\pi} + \Sigma_{\pi\delta}) + \beta_0^2\Sigma_{\pi\pi}$$

  - Holds regardless of instrument strength

AR Statistic

- Building on this observation, can introduce AR statistic

$$AR(\beta_0) = \left(\hat{\delta} - \hat{\pi}\beta_0\right)' \Omega(\beta_0)^{-1} \left(\hat{\delta} - \hat{\pi}\beta_0\right)$$

  - Originally introduced by Anderson and Rubin (1949) for homoskedastic normal case
  - Here, generalization to non-homoskedastic case
- Under $H_0 : \beta = \beta_0$, $AR(\beta_0) \sim \chi^2_k$ for all π
  - Recall $k = \dim(Z_i)$
- AR test $\phi_{AR}(\beta_0) = 1\left\{AR(\beta_0) > \chi^2_{k,1-\alpha}\right\}$
  - $\chi^2_{k,1-\alpha}$ the $1-\alpha$ quantile of a χ^2_k distribution
- AR Confidence set $CS_{AR} = \left\{\beta : AR(\beta) \leq \chi^2_{k,1-\alpha}\right\}$
- AR test and CS fully robust to weak instruments

The Form of AR Confidence Sets

- CS_{AR} can behave in counterintuitive ways
- In just-identified setting ($k = \dim(X) = 1$) can take form of:
  - bounded interval: $CS_{AR} = [a, b]$
  - real line: $CS_{AR} = (-\infty, \infty)$
  - real line excluding bounded interval: $CS_{AR} = (-\infty, a] \cup [b, \infty)$
- In over-identified settings can also be empty, $CS_{AR} = \emptyset$
  - In overidentifed non-homoskedastic settings, can take additional forms

The Form of AR Confidence Sets

- Infinite confidence sets strange-looking...
  - but have natural explanation
- Unboundedness consistent with Gleser and Hwang (1989), Dufour (1997)
- Moreover, can show that

$$\lim_{\beta_0 \to \pm\infty} AR(\beta_0) = \hat{\pi}' \Sigma_{\pi\pi}^{-1} \hat{\pi} = k \cdot F^R$$

  - Implies that CS_{AR} unbounded if and only if first-stage F-test cannot reject $\pi = 0$ at level α
- Unbounded AR confidence sets arise only when cannot reject that model totally unidentified

The Form of AR Confidence Sets

- Empty confidence sets more awkward
- Arise from fact that AR tests $H_0 : \delta = \pi\beta_0$. Can be decomposed into
  - Parameteric restriction $\beta = \beta_0$
  - Overidentifying restrictions $\delta \propto \pi$ if $k > 1$
    - If $k = 1$, no overidentifying restrictions to test
- Empty AR confidence sets can be interpreted as a rejection of the overidentifying restrictions
  - Unfortunate feature: how to interpret a small CS?
  - Confidence sets non-empty with probability one in just-identified case

Optimality of AR in Just-Identified Models

- In just-identified case with single endogenous regressor, AR is optimal
  - 101 out of 230 specifications in our AER sample are just-identified with a single endogenous regressor
- Moreira (2009) shows that AR test uniformly most powerful unbiased
- AR equivalent to two-sided t-test when instruments are strong
- In just-identified settings, strong case for using AR CS
  - Optimal among CS robust to weak instruments
  - No loss of power relative to t-test if instruments strong

AR Tests in Applications

- To examine practical impact of using CS_{AR}, return to our AER sample
- Limit attention to just-identified specifications with single endogenous regressor where can estimate variance-covariance matrix of $\left(\hat{\delta}, \hat{\pi}\right)$
  - Yields 36 specifications
- Comparing 95% t and AR confidence sets, find infinite AR CS in two cases. In remaining cases:
  - AR confidence sets 56.5% longer on average in all specifications
  - 20.3% longer on average in specifications that report F>10
  - 0.04% longer on average in specifications that report F>50

AR Tests in Overidentified Models

- Strong argument for using AR in just-identified settings
- AR tests and CS perform worse in over-identified settings
  - As already noted, CS_{AR} may be empty
- Also inefficient under strong instruments
  - Tests violations of $H_0 : \beta = \beta_0$, and of overidentifying restrictions
  - If only care about parametric restrictions, "wastes" degrees of freedom

