Data and Code for: Using Multiple Outcomes to Improve the Synthetic Control Method
Liyang Sun, Eli Ben-Michael, and Avi Feller

Overview

The code in this replication package replicates the empirical illustration and the simulation study using R. The same R markdown file runs all of the code to generate the figures and tables in the Section 5 and Online Appendix E of the paper. Another R file generates figures in Online Appendix D of the paper. The replicator should expect the code to run for less than 1 hour.

Dataset list

For the empirical illustration based on Trejo et al. (2024), we did not generate the data and do not have access to the data. Rather, Trejo et al. (2024) compiled district-level student outcomes data from the individual-level student outcome data provided by Michigan Education Data Center under restricted use license and conducted their original synthetic control analysis. For this paper, one of the authors of Trejo et al. (2024), Brian Jacob, ran the code we provided in this replication package on their data, which generated the results presented in this paper. This allowed us to examine the value of our proposed synthetic control approach in a realistic context without having to access restricted data ourselves. Detailed documentation on how to obtain access to the Michigan Education Data Center and construct the district-level data can be found in the replication packages of Trejo et al. (2024).

For the simulation study, we generated the data using seeds described in controlled randomness.

Data.Name	Data.Files	Provided	Citation
"Michigan District-Level Student Outcomes 2006-2019"	`augsyn_nonfix_60_2006_2020_07_30.csv`	FALSE	Trejo et al. (2024)

Computational requirements

Software Requirements

- R 3.4.3
  - `tidyverse` (2.0.0)
  - `augsynth` (0.2.0)
  - `lubridate` (1.9.3)
  - `ggrepel` (0.9.5)

- `ggpubr` (0.6.0)
- `dplyr` (1.1.4)
- `ggplot2` (3.3.5)
- `MASS` (7.3.54)
- `xtable` (1.8-4)

Controlled Randomness

No simulation is involved in the Flint analysis, and thus, there is no need for specifying a random seed.

For simulation, line 165, 172 and 219 of `simulation.R` set the seeds.

Memory and Runtime Requirements

Summary Approximate time needed to reproduce the analyses on a standard 2021 desktop machine is 1-5 minutes.

Details The code was last run on a **6-core Intel-based laptop with MacOS version 11.7.3**.

Description of programs/code

- Program `flint_analysis_replication.rmd` will generate all figures in the manuscript and the online appendix.

Instructions to Replicators

- Download the data file referenced above. It should be stored in the same directory as `flint_analysis_replication.rmd`. Create the `/figure` folder. No further action is needed on the replicator's part.
- Run `flint_analysis_replication.rmd` to generate all figures.
- Run `simulation.R` to generate all figures. Comment out blocks corresponding to different ρ in line 17-27 to generate different figures

Figure/Table #	Program	Output file
Figure 1	`flint_analysis_replication.rmd`	`/figure/all_scm_weights-1.pdf`
Figure 2	`flint_analysis_replication.rmd`	`/figure/figure_flint_summary_gap_annual_fixedeff_sped_orig`
Figure D.1(a)	`simulation.R`	`/figure/all_bias_sims_n50_rho1.pdf`
Figure D.1(b)	`simulation.R`	`/figure/all_imbalance_sims_n50_rho1.pdf`

Figure/Table #	Program	Output file
Figure D.2(a)	`simulation.R`	`/figure/all_bias_sims_n50_rho05.pdf`
Figure D.2(b)	`simulation.R`	`/figure/all_bias_sims_n50_rho0.pdf`
Table E.1	`flint_analysis_replication.rmd`	`xtable`
Figure E.1	`flint_analysis_replication.rmd`	`/figure/pre_fit_table-1.pdf`
Figure E.2	`flint_analysis_replication.rmd`	`/figure/flint_nosped-1.pdf`
Figure E.3	`flint_analysis_replication.rmd`	`/figure/frontier_plot-1.pdf`
Figure E.4	`flint_analysis_replication.rmd`	`/figure/p_val_plot-1.pdf`

References

Trejo, Sam, Gloria Yeomans-Maldonado, and Brian Jacob, "The Effects of the Flint water crisis on the educational outcomes of school-age children," *Science Advances*, 2024, *10* (11), eadk4737.