Sanjay Satish

Duke University – Economics Honors Thesis

Regression Equation & Preliminary Empirical Strategy

Note: I will turn in a detailed theoretical and empirical framework next week, this simply outlines my current regression strategy. For now, I have limited my topic to understanding the affect of attrition writ large. That is, I will examine the peer effects of attrition from STAR across all students who left the experiment – not just understanding the impact of those that left for private schools. I am not sure how many observations I have given that I still have some merging to do between school-level controls and student-level data; however, I believe I will have more than enough to satisfy robustness checks. I also aim to center my thesis around the measurement of peer effects between grades 1 and 2 (though I will also run regressions across other grades depending on the extent of the attrition) as this is somewhat isolated from the effect of class-switching/re-randomization which occurred starting in the 1st grade. That being said, I am to run some form of the following regression:

$$T_{i,g,c,t,s} = \beta_0 + \beta_1(L_{i,g,c,t-1,s}) + \beta_2(A_{i,g,c,t-1,s}) + \beta_3(\overline{CA}_{i,g,c,t-1,s}) + \beta_4(\overline{PA}_{i,g,c,t-1,s}) + \beta_5(X_{i,g,c,t-1,s})$$
$$+ \beta_6(T_{i,g,c,t-1,s}) + \alpha_s + \gamma_{t-1} + \epsilon$$

Where:

- $T_{i,g,c,t,s}$ is the test score for student i in grade g in classroom c in school s at time t.

- $L_{i,g,c,t-1,s}$ is the proportion of students in student i's class that left for private schools at time $t-1$.

- $A_{i,g,c,t-1,s}$ is student i's ability, measured by their test score in time $t-1$.

- $\overline{CA}_{i,g,c,t-1,s}$ is classmate ability for student i. That is, the average test score of the students in student i's class that don't leave the experiment, measured by their test score in time $t-1$.

- $\overline{PA}_{i,g,c,t-1,s}$ is the peer ability for student i. That is, the average test score of the students that left for private school in $t-1$. This is perhaps my main coefficient of interest as I am interested in how the leavers' ability affects students.

- $X_{i,g,c,t-1,s}$ is a vector of controls for each student including observable student, teacher, and school data. Examples of this include student attendance, teacher's highest education level, student's race, percentage of school on free and reduced lunch, school geography (e.g. urban, rural, etc.), and others.

- $T_{i,g,c,t-1,s}$ is the class type student i was enrolled in during time $t-1$ (e.g. small, regular-sized, or regular sized with teachers aide).

- α_s is a school fixed effect.

- γ_{t-1} is a year fixed effect.

- ϵ is the regression error term.