Improving Efficiency in Over-identified Settings

- To obtain efficiency under strong instruments, need alternative tests
- For example, tests based on t-statistic

$$t\left(\beta_0\right) = \frac{\left|\hat{\beta} - \beta_0\right|}{\hat{\sigma}_{\hat{\beta}}}$$

- Problem: distribution of $t\left(\beta_0\right)$ under $H_0 : \beta = \beta_0$ depends on π
  - Already know this: distribution of t-statistic depends on instrument strength
  - Since π unknown, not clear what critical values to use with $t\left(\beta_0\right)$

Conditional Critical Values

- Moreira (2003): conditional critical values
  - Originally for homoskedastic case. Here discuss generalization
- Idea: Find a sufficient statistic $D(\beta_0)$ for π under $H_0 : \beta = \beta_0$
  - Means conditional distribution of $t(\beta_0) | D(\beta_0)$ doesn't depend on π under H_0
  - Once condition on $D(\beta_0)$, can compute data-dependent critical values $c_\alpha(D(\beta_0))$
- Question: how to find $D(\beta_0)$

Conditional Critical Values

- Idea for sufficient statistic: separate parts of $\left(\hat{\delta}, \hat{\pi}\right)$ that do/don't depend on π
- Define
$$g(\beta) = \hat{\delta} - \hat{\pi}\beta$$
- Let
$$D(\beta) = \hat{\pi} - (\Sigma_{\pi\delta} - \Sigma_{\pi\pi}\beta)\,\Omega(\beta)^{-1} g(\beta),$$
denote $\hat{\pi}$ orthogonalized with respect to $g(\beta)$
- Under $H_0 : \beta = \beta_0$
$$\left(\begin{array}{c} g(\beta_0) \\ D(\beta_0) \end{array}\right) \sim N\left(\left(\begin{array}{c} 0 \\ \pi \end{array}\right), \left(\begin{array}{cc} \Omega(\beta_0) & 0 \\ 0 & \Psi(\beta_0) \end{array}\right)\right)$$
- Conditional distribution of $g(\beta_0)$ given $D(\beta_0)$ doesn't depend on π under $H_0 : \beta = \beta_0$
$$g(\beta_0)\,|D(\beta_0) \sim N(0, \Omega(\beta_0))$$
    - $(g(\beta_0), D(\beta_0))$ one-to-one transformation of $\left(\hat{\delta}, \hat{\pi}\right)$
    - $\Rightarrow D(\beta_0)$ is sufficient statistic for π

Conditional Critical Values

- To construct conditional distribution of $t(\beta_0)\,|D(\beta_0)$:
  1. Fix $D(\beta_0)$ at observed value
  2. Repeatedly draw $g^*(\beta_0) \sim N(0, \Omega(\beta_0))$
  3. Construct $\left(\hat{\delta}^*, \hat{\pi}^*\right)$ from $\left(g^*(\beta_0), D(\beta_0)\right)$
  4. Calculate $t^*(\beta_0)$ based on $\left(\hat{\delta}^*, \hat{\pi}^*\right)$
- Conditional critical value $c_\alpha(D(\beta_0))$: $1-\alpha$ quantile of $t^*(\beta_0)$

Conditional Critical Values

- Conditional t-test that rejects when $t(\beta_0)$ exceeds $c_\alpha(D(\beta_0))$

$$\phi(\beta_0) = 1\{t(\beta_0) > c_\alpha(D(\beta_0))\}$$

  is fully robust to weak instruments,

$$E_{\beta_0,\pi}[\phi(\beta_0)] = \alpha \text{ for all } \pi$$

- Conditioning not specific to $t(\beta_0)$, works for any statistic $s(\beta_0)$
  - In each case construct data-dependent critical value $c_\alpha(D(\beta_0))$
  - Yields tests that control size
- Question what statistic $s(\beta_0)$ to use

Alternative Test Statistics

Many possible choices of statistic $s(\beta_0)$

- t-statistic

$$t(\beta_0) = \frac{\left|\hat{\beta} - \beta_0\right|}{\hat{\sigma}_{\hat{\beta}}}$$

- Score statistic (Kleibergen 2002, 2005)

$$K(\beta_0) = g(\beta_0)' \Omega(\beta_0)^{-1} D(\beta_0) \times$$

$$\left(D(\beta_0)' \Omega(\beta_0)^{-1} D(\beta_0)\right)^{-1} D(\beta_0)' \Omega(\beta_0)^{-1} g(\beta_0)$$

