Relationships between Oil and GAS

ABSTRACT
This project is to show that CA gasoline prices even with its own special formulation correlate with oil prices. That as oil prices change so do gasoline prices in lock step with some lag, which is attributable to the nature of the market.

Jason Spero
CIS 320

Contents

Introduction

The rapid fall of oil prices and the very soft fall of gas prices not only nationwide but also the even softer fall in the state of California instigated this data analysis project. This begged question such as why this drop occurred? Aare these prices were normal, compared to historic prices. to see if the lock step the correlation between oil and resulting gasoline prices were still in play both in and outside California? If possible, implement forecast data to project future prices.

Oil and Gas prices have not been this low since 2004, before the great recession, during they fluctuated but remained very high, approaching $145/bbl. and $5/gal regular unleaded in CA, then now dropping to $30/bbl. and $2.25/gal in CA and the rest of the US 1.00/gal such as in Missouri. With fuel costs being considered a tax on goods and services, along with a regressive tax on the poor which could make or break profitability for business and family. Making understanding the past, present and possible future of gasoline and oil prices is important for everyone.

Let us begin this analysis with context; for starters, 42 % of every barrel is refined in to gasoline (EIA), making oil the most significant factor in gasolines cost. And as such historically gasoline and oil have been in lock step, as oil rises gasoline rise, and vice versa. The market for oil is global and unlike the doomsayers from the

‘70s, its abundant and new sources are being discovered every day. However, not so plentiful are the refineries to make gasoline, in the last 30 years no new refineries have been built and through attrition fineries have closed, in California 4 alone. On top of current refineries going down due to maintenance or catastrophic malfunctions.

With oil, under current federal law it is illegal to export, but refined products can be so along with California having its own special blend of gasoline, which only a small subset of refineries can make, has constrained refinery capacity on the whole. Which is blamed not only for the US but especially so California as the antidotal reason why even today gasoline prices seem to be higher than they were last time oil was this low.

This analysis with start with line plot comparisons of oil, and reformulated gas prices over time, from that data univariate statistics will be calculated for both products. Then linear and logarithmic regression correlations of gasoline along with oil prices will be plotted and R^2 values will be calculated to further ascertain the importance of the price of oil. From all this data, we hope to answer the question at hand. Then to further show the importance of oil prices, forecasts will be made by using EIA STEO (Shot Term Economic Outlook) data for WTI oil, this will be the basis for future gasoline prices as oil is the foremost component to inform these predications.

Hypothesis

What is the key driver for gasoline prices? Is it the cost of Oil, or is it the other factors?

Since 42% of every barrel of oil goes to gasoline production it is anecdotally assumed that oil will be the key driver, but it is not the only component to the price of Gasoline, if so those other factors will be prominent in the analysis. Linear and Logarithmic regression models will be used to show the amount contributable to Oil and other factors to gasoline price levels.

Methodology

Research Methods

This analysis required on the research on the whole of the oil and gasoline market and two key pieces of data; historic prices of oil from the past to the present, and CA reformulated gas (US all grades is used as proxy) from the years of 1994 to today. This is due to the limitation of gasoline price data.

From research on the relationship between oil and gasoline prices, the EIA had deduced that the petroleum refinery industry as a whole uses the price-pass-through model, where at each stage from oil derrick to gas station value or inputs through additional costs are added along the way to the products final form and purchase. However, from each barrel of oil, 42% goes to gasoline, making it the largest component, so it is logical to use oil as the comparison for gasoline prices, because hypothetically as oil prices change, gasoline should change in lock step.

Even the price-pass-through model, allows for additional costs to associated with other factors besides oil, such as marketing, refining, transport, seasonality and taxes once sold at the pump, but according to the EIA oil is the defining cost.

U.S. all grades reformulated as it the type of Gasoline mandated to use by state law, as it has additives that allow for cleaner burning

WTI (West Texas Intermediate) is used as the average barometer for US oil prices and is routinely quoted in financial news reports.

To understand oil and gas's relationship, I will use these data sets, compare them, and then draw conclusions. To answer the question, is oil sill the key driver in gasoline price? On the other hand, if it is other factors besides oil.

