Gasoline Price of United States

Yingzi XU

Department of Statistics, North Carolina State University

yxu15@ncsu.edu

1. Description of the Data

- Gasoline price data from U.S. energy information administration official website. Retail prices for regular gasoline are collected by telephone from a sample of approximately 800 retail gasoline outlets each month.
- P_{it}: Weighted Regular Conventional Gas price (dollars per gallon) at seven different regions of U.S. each month from January 1995 to January 2014;
- IR_t: Inflation Rate each month from January 1995 to January 2014;
- CO_{it}: Crude oil production(thousand barrels) each month for each region from January 1995 to January 2014.

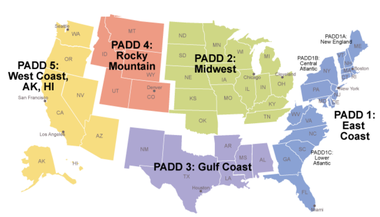

Figure 1: *Petroleum Administration for Defense Districts (PADDs)*

2. Main Goals

- Visualize the price difference in seven regions of U.S.
- Use crude oil production, inflation rate to see the relationship between gas price and these covariates .
- What is the price trend over time in each region and how that pattern varies between different regions?
- Predicting the future gas price for these regions.

3. Data Visualization

- From Figure 2 and 3 , We can see that different regions are having the same trend but has a bit fluctuation due to the different location. West Coast has the highest gas price for almost all time. Also, there seems to be spatial correlation between these regions.

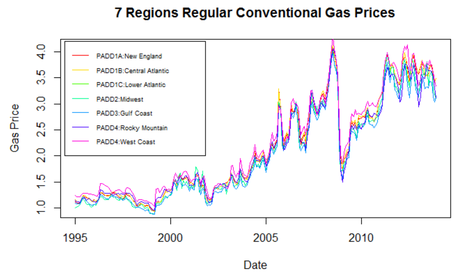

Figure 2: *Regular Conventional Gas Prices for Seven Regions From 1995 to 2014*

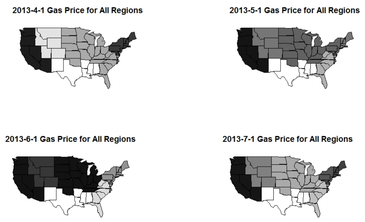

Figure 3: *Spatial Looking of Regular Conventional Gas Price in all Regions*

4. Remove Mean Trend Model

Cubic Model of Gas Price on Time:

$$P_{it} = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + Y_{it},$$

where t is the linear time effect, t^2 denotes the quadratic time effect and t^3 denotes the cubic time effect, we get the estimate of parameters as following,

5. Remove Mean Trend Result

Table 1: *Parameter Estimate for Cubic Model of Gas Price on Time*

Parameter	Estimate	Std. Error	P-value
β_0	$1.922e-01$	$1.537e-02$	$< 2e-16 **$
β_1	$-5.457e-03$	$5.801e-04$	$< 2e-16 **$
β_2	$1.166e-04$	$5.879e-06$	$< 2e-16 **$
β_3	$-3.192e-07$	$1.688e-08$	$< 2e-16 **$

From table 1, all three time effect are significant.
We then check the normality of the residuals Y_{it}. It turns out that this detrend makes residual Yit much more normal by checking its normal Q-Q plot.

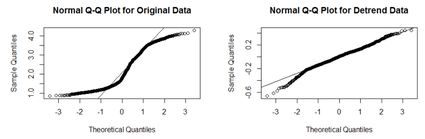

Figure 4: *Normal Q-Q Plot of Original Gas Prices and Residual of Gas Prices for All Regions*

6. Spatial Panel Model

Spatial panel data are typically data containing time series observations of a number of spatial units(zip codes,states,regions,countries,etc.).

Fixed Effect Spatial Lag Model:

$$y = \lambda(I_T \otimes W_N)y + (l_T \otimes I_N)\mu + \beta_{IR} IR + \beta_{CO} CO + \epsilon$$
$$\epsilon = \rho(I_T \otimes W_N)\epsilon + \nu$$

where where λ is the spatial autoregressive coeffcient, W_N a non-stochastic 7×7 spatial weights matrix where 1 denotes two regions are adjacent to each other and 0 denotes otherwise, l_T a column vector of ones of dimension T, I_N an 7×7 identity matrix,μ is a vector of time invariant regional specific effects (not spatially autocorrelated), and $\epsilon \sim N(0, \sigma_\epsilon^2)$, $\nu \sim N(0, \sigma_\nu^2)$.

7. Results

- A maximum likelihood estimation would be used to estimate the parameters. Parameter estimate are shown in Table 2.
- Here, λ and ρ are significant indicating a good capture of spatial autocorrelation. Inflation rate and crude oil production both has a significant positive impact on the gas price.

Table 2: *Parameter Estimate for Spatial Panel Model*

Parameter	Estimate	Std. Error	P-value
λ	$5.7148e-02$	$1.2261e-02$	$3.148e-06 **$
ρ	$2.3914e-01$	$5.2323e-03$	$< 2.2e-16 **$
IR	$8.2573e-02$	$2.4227e-03$	$< 2.2e-16 **$
CO	$1.4467e-05$	$4.4469e-06$	$0.00114 *$

- The final fitted line is shown in Figure 5.it captures both the mean trend and spatial deviation.

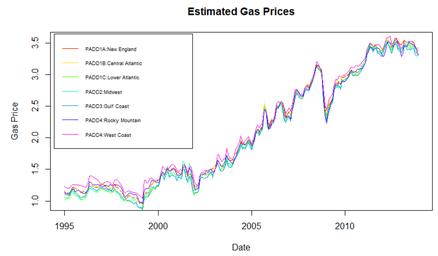

Figure 5: *Fitted Value of Regular Conventional Gas Prices for Seven Regions From 1995 to 2014*

8. Discussion

- Other kinds of weight can also be used and then we can use cross validation to find the best weight for the model.
- For prediction, given the future time, crude oil production and inflation rate, I would be able to do that. However, since we don't know crude oil production and inflation rate in the future, it may need be estimated, in which case, the stand error would be bigger for the prediction of the gas price.