ECON 6160
Econometric Analysis
Assignment 6
Due April 24th, 2020

This assignment picks up where Assignment 5 left off. You will continue to use the same data set, based on Krueger (1999). Moreover, you will make use of the variables that you generated in Assignment 5 so your code from that assignment will give you a good start for this assignment.

You should do this assignment set by writing a program (a *.do file in Stata). This will be turned in to your TA (via email) and should be well annotated. You will also turn in a solution packet in class that includes relevant Stata output and responses to any questions.

As usual, keep in mind that your assignment will be returned to you for revision if it does not run on our machines unaltered or when a mistake in your analysis has been discovered.

0. Getting Started

Create a folder for this problem set (or use the same folder from Assignment 5). Put the data set in this folder along with a do file titled Ass6_LastnameFirstname.do. As before, your do file should begin with the code I initially provided so that it will run on my computer and your TAs computer and so that you have the necessary variables. It should then include the code from Assignment 5 in which you generated additional variables.

1. Getting the data ready for analysis

We did this last time. No need to show these results again in your solution packet. Note that we concluded Assignment 5 by comparing the characteristics of students and teachers assigned to small, regular, and regular+aide classes in order to see whether we believe that they were randomly assigned.

2. Formal consideration of random assignment

a. We will now consider whether any differences in characteristics across the three groups are statistically significant. In order to do this for each characteristic, we will regress the characteristic on an indicator variable taking one for individuals assigned to small classes (and zero otherwise) and an indicator taking one for individuals assigned to regular+aide classes (and zero otherwise). That is, we will estimate regressions of the following form: $x_i = \alpha+\beta Small_i + \gamma RegAide_i +\varepsilon_i$. Based on this regression equation, what is the expected value of x for someone assigned to a small class? For someone assigned to a regular+aide class? For someone assigned to a regular class? (I'm not asking for any statistical analysis here; work this out by hand as we did on the board in class.)

b. Based on your response to (a), provide an intuitive explanation of what β represents. And what γ represents.

c. Now present the regression results (using Stata) for the following characteristics: free or reduced-price lunch (0/1 indicator), white or Asian student race (0/1 indicator), age in 1985 assuming the individual already had their birthday, student's teacher is white or Asian (0/1 indicator), teacher's number of years of experience. In order to estimate the regression, you will use the command "reg." Also, please use the ", robust" option to this command.

d. Are there any statistically significant differences in characteristics between those assigned to small classes and

those placed into regular classes? Or between those placed into regular+aide classes and those placed into regular classes?

e. Explain what you could do to determine whether there are any statistically significant differences between those assigned to small classes and those assigned to regular+aide classes?

f. In actuality, Project Star did not involve pure random assignment of children to different class types. Instead, children were randomly assigned *conditional on their school*. This means that all students at the same school have the same probability of being placed into a small class. However, students at a school with resources for several small classes would have a higher probability of being placed into a small class than students at schools with fewer small classes. As such, it would not necessarily be a cause for concern if there were systematic differences in student characteristics across class types in the previous exercises. It **would** be a cause for concern if we saw systematic differences in student characteristics across class types *conditional on schools*. In order to condition on the school a child attends, we need to control for schools with a comprehensive set of indicator variables for each school. That is, school fixed effects. In order to create all of the indicator variables, you can use the command: "xi i.schid, noomit" which will create the set of variables _Ischidk_1 (an indicator taking one for students at school 1), _Ischidk_2 (an indicator taking one for students at school 2), etc. Estimate the same regressions that you did in part (c) but with these controls included in the regression model. Note: As a shortcut for controlling for all of these variables in your regression, you can input "_Ischidk*" as opposed to "_Ischidk_1 _Ischidk_2 … _Ischidk_80." To make your solution packet easier to read, please omit the regression coefficients associated with these variables and just present the regression estimates of β and γ.

g. Do the estimates from (f) support the notion that class types were randomly assigned conditional on schools?