Marissa Reuther

Econ 582 - Fall 2019

10.25.19

Replication Study – First Steps

1. Draft an Introduction – First Steps from Krueger’s “Experimental Estimates of Education Production Functions

The outcome of interest Kreuger’s study is student achievement level, which is specifically measured through student scores on two standardized tests (the SAT and the Tennessee Basic Skills First (BSF)). Studying the effects of these standardized test scores are important because they are a significant part of what dictates students’ future educational attainment, such as getting into better colleges and universities.

Kreuger looks at several factors that could potentially affect student achievement. Class size is the main independent variable of interest and he hypothesizes, based on previous work pertaining to project STAR, that students placed in smaller class sizes tend to have better standardized test scores, holding all else constant.

Kreuger also analyzes the effects of some other independent variables such as teacher ability and whether a classroom had a full-time teacher aid present.

For standardized test scores (variable *Y*), Kreuger uses percentiles as a measure of student achievement level. For his main analysis, he uses percentiles based only on SAT scores. Later in the paper, he alters the model by using percentiles only from BSF scores and did not find much difference in the results. In both the SAT and BSF scores, a higher percentile corresponds to a student with a higher achievement level.

For class size (variable X1) in the main analysis, he uses dummy variables to indicate which class category each student was placed into. The three categories were small (13-17 students), regular (22-25 students) and regular/aid (22-25 students plus a teacher aid). The first dummy variable, *SMALLcs*,represents either a small size or a regular size. The second dummy variable, *REG/Acs*, represents if a teacher aid was present in the class.

If this experiment did not use randomization for assigning students to different class sizes, Krueger would have needed a conditional independence assumption to be able to interpret the effects of class size on student achievement as causal effects. For conditional independence, unobserved characteristics within the error term need to be uncorrelated with our regressors of interest, after holding all other characteristics constant. In the case of interpreting the causal effects of class size on student achievement, there cannot be any unobservable characteristics correlated with class size.

In the case of a non-randomized experiment, omitted variables could violate the conditional independence assumption, such as student’s inherent ability. Students who are ‘over-achievers’ or who may be smarter than others will have inherent ability correlated with standardized test scores. Additionally, their inherent ability can be correlated with class size if they are placed in ‘gifted’ programs or are given some similar type small class size advantage over those with lower inherent ability.

For this experiment, project STAR used randomization for placing students in different class sizes to overcome possible omitted variable bias. If class size was truly determined randomly, then it should also be randomizing the distribution of student’s inherent abilities.

