Homework 9

Economics 7103

I have provided you access to a subsample of the recycling data used in one of my papers (*recycling_hw.dta*). Each observation is for a region i during year t. The data span from 1997-2008. Table 1 describes the variables in the data.

Variable	Description
region	Name of region
nyc	= 1 if region is in New York City
nj	= 1 if region is in New Jersey
ma	= 1 if region is in Massachusetts
year	Year
munipop2000	Population in year 2000
recyclingrate	Fraction of waste recycled
id	Numerical identifier of each region
incomepercapita	Per capita income
collegedegree2000	Percent of adults with bachelors
democratvoteshareYYYY	Democratic party candidate vote share by year
nonwhite	Fraction of nonwhite population

Table 1: Variable descriptions.

You will study the effect of a pause in recycling collection that took place from 2002-2004 in New York City.

1 Stata

1. Produce a yearly plot of the recycling rate for NYC and the controls to examine the effect of the recycling pause and the possibility of parallel trends.

2. Estimate the effect of the pause on the recycling rate in NYC using a TWFE regression and the data from 1997-2004. Cluster your standard errors at the region level. Report the average treatment effect estimate and the standard error.

3. Use the command `sdid` to estimate the synthetic DID version of the TWFE regression in equation 2. Report the estimated average treatment effect and the synthetic DID plot using the `graph` option.

4. Using the full sample, estimate the following event study regression:

$$Y_{i,t} = \alpha_i + \gamma_t + \sum_{\ell \neq 2001} D_i \cdot 1(t = \ell)\beta_\ell + X_{i,t}\gamma + \varepsilon_{i,t} \tag{1}$$

   where D_i is a binary variable equal to one for New York City regions, $1(t = \ell)$ is an indicator function equal to one for year ℓ, and $X_{i,t}$ are any time-varying controls you would like to include. Do *not* use a canned event study regression. Use `reg`, `xtreg`, or `reghdfe`. Report your results as a picture of the coefficient estimates of β_ℓ with confidence intervals derived from standard errors clustered at the region level (use `coefplot`). Note that you will need to generate treatment variables to estimate this regression.

5. Use the commands `synth` and `synth_runner` to generate synthetic control estimates of the dynamic treatment effects. Generate the synthetic control estimates using whichever matching variables you see as most appropriate. Use placebo inference. Report:

   (a) The plot of raw outcomes for treated and control groups over time.

   (b) The plot of raw outcomes for treated group and synthetic control group over time.

   (c) The plot of estimated synthetic control effects and placebo effects over time.

   (d) The plot of final synthetic control estimates over time.

   (e) *Hints*: Note that all of these plots can be generated using postestimation commands that come with `synth_runner`. You will need to collapse all of New York City to one treated unit to use the canned commands. Finally, remember that these estimates might not look that good if the synthetic control does not approximate NYC very well!