MA678 Midterm Project Report

Jingyu Liang

2022-12-08

Abstract

In this report, I want to find the relationship between natural gas prices and different years and different places. I set each of the years between 2010 to 2020 as one group. I also divided the prices in different places into different groups. In this case, I use the multilevel model to fit the data and do some exploratory data analysis(EDA) to visualize the data and help you to understand my analysis. Finally, I found the coefficients of different places affecting the price and the price was decreasing during 2010 to 2020.

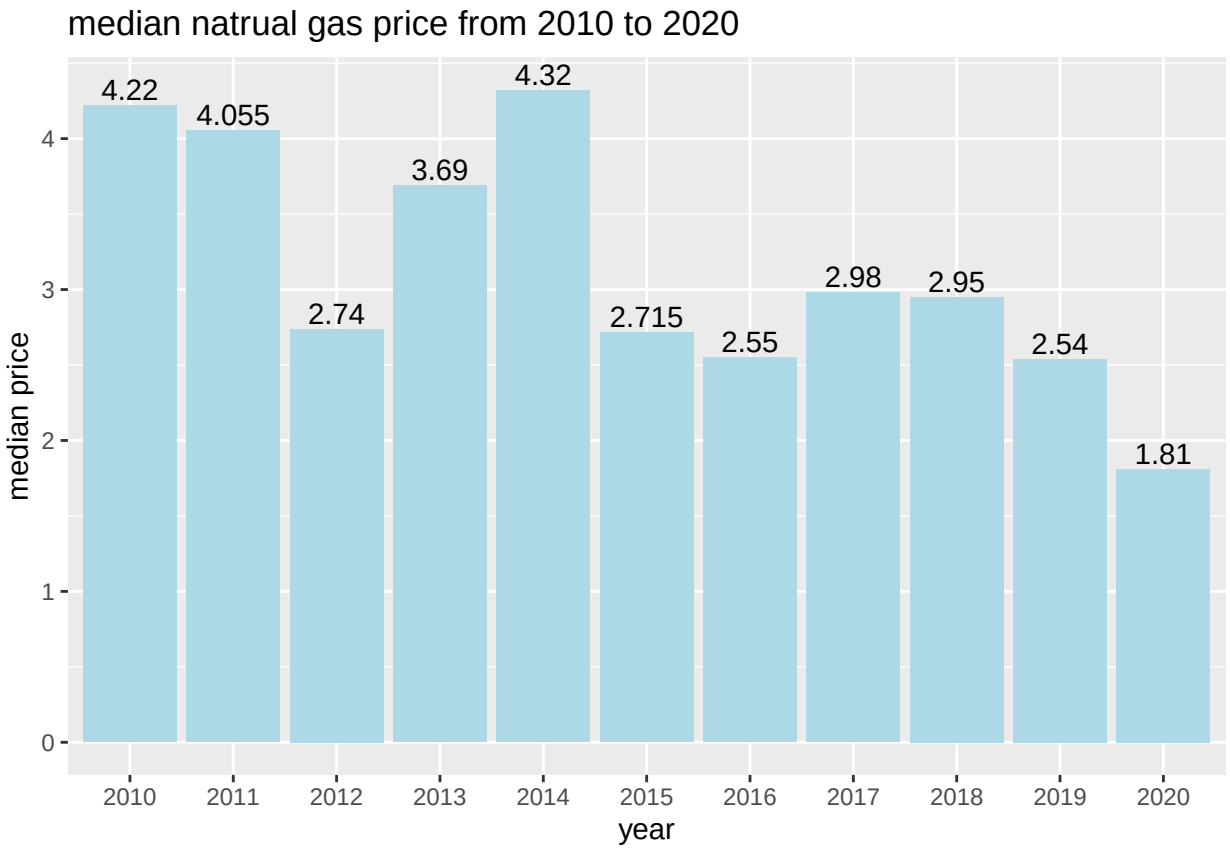
Introduction

Natural gas is one of the most important types of fossil fuels in the world. It's a clean-burning source of energy, which means it is harmless to the natural environment, so natural gas is used in many places in human life. For example, heating, cooling, and electricity generation rely on natural gas every day. In this case, the price of natural gas means a lot to humans. The goal of my analysis is to find the variation in the natural gas price over the years and the variation of natural gas price in different places.

