Columbia University SIPA
Fall 2020

**Empirical Analysis of Energy Markets - U6616**
Empirical Exercise 3

Prof. Ignacia Mercadal

This problem set is due on **November 16th at 11:59pm**. You can work in pairs and submit a common solution. Please submit the homework as an R notebook (if there are data files, put the code and data files in a zip file). The code must run without errors. I recommend using the function *felm* from the package *lfe* to run the fixed effects regression and compute standard errors easily.

The goal of this exercise is to replicate and understand the results of the paper "Deregulation, Consolidation, and Efficiency: Evidence from US Nuclear Power" by Lucas Davis and Catherine Wolfram.

1. First, load the data from file data_panel.csv.
2. Do some preliminary data cleaning. To deal with dates, I recommend the package *lubridate*.
   - Rename capacity_design_mwe as capacity and multiply capacity_factor by 100.
   - Combine month and year to create a **date** variable.
   - Combine date and commercial_operation (date of initial operation) to create an **age** variable (in years)
   - Make a variable for the month-year of divestiture using divest_date. Also create a dummy that indicates if the power plant was ever divested, name it **ever_divest** and make it a factor.
   - Create indicator dummies for type (pwr_bwr) and manufacturer.
   - Create census region indicators:
     - $regionNE = 1$ if censusregion is MAT or NEW
     - $regionMW = 1$ if censusregion is WNC or ENC
     - $regionSO = 1$ if censusregion is WSC or ESC or SAT
     - $regionWE = 1$ if censusregion is PACC or MTN
   - Create a reactor id for each reactor (reactor_name) and a plant id for each power plant (facilityname). Define them as factors to be able to use them as fixed effects.
   - Save the clean dataset as a .Rds file.
3. Now you will compute some summary statistics and compare divested and non-divested plants.

(a) Compute the share of total capacity that has been divested (using the dummy *divested*) every year. For this, notice that you have monthly data but capacity does not change so you may want to restrict your sample to e.g. December. Alternatively, you can compute the mean.

(b) How does the capacity factor vary over time?

(c) How does the capacity factor vary over time for divested and non-divested plants (After 1999)?

4. Reproduce Figure 1 in the paper. For this, you have to first aggregate the data to a yearly level (instead of monthly) by computing the mean capacity factor for each year separately for the two groups (never divested and ever divested). Plot the number of operating reactors in a different panel (not in the same graph as in the paper). Make sure you label everything properly.

5. Reproduce Table 2, except for the specification in which observations are weighted by reactor capacity, and answer the following questions:

   (a) What do the reactor fixed effects do? How can you interpret the different results with and without reactor fixed effects?

   (b) What happens if you do not include month_year fixed effects? How do you interpret this?

   Notice that you should cluster the standard errors at the plant level.

6. Reproduce Table 1. Are the two groups different or similar? How does this matter for the causal interpretation of the estimated coefficients? Notice here that the authors compare the values in December 1998.

7. What happens if you answer this question using only data on the divested generators? And if your sample compares divested and non-divested plants after 2002? Run these regressions and interpret the results. What's the advantage of the strategy used in the paper compared to these two?

8. (for extra points) Reproduce Table 2 and interpret the results.