TA Session 1: Weak Instruments

Shuowen Chen[1]

EC708: PhD Econometrics I (Spring 2020)

[1] Parts of the materials are borrowed from the textbook by Bruce Hansen and teaching slides by Iván Fernández-Val.

Outline

- Motivating Example: Identification Failure
- 2SLS under Weak Instruments
- Weak Instruments Asymptotics
- Detecting Weak Instruments
- Robust Inference: Inverting Anderson-Rubin Statistics
- Bonus: Many Instruments Asymptotics

Irrelevant Instrument

Consider the model

$$y = X\beta + u,$$
$$X = Z\pi + v$$

where y and X are $T \times 1$, Z is $T \times 1$ and

1. $\Pi = 0$.
2. Endogeneity and conditional homoskedasticity

$$Var\left(\begin{pmatrix} u_i \\ v_i \end{pmatrix} \mid Z_i\right) = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \quad \rho = corr(u_i, v_i) \neq 0$$

CLT:

$$\frac{1}{\sqrt{T}} \sum_{i=1}^{T} \begin{pmatrix} Z_i u_i \\ Z_i v_i \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

Inconsistent OLS and IV Estimators

- $\widehat{\beta}_{OLS}$ is inconsistent due to endogeneity

$$\widehat{\beta}_{OLS} - \beta = \frac{\frac{1}{T}\sum_{i=1}^{T} u_i v_i}{\frac{1}{T}\sum_{i=1}^{T} v_i^2} \xrightarrow{p} \rho \neq 0$$

- $\widehat{\beta}_{IV}$ is inconsistent (converges to Cauchy distribution)

$$\widehat{\beta}_{IV} - \beta = \frac{\frac{1}{\sqrt{T}}\sum_{i=t}^{T} Z_i u_i}{\frac{1}{\sqrt{T}}\sum_{i=t}^{T} Z_i v_i} \xrightarrow{d} \frac{\xi_1}{\xi_2} = \rho_0 + \frac{\xi_1 - \rho\xi_2}{\xi_2}$$

- $\widehat{\beta}_{IV}$ is not normally distributed under irrelevant IV, so inference based on normal distribution is not reliable.

Unreliable t-test: Size Distortion

Note that

$$
\begin{aligned}
\widehat{\sigma}_u^2 &= \frac{1}{T}\sum_{i=1}^{T}(y_i - X_i\widehat{\beta}_{IV})^2 \\
&= \frac{1}{T}\sum_{i=1}^{T} u_i^2 - \frac{2}{T}\sum_{i=1}^{T} u_i X_i(\widehat{\beta}_{IV} - \beta) + \frac{1}{T}\sum_{i=1}^{T} X_i^2(\widehat{\beta}_{IV} - \beta)^2 \\
&\xrightarrow{p} 1 - 2\rho\frac{\xi_1}{\xi_2} + \left(\frac{\xi_1}{\xi_2}\right)^2
\end{aligned}
$$

Therefore

$$
t = \frac{\widehat{\beta}_{IV} - \beta}{\sqrt{\widehat{\sigma}_u^2 \sum_{t=1}^{T} Z_i^2}\Big/\left|\sum_{i=1}^{T} Z_i X_i\right|} \xrightarrow{d} \frac{\xi_1/\xi_2}{\sqrt{1 - 2\rho\frac{\xi_1}{\xi_2} + \left(\frac{\xi^1}{\xi^2}\right)^2}}
$$

As $\rho \to 1$, $\xi_1/\xi_2 \to 1$, t stat diverges as $T \to \infty$. Always reject the null!

A Somehow Stronger Instrument

Now suppose $\Pi = 1/\sqrt{T}$ and $\mathbb{E}(Z_i^2) = \sigma_Z^2 > 0$. Then

$$\widehat{\beta}_{IV} - \beta \xrightarrow{d} \frac{\xi_1}{\sigma_Z^2 + \xi_2}$$

If instrument becomes irrelevant asymptotically, IV estimator is inconsistent. Suppose $\Pi = T^{-1/2+\kappa}$.

