

National Retail Sales

Forecast



Matt Burns
DSC 630
Bellevue University
2/29/2020

Abstract

This project will identify a methodology to predict National Retail Sales using publicly available data back to 2002. The forecast is a linear regression model in RStudio.



Background) The United States Census Bureau published monthly retail sales. It is used as a barometer of retail health and is used throughout the business press.



Problem Statement) Since National Retail Sales are widely reported by the media, its reported figure impacts the markets. Using other metrics like population growth, consumer debt burden, energy prices and Consumer Price Index, can we predict the National Retail Sales?



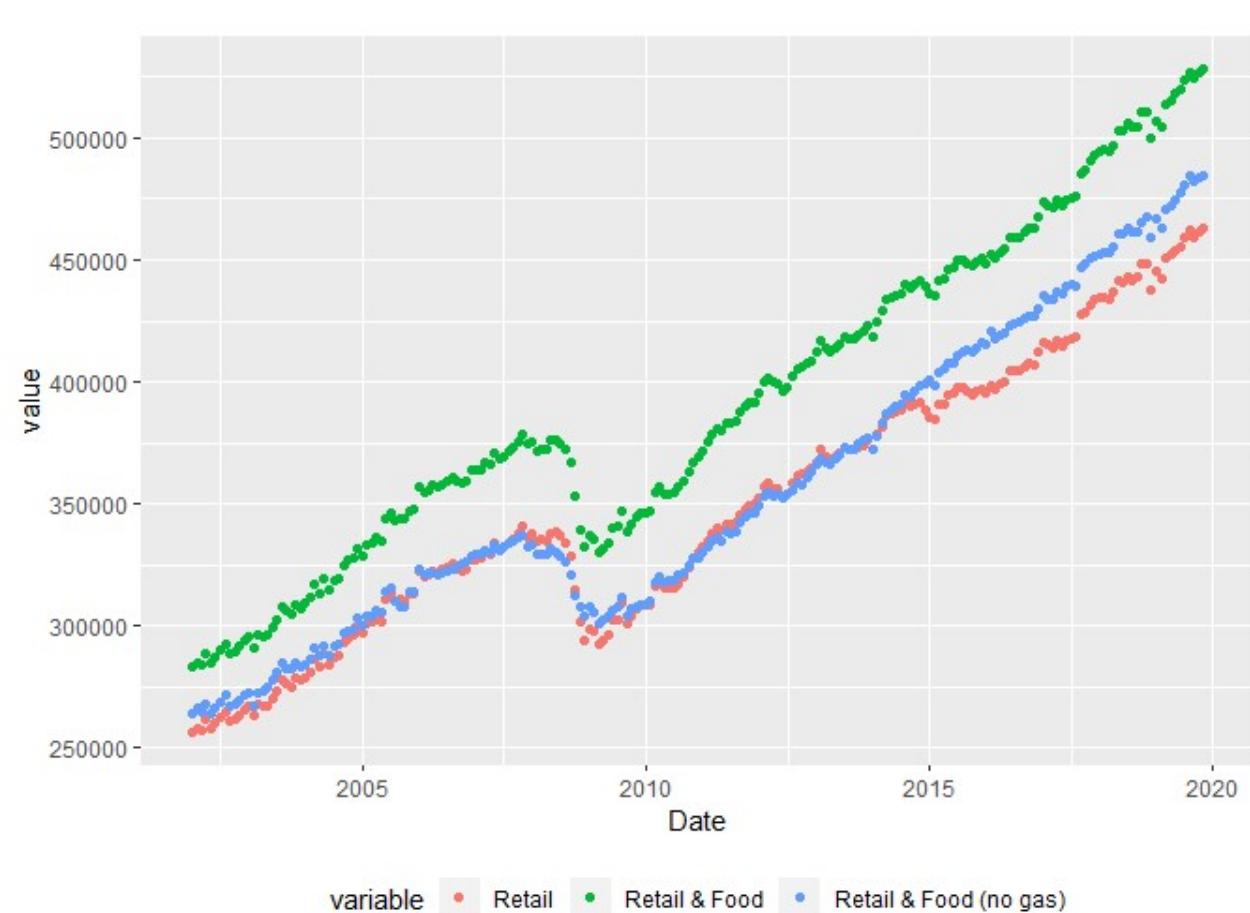
Scope) This analysis will mainly include features derived from structured data. 20 economic features will be included in the analysis and three different versions of National Retail Sales will be investigated. It is not advantageous to include unstructured data and it was not included. We will be looking for leading indicators when we evaluate features.

Retail Sales

3 OPTIONS

- › Retail Sales includes retail sales (including online sales), gasoline sales and auto sales. It does not include food & beverage sales.
- › Retail and Food includes retail sales (including online sales), gasoline sales and auto sales as well as food & beverage sales.
- › Retail and Food without gas includes retail sales (including online sales) and auto sales as well as food & beverage sales.) Gasoline sales have been stripped out.

