Problem Set 3 PAM 6090, Fall 2023
Due Monday November 13, 2023, 11:55pm (Turn in via canvas)

General instructions:

- Please work in your assigned teams.
- Please present your answers in a clear, concise fashion. Please submit a single main PDF document with your answers. In your solution packet, include relevant Stata output (e.g., key regression output, key graphs, etc.) and well-annotated Stata .do commands. (Do NOT include pages and pages of "undigested" Stata log files in the main problem set answers.) Make clear reference to regression output and figures in your written answers.
- Please upload the PDF file to canvas. Only one upload per team is needed (but clearly indicate the members of your team in your writeup). In addition to the main answers, please also upload .do and .log files (one .do file; one .log file) as separate files.

Problem 1 What drives variation in treatment in a DnD setting?

In this problem we will revisit a topic covered in class: in a 2xT Difference-in-Difference framework, where is the variation in treatment coming from?

1.1 Create a dataset with 20 observations: 1 treated unit and 1 control unit, and time = 1,…,10 periods. Create a treatment variable = 0 for the control unit, and which turns from 0 to 1 for the treated unit when time >= 6.

Graph 1: Plot out using a "line" or "connected" graph the time series of this treatment variable for each unit. [hint: to be able to clearly see both lines, you can either (1) play around with "connected" instead of "line"; (2) you can create a small "offset" to one of the units (ie add +0.01 to the treatment variable; (3) some other creative solution of your own design.]

1.2 Now in 4 additional graphs, show the "variation in the treatment variable, after controlling for [fixed effect etc.]"

1.2.1 Graph 2: control for only unit-specific fixed effects
2.2.1 Graph 3: control for only time-specific fixed effects
2.2.2 Graph 4: control for both unit and time fixed effects
2.2.3 Graph 5: control for unit and time fixed effects, and a unit-specific time trend

2.3 Graph 4 gives us the variation in treatment that is used to identify the DnD impact. In this graph, does the control unit have less variation than the treated unit?

Problem 2 - Bretrand/Duflo/Mullinaithan: "How much should we trust Differences-in-Differences" replication and extension

This exercise replicates part of the Paper in the QJE, February 2004. The code "bdm.do" (along with the data file "bdm_loaded.dta") sets up a Monte Carlo exercise like the one in the paper by Bretrand, Duflo, and Mullinaithan. In particular, it shows the "over-rejection" of a differences in differences model.

The code bdm.do computes numbers analogous to the numbers they present in Table 8, (pg 272) Panel A, Rows 5 and 7 ("OLS"), Column "no effect".

a. Briefly explain why the ideal rejection rate for this exercise is 5%, and what the numbers given by the simulation mean for the applied researcher.

b. Modify the code to fill out the rest of the "No effect" column for Table 8, Panel A. You will need to make a few changes to enact this. First, you will likely want to use more than 50 Monte Carlo replications. (For your final run, 1000 or more is a good number, if you have the computational time.) Second, you will need to tell the program to run using extra states. These changes will be easy, and just involve changing the numbers you supply to the program. You will also need to add in (or change the program to use) a "cluster-robust" (what they call "Arbitrary Variance-Covariance Matrix") estimator. In doing this, you will need to be careful to cluster on the "new state IDs" within each Monte Carlo replication, not the original state IDs.

Your numbers will likely be slightly different than theirs. Discuss any substantive differences between your results and theirs – are the "punch lines" similar?

c. One of their punch lines is that the cluster-robust estimator does good at large number of states, but does poorly with 6 or 10 number of states. But this is based on a rejection rule of "T > 1.96". Some (Cameron, Gelbach, Miller, ReStat August 2008, among others!) have argued that when clustering with few numbers of clusters, it is more appropriate to use a critical value base on the T distribution with (N-1) degrees of Freedom, where N is the number of clusters. Add in results for the "Cluster" rows using these T-distribution critical values. How does this do?

d. In Table 5, BDM examine the "Block Bootstrap" (or Cluster bootstrap) to get standard errors. They find that it does much worse than "cluster robust" (From Table 8) for N = 6 or 10. I think they made a mistake. Re-do the exercise using cluster-bootstrapped standard errors, and report the rejection rates you get. Do they correspond to those reported in BDM Table 5?

e. Cameron, Gelbach, Miller, ReStat August 2008 claims that the "Wild cluster bootstrap-t" procedure can give good rejection rates even with as few as 6 clusters. See if you can replicate row 11 of table 5 in that paper. The Stata add-on command "boottest" might help to make the Wild-percentile-T bootstrap easier to implement. (See https://doi.org/10.1177%2F1536867X19830877 for more information on this add-on command.)

Problem 3 – Dehejia and Wahba (DW) replication exercise, Part 1

To do this you will need the stata dataset "lalonde_exper.dta". The data set has the following variables that we will use: t (treatment, = 1 if treated, =0 if control); age; educ (education in years); indicators for Black, Hispanic, and married; re74 (real earnings in 1974); re75 (real earnings in 1975); and re78 (real earnings in 1978). You will need to create "age squared" and "high school droput" (educ < 12). The "reo74" and "reo75" variables should correspond to the "u74" and "u75" variables in DW. You can see this by the command "bysort reo74 reo75: summ re74 re75"

a. Check the means of the key variables against DW Table 1. How do they check out? Do we see evidence of an "Ashenfelter dip" for the treatment? For the controls? What explains the similarity or difference between these two?

b. How well do the data do at replicating the first row of panels B and C of Table 2?

c. I was unable to replicate column 5 exactly. Try a little to see if you can get it (you can choose panel B or panel C). What specification gets you the closest (eyeballing beta-hat and s.e.) to his column 5? How close is your estimate?

d. Now, let's look at the PSID-1 control group. This involves loading up a new dataset ("nswpsiddata.txt"). (the file "readin_psid.do" shows how to use the insheet command to read in raw text files. The variable definitions are all basically the same. Estimate a few sample means (age, educ, hsdrop, re74, and re75) – how do these look, compared against Table 1? Check that the race/ethnicity and marriage variables line up too.

e. Check that the regressions capture the results (row: PSID-1) in panels B and C of table 2.

f. Focus for a minute on Panel B, column 3. How do the over-time dynamics compare across the PSID-1 sample and the NSW sample? What do you make of this?