- If $\kappa > 0$
$$T^{\kappa}(\widehat{\beta}_{IV} - \beta) \xrightarrow{d} \mathcal{N}\left(0, 1/\sigma_Z^4\right)$$
- If $\kappa < 0$
$$\widehat{\beta}_{IV} - \beta \xrightarrow{d} \frac{\xi_1}{\xi_2}$$

We need big Π to have IV estimator well behaved. This is related to the strength of the instrument.

Multiple Instruments: 2SLS

Consider the model

$$y = X\beta + u,$$
$$X = Z\pi + v$$

where y and X are $T \times 1$, Z is $T \times K$ and Π is $K \times 1$. ($K \geq 1$)

$$\widehat{\beta}_{2SLS} - \beta = \frac{X'P_Z u}{X'P_Z X} = \frac{\Pi' Z' u + v' P_Z u}{\Pi' Z' Z \Pi + 2\Pi' Z' v + v' P_Z v}.$$

Define concentration parameter:

$$\mu^2 = \Pi' Z' Z \Pi / \sigma_v^2$$

This parameter plays the role of sample size, is a measure of the quality of the instruments, and is related to F statistics.

Rothenberg (1984) Form

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

where $Z_u = \frac{\Pi' Z' u}{\sigma_u \sqrt{\Pi' Z' Z \Pi}}$, $Z_v = \frac{\Pi' Z' v}{\sigma_v \sqrt{\Pi' Z' Z \Pi}}$, $S_{uv} = \frac{v' P_Z u}{\sigma_u \sigma_v}$, $S_{vv} = \frac{v' P_Z v}{\sigma_v^2}$, whose distributions don't depend on sample size T.

- Larger T means larger μ^2
- If μ^2 is large, $\mu(\widehat{\beta}_{2SLS} - \beta)$ well approximated by $\mathcal{N}(0, \sigma_u^2/\sigma_v^2)$
- If μ^2 is small, distribution is non-normal
  - Extreme case: $\Pi = 0$. Refer to the irrelevant instrument example.
- First stage test statistic of $H_0 : \Pi = 0$ is[2]

$$F = \frac{\Pi'[\sigma_v^2 (Z'Z)^{-1}]\Pi}{K} = \frac{\Pi' Z' Z \Pi}{\sigma_v^2} \frac{1}{K} = \frac{\mu^2}{K}.$$

Takeaway: conventional asymptotics inappropriate for 2SLS under weak IV, especially if endogeneity is high.

[2] Recall relation between F and chi-square distributions: $F_{K,\infty} = \chi_K^2 / K$.

Weak Instrument Asymptotics (Staiger & Stock, 1997)

Consider the model

$$y = X\beta + u,$$
$$X = Z\pi + v$$

where y and X are $T \times 1$, Z is $T \times K$ and Π is $K \times 1$. ($K \geq 1$)

1. $\Pi = c/\sqrt{T}$
2. Conditional Homoskedasticity

$$\begin{pmatrix} u_i \\ v_i \end{pmatrix} \mid Z_i \sim \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{pmatrix}\right)$$

Drifting Sequence Assumption

$\Pi = \frac{c}{\sqrt{T}}$ measures the quality of instruments. Why?

- Recall concentration parameter $\mu^2 = \Pi' Z' Z \Pi / \sigma_v^2$
- Suppose Π is fixed. Then as $T \to \infty$, $\mu^2 \to \infty$ regardless of Π's magnitude. Therefore F-test will reject $\Pi = 0$ for large T, even though Π can be very small
- By setting $\Pi = c/\sqrt{T}$, $\Pi \to 0$ at rate $\sqrt{T}$, can show that $F_{\Pi=0} \xrightarrow{d}$ bounded r.v., hence won't reject for large T with probability 1.