20 YEAR TIME SERIES

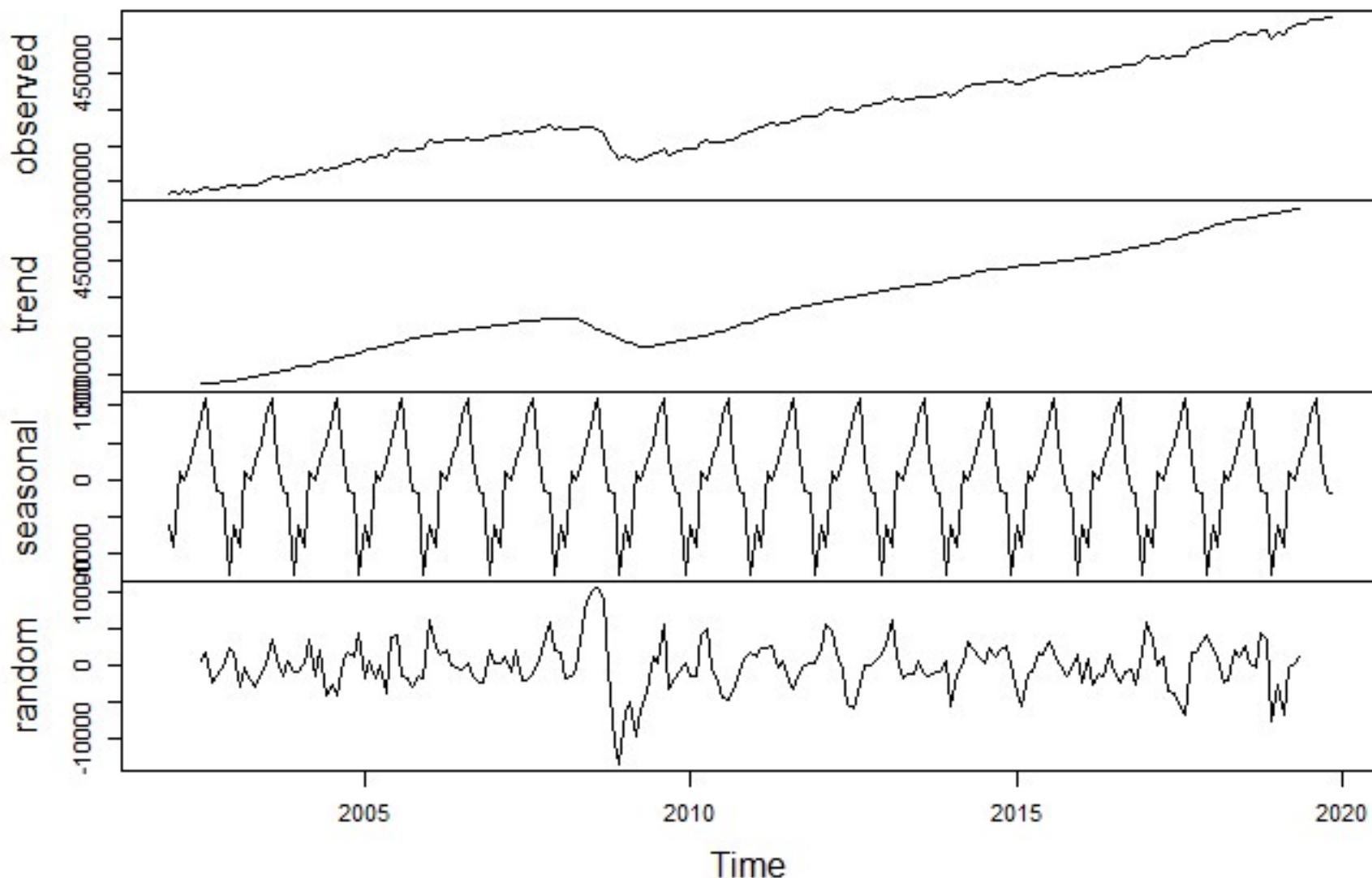


SOURCE
U.S. Census Bureau

UNITS
Data in Millions of Dollars

Decomposing Retail Sales

Decomposition of additive time series



Candidate Predictors

20 Macro Candidates

Overview

- › The St. Louis Federal Reserve provided many of the candidate predictors.
- › The data was organized and without blanks
- › 215 Observations were used
- › Only Debt Service Burden had a native frequency that wasn't monthly

Illustrative Predictors

- › Gasoline Sales
- › Non Farm Payrolls
- › Consumer Price Index
- › Price of Brent Crude Oil
- › Mortgage Service Burden

