EC708 Discussion 2
Weak IV

Yan Liu

Department of Economics
Boston University

January 27, 2023

Outline

Table of Contents

Weak Instruments Asymptotics

Staiger and Stock (1997): Just-Identified Case

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 and Z_t are scalars. Suppose

- $\pi = \delta/\sqrt{T}$, where $\delta \neq 0$;
- conditional homoskedasticity:

$$
\operatorname{Var}\left(\begin{bmatrix} U_t \\ V_t \end{bmatrix} \middle| \, Z_t \right) = \begin{bmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{bmatrix}.
$$

Weak Instruments Asymptotics

Staiger and Stock (1997): Just-Identified Case

In the just-identified case, the 2SLS estimator coincides with the IV estimator. We can write

$$\hat{\beta}_{2SLS} - \beta = \frac{\mathbf{Z}'U}{\mathbf{Z}'\mathbf{X}} = \frac{\mathbf{Z}'U}{\pi\mathbf{Z}'\mathbf{Z} + \mathbf{Z}'V} = \frac{\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t U_t}{\delta\frac{1}{T}\sum_{t=1}^{T} Z_t^2 + \frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t V_t}.$$

By CLT,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\begin{bmatrix} Z_t U_t \\ Z_t V_t \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \Psi_{ZU} \\ \Psi_{ZV} \end{bmatrix} \sim N\left(0, E[Z_t^2]\begin{bmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{bmatrix}\right).$$

Weak Instruments Asymptotics

Staiger and Stock (1997): Just-Identified Case

Hence,

$$\hat{\beta}_{2SLS} - \beta \xrightarrow{d} \frac{\Psi_{ZU}}{\delta E[Z_t^2] + \Psi_{ZV}}.$$

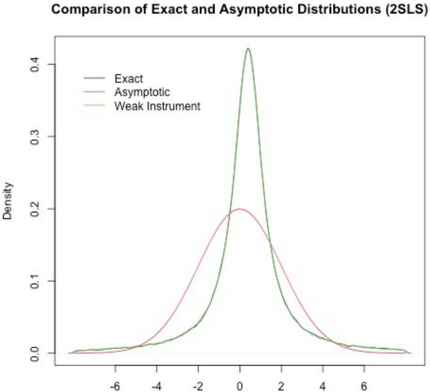
- It is a random mixture of normal distributions $\Rightarrow$ heavy tails
- We can simulate it.
- We cannot use it for inference because the estimation of σ_u^2 and σ_{uv} depends on $\hat{\beta}_{2SLS}$, which is inconsistent:

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

Weak Instruments Asymptotics

A Simulation

$$T = 100, \delta = 0.5, \beta = 0, Z_t \overset{\text{i.i.d.}}{\sim} N(0, 1),$$
$$\begin{bmatrix} U_t \\ V_t \end{bmatrix} \overset{\text{i.i.d.}}{\sim} N\left(0, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}\right), Z_t \perp (U_t, V_t).$$



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

$$\hat{\beta}_{2SLS} - \beta = \frac{\frac{1}{\sqrt{T}} \sum_{t=1}^{T} Z_t U_t}{\delta T^{\kappa} \frac{1}{T} \sum_{t=1}^{T} Z_t^2 + \frac{1}{\sqrt{T}} \sum_{t=1}^{T} Z_t V_t}.$$

Note that

$$T^{\kappa} \frac{1}{T} \sum_{t=1}^{T} Z_t^2 \xrightarrow{p} 0.$$

Hence,

$$\hat{\beta}_{2SLS} - \beta \xrightarrow{d} \frac{\Psi_{ZU}}{\Psi_{ZV}} \quad \Rightarrow \quad \hat{\beta}_{2SLS} \text{ is inconsistent.}$$

Weak Instruments Asymptotics

Other Cases: “Stronger” than $O(T^{-1/2})$

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

$$T^{\kappa}(\hat{\beta}_{2SLS} - \beta) = \frac{\frac{1}{\sqrt{T}} \sum_{t=1}^{T} Z_t U_t}{\delta \frac{1}{T} \sum_{t=1}^{T} Z_t^2 + T^{-\kappa} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} Z_t V_t}.$$

Note that

$$T^{-\kappa} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} Z_t V_t \xrightarrow{p} 0.$$

Hence,

$$T^{\kappa}(\hat{\beta}_{2SLS} - \beta) \xrightarrow{d} N\left(0, \frac{\sigma_u^2}{\delta^2 E[Z_t^2]}\right) \quad \Rightarrow \quad \hat{\beta}_{2SLS} \text{ is consistent.}$$

Table of Contents

Detecting Weak Instrument

Stock and Yogo (2005): Weak Instrument Set

Consider the linear IV model:

$$Y_t = X_t'\beta + U_t,$$
$$X_t' = Z_t'\pi + V_t,$$

where X_t is $k \times 1$ and Z_t is $\ell \times 1$ with $\ell \geq k$. 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 $\alpha = 5\%$;

Detecting Weak Instrument

Stock and Yogo (2005): Test Statistic

- Null hypothesis: π lies in the weak instrument set
- Stock and Yogo (2005) develop tests in cases with homoskedastic errors. With a single endogenous regressor, the test reduces to the first-stage F-statistic:

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

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

Detecting Weak Instrument

Stock and Yogo (2005): Critical Value

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): Multiple Endogenous Regressors

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

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

where mineval denotes the minimum eigenvalue.