LLN and CLT Assumptions

1. $\frac{1}{T}\sum_{i=1}^{T} Z_i Z_i' \xrightarrow{p} \mathbb{E}(ZZ') = \Sigma_{ZZ}$,
2. $\frac{1}{T}\sum_{i=1}^{T} Z_i u_i \xrightarrow{p} \mathbb{E}(Zu) = 0$,
3. $\frac{1}{T}\sum_{i=1}^{T} Z_i v_i \xrightarrow{p} \mathbb{E}(Zv) = 0$,
4. $\frac{1}{\sqrt{T}}\sum_{i=1}^{T} Z_i u_i \xrightarrow{d} \mathcal{N}(0, \sigma_u^2 \Sigma_{ZZ}) = \psi_{Zu}$,
5. $\frac{1}{\sqrt{T}}\sum_{i=1}^{T} Z_i v_i \xrightarrow{d} \mathcal{N}(0, \sigma_v^2 \Sigma_{ZZ}) = \psi_{Zv}$

Remark: ψ_{Zu} and ψ_{Zv} are correlated (having $\rho = \sigma_{uv}/\sigma_u\sigma_v$).

2SLS Asymptotics Under Weak Instrument

$$
\begin{aligned}
\widehat{\beta}_{2sls} - \beta &= (X'P_Z X)^{-1} X'P_Z u \\
&= \Big[\frac{X'Z}{T}\Big(\frac{Z'Z}{T}\Big)^{-1}\frac{Z'X}{T}\Big]^{-1}\Big(\frac{X'Z}{T}\Big)\Big(\frac{Z'Z}{T}\Big)\Big(\frac{Z'u}{T}\Big)
\end{aligned}
$$

Note that

$$
\frac{Z'X}{T} = \frac{Z'(Z\Pi + v)}{T} = \frac{Z'Z}{T}\frac{c}{\sqrt{T}} + \frac{Z'v}{T} \xrightarrow{p} 0
$$

Conclusion: normalization by $\frac{1}{T}$ inappropriate for weak instrument asymptotics, use $\frac{1}{\sqrt{T}}$ instead

2SLS Asymptotics Under Weak Instrument

$$\widehat{\beta}_{2sls} - \beta = \Big[\frac{X'Z}{\sqrt{T}}\Big(\frac{Z'Z}{T}\Big)^{-1}\frac{Z'X}{\sqrt{T}}\Big]^{-1}\Big(\frac{X'Z}{\sqrt{T}}\Big)\Big(\frac{Z'Z}{T}\Big)\Big(\frac{Z'u}{\sqrt{T}}\Big)$$

Note that

$$\frac{Z'X}{\sqrt{T}} = \frac{c}{T}Z'Z + \frac{Z'v}{\sqrt{T}} \xrightarrow{d} c\Sigma_{ZZ} + \mathcal{N}(0, \sigma_v^2\Sigma_{ZZ}) = c\Sigma_{ZZ} + \psi_{Zv}$$

Therefore

$$\widehat{\beta}_{2sls} - \beta \xrightarrow{d} \big[(c\Sigma_{ZZ} + \psi_{Zv})'\Sigma_{ZZ}^{-1}(c\Sigma_{ZZ} + \psi_{Zv})\big]^{-1}(c\Sigma_{ZZ} + \psi_{Zv})\Sigma_{ZZ}^{-1}\psi_{Zu},$$

converges in distribution to a random variable dependent on c and ρ

How Useful is this Asymptotics in Practice?

A better finite-sample distribution approximation than the normal distribution ▸ Simulations

Not useful for inference because ρ cannot be consistently estimated:

$$\widehat{\rho} = \frac{\widehat{\sigma}_{uv}}{\widehat{\sigma}_u \widehat{\sigma}_v} = \frac{\widehat{u}'\widehat{v}}{\sqrt{\widehat{u}'\widehat{u}\widehat{v}'\widehat{v}}} \nrightarrow \rho$$

Why?

$$\widehat{u} = y - X\widehat{\beta}_{2sls}, \quad \widehat{v} = Z - X\widehat{\Pi},$$

but $\widehat{u} \neq u + o_p(1)$ since $\widehat{\beta}_{2sls} - \beta \xrightarrow{d} r.v$, hence inconsistent.

