Columbia University SIPA
Fall 2020

Empirical Analysis of Energy Markets - U6616
Empirical Exercise 4

Prof. Ignacia Mercadal

This problem set is due on **December 7th 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.

The goal of this exercise is to evaluate the consequences of electric deregulation on the prices paid by consumers for electricity. For this, we will use data on prices that I downloaded from the EIA website (you can find the original data here).

1. First, load the price data in prices.csv and do the following to prepare the data for the analysis:

    - Change variable names and remove the first row.
    - Convert relevant values to numerical, removing commas if necessary.
    - The table in the last page contains information about the dates in which the wholesale and retail markets were deregulated in each state. For the wholesale market, deregulation means that the market was opened to entry by any firm owning a generator. For the retail market, deregulation means that any firm can sell electricity to final consumers, and use the former regulated utility's distribution network to deliver it to its customers. Based on this table, create two dummies.

        - *treated*: takes the value 1 when the market was deregulated in that state and year. For example, the dummy will have the value 1 for NJ in 2004, and a value of 0 for ME in 1996.
        - Create a dummy *ever_dereg* that takes the value 1 if that state's market was ever deregulated.

2. Select the data where the industry sector category is the "Total Electric Industry" and plot the evolution of prices over time using a separate line for states that were deregulated and states that were not deregulated (you can use the dummy *ever_wdereg* created above). Do a separate plot for residential, commercial, and industrial customers. What do you observe?

3. Run a regression of price on deregulation status, separately for residential, commercial, and industrial customers. What do you find?

4. What is the underlying assumption for the above estimates to be interpreted causally? Is this assumption reasonable? Explain and use data to support your answer if you can.

5. Now run the same regressions but adding state fixed effects. What do you find? (Make sure to cluster your standard errors at the state level)

6. What is the underlying assumption for the above estimates to be interpreted causally? Is this assumption reasonable? Explain and use data to support your answer if you can.

7. Now run the same regressions but adding state and year fixed effects. What do you find? (Make sure to cluster your standard errors at the state level)

8. What is the underlying assumption for the above estimates to be interpreted causally? Is this assumption reasonable? Explain and use data to support your answer if you can.

9. **(Extra credit)** Use the data in the supplementary files to add additional controls that could help to address some of the concerns you have explained above. For example, pre-deregulation emission levels or fuel mix could help with differential effects of environmental regulation. Interpret your results, are they causal? Why? (This is an open question, you do not need to use all the additional data, one or two additional controls are enough.)

Table 1: Deregulation dates by state

State	Implementation Year
NY	1998
RI	1998
CA	1999
NH	1999
MA	1999
ME	1999
CT	2000
DE	2000
MD	2000
NJ	2000
PA	2000
IL	2001
OH	2001
MI	2002
OR	2002
TX	2002
VA	2002