EC708 Discussion 2
Weak IV

Yan Liu

Department of Economics
Boston University

January 28, 2022

Outline

1. Toy Example: Totally Irrelevant Instruments
2. Weak Instruments Asymptotics
3. Detecting Weak Instruments
4. Robust Inference against Weak Instruments

Throughout, suppose one observes i.i.d. data. For a random scalar or vector w_t, let $W = (w_1, \ldots, w_T)'$.

Toy Example: Totally Irrelevant Instruments

Consider the linear IV model:

$$y_t = x_t\beta + u_t,$$
$$x_t = z_t\pi + v_t,$$

where x_t and z_t are scalars. Suppose

- $\pi = 0$;
- conditional homoskedasticity:

$$\mathsf{Var}\left(\begin{pmatrix} u_t \\ v_t \end{pmatrix} \middle| z_t \right) = \begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{pmatrix}.$$

Toy Example: Totally Irrelevant Instruments

OLS estimator is inconsistent:

$$\hat{\beta}_{OLS} - \beta = \frac{\frac{1}{T}\sum_{t=1}^{T} u_t v_t}{\frac{1}{T}\sum_{t=1}^{T} v_t^2} \xrightarrow{p} \frac{\sigma_{uv}}{\sigma_v^2} \neq 0.$$

On the other hand, by CLT,

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \begin{pmatrix} z_t u_t \\ z_t v_t \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \xi_u \\ \xi_v \end{pmatrix} \sim N\left(0, E(z_t^2)\begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{pmatrix}\right).$$

Toy Example: Totally Irrelevant Instruments

We can decompose $\xi_u = \frac{\sigma_{uv}}{\sigma_v^2}\xi_v + \xi$ with $\xi \perp\!\!\!\perp \xi_v$, and

$$\hat{\beta}_{IV} - \beta = \frac{\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t u_t}{\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t v_t} \xrightarrow{d} \frac{\xi_u}{\xi_v} = \frac{\sigma_{uv}}{\sigma_v^2} + \frac{\xi}{\xi_v}.$$

What is the distribution of $\frac{\xi}{\xi_v}$?

- Conditional on ξ_v, $\frac{\xi}{\xi_v}|\xi_v \sim N\left(0, \frac{\mathsf{Var}(\xi)}{\xi_v^2}\right)$.
- Hence, the unconditional distribution of $\frac{\xi}{\xi_v}$ is

$$\frac{\xi}{\xi_v} \sim \int N\left(0, \frac{\mathsf{Var}(\xi)}{\xi_v^2}\right) f(\xi_v) d\xi_v.$$

$\Rightarrow$ A random mixture of normal distributions (Cauchy distribution)

Toy Example: Totally Irrelevant Instruments

Observations:

- $\hat{\beta}_{IV}$ is inconsistent and centered around the limit of $\hat{\beta}_{OLS}$.
- Asymptotically $\hat{\beta}_{IV}$ has heavy tails.
- Under relevant ($\pi \neq 0$) but weak ($\sqrt{T}\pi \to \alpha \in [0, \infty)$) instruments, the asymptotic distribution of $\hat{\beta}_{IV}$ is between normal and Cauchy distributions.

Weak Instruments Asymptotics

Staiger and Stock (1997)

Consider the linear IV model:

$$
\begin{aligned}
y_t &= x_t\beta + u_t, \\
x_t &= z_t'\pi + v_t,
\end{aligned}
$$

where x_t is scalar and z_t is $k \times 1$. Suppose

- $\pi = C/\sqrt{T}$, where C is a fixed $k \times 1$ vector;
- $\left(\frac{U'U}{T}, \frac{U'V}{T}, \frac{V'V}{T}\right) \xrightarrow{p} (\sigma_u^2, \sigma_{uv}, \sigma_v^2)$, $\frac{Z'Z}{T} \xrightarrow{p} Q_{zz}$;
- $\left(\frac{Z'U}{\sqrt{T}}, \frac{Z'V}{\sqrt{T}}\right) \xrightarrow{d} (z_u, z_v)$, where $(z_u, z_v) \sim N\left(0, \begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{pmatrix} \otimes Q_{zz}\right)$.
  (conditional homoskedasticity)