These will be informed by;

Drawing historical line plots for all price level data sets.

Extracting univariate statistics from both oil and gasoline price data sets.

Developing linear models, and regression line models for both raw and logarithmic numbers, then comparing the R^2 values in determining the magnitude of correlation, and plotting correlations.

For monthly Oil price levels, I used the WTI (West Texas Intermedia) price history due this the primary one quoted in newspapers, there data goes back to 1986 to today, from the EIA.

For CA Reformulated Gasoline price levels, I used data sets for US all grades Reformulated, from the EIA form 1994-today. For this reason, the analysis will consist of observation for this period.

All data will be read in from website and the processed in R.

Data sources

Burdette, Michael, and John Zyren. "Gasoline Price Pass-through." Gasoline Price Pass-through. U.S. Energy Information Agency, Jan. 2003. Web. 29 Feb. 2016. http:/www.eia.gov/pub/oil_gas/petroleum/feature_articles/2003/gasolinepass/gasolinepass.htm%3e.

U.S. All Grades Reformulated Retail Gasoline Prices (US Energy Information Administration)

http://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=EMM_EPM0R_PTE_NUS_DPG&f=W

Crude Oil Prices: West Texas Intermediate (WTI) (Federal Reserve Bank of St. Louis)

http://research.stlouisfed.org/fred2/series/DCOILWTICO/downloaddata

Short-Term Energy Outlook (U.S. Energy Information Administration)

http://www.eia.gov/forecasts/steo/report/prices.cfm

Project Example: What a Gas! The Falling Price of Oil and Ontario Gasoline Prices

http://www.r-bloggers.com/what-a-gas-the-falling-price-of-oil-and-ontario-gasoline-prices/

Code Samples

```
library(ts)
library(forecast)

# Read in the WIT Oil  Price Data
WTI_data <-
read.csv('https://www.quandl.com/api/v3/datasets/EIA/PET_RWTC_D.csv?auth_token=2P8RfNEKSfAsU
Ja2js9g&collapse=monthly&end_date=2016-02-21')
```

```
# Read in the California proxy Gas Price Data
data <-
read.csv('https://www.quandl.com/api/v3/datasets/EIA/PET_EMM_EPM0R_PTE_SCA_DPG_W.csv?auth
_token=2P8RfNEKSfAsUJa2js9g&collapse=monthly')

# Create a time series object for the WTI and California Avg
WTI <- ts(data=WTI_data$Value, frequency=12, start=c(1995,1), end=c(2015,12))
CA <- ts(data=data$Value, frequency=12, start=c(1995,1), end=c(2015,12))

# Create linear models (normal and log-log)
l1 <- lm(CA ~ WTI, data=combined)
l2 <- lm(log(CA) ~ log(WTI), data=combined)

# Compare relative performance
summary(l1)
summary(l2)
plot(l1)
plot(l2)

# Plot
plot(CA ~ WTI, data=combined, pch=16, cex=0.3)
abline(l1)
plot(log(CA) ~ log(WTI), data=combined, pch=16, cex=0.3)
abline(l2)

# Read in WTI forecast data
WTI_forecast <- read.csv
('https://www.quandl.com/api/v3/datasets/EIA/STEO_NYWSTEO_M.csv?auth_token=2P8RfNEKSfAsUJa
2js9g')

fit <- forecast(l2, newdata=data.frame(WTI=WTI_forecast$Value))

# Unlog
fit$mean <- exp(fit$mean)
fit$lower <- exp(fit$lower)
fit$upper <- exp(fit$upper)
fit$x <- exp(fit$x)

# Plot
plot(fit, ylab='California Average gas Price ($/gal)')
```