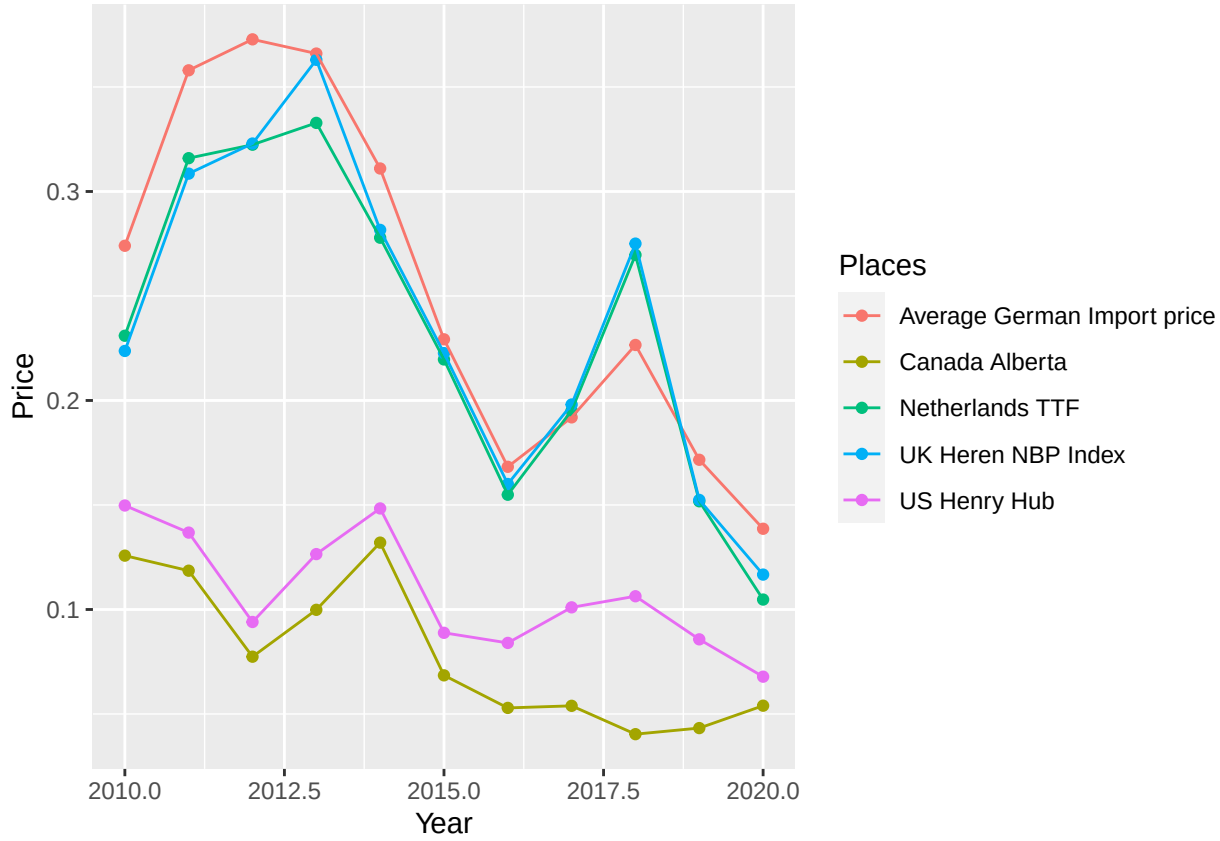
Method

Exploratory Data Analysis (EDA)

Here are some EDA I made to visualize the data. The first plot displays the median natural gas price during 2010 to 2020. The median price is basically going down in these 11 years. In 2020, the median price of natural gas reached the lowest point.



