Econometrics Lab 5

Dealing with Endogeneity

1. IV Estimation of Return to Education. This exercise examines the effect of education on earnings using instrumental variables. We use the dataset wage2.csv from Wooldridge. The variables in the datasheet are as follows:

lwage	natural log of wage
edu	year of education
sibs	number of siblings
brthord	birth order (must be greater or equal to 1, 0 stands for missing value)

(a) Eliminate observations with missing values ($brthord = 0$), and estimate the following model using OLS,

$$\log(wage) = \beta_0 + \beta_1 edu + u.$$

(b) It is well known that *edu* may be endogenous in the above model. Estimate the model using the instrument *sibs* for *edu*, Report parameter estimates, standard errors, and R^2. Compare your result with that obtained from OLS. (Note that R^2 can be negative in IV estimation, think about why.)

(c) The variable *brthord* is birth order (brthord is one for a first-born child, two for a second-born child, and so on). Explain why *edu* and *brthord* might be negatively correlated. Regress *edu* on *brthord* to determine whether there is a statistically significant negative correlation.

(d) Re-estimate the model with both instruments, *sibs* and *brthord*, using TSLS (two-stage least square). Compare your results with those obtained from OLS and IV using *sibs* only.

2. Panel Data Analysis of Deterrence of Murder Use panel data murder.csv (from Wooldridge), which contains

id	id numbers for states (in the United States)
year	year
mrdrte	murders per 100,000 population
exec	total executions, past 3 years
unem	annual unemployment rate

(a) Estimate the following linear regression using pooled OLS,

$$mrdrte_{it} = \beta_0 + \beta_1 exec_{it} + \beta_2 unem_{it} + u_{it}.$$

Discuss the sign of your estimates. Do they make sense? If there are "wrong" signs, explain what may have gone wrong.

(b) Estimate the following fixed-effect panel data regression using LSDV (least square with dummy vairables),

$$mrdrte_{it} = \beta_1 exec_{it} + \beta_2 unem_{it} + \mu_i + u_{it}.$$

Discuss your result and discuss how panel data may have resolved the problem using pooled OLS.