Weak Instruments Asymptotics

Staiger and Stock (1997)

We have

$$\hat{\beta}_{2SLS} - \beta = \frac{X'P_Z U}{X'P_Z X} \quad (P_Z = Z(Z'Z)^{-1}Z' \quad \text{“projection matrix”})$$
$$= \left[\frac{X'Z}{\sqrt{T}}\left(\frac{Z'Z}{T}\right)^{-1}\frac{Z'X}{\sqrt{T}}\right]^{-1}\frac{X'Z}{\sqrt{T}}\left(\frac{Z'Z}{T}\right)^{-1}\frac{Z'U}{\sqrt{T}}.$$

Note that

$$\frac{Z'X}{\sqrt{T}} = \frac{Z'(Z \cdot C/\sqrt{T} + V)}{\sqrt{T}} = \frac{Z'Z}{T}C + \frac{Z'V}{\sqrt{T}} \xrightarrow{d} Q_{zz}C + z_v.$$

Hence,

$$\hat{\beta}_{2SLS} - \beta \xrightarrow{d} \frac{(Q_{zz}C + z_v)' Q_{zz}^{-1} z_u}{(Q_{zz}C + z_v)' Q_{zz}^{-1} (Q_{zz}C + z_v)}.$$

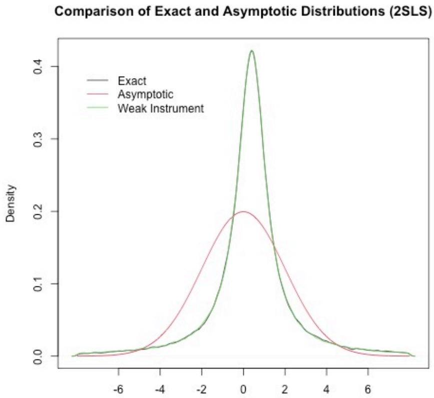
- We can simulate it.
- We cannot use it for inference because the estimation of σ_u^2 and σ_{uv} depends on $\hat{\beta}_{2SLS}$, which is inconsistent:

$$\hat{\sigma}_u^2 = \frac{\hat{U}'\hat{U}}{T}, \quad \hat{\sigma}_{uv} = \frac{\hat{U}'\hat{V}}{T}, \quad \hat{u}_t = y_t - x_t \hat{\beta}_{2SLS}.$$

Weak Instruments Asymptotics

A Simulation

$$T = 100, C = .5, \beta = 0, z_t \overset{i.i.d.}{\sim} N(0,1),$$
$$\begin{pmatrix} u_t \\ v_t \end{pmatrix} \overset{i.i.d.}{\sim} N\left(0, \begin{pmatrix} 1 & .5 \\ .5 & 1 \end{pmatrix}\right), z_t \perp\!\!\!\perp (u_t, v_t).$$



Suppose $\pi = C \cdot T^{-1/2+\kappa}$ with $\kappa < 0$. Then,

$$\hat{\beta}_{2SLS} - \beta = \left[\frac{X'Z}{\sqrt{T}} \left(\frac{Z'Z}{T} \right)^{-1} \frac{Z'X}{\sqrt{T}} \right]^{-1} \frac{X'Z}{\sqrt{T}} \left(\frac{Z'Z}{T} \right)^{-1} \frac{Z'U}{\sqrt{T}},$$

where

$$\frac{Z'X}{\sqrt{T}} = \frac{Z'(Z \cdot C \cdot T^{-1/2+\kappa} + V)}{\sqrt{T}} = \frac{Z'Z}{T} C \cdot T^{\kappa} + \frac{Z'V}{\sqrt{T}} \xrightarrow{d} z_v.$$

Hence,

$$\hat{\beta}_{2SLS} - \beta \xrightarrow{d} \frac{z_v' Q_{zz}^{-1} z_u}{z_v' Q_{zz}^{-1} z_v}.$$