Detecting Weak Instrument

Non-Homoskedastic Errors

- With $k = 1$ and non-homoskedastic errors, the ivreg2 command in Stata automatically reports a "robust" F-statistic $F^R = \frac{\hat{\pi}'\hat{\Sigma}_{\pi\pi}^{-1}\hat{\pi}}{\ell}$ with Stock and Yogo (2005) critical values. (not justified!)
- Montiel Olea and Pflueger (2013) propose using the effective first-stage F-statistic:

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

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

Table of Contents

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 $\phi(\beta)$ 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

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

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

$$\inf_{\beta,\pi} \Pr_{\beta,\pi}\{\beta \in CS\} \geq 1 - \alpha.$$

Consider the linear IV model:

$$Y_t = X_t\beta + U_t,$$
$$X_t = Z_t'\pi + V_t,$$

where X_t is scalar and Z_t is $\ell \times 1$.

- (U_t, V_t) is homoskedastic conditional on Z_t.
- Let $P_Z = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ (projection matrix) and $M_Z = I_T - P_Z$ (annihilator matrix).
- Null hypothesis $H_0 : \beta = \beta_0$.

Anderson-Rubin statistic:

$$AR(\beta) = \frac{(Y - \mathbf{X}\beta)' P_Z (Y - \mathbf{X}\beta)}{(Y - \mathbf{X}\beta)' M_Z (Y - \mathbf{X}\beta)/(T - \ell)}.$$

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

$$AR(\beta) \xrightarrow{d} \frac{\Psi'_{ZU} Q_{ZZ}^{-1} \Psi_{ZU}}{\sigma_u^2} \sim \chi^2_\ell.$$

We can form an (asymptotically) size-α test and an (asymptotically) level $1 - \alpha$ confidence set as

$$\begin{aligned} \phi_{AR}(\beta_0) &= 1\{AR(\beta_0) > \chi^2_{\ell,1-\alpha}\}, \\ CS_{AR} &= \{\beta : AR(\beta) \leq \chi^2_{\ell,1-\alpha}\}, \end{aligned}$$

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

Just-identified case $(\ell = 1)$:

- 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 the Uniformly Most Powerful Unbiased (UMPU)/efficient test (Moreira, 2009): no power loss relative to t-test even when instruments are strong.

Over-identified case ($\ell > 1$):

- $H_0 : \beta = \beta_0$ 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, especially under strong instruments
  - Use ℓ degrees of freedom for one parameter $\Rightarrow$ loss of power

Robust Inference against Weak Instruments

Power Improvement in the Over-Identified Case

Kleibergen (2002) proposes the K-statistic

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

where $\tilde{X}_t(\beta) = Z_t' \tilde{\pi}(\beta)$ such that under $H_0 : \beta = \beta_0$,

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

Caveat: $K(\beta) = 0$ has an extraneous root, leading to non-monotonic power and disconnected confidence intervals in finite samples.

- Both AR and K-statistics are pivotal: their null distributions do not depend on π (nuisance parameter)
- We can also construct a test statistic $s(\beta)$ whose null distribution depends on π but only through some sufficient statistic $D(\beta_0)$
  - Find the largest possible $1 - \alpha$ quantile over some set of π (conservative)
  - Use conditional critical values $c_\alpha(D(\beta_0)) \Rightarrow$ **conditional tests**

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 some sufficient statistics.

- When $\ell = 1$, $LR(\beta_0)$ 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
  - optimizing weighted average power
- multiple endogenous regressors
  - inference on a subvector of $\beta \Rightarrow$ improve power of projection method

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

Table of Contents

Truncated Normal

Heckman Selection Model

Heckman (1979) models the wage determining process as:

$$
\begin{aligned}
Y_t^* &= X_t'\beta + U_t, \\
D_t^* &= Z_t'\gamma + V_t, \\
D_t &= 1\{D_t^* \geq 0\}, \\
Y_t &= Y_t^* \cdot D_t.
\end{aligned}
$$

- Y_t^*: offered market wages
- D_t^*: latent variable representing the propensity to be employed
- (U_t, V_t) are jointly dependent of (X_t, Z_t)
- $(U_t, V_t) \overset{\text{i.i.d.}}{\sim} N(0, \Sigma)$, where $\Sigma = \begin{bmatrix} \sigma_u^2 & \rho\sigma_u\sigma_v \\ \rho\sigma_u\sigma_v & \sigma_v^2 \end{bmatrix}$.

Sample selection bias:

$$E[Y_t|X_t, D_t = 1] = X_t'\beta + E[U_t|V_t \geq -Z_t'\gamma].$$

In general, we are often interested in $E[u|a \leq v \leq b]$, where

$$(u, v) \sim N(\mu, \Sigma), \quad \mu = \begin{bmatrix} \mu_u \\ \mu_v \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \sigma_u^2 & \rho\sigma_u\sigma_v \\ \rho\sigma_u\sigma_v & \sigma_v^2 \end{bmatrix}.$$

Truncated Normal

First Moment

The truncated joint density of u and v is

$$f(u, v|a \leq v \leq b) = \frac{f(u, v)}{\Pr(a \leq v \leq b)}.$$

Hence,

$$E[u|a \leq v \leq b] = \frac{\int_{-\infty}^{\infty} \int_a^b u f(u, v) dv du}{\Pr(a \leq v \leq b)}.$$

Using the transformation $\xi\sqrt{1-\rho^2} = \frac{u-\mu_u}{\sigma_u} - \rho\frac{v-\mu_v}{\sigma_v}$, some algebra shows

$$E[u|a \leq v \leq b] = \mu_u - \rho\sigma_u \frac{\phi(\frac{b-\mu_v}{\sigma_v}) - \phi(\frac{a-\mu_v}{\sigma_v})}{\Phi(\frac{b-\mu_v}{\sigma_v}) - \Phi(\frac{a-\mu_v}{\sigma_v})}.$$