Columbia University SIPA
Fall 2019

Empirical Analysis of Energy Markets - U6616
Empirical Exercise 2

Prof. Ignacia Mercadal

This problem set is due on **October 17th at 11:59pm**. Students are allowed to discuss with others, but each student must submit his or her own 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. If there are errors, it will cost you 10 points and you will be given 24 hours to correct it, after which you will have to indicate clearly where was the error.

1. (October 5) Go to the energy dataset provided by S&P Global available here.[1] Go to Power/Generation Supply Curve and download the dataset for one of the following ISOs: MISO, PJM, ERCOT, or New England ISO, for either 2009 or 2018 (please sign up here **before 10/4** so each student does a different dataset).[2]

   Select the following variables:

   - Plant Unit key
   - Primary fuel type
   - Generation technology
   - Summer capacity MW
   - Variable O&M cost per MWh (this is the variable cost)
   - Total fuel cost per MWh (this is part of the variable cost)
   - Emission allowances costs (this is part of the variable cost)
   - Fixed O&M cost
   - Heat rate btu/ kwh
   - Heat input
   - Net generation MWh
   - Capacity factor
   - NO_X Emissions Rate lbs MMBtu
   - SO_2 Emissions Rate lbs MMBtu
   - CO_2 Emissions Rate lbs MMBtu

---

[1]If you don't have access using your Columbia email, send an email to the professor.

[2]Notice you can only access this dataset from within campus. You can email S&P support if you have any questions about the data, they usually respond quick. You can also email the TA or the professor.

Start by cleaning and understanding your data. For this, do the following:

(a) What does each variable represent?

(b) Assign convenient yet meaningful names to each variable in the dataset.

(c) What is the class of each variable? Make sure to convert them to the proper class before doing this. For example, if net generation is a character, make it numeric.

(d) Describe each variable: what values does it take? Do you have any concerns about some variable (extreme values, missing values)?

2. (October 5) Now let's look at the importance of each fuel in this market.

(a) What is the fuel composition of this market according to capacity (i.e. how much capacity for each fuel)? Show it in a pie chart.

(b) What is the fuel composition of this market in terms of net generation? Show it in a pie chart.

(c) Why are they different or similar?

3. (October 5) Organize the data and plot the generation supply curve using a different color for each fuel (reproduce the plot you can build online). For this, you have to order generators according to variable cost, and calculate the cumulative capacity of the system. Use geom_point such that each plant is a dot, but do not connect the dots. Label the plot properly, add a title and a legend.

(a) Are fuels ordered by cost? What do you think is the role of cost in explaining the differences between the capacity and net generation shares of each fuel?

4. (October 12) Now go to the same website and download load data for the same ISO and year. You can find this by going to Power/Regional Summary Statistics.

(a) You will use three load values: summer peak, winter peak, and average load. To calculate average load, take the Actual Net Energy for Load (you can also use net energy from the power plant dataset you've been using so far), divide by the number of hours in a year (365×24) and multiply by 1000 to go from GWh to MWh.

(b) Add the load values to the supply curve plot as vertical lines. Save this plot as a pdf file using ggsave.

(c) For each of these three load levels, find the price that would have cleared the market if price were equal to cost, i.e. find the point in the supply curve intersects the load curve (vertical line) in the plot.

5. (October 12) Now you will calculate the profits that each generator would have made if the price had been the price you find assuming average load (the price on an average hour).

(a) First, calculate profits, which are given by $(P-mc)Q$. Use variable cost as marginal cost, quantity is net generation. Describe profits by fuel type. For this, create a table that includes minimum, percentile 25, mean, median, percentile 75, and max value for each fuel type (fuel types are rows).

(b) Now compute total profits considering fixed costs $(P-mc)Q-F$, and create the same table as above. Do firms cover their costs?

6. (October 16) Now we will repeat the same exercise but with social cost instead of private cost.

(a) We will use \$50 for the social cost of carbon, but write your code using it as a parameter such that you can easily change it (Define scc as a variable at the beginning, and use scc instead of the actual value in the code.).

(b) Compute the social variable cost for each generation. First, combine heat rate and CO_2 emissions rate to obtain CO_2 tons emitted per MWh. Then, multiply this value by the social cost of carbon.

(c) Plot the social cost supply curve and load. Find the price that would clear if load was equal to average load (call this P^S), and the corresponding quantity.

7. (October 16) Now we will compare total emissions in a world that dispatches plants according to private variable cost (as it is today), with a world in which plants are dispatched according to social cost (which would happen with a tax on carbon, for example).

(a) Calculate total emissions during an average hour assuming firms are dispatched according to social cost. For this, you have to follow these steps:

    i. Create variable that indicates whether a plant operates or not in an average hour (it operates if its social cost is lower than or equal to P^S).

    ii. Create another variable with each plant's emissions assuming they produce at capacity (why capacity? Think of the supply curve).

    iii. Then add up emissions for all plants that operate when the price is lower than or equal to P^S.

(b) Find total emissions for the case in which plants are dispatched according to private variable cost, not social cost, using the procedure above. Notice that now a plant operates if its variable cost is at or below the clearing price calculated in 4c.

(c) How many tons could we save in an hour if the social cost of carbon were internalized? And in a day? And in a year? Comment about how the fuel shares of this particular market affect the results of this exercise.

(d) (Bonus) How does the above answer change with different values for the social cost of carbon?

(e) (Extra bonus) Can you write a function that helps you to compute this?