Weak Instruments Asymptotics

Other Cases

Suppose $\pi = C \cdot T^{-1/2+\kappa}$ with $\kappa > 0$. Then,

$$T^{\kappa}(\hat{\beta}_{2SLS} - \beta) = \left[\frac{X'Z}{T^{1/2+\kappa}}\left(\frac{Z'Z}{T}\right)^{-1}\frac{Z'X}{T^{1/2+\kappa}}\right]^{-1}\frac{X'Z}{T^{1/2+\kappa}}\left(\frac{Z'Z}{T}\right)^{-1}\frac{Z'U}{\sqrt{T}},$$

where

$$\frac{Z'X}{T^{1/2+\kappa}} = \frac{Z'(Z \cdot C \cdot T^{-1/2+\kappa} + V)}{T^{1/2+\kappa}} = \frac{Z'Z}{T}C + \frac{Z'V}{T^{1/2+\kappa}} \xrightarrow{p} Q_{zz}C.$$

Hence,

$$T^{\kappa}(\hat{\beta}_{2SLS} - \beta) \xrightarrow{d} \frac{C'z_u}{C'Q_{zz}C} \sim N(0, (C'Q_{zz}C)^{-1}).$$

Detecting Weak Instrument

Stock and Yogo (2005): Homoskedastic Errors

Stock and Yogo (2005) provide two characterizations of a weak instrument set:

1. The squared bias of $\hat{\beta}_{2SLS}$ relative to the squared bias of $\hat{\beta}_{OLS}$ exceeds a certain threshold b, for example $b = 10\%$;
2. The conventional α-level Wald test based on $\hat{\beta}_{2SLS}$ has an actual size that exceeds a certain threshold r, for example $r = 15\%$ when $\alpha = 5\%$;

They develop tests for the null hypothesis that π lies in the weak instrument set in cases with homoskedastic errors.

Detecting Weak Instrument

Stock and Yogo (2005): Homoskedastic Errors

With a single endogenous regressor, Stock and Yogo (2005)'s test reduces to the first-stage F-statistic:

$$F = \frac{\hat{\pi}'[\hat{\sigma}_v^2(Z'Z)^{-1}]^{-1}\hat{\pi}}{k} = \frac{\hat{\pi}'Z'Z\hat{\pi}'}{\hat{\sigma}_v^2}\frac{1}{k}.$$

$\pi'Z'Z\pi/\sigma_v^2$ is called the concentration parameter (Rothenberg, 1984).

Detecting Weak Instrument

Stock and Yogo (2005): Homoskedastic Errors

The critical values depend on the number of instruments and how the weak instrument set is characterized.

- If we define instruments as weak when the worst-case bias of $\hat{\beta}_{2SLS}$ exceeds 10% of the worst-case bias of $\hat{\beta}_{OLS}$:
  for $3 \sim 30$ instruments, the critical value for a 5% test is $9 \sim 11.52$, which is close to the Staiger and Stock (1997) rule of thumb cutoff of 10.
- If we define instruments as weak when the worst-case size of a nominal 5% Wald test based on $\hat{\beta}_{2SLS}$ exceeds 15%:
  the critical value depends strongly on the number of instruments
  - a single instrument: 8.96;
  - 30 instruments: 44.78.

Detecting Weak Instrument

Stock and Yogo (2005): Homoskedastic Errors

With multiple endogenous regressors, Stock and Yogo (2005)'s test is based on the Cragg and Donald (1993) statistic:

$$g_{\min} = \text{mineval}\left(\frac{\hat{\pi}' Z' Z \hat{\pi}'}{\hat{\sigma}_v^2} \frac{1}{k}\right),$$

where mineval denotes the minimum eigenvalue.

Detecting Weak Instrument

Non-Homoskedastic Errors

- When run without assuming homoskedastic errors, the `ivreg2` command in Stata automatically reports a robust F-statistic $F^R = \frac{\hat{\pi}'\hat{\Sigma}_{\pi\pi}^{-1}\hat{\pi}}{k}$ with Stock and Yogo (2005) critical values. (not justified!)
- Montiel Olea and Pflueger (2013) propose using the effective first-stage F-statistic:

$$F^{Eff} = \frac{\hat{\pi}'Z'Z\hat{\pi}'}{\mathrm{tr}(\hat{\Sigma}_{\pi\pi}Z'Z)}.$$

  - In cases with homoskedastic errors, F^{Eff} reduces to F.
  - When $k = 1$, $F^{Eff} = F^R$.
  - Stata package `weakivtest` implements this test.