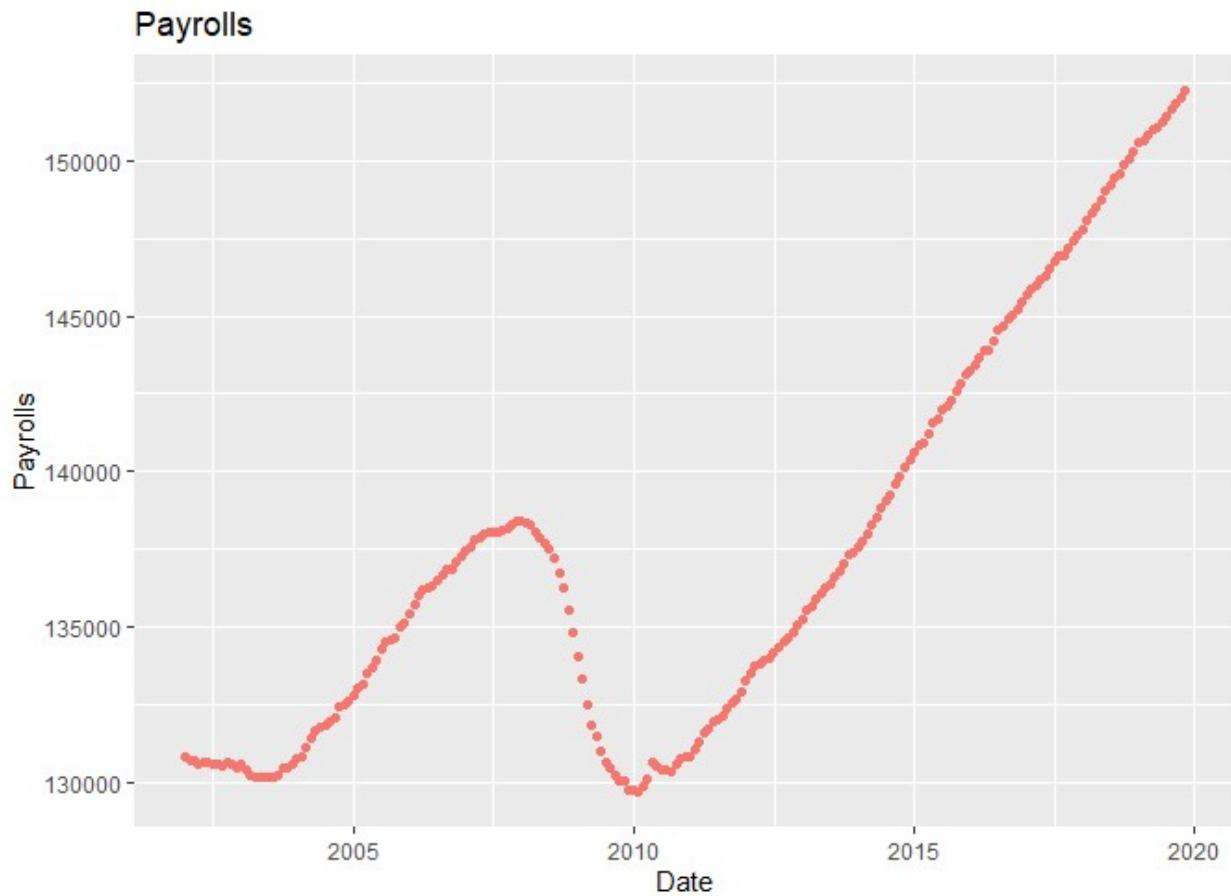
Brent

Sample Candidate EDA Plot

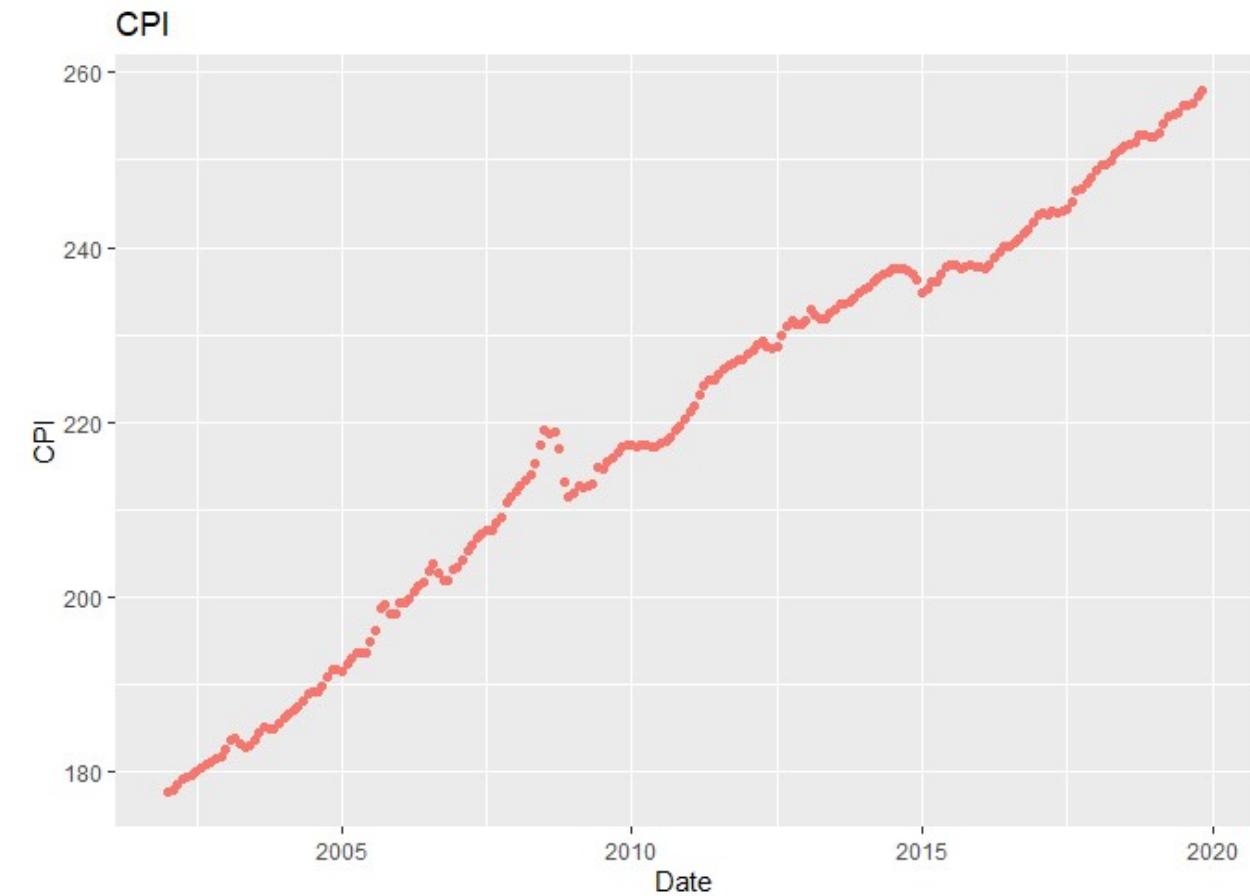


Candidate Predictors

Payrolls

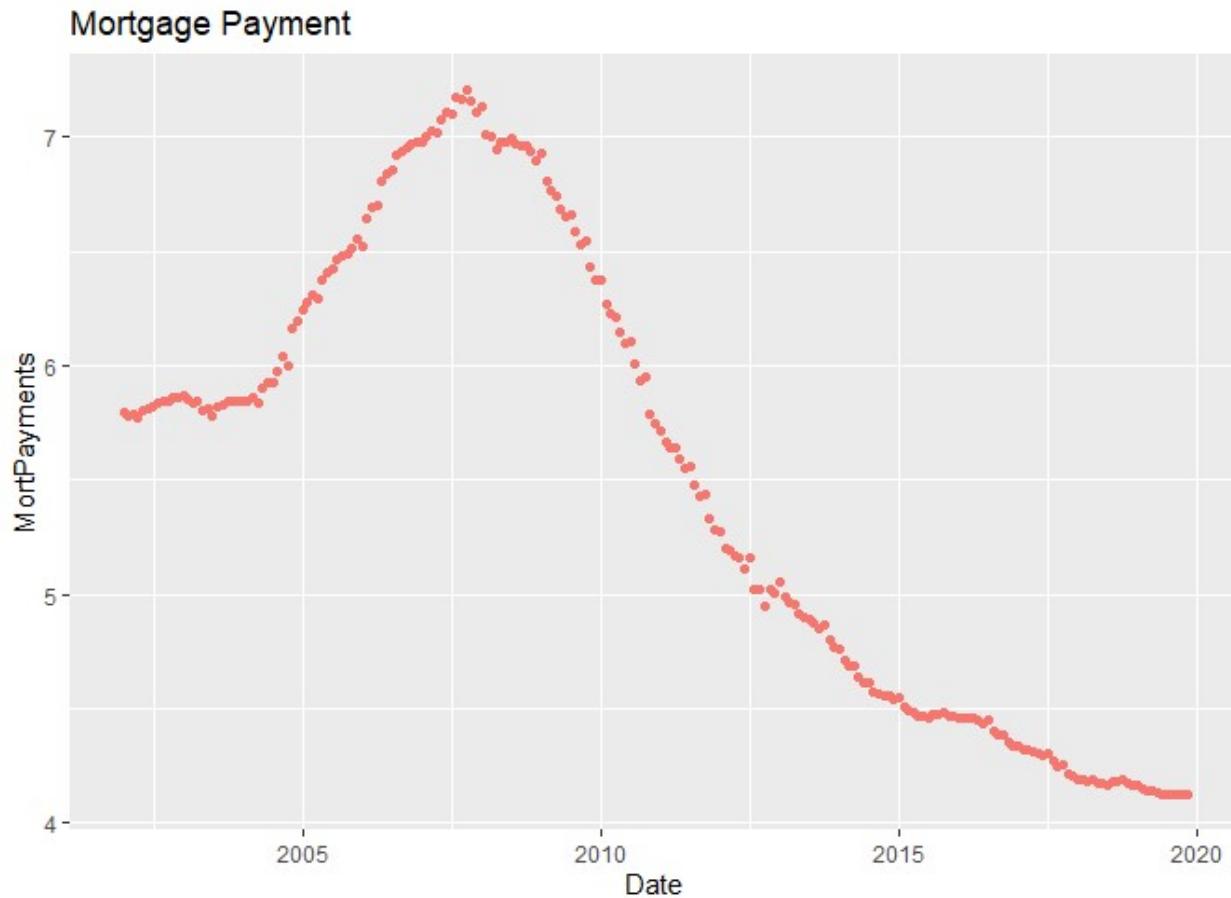


CPI

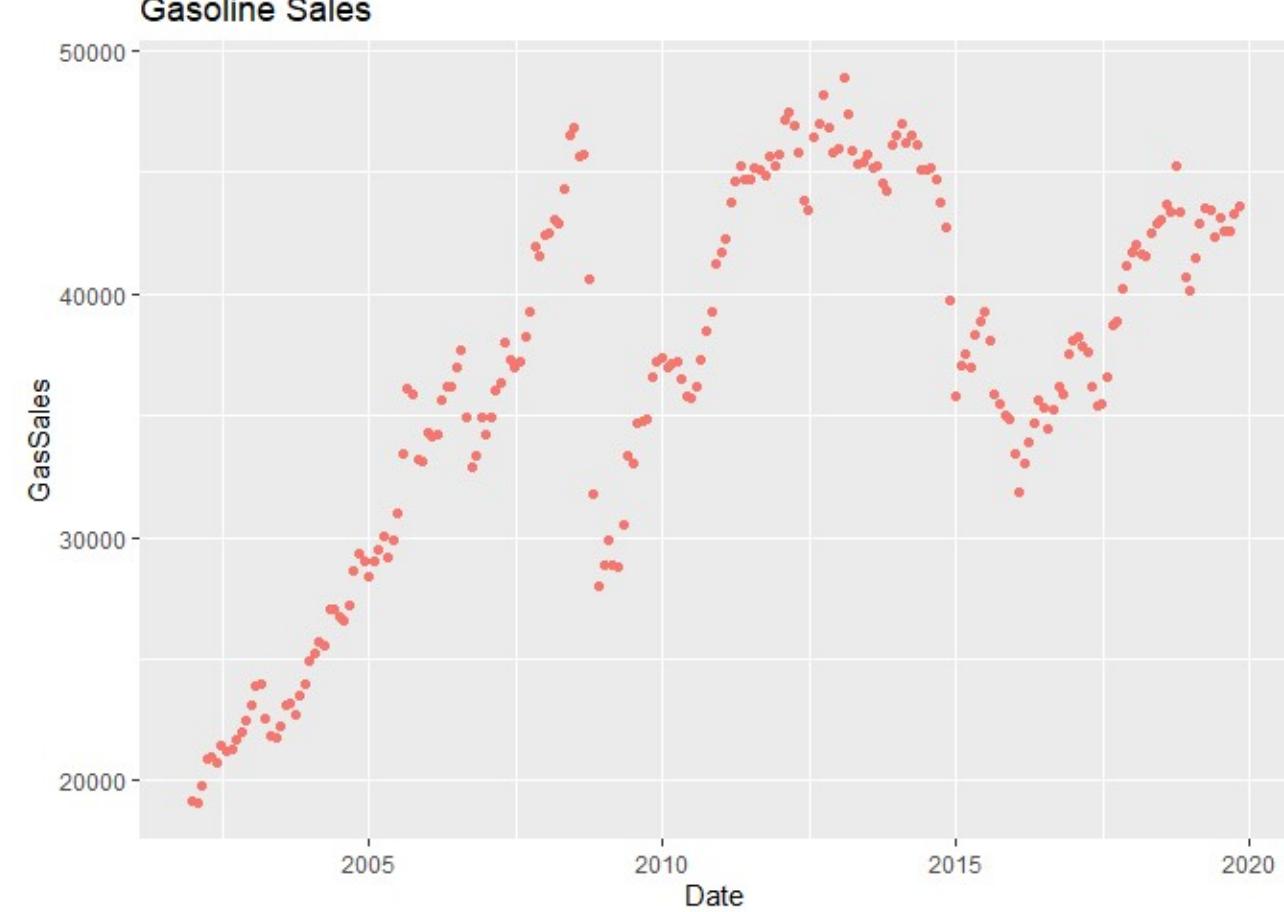


Candidate Predictors

Mortgage Payments



Gasoline Sales



3 Candidate Models

Payrolls, Brent, Mortgage Payments, Gas Sales and CPI

```
> Call:  
> lm(formula = Retail_and_Food ~ Payrolls + Brent +  
MortPayments + GasSales + CPI, data = df)  
  
> Residuals:  
> Min 1Q Median 3Q Max  
