Last time: $\hat{\beta}_{OLS} = (X'X)^{-1}X'Y \not\xrightarrow{p} \beta$ if $E[X_i \varepsilon_i] \neq 0$

Normal equations: $\hat{\beta}_{OLS}$ solves

$$\frac{1}{n}\sum X_i (Y_i - X_i'\hat{\beta}_{OLS}) = 0$$

Method of moments estimator (MM)

In the population:

$$E[X_i \underbrace{(Y_i - X_i'\beta)}_{\varepsilon}] \neq 0$$

analogy principle: you do in the sample what you would do in the population:

$E[Y] = \sum_y y \Pr[Y=y]$. In the sample: $\sum_y y \cdot \frac{1}{n} = \frac{1}{n}\sum_y y$

Suppose we have $\underset{k\times 1}{X_i}$ and $\underset{k\times 1}{Z_i}$ with

1. $E[Z_i \varepsilon_i] = 0$
2. $E[Z_i X_i'] \neq 0$ and nonsingular

Then we say that Z_i is a <u>valid</u> instrument for X_i.

If Z_i is a valid instrument for X_i, $\hat{\beta}_{IV} = (Z'X)^{-1}Z'Y$ is consistent for β_0.

$$\hat{\beta}_{IV} = (Z'X)^{-1}Z'Y = (Z'X)^{-1}Z'(X\beta + \varepsilon) = \beta + (Z'X)^{-1}Z'\varepsilon$$

$$= \beta + \left(\frac{1}{n}\sum_{i=1}^{n} Z_i X_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} Z_i \varepsilon_i\right)$$

$$\xrightarrow{p} \underbrace{E[Z_i X_i]^{-1}}_{\text{exists since nonsingular}} \xrightarrow{p} \underbrace{E[Z_i \varepsilon_i]}_{=0}$$

$$\xrightarrow{p} \beta_0 + 0 = \beta_0$$

Since $\text{plim} \frac{1}{n}\sum_{i=1}^{n} X_i(Y_i - X_i'\hat{\beta}_{OLS}) \neq 0$, we use $\hat{\beta}_{IV}$ and have

$\text{plim} \frac{1}{n}\sum_{i=1}^{n} Z_i(Y_i - X_i'\hat{\beta}_{IV}) = 0$ By analogy princ. it is as if we are solving:

$$E[Z_i(Y_i - X_i'\beta)] = 0$$

What is the asymptotic distribution of $\hat{\beta}_{IV}$?

$\sqrt{n}(\hat{\beta}_{IV} - \beta_0) \xrightarrow{d} ?$

In general, we want to find δ s.t. $n^{\delta}(\hat{\theta} - \theta_0) \xrightarrow{d} D$, where D is a "nice" distribution.

$$\hat{\beta}_{IV} = \beta_0 + \left(\frac{1}{n}\sum_{i=1}^{n} Z_i X_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} Z_i \varepsilon_i$$

$$\Rightarrow \hat{\beta}_{IV} - \beta_0 = \left(\frac{1}{n}\sum_{i=1}^{n} Z_i X_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} Z_i \varepsilon_i$$

$$\Rightarrow \sqrt{n}(\hat{\beta}_{IV} - \beta_0) = \underbrace{\left(\frac{1}{n}\sum_{i=1}^{n} Z_i X_i'\right)^{-1}}_{\xrightarrow{p} [E[Z_i X_i']]^{-1}} \underbrace{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} Z_i \varepsilon_i\right)}_{\xrightarrow{d} N(0, V(Z_i\varepsilon_i))}$$

By Jin's theorem by CLT

$$\text{where } V(Z_i\varepsilon_i) = E[Z_i\varepsilon_i\varepsilon_i' Z_i'] - E[Z_i\varepsilon_i](E[Z_i\varepsilon_i])'$$
$$= E[Z_i Z_i' \varepsilon_i^2]$$

$$\Rightarrow \sqrt{n}(\hat{\beta}_{IV} - \beta_0) \xrightarrow{d} N\Big(0, \underbrace{E[Z_i X_i']^{-1}}_{\Sigma_{ZX}^{-1}} \underbrace{E[Z_i Z_i' \varepsilon_i^2]}_{\Phi} \underbrace{(E[Z_i X_i']^{-1})'}_{(\Sigma_{ZX}^{-1})'}\Big)$$

Define $\Delta \equiv \Sigma_{ZX}^{-1} \Phi (\Sigma_{ZX}^{-1})'$. But we don't know Δ.

Let us use the analogy principle to consistently estimate Δ.

Want to estimate $\Sigma_{ZX} = E[Z_i X_i']$.

$$\frac{1}{n}\sum_{i=1}^{n} Z_i X_i' \xrightarrow{p} E[Z_i X_i']$$

How about $\Phi = E[Z_i Z_i' \varepsilon_i^2]$

$\hat{\Phi} = \frac{1}{n}\sum_{i=1}^{n} Z_i Z_i' \hat{\varepsilon}_i^2$ Even if ε_i^2 are iid, $\hat{\varepsilon}_i^2$ are <u>not</u> iid, since $\hat{\varepsilon}_i = Y_i - X_i'\hat{\beta}_{IV}$

Nevertheless, $\hat{\Phi} \xrightarrow{P} \Phi$.

Let $\hat{\Delta} = \hat{\Sigma}_{ZX}^{-1} \hat{\Phi} (\hat{\Sigma}_{ZX}^{-1})'$. Then, by Jin's theorem, $\hat{\Delta} \xrightarrow{P} \Delta$. In doing statistical inference, we are doing two levels of approximation: $\sqrt{n}(\hat{\beta}_{IV} - \beta_0) \overset{A}{\sim} N(0, \Delta)$ and $\hat{\Delta}$ for Δ.