In this second plot, I wanted to show the comparison of the natural gas price variation between different countries. I used another dataset that has the natural gas prices of different countries from 2010 to 2020. We can clearly see that the prices of natural gas in German, the Netherlands, and the UK are greater than prices in the US and Canada. Therefore, north America basically has a cheaper natural gas price. Overall, the price of natural gas is going down for 11 years.



Model Fitting

From the above EDA, I found that the price is different between different years, so I fit a multilevel model with 11 groups, and each group indicates one year. Below is a simplest multilevel model.

```
M0 <- lmer (formula = Price ~ 1 + (1 | Year), data = outputdata)
```

We can see the estimated model coefficients with each year.

```
coef(M0)
```

```
## $Year
##      (Intercept)
## 2010    4.368054
## 2011    3.995163
## 2012    2.755072
## 2013    3.730494
## 2014    4.371026
## 2015    2.624742
## 2016    2.516866
## 2017    2.988285
## 2018    3.152693
## 2019    2.561785
## 2020    1.864578
```

```
##
## attr(,"class")
## [1] "coef.mer"
```

According to the 2nd plot in the above, I found place also can affect the price of natural gas. Then, I used the year as predictor and places as groups to make another multilevel model. This model is a varying-intercept regression with a single predictor Year, which means the intercept can vary by Places.

```
M1 <- lmer(formula = Price ~ Year + (1 | Places), data = natgas1)
```

We can see the estimated model coefficients with each place.

```
coef(M1)
```

```
## $Places
##                              (Intercept)        Year
## Average German Import price    28.04313 -0.01379156
## Canada Alberta                 27.87183 -0.01379156
## Netherlands TTF                28.02265 -0.01379156
## UK Heren NBP Index             28.02695 -0.01379156
## US Henry Hub                   27.90030 -0.01379156
##
## attr(,"class")
## [1] "coef.mer"
```

Result

The first model just includes a constant term and it can vary by years. From the 2nd model, we can get these formulas of the result:

$Price = 28.04 - 0.01 \times Year$ is for Average German Import price

$Price = 27.87 - 0.01 \times Year$ is for Canada (Alberta)

$Price = 28.02 - 0.01 \times Year$ is for the Netherlands TTF

$Price = 28.03 - 0.01 \times Year$ is for UK (Heren NBP Index)

$Price = 27.90 - 0.01 \times Year$ is for US Henry Hub

So, with the above formula, we can see the price of natural gas was decreasing over the years, which is the same as the result I saw in the EDA. In addition, a different place has a different effect on the price, as we can see the intercepts in different places are different.

Discussion

After getting the above results, I think there may exist some other factor that affects the price of natural gas, such as the temperature in a place. Temperature and place may have interactions because if this place is far to the north, the temperature must be lower than in south places, which will affect the price of natural gas. It is interesting to dig deeper in this direction.

Appendix

The data are downloaded from these two websites. https://datahub.io/core/natural-gas#data-cli https://ourworldindata.org/grapher/natural-gas-prices ##Citations Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel: Hierarchical models.* Cambridge University Press.

Holtz, Y. (n.d.). *Help and inspiration for R charts.* The R Graph Gallery. Retrieved December 11, 2022, from https://r-graph-gallery.com/

Why is natural gas important for society? TC Energy. (n.d.). Retrieved December 11, 2022, from https://www.tcenergy.com/stories/2020/2020-08-06why-is-natural-gas-important-for-society/#:~:text=It%20means%20homes%2C%20businesses%2C%20cities,steel%20and%20concrete)%20and%20more.

Supplement

The code can be found in this GitHub link: https://github.com/JingyuClaire/midterm_project