- AR statistic

$$AR(\beta_0) = g(\beta_0)' \Omega(\beta_0)^{-1} g(\beta_0)$$

- LR statistic

$$LR(\beta_0) = 2\left(\max_{\beta,\pi} \ell(\beta,\pi) - \max_{\pi} \ell(\beta_0,\pi)\right)$$

Properties

- Different test statistics imply different $c_\alpha(D(\beta_0))$
  - For t and LR, need data-dependent critical values
  - Conditional distributions $AR(\beta_0)|D(\beta_0)$ and $K(\beta_0)|D(\beta_0)$ don't depend on $D(\beta_0)$
    - Can use χ^2_k and χ^2_1 critical values, respectively
- Conditional t, K, and LR tests efficient under strong instruments
  - AR test inefficient in overidentified models
- Under weak instruments, also yield different power properties
  - Conditional two-sided t-test poor power
  - K test sometimes poor power
  - Conditional LR (CLR) test performs well (in homoskedastic case)
- See D. Andrews Moreira and Stock (2006), (2007) Plots

Near-Optimality of CLR Test

- D. Andrews, Moreira, and Stock (2006) show that CLR test near-optimal
  - In homoskedastic case with single endogenous regressor
- Power close to upper bound for a natural class of tests over a wide range of parameter values
- Consensus in literature that CLR is a good test for homoskedastic settings
  - ... but homoskedasticity assumption unappealing

Tests for Non-Homoskedastic Models

- Variety of CLR extensions for non-homoskedastic case
  - D. Andrews Moreira and Stock (2004), Kleibergen (2005), D. Andrews and Guggenberger (2015), I. Andrews (2016), I. Andrews and Mikusheva (2016)
  - All efficient with strong instruments, but only simulation evidence on power with weak instruments
- Alternative: tests proposed that maximize weighted average power
  - Maximize integral of power function with respect to some weights
  - Moreira and Moreira (2015), Montiel Olea (2017), Moreira and Ridder (2018)
  - Question: what are "right" weights?
- Many options, but so far no consensus on what tests should be used in over-identified and non-homoskedastic models
  - In just-identified setting, use AR
  - In over-identified settings, use something that's efficient under strong instruments

Two-Step Confidence Sets

- Robust confidence sets not widely used in practice
  - When reported, usually only after authors find evidence of weak instruments
  - In AER sample, reported in 2 papers. Minimal first-stage F of 2.3 and 6.3, respectively
- If only report robust confidence set when F small, can view as constructing confidence set in two steps
  - If $F \geq 10$, report t-statistic confidence set
  - If $F < 10$, report robust CS
- Screening applications on first-stage F can generate very bad behavior. Does two-step CS do the same?
  - Positive results for F^N in homoskedastic case based on Stock and Yogo (2005)
  - Negative result for F^N in non-homoskedastic case based on Montiel Olea and Pflueger (2013), for F^R with conventional critical values based on I. Andrews (2018)
    - Negative results based on extreme forms of non-homoskedasticity: open question how bad in practice

Implementation

- Implementations of some weak-IV tests are in Stata package weakiv, available on SSC
  - Finlay, Magnusson, and Schaffer
  - Versions of CLR, AR, K, and other tests applicable to non-homoskedastic models
  - Can be used with fixed effects, clustered standard errors, etc.
  - Stata Journal article on previous version of package: Finlay and Magnusson (2009)

Summary

- A number of tests and confidence sets are available that are fully robust to weak instruments
    - Avoid pretesting bias, discarding applications
    - Many efficient under strong instruments
- In just-identified models, strong case for using AR CS
    - Covers many applications
- In over-identified models, less clear
    - CLR if assume homoskedastic
    - No consensus for non-homoskedastic case
    - ...other than using something efficient under strong instruments

References

- Anderson T, Rubin H. 1949. Estimators for the parameters of a single equation in a complete set of stochastic equations. *Annals of Mathematical Statistics* 21:570–582
- Andrews D, Guggenberger P. 2015. Identification- and singularity-robust inference for moment condition models. Unpublished Manuscript
- Andrews D, Moreira M, Stock J. 2004. Optimal invariant similar tests of instrumental variables regression. Unpublished Manuscript
- Andrews D, Moreira M, Stock J. 2006. Optimal two-sided invariant similar tests of instrumental variables regression. *Econometrica* 74:715–752
- Andrews D, Moreira M, Stock J. 2007. Performance of conditional wald tests in iv regression with weak instruments . *Econometrica* 74:715–752
- Andrews I. 2016. Conditional linear combination tests for weakly identified models. *Econometrica* 84:2155–2182