> -24467.3 -2195.8 544.2 3065.1 14151.0  
  
> Coefficients:  
> Estimate Std. Error t value Pr(>|t|)  
> (Intercept) -4.377e+05 1.641e+04 -26.679 < 2e-16 ***  
> Payrolls 4.393e+00 1.363e-01 32.220 < 2e-16 ***  
> Brent -2.597e+02 6.250e+01 -4.155 4.78e-05 ***  
> MortPayments -1.070e+04 6.265e+02 -17.077 < 2e-16 ***  
> GasSales 2.286e+00 2.959e-01 7.725 4.79e-13 ***  
> CPI 9.988e+02 5.746e+01 17.384 < 2e-16 ***  
> ---  
> Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '' 1  
  
> Residual standard error: 5134 on 206 degrees of freedom  
> (3 observations deleted due to missingness)  
> Multiple R-squared: 0.9939, Adjusted R-squared: 0.9937  
> F-statistic: 6670 on 5 and 206 DF, p-value: < 2.2e-16
```

Payrolls, Brent, Mortgage Payments and CPI

```
> Call:  
> lm(formula = Retail_and_Food ~ Payrolls + Brent +  
MortPayments + CPI, data = df)  
  
> Residuals:  
> Min 1Q Median 3Q Max  
> -18872.6 -2188.5 983.2 3847.8 12595.8  
  
> Coefficients:  
> Estimate Std. Error t value Pr(>|t|)  
> (Intercept) -5.222e+05 1.386e+04 -37.662 < 2e-16 ***  
> Payrolls 4.937e+00 1.322e-01 37.341 < 2e-16 ***  
> Brent 1.984e+02 2.238e+01 8.866 3.55e-16 ***  
> MortPayments -9.647e+03 6.928e+02 -13.925 < 2e-16 ***  
> CPI 1.256e+03 5.309e+01 23.651 < 2e-16 ***  
> ---  
> Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '' 1  
  
> Residual standard error: 5817 on 207 degrees of freedom  
> (3 observations deleted due to missingness)  
> Multiple R-squared: 0.9921, Adjusted R-squared: 0.9919  
> F-statistic: 6485 on 4 and 207 DF, p-value: < 2.2e-16
```

Payrolls, Brent, Mortgage Payments and CPI (Food and Bev.)

