

## ECONOMICS 582 - ELEMENTS OF ECONOMETRICS I

UNIVERSITY OF TENNESSEE, FALL 2019

Replication Study - Analysis and Writeup

Due 12/3/2019, 6:00PM

Each student will replicate results from Krueger's (1999) study, "Experimental Estimates of Education Production Functions," published in the *Quarterly Journal of Economics*. The full article and source data are posted on Canvas in the "Krueger TNSTAR Replication" subfolder of "Files." I posted all source files affiliated with the STAR experiment,<sup>1</sup> although for this assignment you only need the STAR\_Students.dta file.

The entire replication assignment is worth 60 points, or 15% of your course grade. You have already completed the "first steps" component worth 15 points. This second sub-assignment is worth the remaining 45. For this sub-assignment, you will turn in two files through Canvas: one typed document, and one statistical program (.do, .r, or .py).

1. Write a statistical program in Stata, R, or Python that replicates the following results from Krueger's (1999) study of the Tennessee STAR experiment:

- Table I
- Table II
- Table III
- Figure I
- Table V
- Table VII

With a few minor exceptions (potentially), you should be able to exactly replicate each of these tables and figures.<sup>2</sup> Your program should run smoothly from start to finish and include all necessary steps to format the data and generate results. Ideally, your program should also format figures and tables in a way that makes it easy to transfer them to your writeup document.

2. Write a brief report on your replication. The report should be self-contained, meaning that someone who has never heard of STAR or Krueger's study can understand it from your writeup. You may organize this report however you wish. Here are the pieces I will be looking for:
  - For the introduction, start with the outline that you wrote for question 1 of the "first steps" assignment.
  - Include a "Data" section that describes the source data, citing the Dataverse.
  - Describe Tables I-III, why they are important in the research design, and what they show.
  - Describe what you see in Figure I.
  - Outline the empirical research design. Describe why 2SLS results may be different from OLS results in this setting, and in particular, the problems that 2SLS identification are meant to address here. For each estimating equation, define each variable and parameter. Describe the major identification assumptions necessary to causally interpret the estimated effects of small classes and teacher aides.
  - Discuss results in Tables V and VII. For each table, interpret the magnitude and statistical significance of main results for "small class" and "Regular/aide" class.
  - The report so far has been about following in Krueger's (1999) footsteps. Conclude by describing what you think could be done to improve or expand the analysis without compromising its "internal validity" (meaning, its adherence to the conditional independence assumption).

---

<sup>1</sup>More details on STAR and source data are at the study's Harvard Dataverse site: <https://doi.org/10.7910/DVN/SIWH9F>.

<sup>2</sup>My grateful acknowledgements go to Adrienne Sudbury for development of replication code over the summer.