HCMG 901 Problem Set 3 (Panel data)
Due Apr 7, 2025

In this exercise we will replicate some of the empirical analyses in Stevenson and Wolfers (*QJE*, 2006) as well as the decomposition analysis proposed by Goodman-Bacon (*J Metrics*, 2021). Andrew Goodman-Bacon has graciously shared the data and code he put together to replicate SW2006. I have uploaded the data file "divorce_example.dta" and a code file to the PS3 folder in Canvas. Unfortunately, there is no documentation for the data file, but the code file will help guide you on the key variables to use in the analysis.

1. Skim through SW2006 to orient yourself on the setting, empirical design, treated states, untreated states, etc. This will help when you try to replicate their results. What is their estimating equation? What is the key identification assumption? Do you have any concerns about this assumption? What evidence would you ideally provide to alleviate the concern you mentioned above? (Don't worry about whether you can do this with the data provided here)

2. Remarkably, SW2006 does not have a descriptive statistics table. Prepare a descriptive statistics table that you would normally provide in such a paper. It should describe the sample size, means, and SD of suicide rates for men and women (key outcome) and the policy indicator (i.e., the key explanatory variable).

3. Replicate the analysis presented in SW2006 Table 1 using a TWFE model. The data file does not contain any control variables, so just estimate a specification without controls, equivalent to columns 1f and 1m. Interpret the coefficients.

4. Replicate Figure 5 in GB2021 using a TWFE model. Also produce the equivalent version of this figure for male suicides. Discuss how this event study is helpful in the context of the identification assumption.

5. Discuss the different types of 2x2 DD effects that are averaged together in a generalized D-D model. Following the notation in CH2020 that we discussed in class, why do TWFE estimates get biased in the case of staggered designs? What is the intuition behind the negative weights? Which groups or periods are more likely to get negative weights?

6. Now, produce alternate versions of the event study using Callaway-Santanna and the BJS imputation estimators. Comment on whether these differ from the TWFE estimates in relation to your discussion in point 5 above.

7. Replicate Figure 6 in GB2021 using the *bacondecomp* command that GB has produced (freely available for download). Estimate the model without controls and use the *ddetail* option. Briefly discuss the weights assigned to the different categories of estimators. Comment on the 2x2 estimate when we compare later treated states to early treated states as controls. What role is it playing in the overall estimate?

I also ask you to replicate the simulation on Sant'anna's website which shows that TWFE event studies can be quite biased in the case of staggered designs. The simulation and results are in a blog here: https://psantanna.com/posts/twfe# . He has also released the R code to replicate these results. For the problem set we will start with the runs where the data includes a never treated group and when treatment effects are *heterogeneous* across groups and over time. The is the last scenario under the heading, "Setup with Heterogeneous Treatment Effects."

1. Simulate the data and plot the raw trends in values by group over time. Check that it looks like his figure.
2. Estimate the TWFE event study with all leads and lags (i.e., no binning). Does it approximately recover the true treatment effects in event time?
3. Now use the Callaway Sant'anna estimator to produce the event study. Does this match the true effect?
4. Now, let's change the data generating process so that the treatment effects are constant over time but still heterogeneous across groups. So, the outcomes jump on treatment by different magnitudes across groups but then proceed in parallel. Re-estimate the TWFE estimator on this new dataset. This time does the TWFE recover the true effects?