
Problem Set 5

Instructions: Answer the following questions and submit your results via email to `sean.corcoran@vanderbilt.edu`. Use your name and problem set number as the filename. Working together is encouraged, but all submitted work should be that of the individual student.

This problem will replicate the analysis reported in Bifulco, Rubenstein, and Sohn (2017).¹ That study used a synthetic control design to estimate the impact of Say Yes to Education (a promise scholarship program in Syracuse, New York, which provided free college tuition to any student who graduated from a public high school in Syracuse) on total district enrollment and graduation rates. The program was implemented in 2008.

There are two datasets on Github containing panels of enrollment and graduation data, respectively, that can be opened in Stata using:

```
use https://github.com/spcorcor18/LP0-8852/raw/main/data/nys_data_enroll.dta, clear  
use https://github.com/spcorcor18/LP0-8852/raw/main/data/nys_data_graduation.dta, clear
```

Most of the variables in these datasets should be self-explanatory from their variable names and labels (although I'm not sure what *target_donor* and *small_index* refer to, as they don't appear to align with the paper's selection of potential donor districts).

The authors used two potential donor pools. The comprehensive donor pool included all 275 (non-Syracuse) districts, while the restricted donor pool included 22 districts categorized as "City-Large," "City-Midsize," or "City-Small." (Note these counts are a little smaller for the graduation rate panel). NYC is excluded from the dataset.

Using the synthetic control packages in Stata (`synth`, `synth_wrapper`, and `synth2`), replicate the findings in this paper by reporting the elements listed below. Note you do not need to run all 6 alternative specifications of the pretreatment years (Table 1). Rather, just use their Specification (2)—the first, middle, and last year of pretreatment periods—and Specification (4)—the last pretreatment year and the average of outcomes in all other pretreatment years. They included the percent eligible for free or reduced price lunch, the percent Black, and the percent Hispanic in every specification.

Taken together, you will have eight sets of results: two outcomes (enrollment and graduation rates) \times two specifications (2 and 4) \times two potential donor pools. Brownie points to those who combine these results in a pleasing-to-read format.

¹Thank you to Bob Bifulco and Hosung Song for providing the data used in their paper.

Include these things in your results, and be sure to submit your do-file:

- (a) The weights assigned to donor districts, as in Tables 1 and 5. Write a few sentences summarizing the resulting weighting used. Do they correspond to the weights reported in the paper? **(10 points)**
- (b) The main synthetic control graph showing trends in Syracuse and its synthetic control, as in Figures 2 and 3. Briefly summarize what you see. **(10 points)**
- (c) The “gap” version of the graphs in (b) showing the *gap* in mean outcomes between Syracuse and its synthetic control (these were not shown in the paper). **(5 points)**
- (d) Point estimates of the treatment effect by year (2008, 2009, 2010, and 2011), as in Tables 3 and 6. Note the graduation rate data only include 3 post-treatment years. **(5 points)**
- (e) The graph showing the gap in mean outcomes between Syracuse and its synthetic control, and the placebo gaps. Only use the restricted donor pool here, to keep the number of placebo districts manageable. Briefly summarize what you see. **(10 points)**
- (f) p -values from the placebo-based inference. Explain in words where these come from, and how they should be interpreted. (Note, you only need to do the latter for one set of results, not every specification and outcome). **(10 points)**
- (g) If you have access to Stata 16+ and can use `synth2`, do a “leave-one-out” (loo) robustness test for one of the enrollment outcome specifications, using the restricted donor pool. Interpret the results. **(5 points bonus)**

Note if you have Stata 16+ and can use `synth2` you may find it easier to obtain all of the above items. If not, the other commands can be used. See the in-class exercise do-file for help, and it would (of course) help to refer to the original Bifulco et al paper.