EC708 Discussion 2
Weak IV

Yan Liu

Department of Economics
Boston University

February 5, 2021

Outline

1. Problems of IV and 2SLS Estimators with Weak Instruments
2. Weak Instruments Asymptotics
3. Detecting Weak Instruments
4. Robust Inference against Weak Instruments

Notational remark: For a random scalar or vector w_t, let $w = (w_1, \ldots, w_T)'$.

Problems with Weak Instruments

Toy Example: Totally Irrelevant Instruments

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

where x_t and Z_t are scalars. Suppose

- one observes i.i.d. data $\{(y_t, x_t, Z_t)\}$;
- $\pi = 0$;
- conditional homoskedasticity:

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

Problems with Weak Instruments

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,

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

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

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

where $\frac{\xi}{\xi_v}$ has Cauchy distribution.

Problems with Weak Instruments

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.
- The t-statistic is

$$t_{\hat{\beta}_{IV}} = \frac{\hat{\beta}_{IV} - \beta}{\sqrt{\hat{\sigma}_u^2 \sum_{t=1}^T Z_t^2 / (\sum_{t=1}^T Z_t x_t)^2}}$$

where

$$\hat{\sigma}_u^2 = \frac{1}{T} \sum_{t=1}^{T} (y_t - \hat{\beta}_{IV} x_t)^2 \xrightarrow{d} \sigma_u^2 - 2\sigma_{uv} \frac{\xi_u}{\xi_v} + \sigma_v^2 \left(\frac{\xi_u}{\xi_v} \right)^2 .$$

As $\frac{\sigma_{uv}}{\sigma_u \sigma_v} \to 1$, $\frac{\xi_u}{\xi_v} \to \frac{\sigma_u}{\sigma_v}$, $\hat{\sigma}_u^2 \xrightarrow{d} 0$, and $t_{\hat{\beta}_{IV}} \to \infty$. (rejects very often!)

Talk about 2SLS:

$$y_t = x_t\beta + u_t,$$
$$x_t = Z_t'\pi + v_t,$$

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

- one observes i.i.d. data $\{(y_t, x_t, Z_t)\}$;
- conditional homoskedasticity:

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

Let $P_Z = Z(Z'Z)^{-1}Z'$. Then

$$\hat{\beta}_{2SLS} - \beta = \frac{x'P_Z u}{x'P_Z x} = \frac{\pi'Z'u + v'P_Z u}{\pi'Z'Z\pi + 2\pi'Z'v + v'P_Z v}.$$

Define the concentration parameter $\mu^2 = \pi'Z'Z\pi/\sigma_v^2$ (Rothenberg, 1984).

Assume instruments are fixed and errors are normal. Then

$$\mu(\hat{\beta}_{2SLS} - \beta) = \frac{\sigma_u}{\sigma_v}\frac{\xi_u + S_{uv}/\mu}{1 + 2\xi_v/\mu + S_{vv}/\mu^2},$$

where ξ_u, ξ_v are standard normal, and S_{uv}, S_{vv} are Wishart distributed. $\left(\xi_u = \frac{\pi'Z'u}{\sigma_u\sqrt{\pi'Z'Z\pi}}, \xi_v = \frac{\pi'Z'v}{\sigma_v\sqrt{\pi'Z'Z\pi}}, S_{uv} = \frac{v'P_Z u}{\sigma_u\sigma_v}, S_{vv} = \frac{v'P_Z v}{\sigma_v^2}\right)$.

Problems with Weak Instruments

Concentration Parameter

μ^2 plays the role of the sample size.

- If μ is large, $\mu(\hat{\beta}_{2SLS} - \beta)$ is well approximated by $N(0, \frac{\sigma_u^2}{\sigma_v^2})$
  $\Rightarrow$ "strong instruments asymptotics":
  $\mu \to \infty$ as $T \to \infty$;
- if μ is small, $\mu(\hat{\beta}_{2SLS} - \beta)$ is a non-normal random variable
  $\Rightarrow$ "weak instruments asymptotics":
  $\pi = C/\sqrt{T}$ so that $\mu = C' \left(\frac{Z'Z}{T}\right) C \xrightarrow{p} C' Q_{ZZ} C$.

Weak Instruments Asymptotics

Staiger and Stock (1997)

$$y_t = x_t\beta + u_t,$$
$$x_t = Z_t'\pi + v_t,$$

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)

$$\begin{aligned}
\hat{\beta}_{2SLS} - \beta &= \frac{x' P_Z u}{x' P_Z x} \\
&= \left[\frac{x'Z}{\sqrt{T}} \left(\frac{Z'Z}{T} \right)^{-1} \frac{Z'x}{\sqrt{T}} \right]^{-1} \frac{x'Z}{\sqrt{T}} \left(\frac{Z'Z}{T} \right)^{-1} \frac{Z'u}{\sqrt{T}}.
\end{aligned}$$

Note that

$$\frac{Z'x}{\sqrt{T}} = \frac{Z'(Z \cdot C/\sqrt{T} + v)}{\sqrt{T}} = \frac{Z'Z}{T} C + \frac{Z'v}{\sqrt{T}} \xrightarrow{d} Q_{ZZ} C + z_v.$$

Weak Instruments Asymptotics

Staiger and Stock (1997)

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:

