Marissa Reuther

Econ 582, Fall 2019

10.25.19

Replication Study – First Steps

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

A – Outcomes of interest in the study. Why are they important?

The outcome of interest is student achievement level, specifically measured through scores on standardized tests (SAT and BSF – Tennessee Basic Skills First)

This is important because class size is something that can be fixed relatively easy in terms of educational policy. Increasing test scores are important because they are part of what dictates students’ future education attainment (like getting into better colleges and universities)

B – What does Kreuger hypothesize affects the outcome you listed in part (A)?

He looks at several things 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 in smaller class sizes tend to have better test scores.

He also looks at some other independent variables such as teacher ability and whether a classroom had a teacher aid present.

C – How does Krueger measure Y (the outcome) and X1 (the primary independent variable)

For test scores, he uses percentiles to measure for student achievement level. For the main analysis, he uses percentiles from SAT scores. He later alters the model to use percentiles from BSF scores and does not find much difference in the results. A higher percentile corresponds to a student with a higher achievement level.

For class size in the main analysis, he uses dummy variables to indicate which class size each student was in. 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 is a teacher aid was present in the class.

D – Describe the conditional independence assumption necessary to interpret causal effects of X1 if we did not randomize.

For conditional independence, unobserved characteristics within the error term need to be uncorrelated with our regressors of interest. In the case of interpreting the causal effects of class size, there cannot be any unobservable characteristics correlated with class size.

E – Describe one potential omitted variable that could violate conditional independence

One omitted variable can be 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.

F – Does randomization overcome the omitted variable bias?

Yes, randomization overcomes possible omitted variable bias. If class size is truly determined randomly, then it should also be randomizing the distribution of student’s inherent abilities.

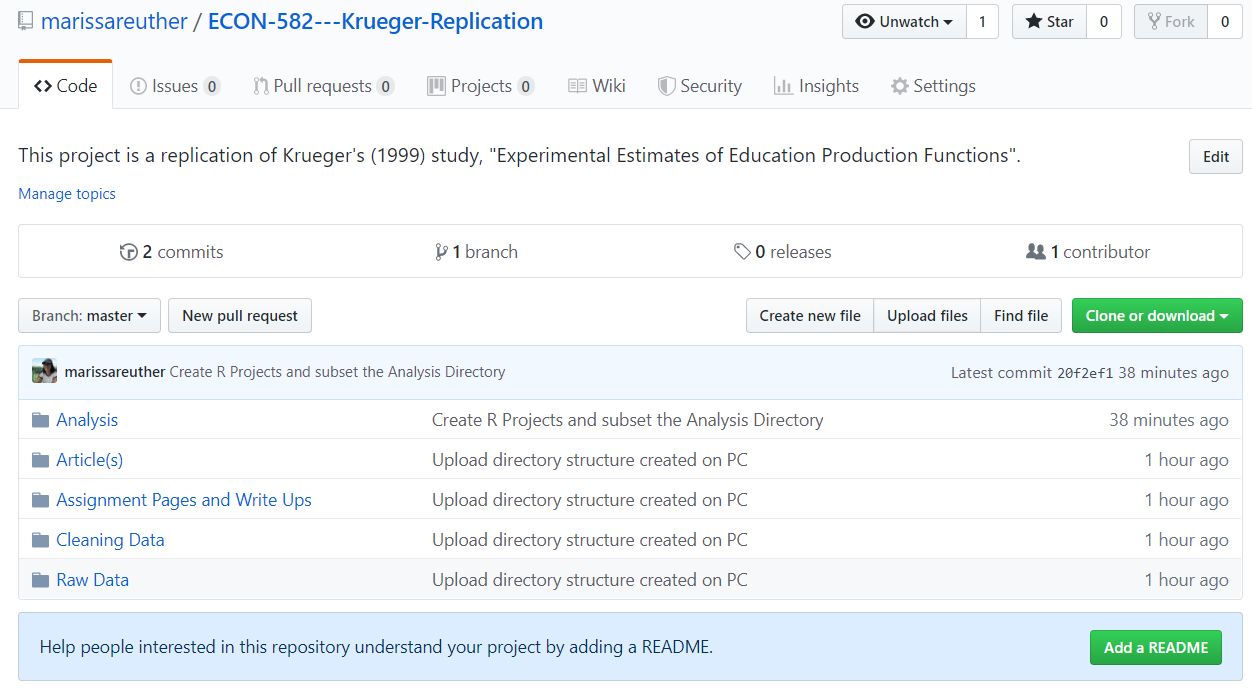
In Table 1, Kreuger reports various mean characteristics across different controls for each class classification. The table shows that the class sizes roughly have the same distribution of these observable characteristics, due to randomization. Because randomization equalized these observable characteristics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

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

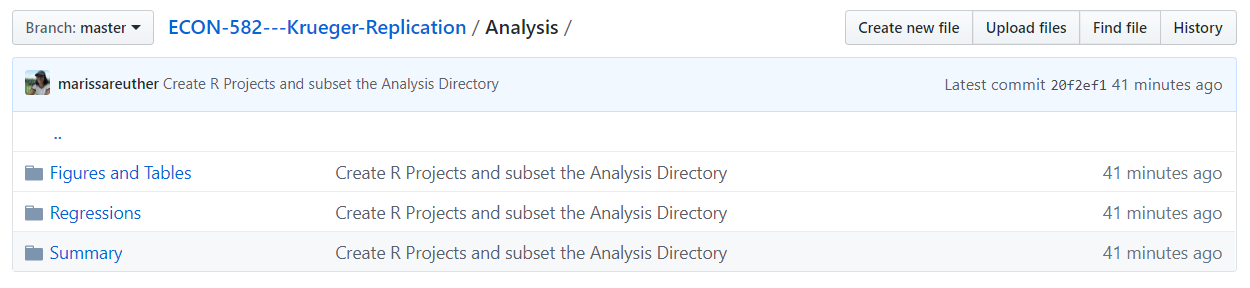
2. Questions pertaining to “Code and Data for Social Sciences” – Chapters 1-4

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 created subfolders within the analysis folder so I don’t have one giant script for all of this portion of the assignment.

 This is the current structure of my analysis folder and I plan on tweaking it as I work.

B – Automate what?

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

C – Plan for version control?

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

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

D – Backup work?

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