Results

Descriptive Statistics

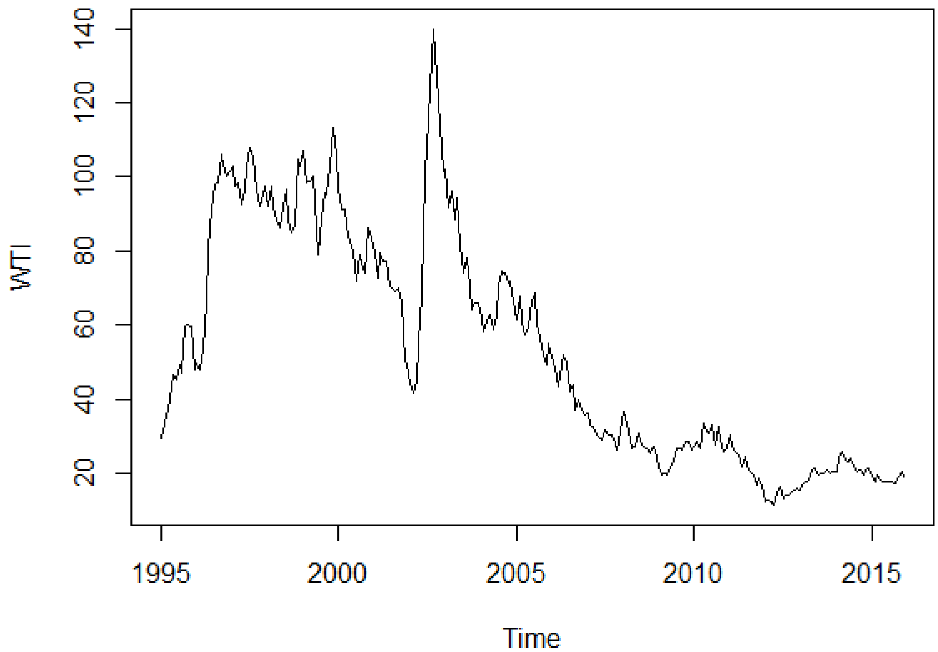
Univariate Statistics

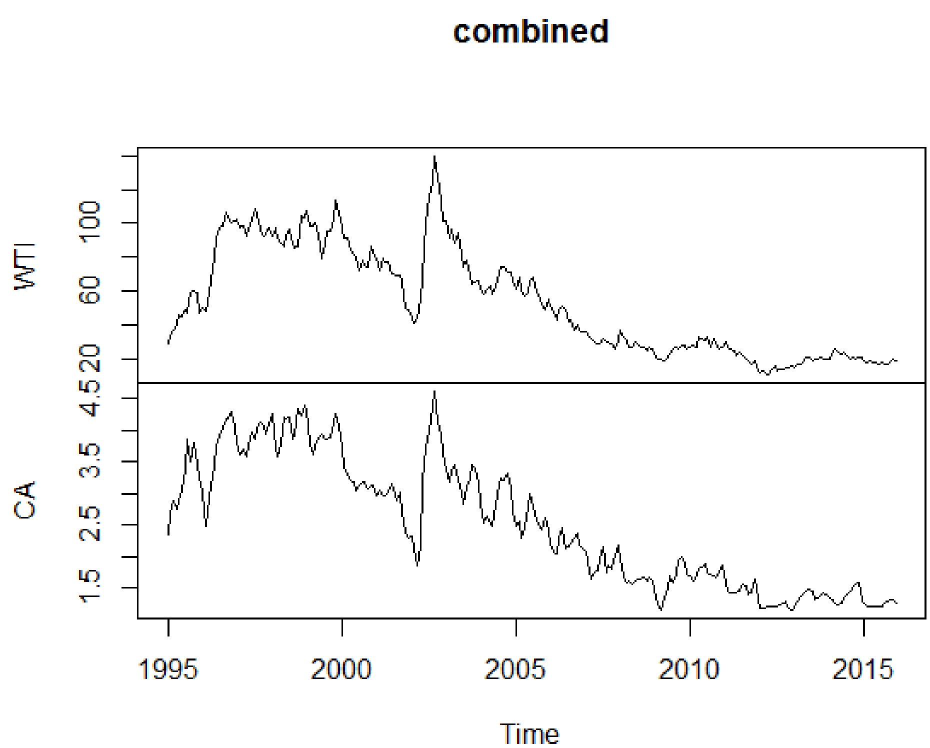
```
summary(WTI)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  11.35   20.19   29.70   46.66   74.44  133.90 
> summary(CA)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
  1.121   1.398   2.007   2.353   3.206   4.707       1 
```

