Columbia University SIPA
Fall 2020

Empirical Analysis of Energy Markets - U6616

Empirical Exercise 2 - Supply curve

Prof. Ignacia Mercadal

This problem set is due on October 26. I strongly recommend to start working on the homework early. You can work in pairs and submit a common solution. Please submit the homework as an R markdown file (if there are data files, they put all the files in a zip file). The code must run without errors. To make this easier, set the working directory at the beginning so it can be easily changed by someone else running the code.

For this exercise you will use a dataset collected by S&P Global. Choose one of the datasets available, which have data for either 2009 or 2018 for one of the following ISOs: MISO, PJM, ERCOT, or New England ISO. The goal of the exercise is to build the supply curve for a wholesale electricity market and to analyze how costs determine the composition of fuels and emissions. We will also use the exercise to see how things would change with a carbon tax.

Select the following variables from your dataset:

- 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 (MMBTU)
- Net generation MWh
- Capacity factor
- NO_X Emissions Rate (lbs/MMBtu)
- SO_2 Emissions Rate (lbs/MMBtu)
- CO_2 Emissions Rate (lbs/MMBtu)

1. 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. 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.[1]
   (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?
   (d) How much does each fuel contribute to NO_x, SO_2, and CO_2 emissions? Choose an appropriate plot type to answer this.

3. Organize the data and plot the generation supply curve using a different color for each fuel (Check here for a reference about supply curves.). The idea is to have a plot in which each plant is a dot, its height is its variable cost and its x-coordinate is the capacity of the system at a cost equal or lower than the plant's. 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.

4. In the supply curve, 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?

5. Now you will create three values that we will use to represent load. Let's assume average load is 60% of capacity, winter peak is 80% of capacity, and summer peak is 90% of capacity.
   (a) Compute these three values of load.
   (b) Add the load values to the supply curve plot as vertical lines. Save this plot as a pdf file using ggsave.

[1] Net generation is the amount of energy produced by a power plant, net of the energy used to produce. Basically, the amount of energy that comes out of the plant. Capacity is the maximum amount of energy that a plant can produce in a given hour. For this reason, net generation is measured in MWh over a certain period and capacity in MW.

(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.

6. Suppose we want to know if the dispatch of power plants is efficient, i.e. if cheaper plants are dispatched first. Do cheaper power plants produce more? To check this, do the following:

   (a) Run an OLS regression of net generation on cost. What cost is the most relevant here? Try total cost and variable cost and argue why/how results vary with the cost definition. Briefly discuss.

   (b) Now control for capacity, how do results change?

   (c) What else could you be missing that may lead to bias? Can you control for it? Add some control that you consider relevant and discuss how it changes the results.

7. (Extra credit) 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?

8. (Extra credit) 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, heat input, and CO_2 emissions rate to obtain CO_2 tons emitted per MWh. In the data we have the emission rates in lbs/MMBTU for CO_2. We also have the heat input in MMBTU. Therefore, by multiplying both variables we can construct the emission in lbs. Then, multiplying this variable by a constant $k = 0.00045359237$ we can convert from lbs to tons. Finally, we know the generation in MWh. So, we can calculate the emissions in tons per MWh.

   CO_2 Emissions rate$_i$ $(tons/MWH) =$

   $$\frac{\text{Emission rate}_i\ (lbs/MMBTU)\ \cdot\ \text{Heat input}_i\ (MMBTU)\ \cdot\ k\ (ton/lbs)}{Generation_i\ (MWh)}$$

(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.

9. (Extra credit) 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 5c.

   (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?