ECON 6200: Section 4 2/12/25

Office hours next week will be moved to:

Thursday (2/20) 12pm – 2pm

Agenda

- Review of IV
  - Binary IV
- Useful Things to Know
  - Dimensions
  - Identification
- Deriving IV from GMM
- Deriving IV from TSLS
- IV Asymptotics

Review of IV

Consider a linear model

$$Y = X'\beta + \varepsilon \qquad \text{(structural equation)}$$

(structural param)

if β is the parameter of interest and $E[X\varepsilon] \neq 0$

$\Rightarrow$ there is endogeneity in the model

We can still perform a linear projection

$$Y = X'\beta^* + \underbrace{e^*}_{\text{projection error}}$$

where our projection coefficient is $\beta^* = E[XX']^{-1}E[XY]$
and $E[Xe^*] = 0$ by construction.

However, the projection coefficient (β^*) $\neq$ structural parameter β

$$\begin{aligned}
\beta^* &= \left(E[XX']\right)^{-1} E[XY] \\
&= \left(E[XX']\right)^{-1} E[X(X'\beta + e)] \\
&= \beta + \left(E[XX']\right)^{-1} \underbrace{E[Xe]}_{\substack{\neq 0 \\ \text{under structural} \\ \text{model}}} \neq \beta \quad \text{(structural parameter)}
\end{aligned}$$

$\Rightarrow$ Hence, OLS estimator is <u>inconsistent</u> for the structural parameter

$$\hat{\beta} \xrightarrow{p} (E[XX'])^{-1} E[XY] = \beta^* \neq \beta$$

For simplicity, consider the simple linear regression

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

where we cannot assert $E[X\varepsilon] = 0$

Now suppose we observe RV Z w/ the following properties:

$$\text{cov}(Z, X) \neq 0 \qquad \text{(relevance)}$$

$$\text{cov}(Z, \varepsilon) = 0 \qquad \text{(validity)}$$

Then $\beta_{1IV} = \dfrac{\text{cov}(Z,Y)}{\text{cov}(Z,X)} = \dfrac{\text{cov}(Z, \beta_0 + \beta_1 X + \varepsilon)}{\text{cov}(Z,X)}$

$$= \frac{\beta_1 \text{cov}(Z,X) + \cancelto{0}{\text{cov}(Z,\varepsilon)}}{\text{cov}(Z,X)}$$

$$= \beta_1$$

Binary IVs

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

where X is now a binary variable

If Z is a binary instrument for X, then

$$\beta_{1IV} = \frac{Cov(Z,Y)}{Cov(Z,X)} = \frac{E[Y|Z=1] - E[Y|Z=0]}{E[X|Z=1] - E[X|Z=0]}$$

Proof

$$Cov(Z,Y) = \underbrace{E[ZY]}_{A} - \underbrace{E[Z]}_{B}\,\underbrace{E[Y]}_{C}$$

A $\quad E[ZY] \overset{LIE}{=} E[(E[ZY|Z])]$

$E(f(x)) = \sum f(x) p(x)$

Def of $E(x)$

$$= P(Z=1)\,E[ZY|Z=1] + P(Z=0)\,E[ZY|Z=0]$$

0 since $Z=0$

$$= P(Z=1)\,E[Y|Z=1]$$

B $\quad E[Z] = 1 \cdot P(Z=1) + 0 \cdot P(Z=0)$

$$= P(Z=1)$$

(c) $$E[Y] = E[E[Y|Z]]$$
$$= P(Z=1)\,E[Y|Z=1] + P(Z=0)\,E[Y|Z=0]$$

Then $$\mathrm{Cov}(Z,Y) = P(Z=1)\,E[Y|Z=1]$$
$$- P(Z=1)\big[P(Z=1)\,E[Y|Z=1] + P(Z=0)\,E[Y|Z=0]\big]$$

$$= P(Z=1)\Big[E[Y|Z=1]\underbrace{\big(1-P(Z=1)\big)}_{P(Z=0)} - P(Z=0)\,E[Y|Z=0]\Big]$$

$$= P(Z=1)\,P(Z=0)\big[E[Y|Z=1] - E[Y|Z=0]\big]$$

If we repeat the above steps for $\mathrm{Cov}(Z,X)$ and replace "Y" with "Z":

$$\mathrm{Cov}(Z,X) = P(Z=1)\,P(Z=0)\big(E[X|Z=1] - E[X|Z=0]\big)$$

$$\Rightarrow \beta_{1IV} = \frac{\mathrm{Cov}(Z,Y)}{\mathrm{Cov}(Z,X)} = \frac{P(Z=1)\,P(Z=0)\big(E[Y|Z=1] - E[Y|Z=0]\big)}{P(Z=1)\,P(Z=0)\big(E[X|Z=1] - E[X|Z=0]\big)}$$

$$= \frac{\big(E[Y|Z=1] - E[Y|Z=0]\big)}{\big(E[X|Z=1] - E[X|Z=0]\big)}$$

How can we think of binary β_{1IV}?