Linear and Log-log Regression Statistics

```
summary(l1)

Call:
lm(formula = CA ~ WTI, data = combined)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.62951 -0.18536 -0.05619  0.10141  1.55799 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.8646517  0.0380061   22.75   <2e-16 ***
WTI         0.0307017  0.0006131   50.08   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3062 on 250 degrees of freedom
Multiple R-squared:  0.9093,	Adjusted R-squared:  0.909 
F-statistic:  2508 on 1 and 250 DF,  p-value: < 2.2e-16

> summary(l2)

Call:
lm(formula = log(CA) ~ log(WTI), data = combined)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.27608 -0.07775 -0.01123  0.05739  0.47707 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -1.53966    0.04196  -36.69   <2e-16 ***
log(WTI)     0.62702    0.01094   57.34   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1129 on 250 degrees of freedom
Multiple R-squared:  0.9293,	Adjusted R-squared:  0.929 
F-statistic:  3288 on 1 and 250 DF,  p-value: < 2.2e-16
```

Illustrative graphics

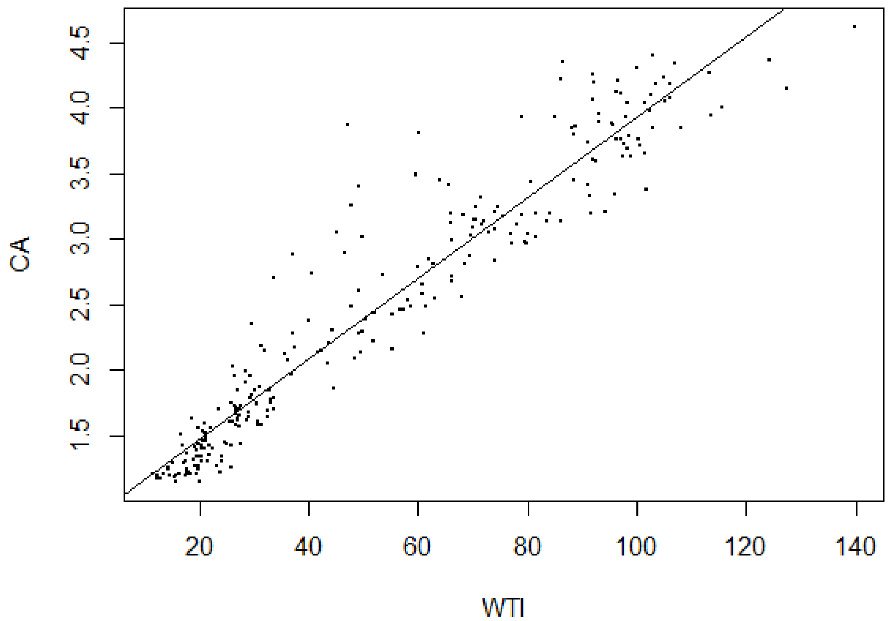




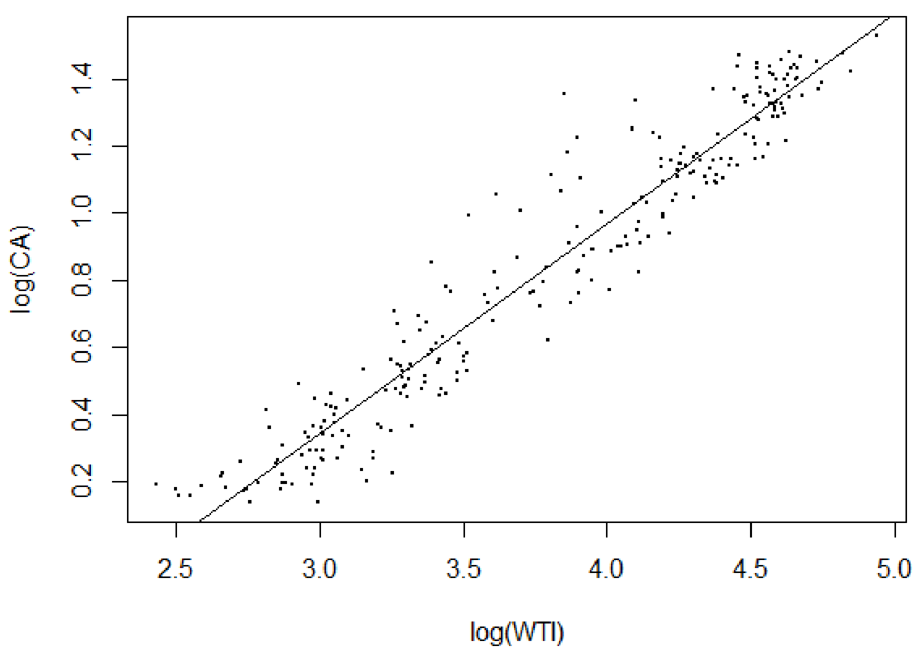
Hypothesis Testing

Regression models

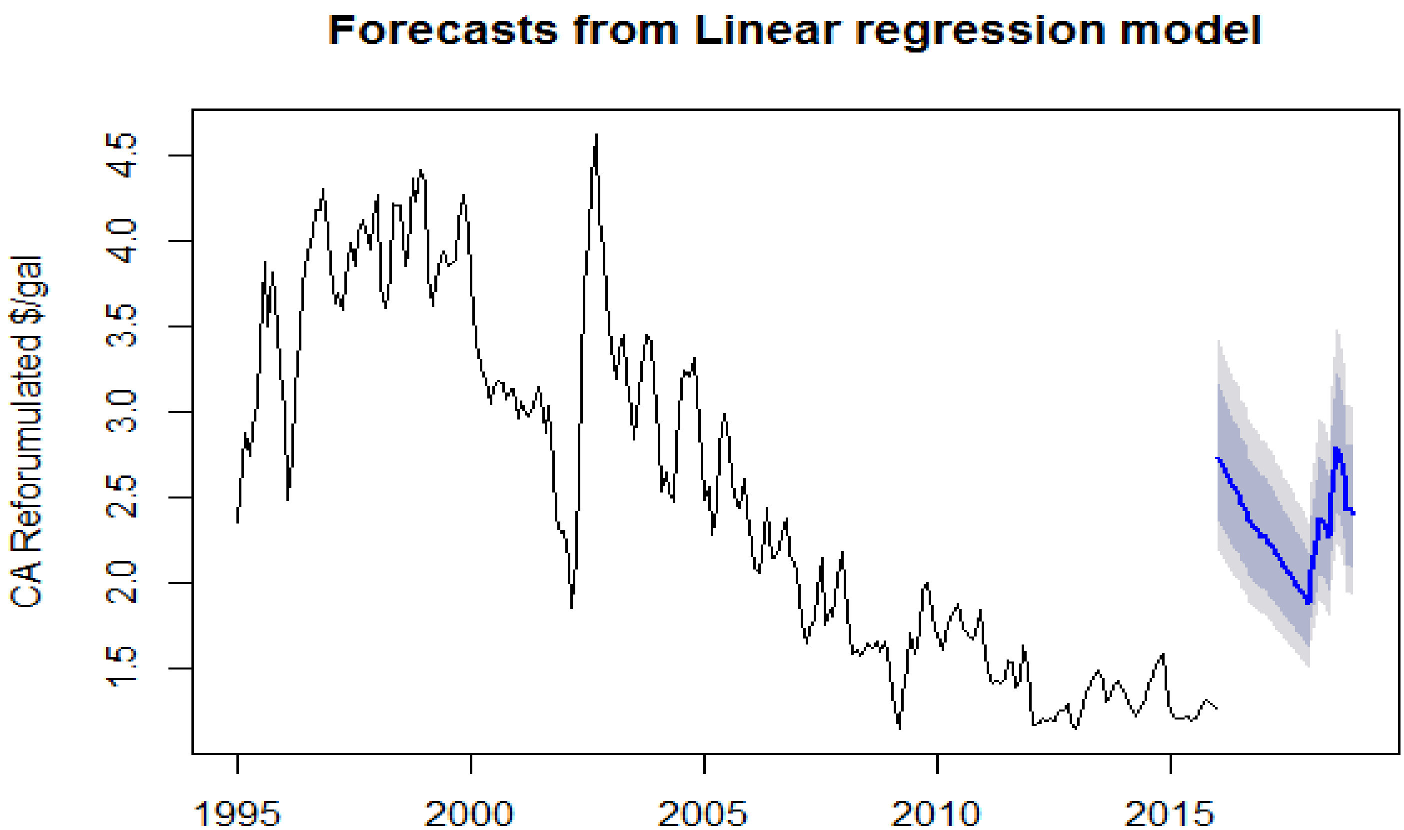
Linear



Log-log



Forecasting



Future gasoline prices are predicted using STEO derived oil prices, from 2016-2017.

Discussion

(craft your main arguments
by building on the results you have presented earlier)

The argument is that oil is a key driver for gasoline costs, and that as oil prices change so do gasoline prices change directly but somewhat lagged as the price difference moves through the market. We can see from the results just from the combined line plots of WTI oil prices v CA gas prices that they are an almost mirror each other, with minor divergences in later years.