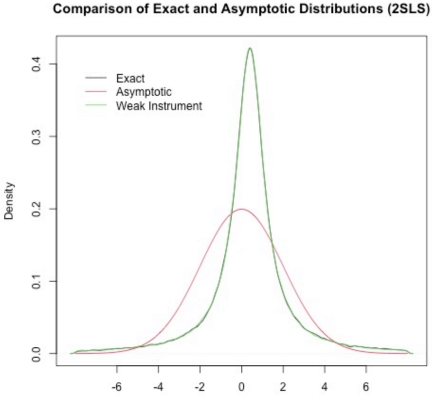
$$\hat{\sigma}_u^2 = \frac{\hat{u}'\hat{u}}{T}, \quad \hat{\sigma}_{uv} = \frac{\hat{u}'\hat{v}}{T}, \quad \hat{u}_t = y_t - x_t\hat{\beta}_{2SLS}.$$

Weak Instruments Asymptotics

A Simulation

$$n = 100, C = .5, \beta = 0, Z_t \overset{i.i.d.}{\sim} N(0,1),$$

$$\begin{pmatrix} u_t \\ v_t \end{pmatrix} \overset{i.i.d.}{\sim} N\left(0, \begin{pmatrix} 1 & .5 \\ .5 & 1 \end{pmatrix}\right), Z_t \perp\!\!\!\perp (u_t, v_t).$$



Detecting Weak Instrument

Stock and Yogo (2005)

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 $\alpha = 5\%$;

(Both characterizations look at the "worst case" with respect to the degree of simultaneity between u_t and v_t.)

They then develop tests for the null hypothesis that π lies in the weak instrument set.

Detecting Weak Instrument

Stock and Yogo (2005)

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

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

kF is asymptotically distributed as a non-central χ^2 with a non-centrality parameter $C'Q_{ZZ}C = \text{plim}_{T\to\infty}\mu^2$.

Detecting Weak Instrument

Stock and Yogo (2005)

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)

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

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

but with different critical values.

Detecting Weak Instrument

Non-Homoskedastic Errors

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

$$F^{Eff} = \frac{\hat{\pi}'Z'Z\hat{\pi}'}{\text{tr}(\hat{\Sigma}_{\pi\pi}(Z'Z)/T)}$$

with their critical values or the rule-of-thumb value of 10. Stata package `weakivtest` implements this test.

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. Let

$$\phi(\beta_0) = \begin{cases} 1 & \text{reject} \\ 0 & \text{do not reject} \end{cases}.$$

$\phi(\beta_0)$ 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 case of one 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). 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

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

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

In just-identified models:

- $CS_{AR}(\beta)$ can take one of three forms: (i) $[a, b]$, (ii) $(-\infty, a] \cup [b, \infty)$, (iii) the real line $(-\infty, \infty)$ (iff a robust F-test cannot reject $\pi = 0$).
- AR test has (weakly) higher power than any other size-α unbiased test. (We say a size-α test ϕ is unbiased if $\mathbb{E}_{\beta,\pi}[\phi(\beta_0)] \geq \alpha$ for all $\beta \neq \beta_0$ and all π.) In other words, the AR test is efficient.

In over-identified models with homoskedastic errors:

- $CS_{AR}(\beta)$ can take one of four forms: (i) $[a, b]$, (ii) $(-\infty, a] \cup [b, \infty)$, (iii) the real line $(-\infty, \infty)$, (iv) empty set.
- AR test is inefficient under strong identification with $\|\pi\|$ large
  - Use k degrees of freedom for one parameter: loss of power
  - Failure of overidentifying restrictions can lead to empty $CS_{AR}(\beta)$

Kleibergen (2002) proposes a modified score test:

$$K(\beta) = \frac{(y - x\beta)' P_{\tilde{X}(\beta)} (y - x\beta)}{(y - x\beta)' M_Z (y - x\beta)/(T - k)} \xrightarrow{d} \chi^2_1 \quad \text{when } \beta = \beta_0,$$

where $\tilde{X}(\beta) = Z\tilde{\pi}(\beta)$. Under $H_0 : \beta = \beta_0$, $\tilde{\pi}(\beta)$ is both a consistent estimator of π and asymptotically independent of $(y - x\beta_0)'Z$.

- Take a linear combination of $(y - x\beta_0)'Z$ to reduce dimension and improve power.
- $K(\beta) = 0$ has an extraneous root, leading to non-monotonic power and disconnected confidence intervals in finite samples.

Robust Inference

Conditional Likelihood Ratio (CLR) Test

With homoskedastic errors, Moreira (2003) considers a test statistic ϕ conditional a sufficient statistic $Z'(y, x)$ for β and π.

- $Z'(y, x)$ can be represented by (S, T), where under $H_0 : \beta = \beta_0$, $S = Z'(y - x\beta)$ is independent of π while T depends on π, and $S \perp\!\!\!\perp T$.
- The conditional null distribution of ϕ given $T = t$ does not depend on π. Hence, one can use critical values depending on realization of T.
- In fact, any test that has exact size α for all π is a conditional test.
- Moreira (2003) recommends using the conditional test based on the LR statistic. It is implemented in Stata (command `condivreg`).

However, the literature has not yet converged on a recommendation in the non-homoskedastic case. 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*, pages 222–240.

Kleibergen, F. (2002). Pivotal statistics for testing structural parameters in instrumental variables regression. *Econometrica*, 70(5):1781–1803.

Montiel Olea, J. L. and Pflueger, C. (2013). A robust test for weak instruments. *Journal of Business & Economic Statistics*, 31(3):358–369.

Moreira, M. J. (2003). A conditional likelihood ratio test for structural models. *Econometrica*, 71(4):1027–1048.

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*, pages 557–586.