Suppose

$$\text{Earnings} = \beta_0 + \beta_1 \cdot (\text{Private School}) + \varepsilon$$

↳ use Z = lottery to attend

$$Z = \begin{cases} 1 & \text{won lottery} \\ 0 & \text{lost lottery} \end{cases}$$

↳ can assume lottery is random

$$\beta_{1IV} = \frac{(E[Y|Z=1] - E[Y|Z=0])}{(E[X|Z=1] - E[X|Z=0])}$$ (Wald Estimator)

$$= \frac{\text{Difference in earnings for those who won vs lost lottery}}{\text{Difference in probability of attending private school between winners and losers}}$$

$\Longrightarrow$ can be interpreted as LATE (ie: only for compliers)

Useful Things to Know

- Dimensions

  For the model

  $$Y = X'\beta + e$$

  with assumptions $E[Ze] = 0$ & $\text{rank}(E[ZX']) = k$

  vectors $\begin{cases} - Z : (\ell \times 1) \\ - X : (k \times 1) \end{cases}$ where ℓ = # instruments
  k = # regressors

- Identification

  - Just-identified ($\ell = k$)
    ↳ our assumption when dealing w/ IV

  - Overidentified
    ↳ often dealt with in TSLS & GMM

Deriving IV as GMM Estimator

- Setup moment conditions

$$E[Z\varepsilon] = E[Z(Y - X'\beta)] = 0$$ (by validity assumption)

- Solve for $\hat{\beta}_{IV}$

$$E[ZY] - E[ZX'\beta] = 0$$

$$\Rightarrow \hat{\beta}_{IV} = \left(\mathbb{E}_n[ZX']\right)^{-1}\left(\mathbb{E}_n[ZY]\right)$$

Note: This is only possible because $l = k$, so ZX' is square

Deriving IV from TSLS

(Hansen p355)

- What is TSLS?

Exploiting variation in X due to Z!

1) Regress X on Z to obtain the fitted $\hat{X}$

$$X = Z\gamma + u \quad \text{where} \quad \hat{\gamma} = (Z'Z)^{-1}(Z'X)$$

2) Regress Y on $\hat{X}$

$$Y = \tilde{\beta}\hat{X}$$

Note: $\hat{X} = Z\hat{\gamma}$
$= Z(Z'Z)^{-1}(Z'X)$

Distributed transpose

$$\begin{aligned}
\tilde{\beta}^{TSLS} &= (\hat{X}'\hat{X})^{-1}\hat{X}'Y \\
&= ((Z(Z'Z)^{-1}Z'X)'Z(Z'Z)^{-1}Z'X)^{-1}(Z(Z'Z)^{-1}Z'X)'Y \\
&= (X'Z(Z'Z)^{-1}\cancel{Z'Z(Z'Z)}^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'Y \\
&= (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'Y \\
&= (Z'X)^{-1}\cancel{Z'Z(X'Z)^{-1}X'Z(Z'Z)^{-1}}Z'Y \\
&= (Z'X)^{-1}Z'Y = \hat{\beta}_{IV}.
\end{aligned}$$

⟹ Hence, we've shown that IV is a special case of TSLS when $l = k$!

Note: In TSLS, it is possible to have $l \geq k$!

IV Asymptotics

1) IV estimator is consistent

$$\hat{\beta}_{IV} = (\mathbb{E}_n[ZX'])^{-1} \mathbb{E}_n[ZY]$$
$$= (\mathbb{E}_n[Z'X])^{-1} \mathbb{E}_n[Z(X'\beta + \varepsilon)]$$
$$= (\mathbb{E}_n[\cancel{Z'X}])^{-1} \mathbb{E}_n[\cancel{ZX'}\beta] + (\mathbb{E}_n[Z'X])^{-1} \mathbb{E}_n(Z\varepsilon)$$
$$= \beta + \underbrace{(\mathbb{E}_n[Z'X])^{-1} \mathbb{E}_n(Z\varepsilon)}_{\xrightarrow{p} 0}$$

$$\Rightarrow \hat{\beta}_{IV} \xrightarrow{p} \beta$$

2) Cannot claim IV is unbiased, even if we assume $E[\varepsilon | Z] = 0$. Why?

$$E[\hat{\beta}_{IV} | Z] = E[\beta + (\mathbb{E}_n[ZX'])^{-1} \mathbb{E}_n[Z\varepsilon] \,|\, Z]$$
$$= \beta + E[\underbrace{\mathbb{E}_n[ZX']^{-1} \mathbb{E}_n[Z\varepsilon] | Z}]$$