Robust Inference against Weak Instruments

Test Inversion

Idea: Given a size-α test of $H_0 : \beta = \beta_0$, we can construct a level $1-\alpha$ confidence set for β by collecting the set of non-rejected values.

- Represent the test by $\phi(\beta_0)$ with $\phi(\beta_0) = 1$ if H_0 is rejected and $\phi(\beta_0) = 0$ otherwise.
- $\phi(\beta_0)$ is a size-α test of $H_0 : \beta = \beta_0$ if

$$\sup_{\pi} \mathbb{E}_{\beta_0,\pi}[\phi(\beta_0) = 1] \leq \alpha.$$

- $CS = \{\beta : \phi(\beta) = 0\}$ is a level $1-\alpha$ confidence set if

$$\inf_{\beta,\pi} \mathbb{P}_{\beta,\pi}\{\beta \in CS\} \geq 1-\alpha.$$

We focus on the homoskedastic case with a single endogenous regressor.

Anderson-Rubin statistic:

$$AR(\beta) = \frac{(Y - X\beta)' P_Z (Y - X\beta)}{(Y - X\beta)' M_Z (Y - X\beta)/(T - k)},$$

where $M_Z = I - P_Z$ (annihilator matrix).

Under $H_0 : \beta = \beta_0, Y - X\beta = U$, and so

$$AR(\beta) \xrightarrow{d} \frac{z_u' Q_{zz}^{-1} z_u}{\sigma_u^2} \sim \chi_k^2.$$

We can form a size-α test and a level $1 - \alpha$ confidence set as

$$\phi_{AR}(\beta_0) = 1\{AR(\beta_0) > \chi^2_{k,1-\alpha}\}, \quad CS_{AR} = \{\beta : AR(\beta) \leq \chi^2_{k,1-\alpha}\},$$

where $\chi^2_{k,1-\alpha}$ is the $1 - \alpha$ quantile of a χ_k^2 distribution.

In **just-identified** models:

- CS_{AR} can take one of three forms:
  1. $[a, b]$,
  2. $(-\infty, a] \cup [b, \infty)$,
  3. the real line $(-\infty, \infty)$
     (iff a robust F-test cannot reject $\pi = 0 \Rightarrow \beta$ is totally unidentified).
- AR test is unbiased: $\mathbb{E}_{\beta,\pi}[\phi_{AR}(\beta_0)] \geq \alpha$ for all $\beta \neq \beta_0$ and all π.
- AR test has (weakly) higher power than any other size-α unbiased test. In other words, the AR test is efficient. (Moreira, 2009)
  - In particular, the AR test does not sacrifice power relative to t-test even when instruments are strong.

In **over-identified** models:

- $H_0 : \beta = \beta_0$ could also fail because the IV model's over-identifying restrictions fail, e.g. invalid instruments.
- CS_{AR} can take one of four forms:
  1. $[a, b]$,
  2. $(-\infty, a] \cup [b, \infty)$,
  3. the real line $(-\infty, \infty)$,
  4. empty set (rejection of over-identifying restrictions).
- AR test is inefficient under strong instruments
  - Use k degrees of freedom for one parameter $\Rightarrow$ loss of power

Robust Inference against Weak Instruments

K-statistic

Kleibergen (2002) proposes

$$K(\beta) = \frac{(Y - X\beta)' P_{\tilde{X}(\beta)} (Y - X\beta)}{(Y - X\beta)' M_Z (Y - X\beta)/(T - k)},$$

where $\underset{n\times 1}{\tilde{X}(\beta)} = \underset{n\times k}{Z}\,\underset{k\times 1}{\tilde{\pi}(\beta)}$ such that under $H_0 : \beta = \beta_0$,

- $\tilde{\pi}(\beta)$ is a consistent estimator of π;
- $\tilde{\pi}(\beta)$ is asymptotically independent of $(Y - X\beta_0)'Z$.