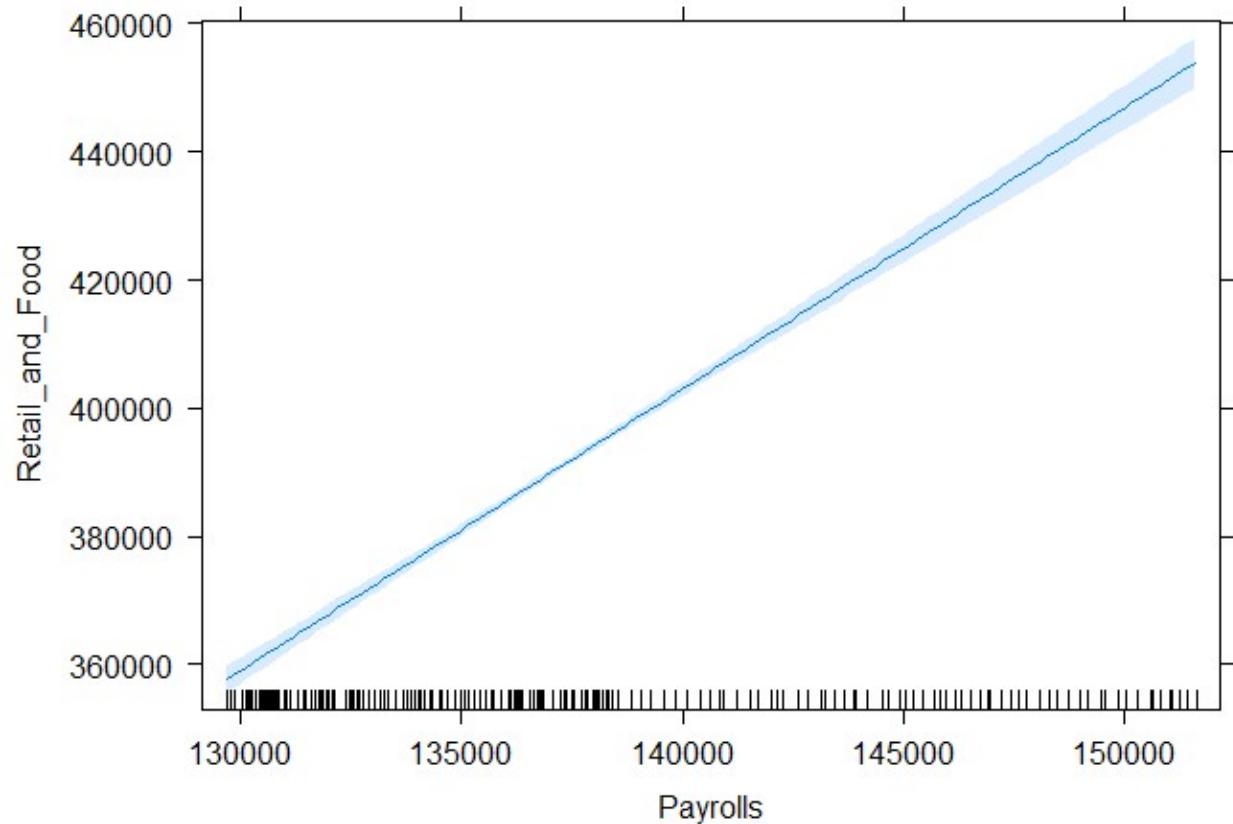
```
> lm(formula = Retail_and_Food ~ Payrolls + Brent +  
MortPayments + CPIfoodbev, data = df)  
  
> Residuals:  
> Min 1Q Median 3Q Max  
> -21965.1 -2355.4 695.8 4366.9 11268.1  
  
> Coefficients:  
> Estimate Std. Error t value Pr(>|t|)  
> (Intercept) -5.542e+05 1.554e+04 -35.66 < 2e-16 ***  
> Payrolls 5.720e+00 1.250e-01 45.77 < 2e-16 ***  
> Brent 3.039e+02 2.198e+01 13.83 < 2e-16 ***  
> MortPayments -9.742e+03 7.921e+02 -12.30 < 2e-16 ***  
> CPIfoodbev 8.755e+02 4.378e+01 20.00 < 2e-16 ***  
> ---  
> Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '' 1  
  
> Residual standard error: 6536 on 207 degrees of freedom  
> (3 observations deleted due to missingness)  
> Multiple R-squared: 0.99, Adjusted R-squared: 0.9898  
> F-statistic: 5125 on 4 and 207 DF, p-value: < 2.2e-16
```

The first option will be selected due to its lower standard error and its R^2 of 99.37%.

Payrolls Influence

- › Payrolls has a large influence on the Retail Sales. The values for Payrolls result in \$360Bn on the low end and just under \$460MM on the high end. Also the T-Value is 32.2 which tells us that Payrolls' impact is large in comparison to its standard error.
- › Additionally the slope of the relationship makes sense. As more people have jobs, more retail good will be purchased.

