Capstone 1 Experimental Design - Research Proposal

May 31, 2020

Michael Dwyer

1 Introduction

I will be examining average retail electricity prices throughout the United States over nineteen (19) years to understand any significant pricing differences in geographies and industry sectors. The power generation industry is a very long-term industry with most company revenue contracts and fixed asset investment horizons ranging from ten (10) to thirty (30) years long. When making such large, long-term investments and commitments, power generators expose themselves to significant known and unknown business risks. With the glut of natural gas maintaining low power prices in certain regions, generators which are renewing their contracts end up repricing their contracts and receiving less income or sometimes decide to shutdown due to the negative, inverted economics. Today, generators are exposed more and more to fluctuating energy market prices with the continued establishment of idependent system operators (ISOs).

1.1 What energy market geographies within the US, if any, have held a consistently higher power price compared to the national average?

This information could start the process of understanding where to target new opportunities for energy generators to build new plants, purchase land for future development near key grid sites, and acquire existing power plants. This might result in identifying local areas where supply is constrained but could be helped by a new technology, or where demand is consistently outpacing the speed of current development. Either way, a general direction for further research can be established that will pave the wave to understand regional differences in the power markets and composition type of power plants.

1.2 What energy market geographies within the US, if any, show a statistically significant trend up or down during the period from January 2001 to February 2020?

This also has strategic value for the power generation industry in the way of new opportunities. However, this could also provide clues for understanding what drivers and signals have affected energy prices the most throughout recent history. By examing the statistically significant pricing trends up or down in each energy market geography over a time range of nineteen (19) years, specific points in time and geography could be established for further research. Understanding significant changes and trends in the energy market and generation structure could give clues or context for management teams to better predict potential risks and encourage mitigation of impacts from potential major events. For example, a future analysis could examine the impact that low natural

gas prices and growing solar installations have had on energy prices over periods of time, and use that to predict what impact another future event could have on the power generation industry.

2 Hypothesis

2.1 Comparing energy prices between geographies.

H_o: There is no significant difference in the average retail electricity prices of the geographic regions of Pacific Contiguous, Pacific Noncontiguous, South Atlantic, Middle Atlantic, West South Central, and East North Central when compared to the entire United States.

H_a: There is a significant difference in the average retail electricity prices of the geographic regions of Pacific Contiguous, Pacific Noncontiguous, South Atlantic, Middle Atlantic, West South Central, and East North Central when compared to the entire United States.

2.2 Comparing energy prices within geographies over two separate time periods.

H_o: There is no significant difference in the average retail electricity prices between chronologically separated groups, dated from 2001-01 to 2009-12 and from 2010-01 to 2020-02, of the geographic regions of Pacific Contiguous, Pacific Noncontiguous, South Atlantic, Middle Atlantic, West South Central, East North Central, and the entire United States.

H_a: There is a significant difference in the average retail electricity prices between chronologically separated groups, dated from 2001-01 to 2009-12 and from 2010-01 to 2020-02, of the geographic regions of Pacific Contiguous, Pacific Noncontiguous, South Atlantic, Middle Atlantic, West South Central, East North Central, and the entire United States.

3 Data

Both research topics use the same data set. The dataset has been downloaded from the Energy Information Administration (EIA) website at https://www.eia.gov/electricity/data/browser/, or https://www.eia.gov/opendata/qb.php?category=40. Although the data is a more comprehensive data set including generation statistics, average fuel costs, fuel deliveries, fuel consumption, and fuel inventories; for this project I will only examine average retail electricity prices in the United States on a monthly basis from 2001 to February 2020. The electricity price data is broken down by region, state, and sector, and has been downloaded into a usable .csv file in the local folder.

Within the electricity prices report, there are 7,844 observations and five (5) variables including location, sector, month, year, and price. There are 1,150 observations in the South Atlantic, East North Central, Pacific Contiguous, Middle Atlantic, West South Central, and United States regions, and 944 in the Pacific Noncontiguous region.

The geographic areas of the United States are defined as follows:

1. **New England:** Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont.
2. **Middle Atlantic:** New Jersey, New York, and Pennsylvania.
3. **East North Central:** Illinois, Indiana, Michigan, Ohio, and Wisconsin.

4. **West North Central:** Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, and South Dakota.
5. **South Atlantic:** Delaware, District of Columbia, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, and West Virginia.
6. **East South Central:** Alabama, Kentucky, Mississippi, and Tennessee.
7. **West South Central:** Arkansas, Louisiana, Oklahoma, and Texas.
8. **Mountain:** Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, and Wyoming.
9. **Pacific Contiguous:** California, Oregon, and Washington.
10. **Pacific Noncontiguous:** Alaska and Hawaii.

4 Methods

I will split the data set into seven (7) dataframes, one for each location region, to separately compare each region to the United States and within itself over time. Each dataframe will be reviewed to see if the segmented variables are normally distributed. Normality will be tested by visual inspection of histogram distributions, a review of skewness and kurtosis statistics, and a Shapiro-Wilk test.

4.1 Comparing energy prices between geographies.

To test H_o, I will run a One-way ANOVA test and Tukey's honest significant differences (HSD) test if the data is normally distributed, and a Kruskal-Wallis test if it is not, to see if any stand out from the group. Then, I will proceed to run six (6) independent t-tests if the data is normally distributed comparing each geographic region to the United States. If no clear patterns emerge, then I will expand the analysis to run the same six (6) independent t-tests on a smaller time horizon such as each year, and graph the results to see if any events did occur that were diluted down by or lost in the overall noise of the larger data set.

4.2 Comparing energy prices within geographies over two separate time periods.

To test H_o, I will run seven (7) independent t-tests if the data is normally distributed, and a Kruskal-Wallis test if it is not, comparing each geographic region to itself during different time periods to see if an overall upward or downward trend occurs. As a broad starting point, I will split the time periods into 2001-01 to 2009-12 and 2010-01 to 2020-02. If no clear patterns emerge, then I will expand the analysis to run the same seven (7) independent t-tests on a smaller time horizon such as each year or each month, and graph the results to see if any events did occur that were diluted down by or lost in the overall noise of the larger data set.