Gas Price Micro Markets

Kai Lukowiak

September 16, 2018

Abstract

Gas station prices are dependent on competiton. Due to the nature of gasoline markets this competition is very local. Industry is not very sophisticated in segmenting these micro-markets leaving money on the table.

This proposal shows the roadmap to use variation and similarity in regional prices and geospatial data to identify these micro-markets and find outliers that are close to but otherwise perform differently.

The end goal is to provide information on these stations along with a framework that companies can use to experiment further.

1 Description of the Problem

Retail Gas Prices (RGP) are determined in the competitive market. Gas stations are surprisingly unsophisticated given that the market in Canada is $34 billion. If pricing is granular to the site level, it is almost entirely set by regional or even station managers and national campaigns do not have the granularity to capitalize on market differences.

There are many facets that go into the definition of a competitive market. Cataloguing all features and recording them is a gargantuan task. Instead this study hopes to discern which stations may be different from their surrounding competition. This will allow more targeted experimentation.

The competition is based on many factors like loyalty programs, marketing, customer preferences etc. One major influence on RGP is geographic location. Even within the same distribution market (same base fuel price), a station on a major highway, one in close proximity to three other competitors and one in a rural area will all behave differently.

2   Why the problem is interesting

There is aproximetly \$50 billion L of motor fuels sold in Canada yearly. $0,000,000,000L \cdot \alpha$ where α is small increase in cents per litre (CPL) is still a lot of money.

Beyond the fiscal incentives, this is an interesting study because it will reveal (hopefully) two findings. The first is finding relevant markets in geographic areas. For example, it is easy to believe that gas stations on different sides of a major river or other barrier would price differently.

Secondly, after geographic markets are found, gas stations that are contained within a similar geographic area, but that price differently can be further analyzed. For example, a suburban area may be a well defined market with similar price fluctuation in stations except for one. This will allow an analyst to further investigate this one station to find which features make it more profitable (or less).

These findings could help set prices more efficiently as well as indicate which areas are best for future investment.

3   What other approaches have been tried

Investigating retail gasoline markets is nothing new. Most recently in Canada, Parkland has a large market of retail stations and merged with Pioneer Energy based in Ontario. The Competition Bureau forced Parkland to divest several stations they believed would form regional monopolies.

I am also sure marketing departments have segmented promotional efforts based on location.

Both of these focus on a more macro level geography.

There are many examples of spatial/temporal regression investigating gas prices in the economic literature. Some of these are quite granular, although I was unable to find any on the same scale as I plan. These are listed in the links at the bottom of this proposal.

Further these look at the causal effect of distance on price. My intention is to cluster sites and compare them to other clustering techniques such as KNN. Some examples of spatial regression can be seen below as well.

4 Discussion on your hypothesis is and how your specific solution will improve

My hypothesis is that there are anomalous gas stations and unobserved sub-regions which can be detected through variance spatial/temporal pricing.

Further, due to my disinterest in causality/interpretation I can use also use non-linear techniques that make interpretation difficult.

My first approach will be to use PySal to investigate linear relationships between stores and cluster them by similarity to their prices.

I also plan on investigating other correlation/regression techniques with numbers weighted by distance.

After getting a functioning model, I plan on comparing this to clustering algorithms such as KNN with a similar number of centroids to the previous model. I will then compare the overlap to the model taking into account price. This geographic comparison will shed light onto unique gas stations.

Further, this model can be adopted by industry to perform experiments on individual stations allowing for the discovery of price leaders and even more granular market segments.

4.1 Effects of Edgeworth Cycles

Edgeworth Price Cycles occur in competitive industries where firms find it profit maximizing to cut prices until they reach marginal cost. At this point, they can profit maximize by increases prices well above marginal cost. This cycle repeats, and in the case of retail gas, often on a weekly basis. I plan on using this cycle to determine relevant markets.

4.2 Additional Features

I also plan on taking in historic prices for the area on an anualized basis. This, along with weather, holiday and macroeconomic information will be added to try and remove omitted variable bias.

4.3 Data Acquisition

I plan on using a scraping script, AWS Lambda and S3 to find current gas prices in Western Canada. Other data like historic weather conditions and gas prices, will be found from Stats Can.

Locational data will be scraped from google maps API.

5 Additional Info

- https://link.springer.com/article/10.1007/s00168-007-0206-7

- https://www.jstor.org/stable/20111978?read-now=1&googleloggedin=true&seq=1#page_scan_tab_contents

- https://link.springer.com/chapter/10.1007/978-3-7908-2070-6_12

- http://journals.ama.org/doi/abs/10.1509/jmkr.44.4.622?code=amma-site

- https://www.jstor.org/stable/41323223?seq=1%23page_scan_tab_contents

- http://ses.wsu.edu/wp-content/uploads/2015/03/SpatialDifferences.pdf

- http://www.econ.uiuc.edu/~lab/workshop/Spatial_in_R.html

- http://darribas.org/gds_scipy16/ipynb_md/08_spatial_regression.html