Hence, under $H_0 : \beta = \beta_0$,

$$K(\beta) \xrightarrow{d} \frac{z_u' \pi (\pi' Q_{zz} \pi)^{-1} \pi' z_u}{\sigma_u^2} \sim \chi_1^2$$

- In case of just-identification, $K(\beta)$ is identical to the $AR(\beta)$.
- In case of over-identification, $K(\beta)$ takes a linear combination of $(Y - X\beta_0)'Z$ to reduce dimension and improve power.
- **Caveat:** $K(\beta) = 0$ has an extraneous root, leading to non-monotonic power and disconnected confidence intervals in finite samples.

Consider the reduced-form of the structural IV model:

$$y_t = z_t'\pi\beta + \varepsilon_t,$$
$$x_t = z_t'\pi + v_t.$$

Assume reduced-form errors $(\varepsilon_t, v_t) \sim N(0, \Omega)$.

- When Ω is known, the sufficient statistic for the conditional distribution of (Y, X) given Z can be represented by (S, T), where $S \perp\!\!\!\perp T$ and under $H_0 : \beta = \beta_0, T$ depends on π while S does not.
- A test statistic $\psi(S, T, \Omega, \beta_0)$ may depend on π through T
  - Find the largest possible $1 - \alpha$ quantile over some set of π (conservative)
  - Use conditional critical values $c_\psi(T, \Omega, \beta_0, \alpha) \Rightarrow$ **conditional tests**

Robust Inference against Weak Instruments

Conditional Tests

- In fact, any test that has exact size α for all π is a conditional test.
- Let $S = Z'(Y - X\beta_0)$. AR and K-statistics for known Ω are

$$\begin{aligned} AR(\beta_0) &= S'(Z'Z)^{-1}S/\sigma_0^2, \\ K(\beta_0) &= S'\tilde{\pi}(\tilde{\pi}'Z'Z\tilde{\pi})^{-1}\tilde{\pi}'S/\sigma_0^2, \end{aligned}$$

  where $\sigma_0^2 = \mathrm{Var}(y_t - x_t\beta_0)$.
- Both are pivotal and their critical value functions collapse to constants.

Among conditional tests, Moreira (2003) proposes to use the conditional likelihood ratio (CLR) test:

$$LR(\beta_0) = \overline{S}'\overline{S} - \text{mineval}((\overline{S}, \overline{T})'(\overline{S}, \overline{T})),$$

where $\overline{S}$ and $\overline{T}$ are standardized sufficient statistics.

- When $k = 1$, $LR(\beta_0)$ collapses to $AR(\beta_0) = \overline{S}'\overline{S}$.
- CLR test is preferable because it (in terms of power)
  - dominates AR and K-statistics under weak-instrument asymptotics;
  - is optimal under usual asymptotics.
- CLR test is implemented in Stata (command `condivreg`).

The literature has not yet converged on a recommendation in case of

- non-homoskedastic errors
- multiple endogenous regressors $\Rightarrow$ inference on subsets of parameters

See Andrews et al. (2019) for a survey.

Bibliography

Andrews, I., Stock, J. H., and Sun, L. (2019), "Weak instruments in instrumental variables regression: Theory and practice," *Annual Review of Economics*, 11, 727–753.

Cragg, J. G. and Donald, S. G. (1993), "Testing identifiability and specification in instrumental variable models," *Econometric Theory*, 222–240.

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

Montiel Olea, J. L. and Pflueger, C. (2013), "A robust test for weak instruments," *Journal of Business & Economic Statistics*, 31, 358–369.

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

— (2009), "Tests with correct size when instruments can be arbitrarily weak," *Journal of Econometrics*, 152, 131–140.

Rothenberg, T. J. (1984), "Approximating the distributions of econometric estimators and test statistics," *Handbook of econometrics*, 2, 881–935.

Staiger, D. and Stock, J. H. (1997), "Instrumental variables regression with weak instruments," *Econometrica: journal of the Econometric Society*, 557–586.

Stock, J. H. and Yogo, M. (2005), "Testing for weak instruments in linear IV regression," *Identification and inference for econometric models: Essays in honor of Thomas Rothenberg*, 80, 1.