To be robust against strength of instruments, we will consider inference based on confidence region.

Detecting Weak Instruments

We've seen that the concentration parameter μ^2 plays the role of sample size and 2sls estimators don't have normal distribution when μ^2 is small, but

- how small is small?
- in practice μ^2 is unknown, how to detect weak IV?

Stock & Yogo (2005) provide answers assuming homoskedasticity

- weak if bias of IV rel. to OLS exceeds a threshold, say 10%, or
- if size α Wald test has actual size exceeding a threshold, say 10%

Stock & Yogo's test reduces to first–stage F test when having only one endogenous variable

- Null: IVs are weak.
- Critical values: based on Staiger & Stock (1997) distribution. Do not use F table as we are not testing whether $\Pi = 0$
- Critical values depend on the perspective of weakness. In practice usually work with small number of instruments, and 10 is good from either perspective.

Why F Test?

Recall F statistic

$$F = \widehat{\Pi}' Z' Z \widehat{\Pi} / \widehat{\sigma}_v^2 K$$

and concentration parameter, which contains info about strength of IV

$$\mu^2 = \Pi' Z' Z \Pi / \sigma_v^2$$

therefore F statistic is an indicator of μ^2/K, and thus useful for detecting if instruments are strong.

Reiteration:

1. Do not use F table critical value since we are not testing if $\Pi = 0$ in the first stage.
2. The critical value is based on weak IV asymptotics. Stock and Yogo (2005) avoid the inconsistency issue of $\widehat{\rho}$ and obtained critical values via simulations. 10 is a good critical value in practice from both weakness perspectives.
3. **Do not use Stock and Yogo test unless you impose homoskedastic assumptions**

Effective F-Statistic (Montiel Olea & Pflueger, 2013)

Relaxes homoskedasticity assumption of u and v. In other words, it is **NO LONGER** the case that

$$\begin{pmatrix} u \\ v \end{pmatrix} \sim \begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{pmatrix} \otimes I_T$$

instead it is a general variance–covariance matrix

$$\begin{pmatrix} u \\ v \end{pmatrix} \sim \begin{pmatrix} \omega_1^2 & \omega_{12} \\ \omega_{21} & \omega_2^2 \end{pmatrix}$$

CLT

$$\frac{1}{\sqrt{T}} \sum_{i=1}^{T} \begin{pmatrix} Z_i u_i \\ Z_i v_i \end{pmatrix} \xrightarrow{d} \sim \mathcal{N}\left(0, \begin{pmatrix} W_1 & W_{12} \\ W_{21} & W_2 \end{pmatrix}\right)$$

The authors propose effective F-statistic

$$F^{eff} = \frac{K\widehat{\omega}_2^2}{tr((Z'Z/T)^{-1}\widehat{W_2})} F$$

Weak IV Robust Inference: Anderson–Rubin Test

Even if instruments are weak, can still conduct inference regardless of their strength. We focus on models with one endogenous variable as it is common in applied research

Test Inversion

Given a size α test of $H_0 : \beta = \beta_0$ for any β_0, construct a level $(1-\alpha)$ confidence set for β by collecting the set of non-rejected values[3]

Test of H_0 :

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

is size-α of H_0 if $\sup_{\Pi} \mathbb{E}_{\beta_0,\Pi}[\phi(\beta_0) = 1] \leq \alpha$: max prob of falsely rejecting the null is bounded by α regardless of the value of Π

The set of β not rejected by $\phi : CS = \{\beta : \phi(\beta) = 0\}$, is a level $(1-\alpha)$ **confidence set** (contains the true β $1-\alpha$ of the time):

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

In practice:

1. Specify a grid of potential β's and evaluate the test ϕ at all β's
2. Collect the non-rejected values of β's as an approximation

---

[3] Standard trick in moment inequalities literature (Canay and Shaikh, 2016)

