

Problem Set 4

MaCCS 201 - Fall 2024

2024-11-24

Tentative Due Date: December 19

This will be a pass fail Problem Set.

Please submit markdown file named [last_name]__[first_name]__ps4.Rmd or a pdf with all code and answers.

#Part A: Difference in Differences.

Let's replicate the OG. Card and Krueger. We have been talking about them forever. Plus the consequences of raising the minimum wage will be forever policy relevant. The dataset is in Stata form, which is easy to import (I will show you below). It is called `cardkrueger.dta`.

The interviewed Fast food restaurants in two waves in New Jersey and Pennsylvania 3/1992 and 11/1992). In the middle the minimum wage in NJ went up from \$4.25 to 5.05 per hour but did not change across the border. You will DiD to see whether you detect an effect of raising the minimum wage. If you see a 1 or a 2 at the end of a variable that indicates which survey wave it is from.

The variables you need are:

- `state`: NJ=1, PA=0
 - `wage_st` / `wage_st2`: Starting wage at the restaurant
 - `fte` / `fte2`: Full-time equiv. employment = $\#(\text{Full time employees}) + \#(\text{Part-time Employees})/2$. Excludes managers.
 - `chain`: which fast food chain you are dealing with (there are 4.)
 - `co_owned`: = 1 if restaurant is company-owned, =0 if franchised
 - `sample`: Dummy variable = 1 if wage and employment data are available for both survey waves
1. Dump all observation for which `sample=0`, so you have balance.
 2. Create a variable `treated` which equals one if the state is NJ and zero otherwise.
 3. Create a dummy called `after`, which equals 1 if the observation is from round 2.
 4. Calculate the difference in average (across stores) starting wages before and after for each state and then calculate the difference in difference by differencing the two. What do you get?
 5. Do the same, but with average full time employment. What do you get?
 6. Now set your data up for proper DiD estimation (stacked or long format). This is one of the most pain in the neck (seemingly simple but in practice annoying feats.) Instead of having observations "next" to each other, you want the round 1 observations to be the top block of rows and round 2 observations be the bottom block of rows. You should have the indicator `after`, which is now 0 for the before and 1 for the after periods as a nice column. If you struggle, text me (925) 360-6473.
 7. Run a difference in difference regression on these data. First use `wage` as the outcome. Then use full time employment as the outcome. The unit of observation is the store here!
 8. Do the same thing as in 7, but control for whether the store is a franchise or not. What do you see?
 9. Cluster your standard errors by state. Then cluster by store. What do you see?

Part B: Synthetic Controls

Let's replicate another OG. Abadie and Gardeazabal (2003). They estimate the effect of terrorist conflict in the Basque Country (Spain) on GDP per capita. You essentially construct a synthetic version of Basque Country.

We want to replicate the paper. I will save you the digging through the pages.

1. You want to use the following variables as predictors. `"school.illit"`, `"school.prim"`, `"school.med"`, `"school.high"`, `"school.post.high"`, `"invest"`
2. You want to use the `mean` as the relevant operator.
3. As special predictors you do: `special.predictors = list(list("gdpcap", 1960:1969, "mean"), list("sec.agriculture", seq(1961, 1969, 2), "mean"), list("sec.energy", seq(1961, 1969, 2), "mean"), list("sec.industry", seq(1961, 1969, 2), "mean"), list("sec.construction", seq(1961, 1969, 2), "mean"), list("sec.services.venta", seq(1961, 1969, 2), "mean"), list("sec.services.nonventa", seq(1961, 1969, 2), "mean"), list("popdens", 1969, "mean"))`
4. The dependent variable is `gdpcap`
5. The unit variable is `regionno`
6. The unit names variable is `"regionname"`
7. The time variable is `"year"`
8. The treatment identifier should be obvious. (Basque Country!!!)
9. As controls identifiers use 2:16 and 18. Leave out 1, because that is all of Spain.
10. `time.optimize.ssr = 1960:1969`
11. `time.plot = 1955:1997`
12. As method use `method = "BFGS"`
13. Generate a nice looking plot of the gap between the synthetic control and Basque Country GDP.
14. Create a falsification plot, using all other areas from 9. above as fake units.
15. Optional. Repeat 14, but fake move treatment up a bit.