$\neq 0$ bc X, ε correlated

$\Rightarrow$ can't pass through E

Recall if X, Y indep, then
$E[XY] = E[X]E[Y]$

$\Rightarrow$ However, since $E[X\varepsilon] \neq 0$, we cannot split it into

$$\cancel{E[\mathbb{E}_n[ZX]^{-1} | Z] \, E[\mathbb{E}_n[Z\varepsilon] | Z]}$$

3) Asymptotics of Simple IV

$$\sqrt{n}\left(\hat{\beta}_{1IV} - \beta_1\right) = \sqrt{n}\left(\frac{\sum(z_i - \bar{z})(y_i - \bar{y})}{\sum(z_i - \bar{z})(x_i - \bar{x})} - \beta_1\right)$$

$$= \sqrt{n}\left(\frac{\sum(z_i - \bar{z})\, y_i}{\sum(z_i - \bar{z})(x_i - \bar{x})} - \beta_1\right)$$

$$= \sqrt{n}\left(\frac{\sum(z_i - \bar{z})(\beta_0 + \beta_1 x_i + \varepsilon_i)}{\sum(z_i - \bar{z})(x_i - \bar{x})} - \beta_1\right)$$

$$= \sqrt{n}\left(\beta_0 \underbrace{\cancelto{0}{\frac{\sum(z_i - \bar{z})}{\sum(z_i - \bar{z})(x_i - \bar{x})}}}_{(*)} + \beta_1 \frac{\sum(z_i \cancel{-\bar{z}})x_i}{\sum(z_i - \bar{z}\cancel{)(x_i} - \bar{x})} + \frac{\sum(z_i - \bar{z})\,\varepsilon_i}{\sum(z_i - \bar{z})(x_i - \bar{x})} - \beta_1\right)$$

$$(*) \quad \frac{\frac{1}{n}\sum(z_i - \bar{z})}{\frac{1}{n}\sum(z_i - \bar{z})(x_i - \bar{x})} = \frac{\frac{1}{n}\sum z_i - \frac{1}{n}\cdot n\,\bar{z}}{\frac{1}{n}\sum(z_i - \bar{z})(x_i - \bar{x})}$$

$$= \frac{\bar{z} - \bar{z}}{u} = 0$$

$$= \sqrt{n}\left(\frac{\sum (Z_i - \bar{Z})\varepsilon_i}{\sum (Z_i - \bar{Z})(X_i - \bar{X})}\right)$$

$$= \sqrt{n}\left(\frac{\sum \left((Z_i - E(Z)) + (E(Z) - \bar{Z})\right)\varepsilon_i}{\sum (Z_i - \bar{Z})(X_i - \bar{X})}\right)$$

$$= \sqrt{n}\left(\frac{\sum (Z_i - E(Z))\varepsilon_i}{\sum (Z_i - \bar{Z})(X_i - \bar{X})} + \frac{\sum \overbrace{(E(Z) - \bar{Z})}^{\text{constant}}\varepsilon_i}{\underbrace{\sum (Z_i - \bar{Z})(X_i - \bar{X})}_{\text{constant}}}\right)$$

$$\Downarrow$$

$$\sqrt{n}\left(\frac{\overbrace{(E(Z) - \bar{Z})}^{\to 0 \text{ as } n \to \infty} \overset{\text{finite \#}}{\sum \varepsilon_i}}{\sum (Z_i - \bar{Z})(X_i - \bar{X})}\right) \xrightarrow{P} 0 \text{ as } n \to \infty$$

$$= \sqrt{n}\left(\frac{\overbrace{\frac{1}{n}\sum (Z_i - E(Z))\varepsilon_i}^{A}}{\underbrace{\frac{1}{n}\sum (Z_i - \bar{Z})(X_i - \bar{X})}_{B}}\right) + o_p(1)$$

Note: Assume all necessary moments exist

(A) Numerator looks like:

$$E[(Z-E(Z))\varepsilon] = E[Z\varepsilon] - E[E(Z)\varepsilon]$$

$$= \underbrace{E[Z\varepsilon]}_{=0} - E(Z)\underbrace{E(\varepsilon)}_{=0} = 0$$

$$Var[(Z-E(Z))\varepsilon] = E[([Z-E(Z)]\varepsilon)^2] - 0^2$$

$$= E[[Z-E(Z)]^2\varepsilon^2]$$

$$= \sigma_Z^2\sigma^2 \quad \text{bc } Z, \varepsilon \text{ indep}$$

$$\Rightarrow \text{By CLT, } \sqrt{n}\left(\frac{1}{n}\sum(Z_i - E(Z))\varepsilon_i\right) \xrightarrow{d} N(0, \sigma_Z^2\sigma^2)$$

(B) Denominator by WLLN

$$\frac{1}{n}\sum(Z_i - E(Z))(X_i - E(X)) \xrightarrow{p} E[(Z-E(Z))(X-E(X))]$$

$$= cov(Z, X)$$

$$= \rho_{ZX}\sigma_Z\sigma_X$$

Combining everything

$$\sqrt{n}\left(\hat{\beta}_1^{IV} - \beta_1\right) \xrightarrow{d} N\left(0, \frac{\cancel{\sigma_z^2}\,\sigma^2}{\rho^2 \cancel{\sigma_z^2}\,\sigma_x^2}\right)$$

$$\rightarrow N\left(0, \frac{\sigma^2}{\rho^2 \sigma_x^2}\right)$$