Instrumental Variables Estimation and Two Stage Least Squares

Wooldridge Ch.15 section 15.1-15.4

1 Motivation

Assumption MLR4 or TS3 does not hold! So, $\hat{\beta}_{OLS}$ would be biased.

Because of various possible reasons. For example

1. Omitted Variable

2. Endogeneity or Simultaneity

From problem 1) Omitted Variable bias, if we have a good proxy variable for "ability", we can estimate

Let "z" be an instrumental variable, it has to satisfy two conditions

What might be correlated with *educ* but not *abil*?

Example: Wongsurawat (2004) Pornography and Social Ill

$$crime\_rate = \theta_0 + \theta_1 porn\_magazine + \theta_2 population + \theta_3 dangerous + e$$

- If we cannot find a measurement for "$dangerous$", then the error term becomes "$u = \theta_3 dangerous + e$".
- $Cov(porn\_magazine, u) \neq 0$
- Use $z =$ ________________________________
- This z satisfies both requirements for a valid IV because ..

2 Consistency of the estimators

As long as the two requirements for a valid IV are satisfied, $\hat{\beta}$ would be consistent.

** Note that we don't call $\hat{\beta}$ under the IV method $\hat{\beta}_{OLS}$ anymore.

- $\hat{\beta}_{OLS}$ usually refers to $\hat{\beta}$ obtained from an estimation that does not employ IV to correct the problem.
- $\hat{\beta}_{IV}$ usually refers to $\hat{\beta}$ obtained from an estimation that employs IV to correct the problem.

To prove for consistency of $\hat{\beta}_{IV}$ we start with a simple regression function

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

Inference

3 Properties of IV with a Poor Instrumental Variable

For $\hat{\beta}$ to be consistent, we require $plim\hat{\beta} = \beta$.

4 IV Estimation for the Multiple Regression Model

given

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + u_1$$

Let
y_1 = dependent variable of interest
y_2 = explanatory variable that is endogenous ($corr(y_2, u_1) \neq 0$)
z_1 = explanatory variable that is exogenous ($corr(z_1, u_1) = 0$)
z_2 = instrumental variable (also exogenous because $corr(z_2, u_1) = 0$)
We know that

5 Two-Stage Least Squares (2SLS)

Suppose now we have 1 more IV

$$z_3 = \text{cost of input}$$

$corr(z_3, u_1) = 0$

$corr(z_3, y_2) \neq 0$

We know that if $corr(z_1, y_2) \neq 0$, $corr(z_2, y_2) \neq 0$, $corr(z_3, y_2) \neq 0$, we can write:

The 2SLS method

Stage 1:

Stage 2:

STATA Command: