ECON3360 Causal Inference for Microeconometrics

Tutorial 7: Stata Application of advanced panel data methods

Instructor: Julie Moschion

Problem I: the deterrent effect of execution on murder

Use the data in murder.dta for the following questions. These include state-level data on murder rates and executions for the U.S.
Consider the following model:

$$mrdrte_{it} = theta_t + \beta_1 exec_{it} + \beta_2 unem_{it} + c_i + u_{it} \quad (1)$$

where $theta_t$ denotes different year intercepts and c_i is the unobserved state effect.

(1) If past executions of convicted murderers have a deterrent effect, what should be the sign of β_1? What sign do you think β_2 should have?

(2) Using just the years 1990 and 1993, estimate equation (1) by pooled OLS. Ignore the serial correlation problem in the composite errors. Do you find any evidence of a deterrent effect?

(3) Now, using 1990 and 1993, estimate the equation (1) by fixed effects. Now, is there any evidence of a deterrent effect? What does that tell you about the potential bias in the pooled OLS?

(4) Now, using 1990 and 1993, estimate the equation (1) by first differencing. Use the heteroskedasticity-robust standard error and do not include a constant. Is there any evidence of a deterrent effect?

(5) Find the state that has the largest number of executions in 1993. How much bigger is this value compared with the next highest value?

(6) Estimate the equation using first differencing, while dropping Texas from the analysis. Use the basic and heteroskedasticity-robust standard errors. Now, what do you find?

(7) Use all three years of data and estimate the model by fixed effects with robust standard errors (including Texas in the analysis). Now cluster your standard errors to account for serial correlation. What do you find?

Problem II: a study of the effect of small class size with the Tennessee Project STAR

Background The dataset for this exercise comes from the paper by Krueger (1999) "Experimental Estimates of Education Production Functions", published in the Quarterly Journal of Economics, vol. 114, No.2, p. 497-532. He analyses data on 11,600 students and their teachers who were randomly assigned to different class sizes from kindergarten through 3rd grade. Use the data in starwide.dta for the following questions. The data include students from kindergarten to 3rd grade.

(1) The loaded dataset is in a wide form. "aide*", "small*" and "regular*" are dummy variables indicating whether the child was in an aide, a small, or a regular class in grade * (with *: 0 to 3). Construct four variables for the cumulative sum of being in a small class over the four grades. Generate similar variables for being in an aide class.

(2) Convert the wide data to a long format with *stdntid* the unit id and *grade* the time id.

(3) Now we have four year data for each students. Define data as panel data with *stdntid* the unit id and *grade* the time id.

(4) Replicate the following figures from Krueger (1999): these are densities of SAT scores distinguishing whether students are in a regular or small class, for each grade available in the data.

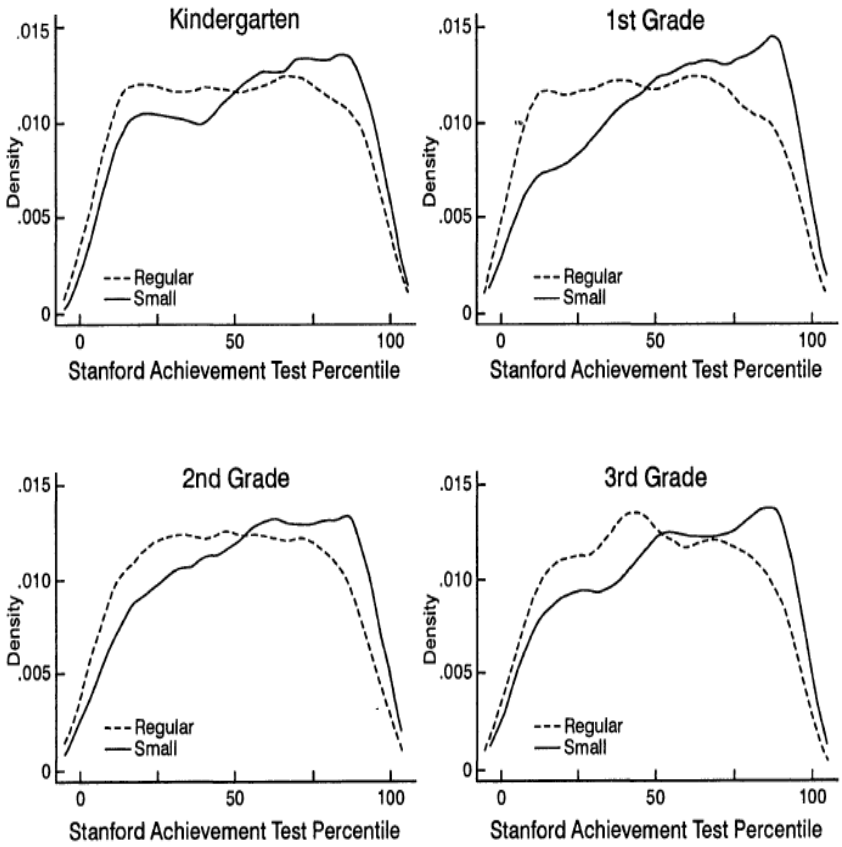

FIGURE I
Distribution of Test Percentile Scores by Class Size and Grade