References

- Andrews I, Mikusheva A. 2016. Conditional inference with a functional nuisance parameter. *Econometrica* 84:1571–1612.
- Dufour J. 1997. Some impossibility theorems in econometrics with applications to structural and dynamic models. *Econometrica* 65:1365–1387
- Finlay K, Magnusson L. 2009. Implementing weak-instrument robust tests for a general class of instrumental-variables models. *Stata Journal* 9, 398-421
- Geser L, Hwang J. 1987. The nonexistence of 100(1-α)% confidence sets of finite expected diameter in errors-in-variables and related models. *Journal of the American Statistical Association* 15:1341–1362
- Kleibergen F. 2002. Pivotal statistics for testing structural parameters in instrumental variables regression. *Econometrica* 70:1781–1803
- Kleibergen F. 2005. Testing parameters in gmm without assuming they are identified. *Econometrica* 73:1103–1123

References

- Montiel Olea J. 2017. Admissible, similar tests: A characterization. Unpublished Manuscript
- Montiel Olea J, Pflueger C. 2013. A robust test for weak instruments. *Journal of Business and Economic Statistics* 31:358–369
- Moreira H, Moreira M. 2015. Optimal two-sided tests for instrumental variables regression with heteroskedastic and autocorrelated errors. Unpublished Manuscript
- Moreira M. 2003. A conditional likelihood ratio test for structural models. *Econometrica* 71:1027–1048
- Moreira M. 2009. Tests with correct size when instruments can be arbitrarily weak. *Journal of Econometrics* 152:131–140
- Moreira M, Ridder G. 2017. Optimal invariant tests in an instrumental variables regression with heteroskedastic and autocorrelated errors. Unpublished Manuscript

Power Comparisons

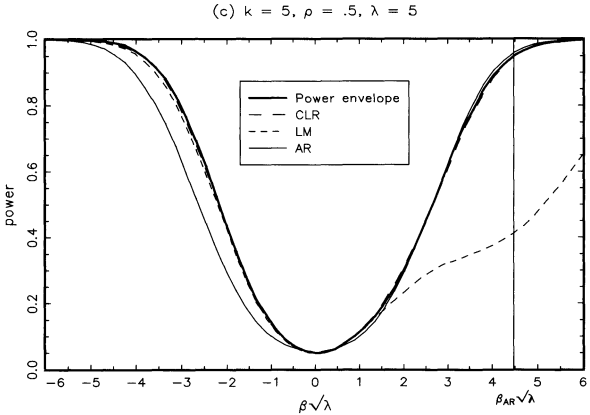


Figure: Power of AR, K, and C LR tests in homoskedastic case (from D. Andrews, Moreira, and Stock (2006))

Power Comparisons

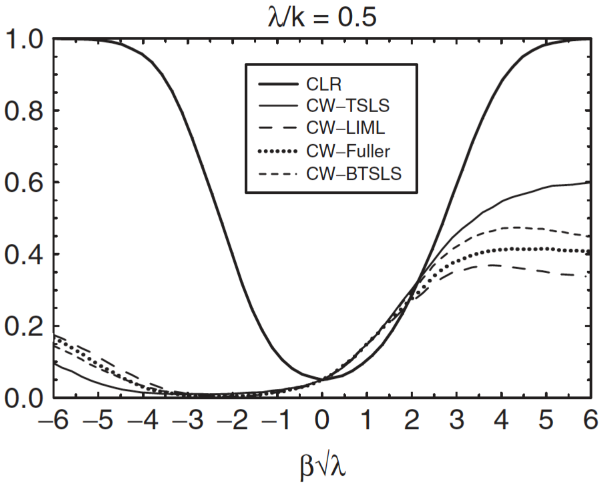


Figure: Power of Conditional t-and LR-tests in homokedastic (from D. Andrews, Moreira, and Stock (2007)) Return

4: Open Issues and Recent Research

July 22, 2018

Outline

Two goals for this section

1. Examine practical importance of issues covered so far
    - Simulations calibrated to specifications published in AER
2. Discuss other open questions and recent research on weak instruments

