Regression II
Vanderbilt University
Prof. Sean P. Corcoran

October 31, 2023

Lecture 6 In-Class Exercise

This exercise will replicate and expand upon the results presented in the *Mixtape* chapter on Synthetic Control. In response to a civil action lawsuit related to prison overcrowding, Texas significantly expanded its state prison capacity in the 1980s. The growth in prisons capacity increased rapidly beginning in 1993 when the state approved $1 billion in new prison construction. In the *Mixtape*, Cunningham examines the effect of new prison construction on the incarceration rates of Black men. There is evidence that Black men are systematically more likely to be incarcerated (even for the same infractions), and that parole decisions are racially biased. One might predict, then, that the expanded capacity of prisons in Texas had a disproportionate effect on Black men. Cunningham uses synthetic control to address this question. (In this working paper, he looks at the effect of higher incarceration rates on drug markets). The dataset he uses can be read into Stata as follows:

```
use https://github.com/scunning1975/mixtape/raw/master/texas.dta, clear
```

See the accompanying do-file for Stata syntax.

1. Part 1 of the accompanying do-file uses the `synth` command to construct a synthetic control and estimate the effect of prison construction in Texas on the incarceration rates of Black men (number incarcerated per 100,000 population). The above dataset has already been "xtset" with *statefip* as the cross-sectional unit and *year* as the time period. The state ID for Texas is 48 and the first treatment year is assumed to be 1993. Cunningham selected 14 predictor variables, a combination of pre-treatment outcomes and covariates:

   Incarceration rate in 1988, 1990, 1991, and 1992

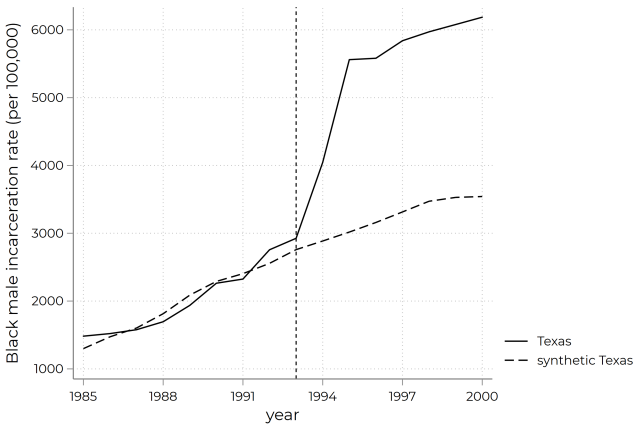
   Alcohol (1990), and AIDS per capita (1990 and 1991) – not sure what these are

   Income, unemployment rate, poverty rate – *averages over the full pre-treatment period*

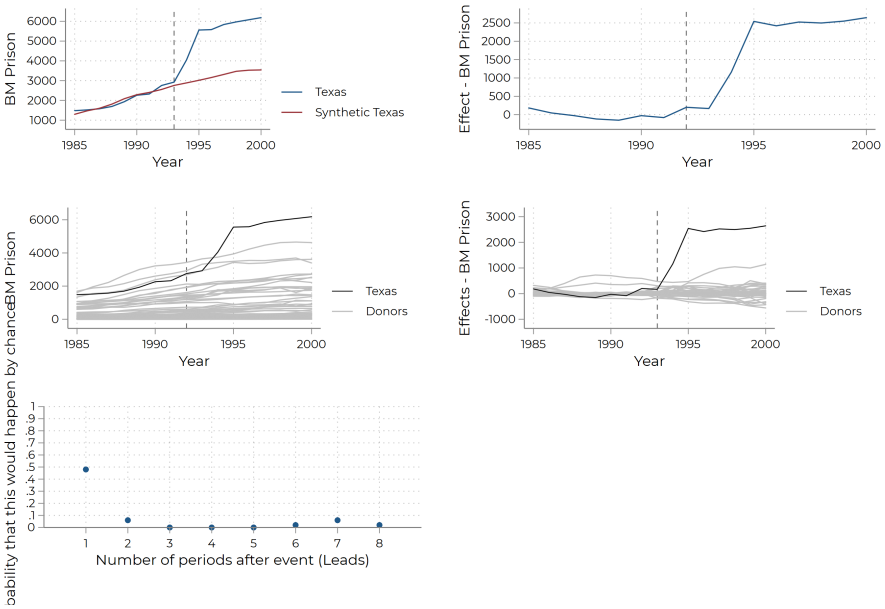
   Percent Black (1990, 1991, and 1992)

   Percent aged 15-19 (1990)

   From `synth` you get the cross-sectional unit weights (i.e., which states are used as the synthetic control and the weights they are given), a pre-treatment balance table fro the predictors, and the standard graph (below). You can also view the V matrix of coefficient weights. Note: the scale of the *bmprison* variable in this dataset appears to be off by a factor of 10. I divided it by 10, but am not 100% confident in this variable.

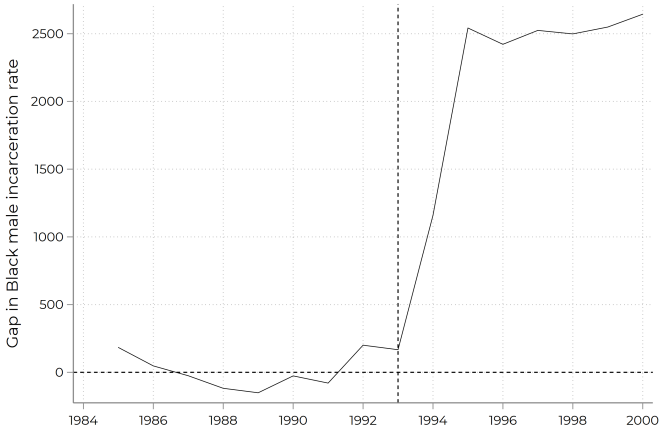


2. Part 2 of the do file uses the `synth_runner` command which builds upon `synth`. With `synth_runner` you get the estimated treatment effects by year, p-values, and some additional graphs (a graph of the *gap* between Texas and its synthetic control, graphs of Texas vs. all of the placebos, and graphs of the p-values by year–see below).



3. Part 3 of the do file uses the `synth2` command which improves a lot upon `synth` and `synth_runner`. Unfortunately, it only works with Stata 16+. With `synth2` you get the standard placebo test, an "in-time" placebo test that sets the treatment year to an earlier "fake" treatment year, and a "leave-one-out" robustness test where the model is repeatedly re-estimated excluding one of the units used in the synthetic control.

4. Part 4 of the do file manually creates a graph of the gap between Texas and its synthetic control, using data saved in Part 1. (With `synth` you can save the synthetic control's means by year).



5. Parts 5-7 of the do file manually run the synthetic control method for placebo states, and graphs the results. Part 7 calculates the pre- and post-RMSPE for each state and calculates the ratio of post-to-pre RMPSE. It then plots a histogram of these ratios, highlighting where Texas sits in relation to other states.