**Assignment 3: Callaway and Sant’anna (2020) modeling of Lott and Mustard (1997)**

**Due end of week 8**

**Directions:**

This week’s exercise *will include programming*. Your goal is to use Callaway and Sant’anna, or CS, to analyze the Lott and Mustard (1997) data. I have put a series of goals below for you to consult.

1. Place the following in a “beautiful table” (I leave that to you – just make it readable and interesting using whatever idea you have).
   1. List all treatment dates along with how many states are in each “group” or “cohort”.
   2. How many “never treated states” are there?
2. Estimate the CS model on the county level data. Be sure to cluster at the state level and use bootstrapping. Default bootstrap of 100 or 1000 is fine (whichever it is). Note that I haven’t done this for county (or even state) so if you find it’s impossible to do, report why you think it was, otherwise complete the following objectives.
   1. Never treated comparison groups
      1. Estimate the overall ATT using the never treated (if any)
      2. Estimate each group’s ATT using the never treated (if any)
      3. Estimate event study graphs for the dynamic model (ie., the figure listing along the horizontal axis -5, -4, etc.)
   2. Not-yet-treated comparison groups
      1. Estimate the overall ATT
      2. Estimate each group’s ATT
      3. Estimate event study graphs for the dynamic model
   3. Repeat (a) and (b) using a select number of covariates that you chose from the pre-registration assignment (Assignment 2). You are probably not going to be able to fit the propensity score using all covariates – that’s been my experience. If you are able to figure out the max variables, do that. I’m sort of curious what happens when each of you ends up using different covariate set. Just use your own theories about why those variables are important for selection into treatment, as well as Assignment 2’s analysis, to pick the covariate set.
3. Repeat step 2 using the state level data. Are there any differences in what you found?
4. Write up what you did. Your write up should include the following.
   1. In your write up, separate your analysis up by “county vs “state” and “never vs not yet treated” and “with and without controls”.
   2. Note where you got the data – the website, the authors who distributed it.
   3. Explain the CS model briefly, including the econometric equation.
   4. Explain the assumptions needed for unbiasedness.
   5. Report your Questions 1 and 2 in a neatly and clearly written document, with section headings numbered in some logical way (in your opinion).

This analysis can ultimately feed directly into the final project if you so desire.