The analysis of approximately 20 years, from 1995-2016 we can see from the 90's to the mid 2000's that oil and gas are in tight lock step, but after the recession in 2008 is when the beginnings of some noticeable lag or divergence, but still move relatively in step with each other. This is probably do to the wild swings in oil prices during this time, caused by rampant speculation and for gas, refinery's either shutting down or suffering catastrophic failures.

The key definitive evidence is the regression analysis. When gasoline and oil prices are scatter plotted together, we see even before plotting the regression line, that there is a positive correlation between the two. Both plots of regressions yielded statistics that over 90% percent of the cause for price changes over time is attributable to oil prices.

Logarithmic Regression shows a graph where data is plotted with Oil prices v. Gasoline prices, but Gas broken up in to categories based on price of gas. From the data we can see a diagonal form, indicating as oil prices change so do gas price indicating a direct relationship. There is not any anomalous data that would disprove the hypothesis, such has data points showing low oil price with high gas price or vice versa. This regression along with linear regression do seem to support the hypothesis that gas and oil prices are directly related, with oil the being the driving factor.

Linear Regression: Adj. R^2= 90.9%, so that leaves approx. 9% attributable to other factors.

Logarithmic Regression: Adj. R^2= 92.93, so that leaves approx. 7% attributable to other factors.

