ECONOMICS 582 - ELEMENTS OF ECONOMETRICS I

UNIVERSITY OF TENNESSEE, FALL 2019

Replication Study - First Steps

Due 10/25/2019, 6:00PM

Each student will replicate several 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."

The entire replication assignment is worth 60 points, or 15% of your course grade. This sub-assignment lays out some of the first steps you should undertake in this project, and it is worth 15 points of the full 60. For this sub-assignment, turn in one document through Canvas.

1. Read Krueger's paper, and draft the introduction for your report. Your introduction should answer the following, but in a narrative form rather than an outline:

   A. Describe the outcomes of interest in this study. Why are they important?

   B. What is it that Krueger hypothesizes affects the outcome you listed for part (A)?

   C. How does Krueger measure the outcome you listed in part (A)? In part (B)? Let's call these variables $\mathbf{Y}$ and $\mathbf{X}_1$.

   D. Suppose that this was not a randomized experiment, but that we were able to measure $\mathbf{Y}$ and $\mathbf{X}_1$ for a large i.i.d. sample, as well as other control variables. This allows us to estimate a least squares regression of the form $\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}$. For this study, describe the conditional independence assumption that is necessary in order to interpret the $\hat{\beta}_1$ estimator as the causal effect of $\mathbf{X}_1$ on $\mathbf{Y}$.

   E. Describe one potential omitted variable that could violate the conditional independence assumption in a hypothetical non-experimental study, and in the non-experimental literature on this particular research question. What can you think of that is hard to measure, affects $\mathbf{Y}$, and also probably affects $\mathbf{X}_1$?

   F. For Krueger, does randomization overcome the omitted variable problem you describe in (E)? Why or why not?

2. Read (or re-read) Gentkow and Shapiro's "Code and Data for Social Sciences," chapters 1-4. This is available online and also posted to Canvas ("Files").

   A. Using words, images, or both, describe the directory system that you will use for your replication project.

   B. What will you automate in your analysis?

   C. What is your plan for version control?

   D. How will you backup your work?