Osama Mahmood

Econometrics

10/May/2016

Review

Endogeneity

Instrumental Variable

I V Estimator

2 SLS

Multiple Endogenous Variables

Weak Instrument

Over-identifying Restriction

Cases of M.E.

Case 1: $y = y^* + e_0 \Longrightarrow Var = \sigma_u^2 + \sigma_{e_0}^2$] Loss of efficiency But unbiased

Case 2: $x = x^* + e_1 \quad Cov(x, e_1) = 0$

Case 3: $x = x^* + e_1 \quad Cov(x^*, e_1) = 0$

Case 2 $\Longrightarrow$ $y = \beta_0 + \beta_1 x + (u - \beta_1 e_1)$

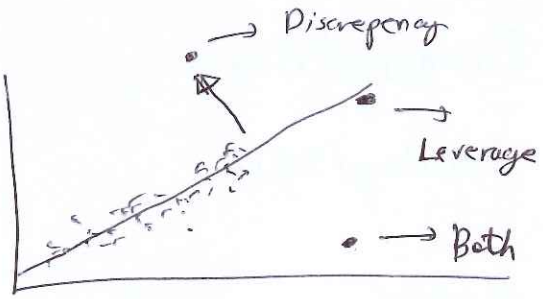
$\hookrightarrow Var = \sigma_u^2 + \beta_1^2 \sigma_{e_1}^2$

Case 3 $\Longrightarrow$ $Cov(x^*, e_1) = 0$

$Cov(x - e_1, e_1) = 0 \Longrightarrow Cov(x, e_1) = Cov(e_1, e_1) = \sigma_{e_1}^2 \neq 0$

$$\Longrightarrow E(\hat{\beta}_1) = \beta_1 \frac{\sigma_{x^*}^2}{\underbrace{\sigma_{x^*}^2 + \sigma_{e_1}^2}_{<1}}$$ (Attenuation Bias)

① Missing observations are a problem if they are a function of the Y variable



Outliers

Influence = Leverage x Discrepancy

To detect : rvfplot

(STATA) lvr2plot (plots lev against dis)

→ reg

→ predict ~~r~~esid, rstudent

Use areg y x1 x2 (based on median)

rreg y x1 x2 (assigns less weights to outliers)

Endogeneity

X is endogenous if

$$Cov(x, u) \neq 0$$

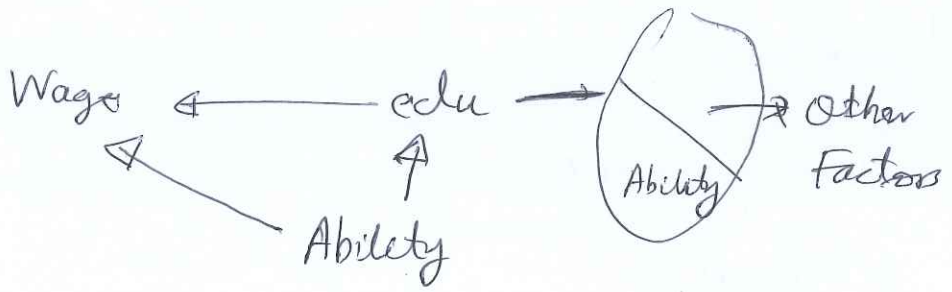
Wage studies have a problem of endogeneity due to the fact that ability is unobservable.

In case of price-quantity relationship

$$P \longrightarrow Q$$
$$P \longleftarrow Q$$

price and quantity both effect each other.

Hence quantity is effected by price and other factors such as ~~what~~ weather or political tensions.



~~Wher~~ Problem of Endogenecty

4. OVB $E[\hat{\beta}_1] = \beta_1 + \beta_2 \delta$

~~2. Missing Data~~

2. Non-random Missing Data/Self-Selection

3. Measurement Error

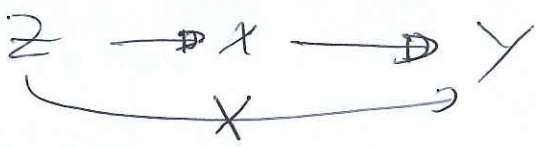
4. Functional form Misspecification

$y = \beta_0 + \beta_1 x + u$

Qualities of an Instrument

1. $Cov(z, x) \neq 0$ (Relevance)

2. $Cov(z, u) = 0$ (Exogenecty)



Z should not directly impact Y.

If z effects Y then $Cov(z, u) \neq 0$

(as u is a part of Y).

$Cov(y, z) = \beta_1 Cov(x, z) + Cov(u, z)$

$\Rightarrow \hat{\beta}_1^{IV} = \frac{Cov(y, z)}{Cov(x, z)} \sim \beta_1 = \frac{Cov(y, x)}{Var(x)}$ ($Cov(u, z) = 0$)

2 Stage Least Squares

1. $x = \gamma_0 + \gamma_1 z + e$

$x = \hat{x} + e$

↳ variation due to z

2. $y = \beta_0 + \beta_1 \hat{x} + u$

$$\gamma_1 = \frac{Cov(x, z)}{Var(z)}$$

$$\beta_1 = \frac{Cov(y, \hat{x})}{Var(\hat{x})} = \frac{Cov(y, \gamma_1 z)}{Var(\gamma_1 z)}$$

$$= \frac{\gamma_1 Cov(y, z)}{\gamma_1^2 Var(z)}$$

$$= \frac{Cov(y, z)}{\gamma_1 Var(z)}$$

$$= \frac{Cov(y, z)}{Var(z)} \times \frac{Var(z)}{Cov(x, z)}$$

$$= \frac{Cov(y, z)}{Cov(x, z)}$$

Multiple Endogeneous Variables

If we have two endogenous variables then we will need two instrumental variables.

If # endogenous $\geq$ # instruments

$>$ $\Rightarrow$ under-identification

$<$ $\Rightarrow$ over identification

$=$ $\Rightarrow$ exact identification

If we have more instruments then we can run different IV regressions and test our results from different regressions if they confirm to each other.

Weak Instrument

$$\beta^{IV} = \frac{Cov(y,z)}{Cov(x,z)}$$

$\rightarrow$ small

to check for weak instruments run regression of x on z's and check for the t-statistic.