With this high correlation between gasoline and oil prices, an attempt to predict future gas prices is calculated. Using linear regression models, and future oil price data from the EIA STEO data set and the fit and forecast functions in R. Future gas prices were predicted and were fitted in to the current line graph of gasoline prices. The blue line is the predicted price,, while the gray and light gray regions represent other possible prices. According to it by approx. 2017, gas could reach 3.00, which if future events do not drastically change. That is the problem with linear regression, it requires variables to be stationary, but this is not true in the real world, which can and will affect the accuracy of these predictions. Nevertheless, regression analysis is used by the EIA in developing there models, so there is at least some validity in using it here to not only derive future gas prices but also find what is most contributable for their change.

Conclusion

Well, from all the data and research collected, regression analysis is is used and can at least point one in the direction of where gasoline prices are going and why do they change. We see her in spite that gasoline only makes up 42% of each oil barrel, it is a key driver of present and future gasoline price. Therefore, as long as oil remains cheap, gasoline should be also even if there is a bit of lag due to the nature of the market.

Proviso… due to the analysis inability to correct an error in reading in the data in to R for processing, all data points are read in backwards. Nevertheless, since all data is read in the same way, regression analysis and forecasting was still possible without any adverse effect to outcome. The only affect is cosmetic in plotting the line plots and forecasting. All code and data sources used in project are listed.