AER Simulation Specifications

- To assess practical importance of weak instrument issues, calibrate simulations to AER data
- Specifications from AER articles (excluding Papers and Proceedings) from 2014-2018 that:
  1. Published in main text
  2. Allow us to estimate variance matrix Σ of $\left(\hat{\delta}, \hat{\pi}\right)$
     - Mostly papers with replication data
     - In one other case, back out Σ from published results
- Yields 124 specifications from 8 papers
  - All specifications have a single endogenous variable

Simulation Design

- To focus attention on weak instrument issues, simulations use normal model

$$\begin{pmatrix} \hat{\delta} \\ \hat{\pi} \end{pmatrix} \sim N\left(\begin{pmatrix} \delta \\ \pi \end{pmatrix}, \Sigma\right)$$

  with Σ known, $\delta = \pi\beta$
- Simulations fix β, π, and Σ at estimated values
- Abstracts away from:
  - Non-normality of $\hat{\delta}$, $\hat{\pi}$
  - Estimation error in Σ
  - Will return to these later
- Any disortions must arise from weak instruments

Distribution of t-Statistics

- Theoretical results show t-tests can perform poorly when instruments weak
  - Distribution of t-statistics may not be centered at zero
  - Rejection probability of 5% t-tests may be much larger
- In each of our 124 AER specifications simulate
  - Median t-statistic

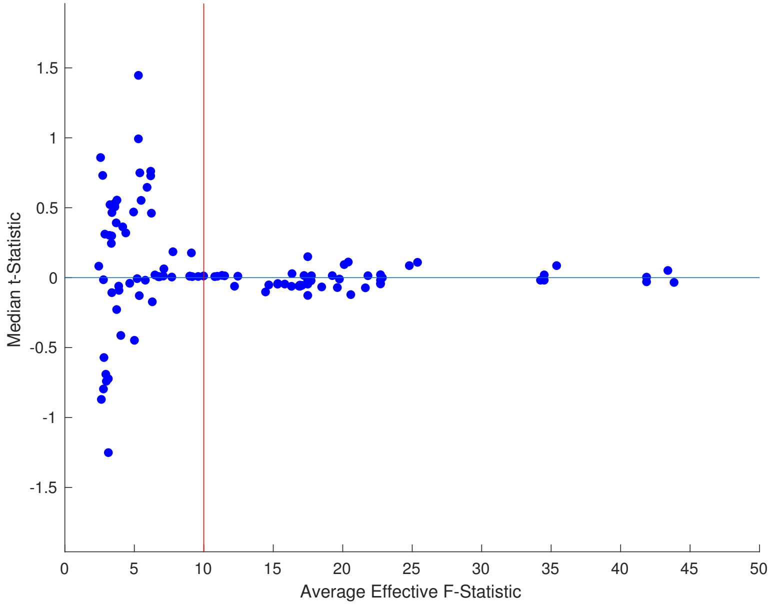
$$Med\left(\frac{\hat{\beta}-\beta_0}{\hat{\sigma}_{\hat{\beta}}}\right)$$

  - Size of 5% t-tests

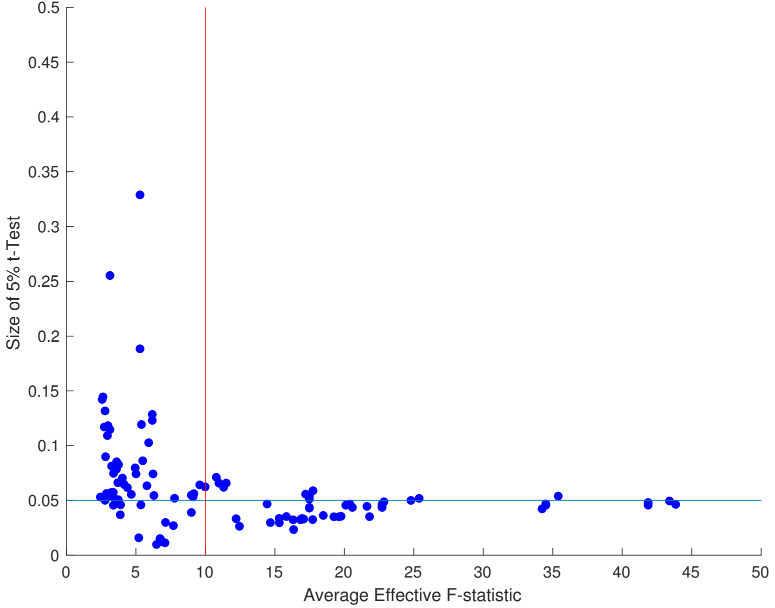
$$Pr\left\{\frac{\left|\hat{\beta}-\beta_0\right|}{\hat{\sigma}_{\hat{\beta}}} > 1.96\right\}$$