<u>Lecture notes 2:</u>

Claim: If Z_i is a valid instrument, then $Z_i^2 = \begin{bmatrix} z_{i1}^2 \\ \vdots \\ z_{ik}^2 \end{bmatrix}$ is a valid instrument. Thus $\underset{2k+1}{w_i} = \begin{bmatrix} Z_i \\ Z_i^2 \end{bmatrix}$ is a valid instrument. Here, $\ell > k$. What do we do in this situation?

<u>Defn:</u> Let $\underset{n\times n}{A}, \underset{n\times n}{B}$. Then $A \leq B$ iff $(A-B)$ is negative semidefinite.

By an appropriate combination of instruments, we can find a more efficient estimator.

$\underset{n\times 1}{Y} = \underset{n\times k}{X}\,\underset{k\times 1}{\hat{\beta}_{IV}} + \underset{n\times 1}{\hat{\varepsilon}}$ where $\hat{\varepsilon} \equiv Y - X\hat{\beta}_{IV}$. Let $\underset{n\times \ell}{Z}$

$\Rightarrow \underset{\ell\times n}{Z'}\,\underset{n\times 1}{Y} = \underset{\ell\times n}{Z'}\,\underset{n\times k}{X}\,\underset{k\times 1}{\hat{\beta}_{IV}} + \underbrace{\underset{\ell\times n}{Z'}\,\underset{n\times 1}{\hat{\varepsilon}}}_{=0 \text{ by normal equations}}$

Let $\underset{\ell\times\ell}{A}$ be s.t. $\underset{\ell\times\ell}{A} \xrightarrow{P} \Psi$ where Ψ is p.d.

Then $A^{-1}Z'Y = A^{-1}Z'X\hat{\beta}_{IV}$

and $X'ZA^{-1}Z'Y = X'ZA^{-1}Z'X\hat{\beta}_{IV}$

$\Rightarrow \hat{\beta}_{IV} = (X'ZA^{-1}Z'X)^{-1}X'ZA^{-1}Z'Y$

This looks like GLS

$$\hat{\beta}_{IV} = (X'ZA^{-1}Z'X)^{-1} X'ZA^{-1}Z'(X\beta_0 + \varepsilon)$$
$$= \beta_0 + (X'ZA^{-1}Z'X)^{-1} X'ZA^{-1}Z'\varepsilon$$

$\Rightarrow \sqrt{n}(\hat{\beta}_{IV} - \beta_0) \xrightarrow{d} N(0, \Delta)$. How do we choose A optimally?

$$\hat{\beta}_{IV} = \beta_0 + \left(\left(\frac{1}{n}\sum_{i=1}^n Z_i X_i'\right)' A^{-1} \left(\frac{1}{n}\sum_{i=1}^n X_i Z_i'\right)\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^n Z_i X_i'\right)' A^{-1} \left(\frac{1}{n}\sum_{i=1}^n Z_i \varepsilon_i\right)$$
$$= \beta_0 + (S_{ZX}' A^{-1} S_{ZX})^{-1} S_{ZX}' S_{Z\varepsilon}$$

$$\Rightarrow \sqrt{n}(\hat{\beta}_{IV} - \beta_0) = \underbrace{(\underset{\xrightarrow{p} \Sigma_{ZX}'}{S_{ZX}'} \; \underset{\xrightarrow{p} \Psi^{-1}}{A^{-1}} \; \underset{\xrightarrow{p} \Sigma_{ZX}}{S_{ZX}})^{-1} \; \underset{\xrightarrow{p} \Sigma_{ZX}}{S_{ZX}} \; \underset{\xrightarrow{p} \Psi^{-1}}{A^{-1}}}_{\xrightarrow{p} (\Sigma_{ZX}' \Psi^{-1} \Sigma_{ZX})^{-1} \Sigma_{ZX} \Psi^{-1}} \; \underset{\xrightarrow{d} N(0, \underset{\equiv \Phi}{\underline{V(Z_i\varepsilon_i)}})}{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n Z_i \varepsilon_i\right)}$$

$$\xrightarrow{d} N(0, (\Sigma_{ZX}' \Psi^{-1} \Sigma_{ZX})^{-1} \Sigma_{ZX}' \Psi^{-1} \Phi (\Psi^{-1})' \Sigma_{ZX} (\Sigma_{ZX}' (\Psi^{-1})' \Sigma_{ZX})^{-1})$$

Choose A s.t. $\Psi = \Phi$. Then,

$$\sqrt{n}(\hat{\beta}_{IV} - \beta_0) \xrightarrow{d} N(0, (\Sigma_{ZX}' \Phi^{-1} \Sigma_{ZX})^{-1} \Sigma_{ZX}' \Phi^{-1} \Phi (\Phi^{-1})' \Sigma_{ZX} (\Sigma_{ZX}' (\Phi^{-1})' \Sigma_{ZX})^{-1})$$
$$\xrightarrow{d} N(0, \underbrace{(\Sigma_{ZX}' \Phi^{-1} \Sigma_{ZX})^{-1}}_{\Delta})$$

Recall: $A \equiv \hat{\Phi} = \frac{1}{n}\sum_{i=1}^n z_i z_i' \hat{\varepsilon}_i^2$ where, say $\hat{\varepsilon}_i = Y_i - X_i'\hat{\beta}_{OLS}$

$A \to \Phi$, as shown earlier.

Claim: Δ is the "smallest possible matrix" for the given set of instruments.