Recitation Eight - IV Regression

Nathaniel Mark

This Recitation is Ordered According To the Accompanying R Analysis.

Topic 1: What is IV Regression?

Set up: True Model

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \ldots + \beta_{k+1} W_{1,i} + \ldots + u_i$$

But SSA #1 fails for the X's.

That is,

$$E[u_i | X_{1,i}, X_{2,i}, \ldots, W_{1,i}, \ldots] \neq 0$$

(But $E[u_i | W_{1,i}, \ldots W_{r,i}] = 0$)

This SSA #1 failure could be caused by anything.

- Reverse causality
- Omitted Variables
- Errors-in-Variables
- ⋮

In this case β's just estimate relationships (such as corr), not causation.

No matter what the cause is,

IV can fix it!

(not will -- be careful)

TSLS Method: First, we run the "FIRST STAGE"

- Regress

$$X_{1,i} = \gamma_0^1 + \gamma_1^1 W_{1,i} + \ldots + \gamma_r^1 W_{r,i} + \gamma_{r+1}^1 Z_{1,i} + \ldots \gamma_{r+m}^1 Z_{m,i} + \varepsilon_i$$

- Calculate

$$\hat{X}_{1,i} = \hat{\gamma}_0^1 + \hat{\gamma}_1^1 W_{1,i} + \ldots + \hat{\gamma}_{r+m}^1 Z_{m,i}$$

Do this for all endogenous regressors:

⋮

- regress

$$X_{k,i} = \gamma_0^k + \gamma_1^k W_{1,i} + \ldots + \gamma_r^k W_{r,i} + \gamma_{r+1}^k Z_{1,i} + \ldots + \gamma_{r+m}^k Z_{m,i} + v_i$$

- calculate

$$\hat{X}_{k,i} = \hat{\gamma}_0^k + \hat{\gamma}_1^k W_{1,i} + \ldots + \hat{\gamma}_{r,m}^k Z_{m,i}$$

What are we doing here? In a sense, estimating the part of the Xs that is uncorrelated with the u_i error term.

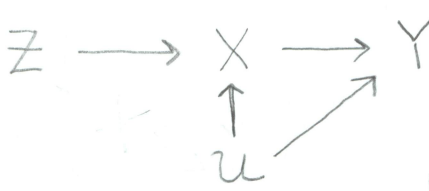
Once these are calculated, we can estimate IV using the "SECOND STAGE" regression

$$Y_i = \beta_0 + \beta_1 \hat{X}_{1,i} + \ldots + \beta_k \hat{X}_{k,i} + \beta_{k+1} W_{1,i} + \ldots + \beta_{k+r} W_{r,i} + \varepsilon_i$$

The $\hat{\beta}$'s will now all be consistent and unbiased.

Now, this is easy to implement. The hard part is getting good instruments

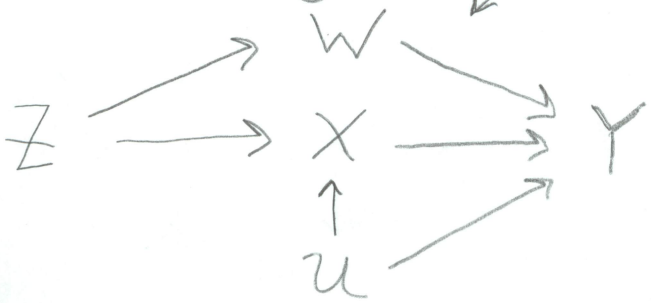
Topic 2: When are instruments valid? PATH DIAGRAMS



→ denotes "is associated with"

without exogenous vars

with exogenous vars



Two Conditions:

- Simple specification
  - of endogenous variable
  - of instrument

1) Relevance Condition
   - $Corr(z_i, X_i) \neq 0$

2) Exogeneity Condition
   - $Corr(z_i, u_i) = 0$

See slide 44 for general case, but think about the above two conditions.

Topic Three: Testing These Conditions

Relevance: As long as $Corr(z_i, X_i) \neq 0$, this condition holds, so it is fairly difficult to test formally. Usually, we proceed in two steps:

1) Eyeball a plot of z_i on X_i and look at $Corr(z_i, X_i)$. Does the correlation look 0?

2) To test if instruments are "weak" -- that is, the correlation is positive but small (or 0), we use the "F-test". Note: only for 1 endogenous var.

F-test for weak instruments

STEP ONE: Run the regression

$$X_i = \beta_0 + \beta_1 W_{1i} + \dots + \beta_r + \beta_{r+1} z_{1i} + \dots + \beta_{r+m} z_{mi} + \varepsilon_i$$

Note this is just the first stage regression.

STEP Two: Conduct an F-test with the following null hypothesis:

$H_0: \beta_{r+1} = 0, \beta_{r+2} = 0, \dots, \beta_{r+m} = 0$

H_a: One or more $\neq 0$

STEP THREE:

Is $\hat{F} > 10$? If so, we say instruments are strong, otherwise, they are weak.

Exogeneity:

We want to test whether $Corr(u_i, z_i) = 0$. Problem: how do we get estimates of u_i without using z_i? Remember your midterm's question! By this logic, we cannot test exogeneity assumption unless we have more

instruments than endogenous vars.

J-TEST

m = # of instruments
k = # of endogenous vars.

STEP ONE:
Run IV model with all Z_i's instrumenting for X_i.

STEP TWO:
Estimate residuals from that model: (OR more than 1 endogenous vars.)

$$\hat{u}_i = Y_i - (\hat{\beta}_0 + \hat{\beta}_1 \hat{X}_{1i} + \hat{\beta}_2 W_{1i} + \beta_3 W_{2i} + \ldots + \beta_{1+r} W_{ri})$$

STEP THREE:
Run a regression of $\hat{u}_i$ on instruments and exogenous vars:

$$\hat{u}_i = \gamma_0 + \gamma_1 W_{1i} + \ldots + \gamma_r W_{ri} + \gamma_{r+1} Z_{1i} + \ldots + \gamma_{r+m} Z_{mi} + \varepsilon_i$$

STEP FOUR: Conduct F-test of null hypothesis that Z_i's have no effect on u_i's
Note: This is the "testing exogeneity part

$H_0: \gamma_{r+1} = 0, \ldots, \gamma_{r+m} = 0$

H_a: One or more $\neq 0$

STEP FIVE: Is $m\hat{F} > \chi^2_{(m-k),.95}$?

↓ critical value of χ^2_{m-k} test at $\alpha = .05$

If so, we reject exogeneity, but this does not tell us which instrument fails.