- Plot against average effective first-stage F-stat
  - Little action for $E\left[F^E\right] > 50$. Limit plot to $E\left[F^E\right] \leq 50$
  - Includes 106 of 124 specifications

Distribution of t-Statistics



Distribution of t-Statistics



Distribution of t-Statistics

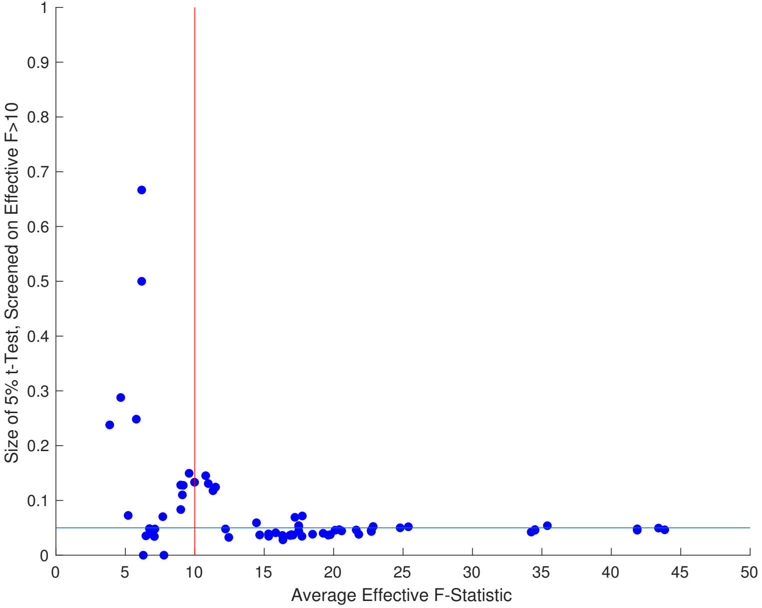
- Weak instrument issues apparent in some specifications
    - Median t-statistic far from zero
    - Size of nominal 5% t-test much larger than 5%
- Problems limited to specifications with a small average effective F-stat
    - No large distortions in specifications with $E\left[F^E\right] > 10$
    - Population rule of thumb seems to work pretty well
        - Not a theorem!

Weak instrument issues appear relevant for some recently-published specifications, but only in cases with $E\left[F^E\right]$ small

Screening on F-Statistics

- Given that average effective F-statistics seem to capture weak instrument issues, tempting to screen applications on F
  - e.g. only pursue applications with $F^E > 10$
- Distribution of F-statistics in AER sample suggests may be common
- As already discussed, can introduce size distortions
- Examine effect in our AER specifications
  - For each specification, calculate size of 5% t-test conditional on $F^E > 10$

Screening on F-Statistics



Screening on F-Statistics

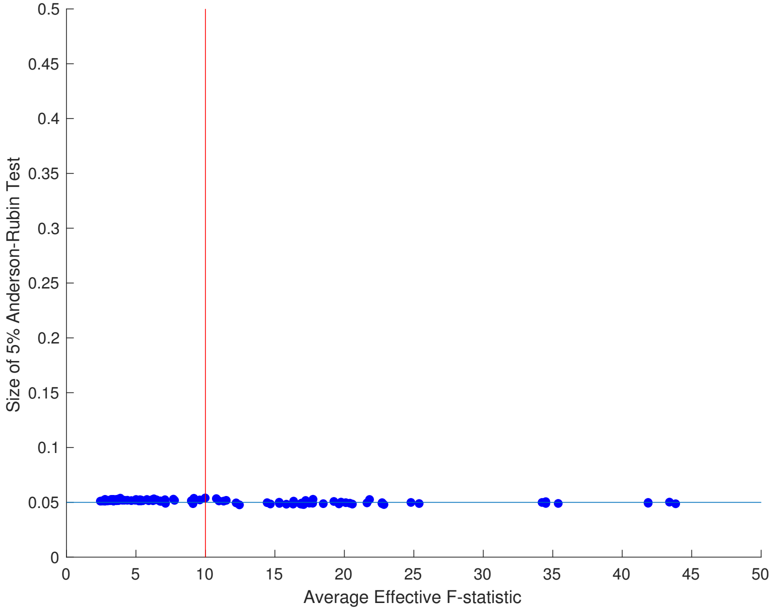
- Screening leads to much larger size for some specifications
  - Not specific to F^E, same issues appear for F^N, F^R
  - Not specific to threshold of 10: if move threshold, get distortions in neighborhood of new cutoff

