

# Econ 758 Homework 2

Ege Can, John Appert

March 23, 2019

## 1 Question 1: Summary Statistics

- a Calculate the share of students receiving free lunch, the share of white Asian students, the average age in 1985, the attrition rate, average class size, and gender by the enter-variables 2generated above and by treatment status (small class, regular class, and regular class with aide). Present your summary statistics for STAR participants in Table 1, structured as in Table I in Krueger (1999). You do not need to provide standard deviations.
- b Calculate the average of the percentile ranks of the math and reading tests for every individual in each year (name the variables testk etc.). (Hint: If one subtest score is missing, take the percentile score corresponding to the only available test as in Krueger (1999), fn.11.) Add the average values by the enter-variables and by treatment status to Table 1.
- c Comment on the characteristics of students assigned to the “small class” treatment, who entered STAR in kindergarten.

## 2 Question 2: Random Assignment

The first question to ask about a randomized experiment is whether the randomization successfully balanced subjects’ characteristics across the different treatment groups.

- a The STAR data does not include any pre-treatment test scores. Do you think that this is a problem? Explain briefly.
- b Compare the student characteristics collected in Table 1 across treatments. Formally test the null hypothesis of no difference across treatment groups using an F-test. Add both F-statistics (rounded to 2 digits after the decimal point) and p-values to Table 1. Do you think that randomization was successful, and why/why not? (Hint: Use the regress and test commands.)
- c In fact, the treatment was randomly assigned to students and teachers within schools. For each of the variables in Table 1, test the null hypothesis that, conditional on school of attendance, there are no significant differences across treatment groups. Once again, give both F-statistics and p-values and add them to Table 1. (Hint: Use school dummy variables.)

Are the results consistent with random assignment conditional on school attendance? Explain.

## 3 Question 3: OLS estimates of Class Size Effects

- a Run OLS regressions with the average percentile test score (testk etc.) constructed in Question 1b as a dependent variable (assuming that the errors are iid). Produce regression results similar to those given in columns (1) to (3) of Table V in Krueger (1999) and present them in Table 2 A-D.
- b Interpret the coefficients on the small class and the regular/aide class indicators for kindergarten children in column (1) of Table 2.

- c Do the coefficients on the small class indicator change if additional explanatory variables are added to the model? What does this tell you about selection on observables?
- d Suppose that test scores are measures of true skills that are noisy (i.e., subject to measurement error). How do you expect this to affect your estimates of the coefficient? Explain briefly.

## 4 Question 4: Instrumental Variable Regression

So far, we have focused on actual class type. As described in Krueger (1999), there might be transitions between class types after the first year of the program. Students were moved between class types because of behavioral problems or parental complaints.

- a Reproduce Table IV in Krueger (1999) showing transitions between class types in adjacent grades. Display your results in Table 3.
- b Generate a variable that contains each student's initial assignment to a class type.
- c Krueger (1999) argues that initial class assignment is highly correlated with actual class assignment in later years. Show that initial class assignment is a good predictor of actual class treatment in each grade.
- d Consider using initial class assignment as an instrumental variable for actual class treatment. Name the two conditions that a valid instrument needs to fulfill. Do you think that the requirements are met by initial class assignment, and why/why not?
- e For each grade, run an instrumental variable regression of average math reading test score on actual class type dummies (small class and regular with aide class), using initially assigned class type dummies as instruments (without further control variables). Report your results in Table 4. How do your results compare to those found in Table 2? Do they suggest that non-random transitions between class types were a problem?

## 5 Code

### 5.1 Python Code from Jupyter Notebook

### 5.2 Stata Code