Anderson–Rubin Test Statistic

Consider a specific test ϕ

$$AR(\beta) = \frac{(y - X\beta)' P_Z (y - X\beta)}{(y - X\beta)'(I_T - P_Z)(y - X\beta)/(T - K)}$$

The distribution of AR doesn't depdent on μ, and under the null asymptotically

$$AR(\beta) \rightarrow \chi^2_K$$

Invert the AR statistic to construct the level $(1 - \alpha)$–confidence set:

$$CS_{1-\alpha} = \left\{ \beta : AR(\beta) < \chi^2_{K,1-\alpha} \right\}$$

Inverting AR is the most efficient method if we work with just–identified models.

Some Other Robust Inference Methods

1. Kleibergen (2002) LM test
   - χ_1^2 distribution irrespective of the instrument's strength
2. Moreira (2003) CLR test
   - Proposes a sufficient statistic of $\mu^2 : Q_T$. Critical value depends on simulated realization of Q_T.
   - Probably the best if we work with one endogenous regressor, multiple instruments and conditional homoskedasticity
   - If Non-homoskedastic, not the best
3. See Andrews, Stock and Sun (2018) for open questions

Bonus: Many Instruments Set Up

Including more instruments reduces variances (Hansen, p371). But in practice bias increases. To model this phenomenon, consider the model

$$\begin{aligned} y &= X\beta + u, \\ X &= Z\pi + v \end{aligned}$$

where y and X are $T \times 1$, Z is $T \times K$ and Π is $K \times 1$. ($K \geq 1$)

1. Homoskedasticity

$$\begin{pmatrix} u_i \\ v_i \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{pmatrix}\right)$$

2. $K/T \to \alpha \neq 0$. Number of instruments is nonnegligible to the sample size
3. $\Pi' Z' Z \Pi \to Q$

Many Instruments Asymptotics

$$\widehat{\beta}_{2sls} - \beta = \Big(\frac{X'P_ZX}{T}\Big)^{-1}\Big(\frac{X'P_Zu}{T}\Big)$$

Note that

$$\begin{aligned}\mathbb{E}\Big(\frac{X'P_ZX}{T}\Big) &= \mathbb{E}\Big[\frac{(\Pi'Z' + v')P_Z(Z\Pi + v)}{T}\Big] = \mathbb{E}\Big[\frac{\Pi'Z'P_ZZ\Pi}{T}\Big] + \mathbb{E}\Big[\frac{v'P_Zv}{T}\Big] \\ &= \frac{\Pi'Z'Z\Pi}{T} + \frac{K}{T}\sigma_v^2\end{aligned}$$

The last equality uses the following observation:
$v'P_Zv = tr(v'P_Zv) = tr(v'vP_Z)$ and hence

$$\mathbb{E}(v'P_Zv) = \mathbb{E}tr(v'vP_Z) = tr\mathbb{E}(v'vP_Z) = \sigma_v^2 tr(P_Z) = \sigma_v^2 K$$

In a similar vein, we have $\mathbb{E}\Big(\frac{X'P_Zu}{T}\Big) = \mathbb{E}\Big(\frac{v'P_Zu}{T}\Big) = \frac{K}{T}\sigma_{uv}$. Therefore

$$\widehat{\beta}_{2sls} - \beta \xrightarrow{p} (Q + \alpha\sigma_v^2)^{-1}\alpha\sigma_{uv}$$

Extreme case: $K = n$, $\widehat{\beta}_{2sls} = (X'X)^{-1}X'y$ since $P_Z = I_T$.

Overlay of Three Distributions with Weak IV

Simulate the following model $S = 20000$ times: Back

$$Y = X\beta_0 + u$$
$$X = Z\Pi + v$$

where $T = 100$, $\beta_0 = 0$, $\sigma_u^2 = \sigma_v^2 = 1$, $\rho = 0.5$, $Z \sim \mathcal{N}(0,1)$, $\Pi = \frac{0.5}{T}$