Screening on F-statistics can make published results less reliable

Robust Confidence Sets

- Rather than screening applications on F^E, can compute robust confidence sets
  - Guaranteed to have correct coverage regardless of instrument strength
- For illustration, here consider Anderson-Rubin (AR)
  - Plot size of AR tests in AER specifications

Robust Confidence Sets



Robust Confidence Sets

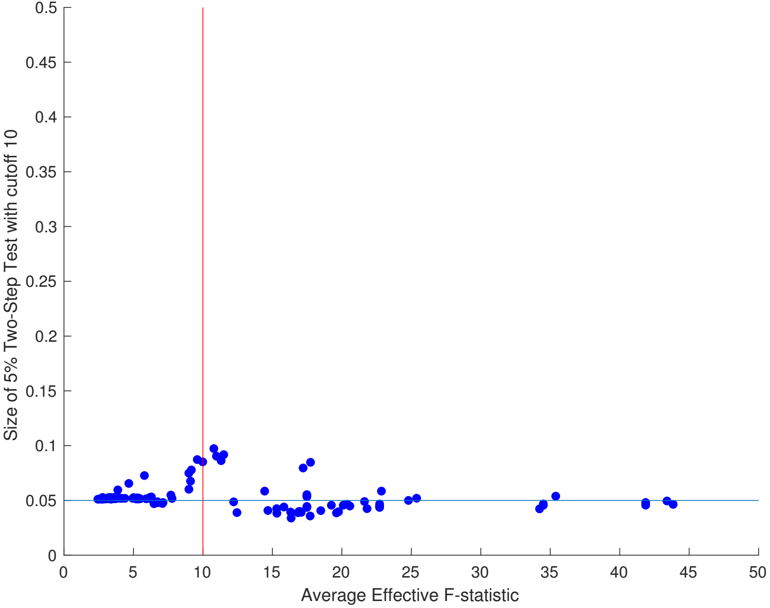
- AR size is flat at 5% regardless of instrument strength
  - AR also efficient in just-identified case
- For over-identified models, variety of robust tests and confidence sets available
  - Many ensure efficiency in strongly-identified case
  - All again ensure correct size regardless of instrument strength

Robust confidence sets eliminate size distortions from weak instruments

Two-Step Confidence Sets

- Robust confidence sets currently little-used in practice
  - When used, often because weak identification is suspected
- When used in this way, can be viewed as alternative to screening on F-statistic. For example
  - If $F^E \geq 10$, report t-statistic
  - If $F^E < 10$, report AR
- Alternatively, could use Montiel Olea and Pflueger (2013) critical values
- May introduce size distortions, but not clear how large
  - Examine performance in our AER specifications

Two-Step Confidence Sets



Two-Step Confidence Sets

- Observe some size distortions for specifications with $E\left[F^E\right] \approx 10$
  - True size never above 10% in these simulations
  - Results similar if instead use Montiel Olea and Pflueger (2013) critical values
- Also not a theorem!

Deciding to use a robust confidence set based on F^E isn't theoretically guaranteed to work, but cost appears small in our AER specifications

Summary of Simulation Results

1. Weak instruments appear to be a problem in some published specifications
2. Bad behavior largely limited to specifications with $E\left[F^E\right] < 10$
3. Screening on F^E can amplify problems
4. Robust confidence sets eliminate size distortions
5. Choosing whether to report robust CS based on F^E introduces some distortions, but small

Questions from Simulations: Performance of F-Statistics

Simulation results suggest some questions

1. Theoretical justification for F^E in Montiel Olea and Pflueger (2013) only concerns bias. Appears to also diagnose size problems reasonably well. Can this be formalized?
2. Two-step confidence sets based on F^E work reasonably well in simulations. Can this be formalized?