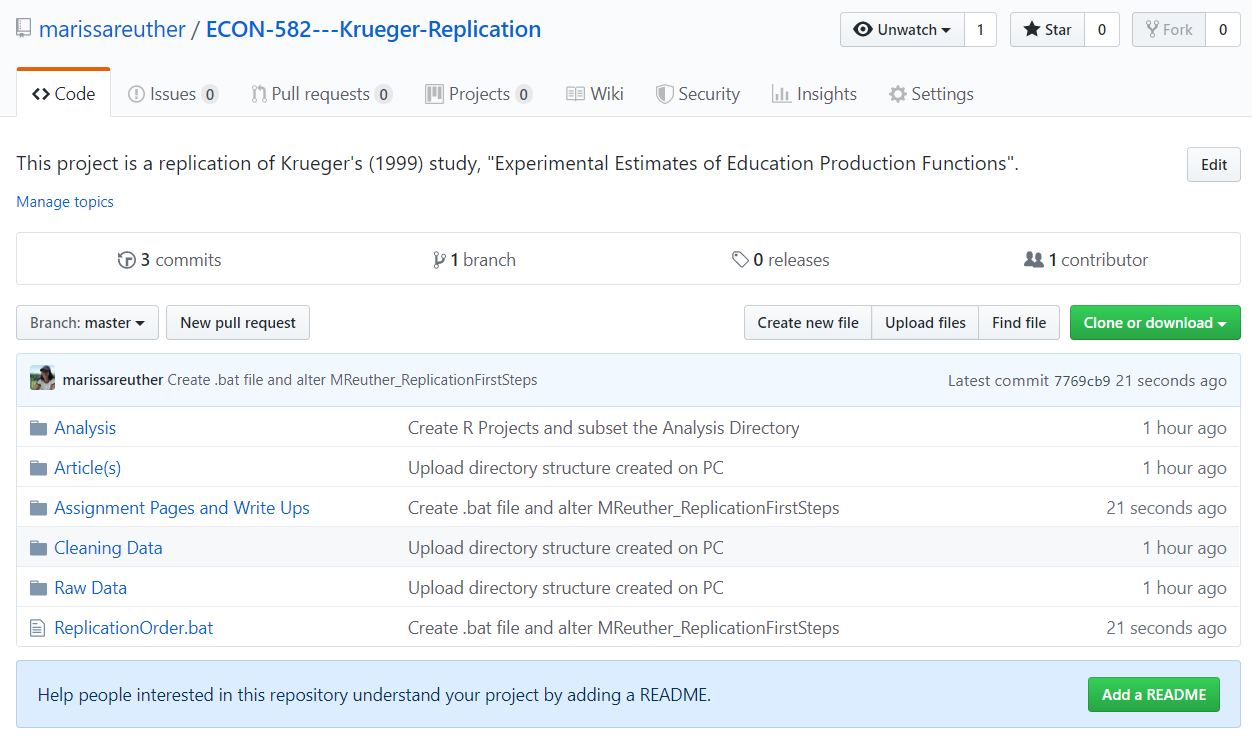
In Table 1, Kreuger reports various mean characteristics across different controls for each class type. The table shows that the class sizes roughly have the same distribution of these observable characteristics, due to randomization. Because randomization equalized these observable characteristics across class types, it likely also equalized unobservable characteristics across class types.

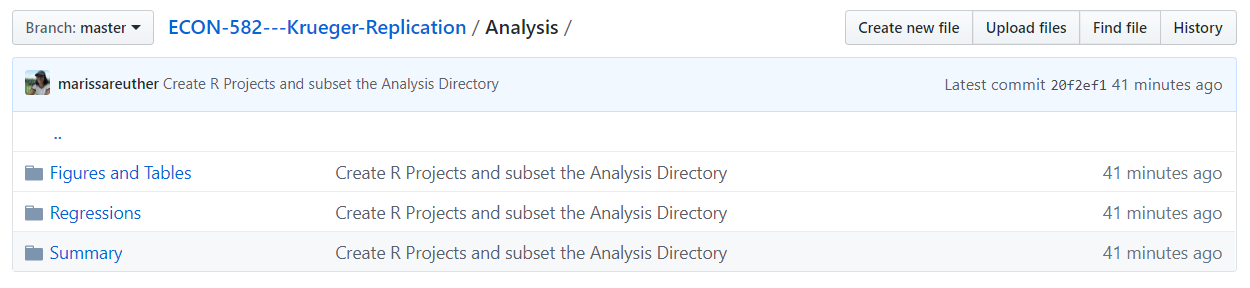
This randomization should make class size uncorrelated with inherent ability and other unobservable characteristics, thus eliminating any potential omitted variable bias.

2. Questions pertaining to “Code and Data for Social Sciences” – Chapters 1 - 4

A – Describe the directory system

The image below shows the overall directory system. I created separate folders for the analysis portion, the articles and readings, the assignment pages/write ups, cleaning data, and the raw data. I also have created a .bat file.



 I also created subfolders within the analysis folder so I don’t have one giant script for all of this portion of the assignment. Below is the current structure of my analysis folder and I plan on altering it as I work.

B – What will be automated?

I plan on using a .bat file which will keep track of the correct order for scripts to run in. It will also run all scripts at once. (I think that is the case; I have never used a .bat file for this so I want to use this assignment as a chance to learn about using .bat files for automation).

C – Plan for version control?

I am using a github repository (<https://github.com/marissareuther/ECON-582---Krueger-Replication>).

This will keep a log of all updated files with descriptions. Github also tracks when new files are created.

D – Backup work?

I will push all local changes to Github so they are backed up remotely. I also have everything backed up on Dropbox in case I forget to push changes or my computer crashes before I can push changes.