Econometrics (3-May-2012)

- Measurement Error
- Instrumental Variable
- Problem of indogeneity
- Violation of Assumption 3

① Measurement : When one variable is measured with error

② ~~Under~~ Case 1: (ignore)
Case 2: (ignore)
Case 3: (cannot ")

eg. calculation of GDP is full of measurement error.

- Selection Effects

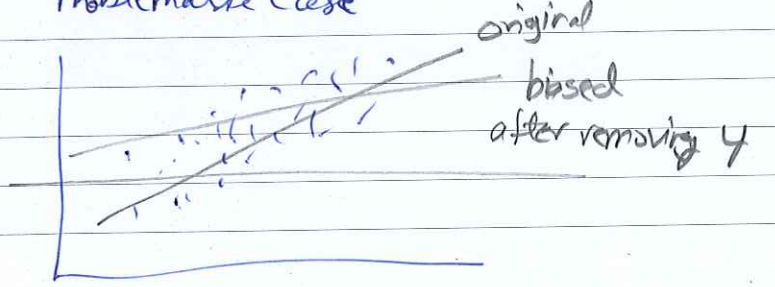
When studying wage-educ relation we might suffer from selection effects bcz smart people go to higher level of educ.

③ Ken Ragoff

④
Data cases: (In example of wage-educ study)

I : Missing based on X (highly educated do not answer)

II : Missing based on Y (highly wealthy people choose not to reply)
↳ Problematic Case



⊕ Do the STATA exercise of showing the previous conclusion by generating numbers.

(?) gen s = rbinomial(1, yp)

⊕ See textbook for Selection Effects.

Proxy Variables

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3^* + u$$

For example, we might take "I.Q." as a proxy for "ability" when doing wage studies.

We should have

$$\underbrace{x_3^*}_{\text{ability}} = \delta_0 + \delta_1 \underbrace{x_3}_{\text{I.Q.}} + v_3$$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 (\delta_0 + \delta_1 x_3 + v_3) + u$$

$$y = (\beta_0 + \beta_3 \delta_0) + \beta_1 x_1 + \beta_2 x_2 + \beta_3 \delta_1 x_3 + (u + \beta_3 v_3)$$

Conditions

① δ_1 is significant and not zero

② v_3 is not correlated with δ's and β's

$$\text{Cov}(x_1, v_3) = 0$$
$$\text{Cov}(x_2, v_3) = 0$$
$$\vdots$$

⊙ Lagged Dependent Variable

$$Conflict = \alpha_0 + \alpha_1 Corruption + \alpha_2 Inst. + \varepsilon$$

Since we cannot measure α_2 we include a lagged dependant variable as it also includes the effect of institutions. We might get

$$Conflict_{00} = \alpha_0 + \alpha_1 Corruption + \alpha_2 Conflict_{90} + \varepsilon$$

Outliers

Influence = Leverage x Discrepency

Outliers that have a high absolute x-axes values

Outliers that have a high absolute y-axis values.

We have an influence when we have both.

Simulation

⊕ Try the simulation

see that only leverage or discrepency reduces significance but not the coefficients

⊖ lvr2plot finds outliers quickly

to change an obs.: replace y = 999 in 25

to drop the obs: drop in 25

to get labels: lvr2plot, mlabel(id) / extremes y x

~~st~~

② Studentized Residual

extremes stdresid

predict rstudent, rstudent (?)

extremes rstudent

Use ~~rreg~~ rreg ~~y~~ x (to eliminate outliers)

Use qreg (as a nonlinear iteration method).