Questions from Simulations: Normal Approximation

- Our simulations take $\left(\hat{\delta}, \hat{\pi}\right)$ to be normally distributed with known variance
  - Focuses attention solely on distortions from weak instruments
- Results from Young (2018) suggest may be problematic
  - Based on sample of papers from AEA journals (larger than, but overlaps with, our AER sample)
- Young (2018) finds that
  1. A small number of observations have a large influence on estimates and p-values
  2. Variance estimates $\hat{\Sigma}$ often extremely noisy in simulation
  3. As a result of noisy estimates $\hat{\Sigma}$, AR tests can have large size distortions in over-identified settings
- Further exploration of interaction between weak instruments and issues discussed by Young (2018) of considerable interest

Other Research: Subvector Inference

- Some applications have more than one endogenous regressor
  - 19 out of 230 specifications in AER sample
- Most tests previously discussed extend to tests of $\dim(X) \times 1$ vector β in settings with multiple endogenous variables
  - Imply joint confidence sets for full vector
- Joint confidence sets rarely reported in strong-instrument settings. Instead, usually report e.g. estimates and standard errors for each element of β separately
  - Write $\beta = (\beta_1, \beta_2)$, and want confidence set for β_1 alone
- If assume instruments strong for β_2, simple solution
  - "Strong for β_2" meaning strong if treat β_1 as known
  - Plug appropriate estimate $\hat{\beta}_2(\beta_1)$ into robust test statistics (see e.g. Stock and Wright (2000))
- If instruments weak for β_2, hard problem

Other Research: Subvector Inference

- One option projection method
  - Form joint confidence set for (β_1, β_2), and collect implied set of values for β_1
  - Can have very low power
- Several recent papers seeking to improve power of projection method
  - Smaller critical values for AR statistic in homoskedastic case: Guggenberger et al. (2012)
  - Modified projection approach to improve power in well-identified case: Chaudhuri and Zivot (2011), D. Andrews (2017)
- Active area of research

Subvector Inference: Implementation

- Stata package weakiv can
  - Compute joint confidence sets for β
  - Plug in estimates for strongly identified β_2
  - Implement projection method
- Package twostepweakiv (also on SSC/Github) implements refined projection based on Chaudhuri and Zivot (2011)
  - Stata Journal article: Sun (Forthcoming)
  - Nearly-efficient inference on β_1 under strong identification
  - Also implements two-step CS with guaranteed coverage

Other Research: Nonlinear Models

- All the results discussed for IV apply directly to linear GMM
  - GMM moments linear in parameters
- Many (though not all) generalize to nonlinear GMM
  - No known analog of first-stage F-statistic
    - Alternative for approach detecting weak identification: I. Andrews (2018)
  - Many procedures for robust inference, e.g. Stock and Wright (2000), Kleibergen (2005), D. Andrews and Guggenberger (2015), I. Andrews and Mikusheva (2016)

The End

Thank you!

References

- Andrews D. 2017. Identification-robust subvector inference. Unpublished Manuscript
- Andrews D, Guggenberger P. 2015. Identification- and singularity-robust inference for moment condition models. Unpublished Manuscript
- Andrews I. 2018. Valid two-step identification-robust confidence sets for gmm. *Review of Economics and Statistics* 100:337–348
- Andrews I, Mikusheva A. 2016. Conditional inference with a functional nuisance parameter. *Econometrica* 84:1571–1612.
- Chaudhuri S, Zivot E. 2011. A new method of projection-based inference in gmm with weakly identified nuisance parameters. *Journal of Econometrics* 164:239–251
- Guggenberger P, Kleibergen F, Mavroeidis S, Chen L. 2012. On the asymptotic sizes of subset anderson–rubin and lagrange multiplier tests in linear instrumental variables regression. *Econometrica* 80:2649–2666

References

- Kleibergen F. 2005. Testing parameters in gmm without assuming they are identified. *Econometrica* 73:1103–1123
- Montiel Olea J, Pflueger C. 2013. A robust test for weak instruments. *Journal of Business and Economic Statistics* 31:358–369
- Sun L. Forthcoming. Implementing valid two-step identification-robust confidence sets for linear instrumental-variables models. *Stata Journal*
- Stock J, Wright J. 2000. Gmm with weak identification. *Econometrica* 68:1055–1096
- Young A. 2018. Consistency without inference: Instrumental variables in practical application. Unpublished Manuscript