ECO3211 Problem Set 2

IV Estimation

A researcher is interested in studying how years of education (s_i) affects health outcomes (H_i). He wants to run the ideal regression model $H_i = \alpha + \beta S_i + A'_i\gamma + u_i$, but can only control for a subset of X'_i of all the variables in A'_i. So there might be omitted variable bias in $\hat{\beta}_{\text{OLS}}$ estimated from the regression equation $H_i = \alpha + \beta S_i + X'_i\gamma + u_i$.

To solve the endogeneity of education, he exploits the 1986 Compulsory Education Law and the schooling system in China. 1986 Compulsory Education Law in China mandates 9 years of education, and all children who have reached the age of six by the start of the academic year shall be enrolled in schools. Meanwhile, China has followed the former Soviet Union and used September as the school opening month since its establishment. Given these features, students born after August (Sep. – Dec.) start school at an older age and are maturer and perform better at school. And then the researcher decides to use the birth quarter as an instrumental variable for total years of schooling. He defines the instrumental variable as:

$z_i = I[birth\ month_i \geq 9]$.

Part I:

a) Write down the first stage, second stage, and reduced form regression equations.
b) Write down the standard 2SLS estimator $\hat{\beta}_{2SLS}$ for using z_i as an instrument for s_i.
c) Show that $\hat{\beta}_{2SLS}$ equals to the ratio of reduced form coefficient to the first stage one.
d) Describe and explain the most important two conditions that would have to hold in order for this IV to "work".
e) Could you test those conditions? If your answer is yes, how would you do it? If not, why?

Part II: Please help the researcher to finish this empirical study using the 2010 China Family Panel Studies (baseline) survey data.

Steps:

1. Download the data and questionnaires from Blackboard or from https:// opendata.pku.edu.cn/dataverse/CFPS?language=en. Keep adult individuals born between 1965 and 1985
2. Divide the birth month into two groups, half=1 if birth month<9, and half=2 if birth month >=9

3. Graph the relationship between total years of schooling and birth cohort
4. IV estimation: use indicator for being born after August as IV for years of schooling. Check the validity of your IV.
5. Investigate the impacts of education on both physical and mental health outcomes, explain your empirical findings

6. Check the heterogeneous effect across the gender, Hukou status (urban or rural), interpret your empirical results

Notes:

Problem set 2 will due on Nov. 10th 2020. You should upload three separate documents electronically: one that contains your typed answers to the problem set questions, one well-organized and well-commented Stata do file, and one Stata log file.