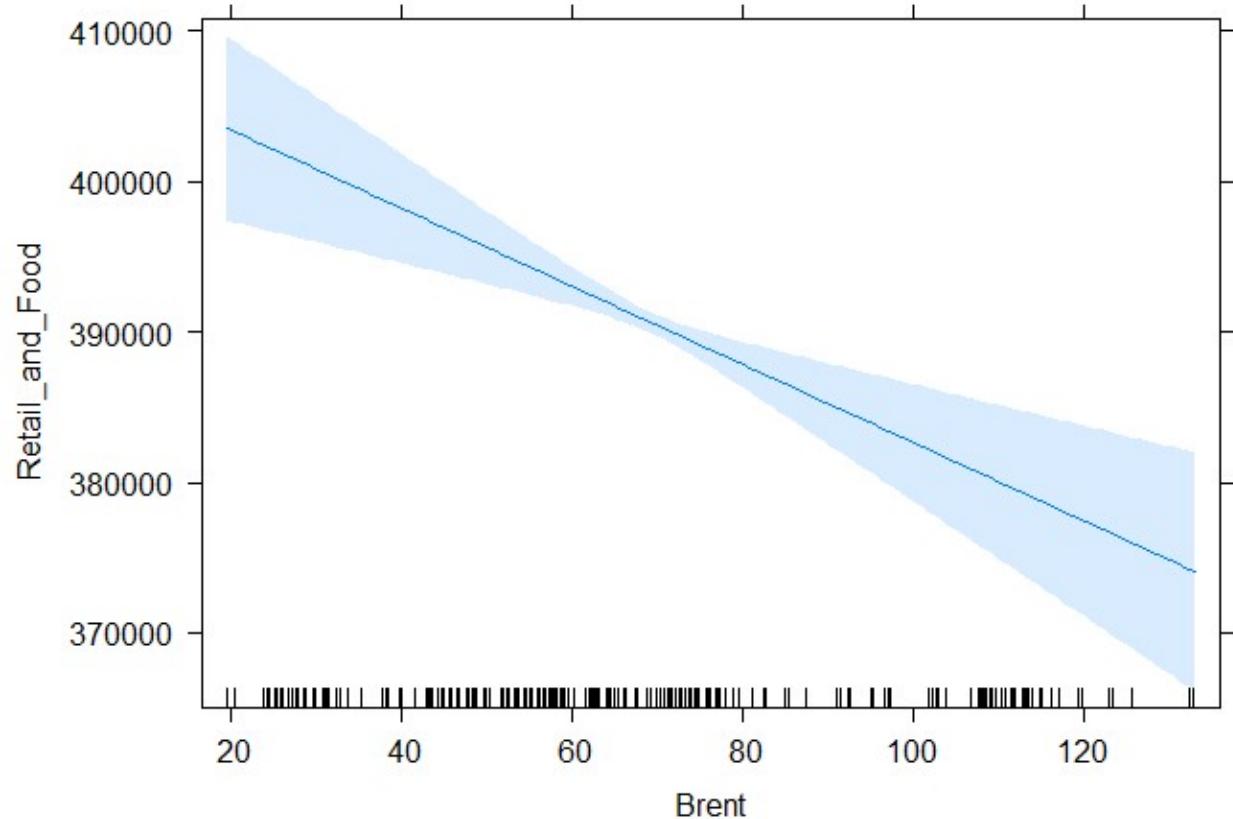
Payrolls effect plot



Brent Price's Influence

- › Brent's Price influences on the Retail Sales inversely. The Retail Sales values resulting from Brent range from \$405Bn on Brent's low end and just under \$380MM on Brent's high end. Also the T-Value is -4.2 which tells us that Brent's impact is large enough in comparison to its standard error to keep in the model.
- › Additionally the slope of the relationship makes sense. As oil prices rise , the economy tends to slow which pulls down retail sales.

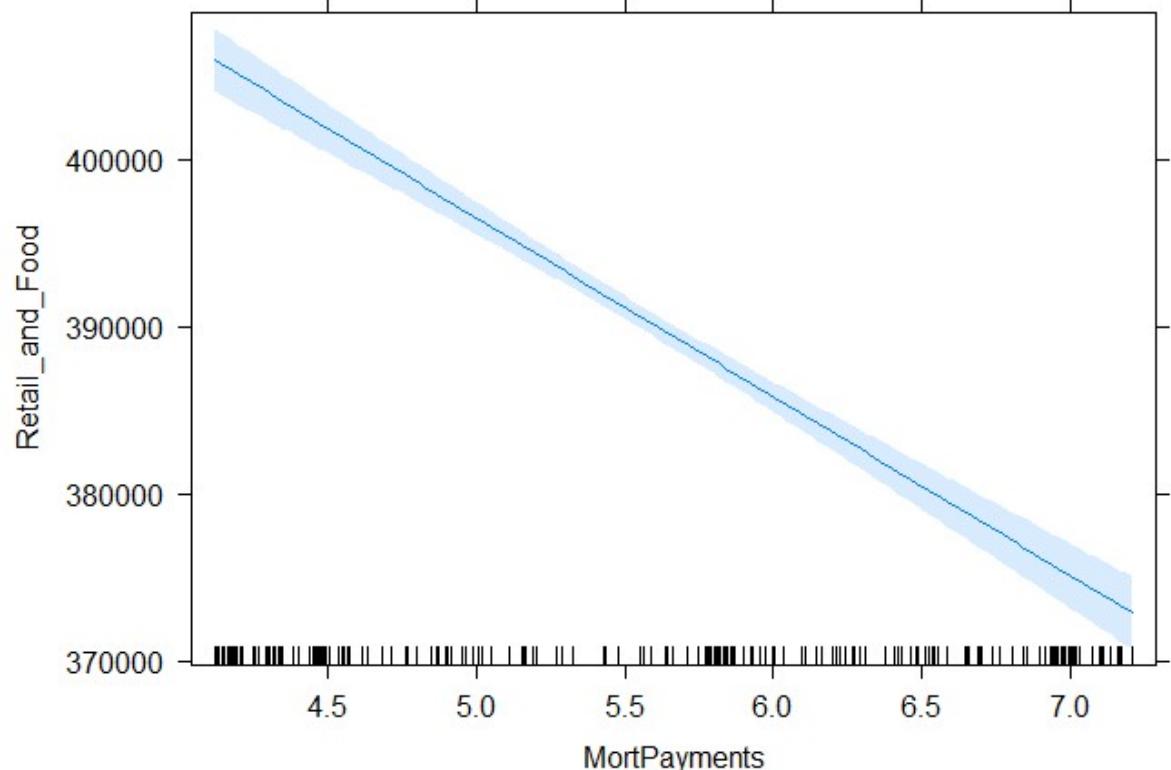
Brent effect plot



Homeowner Mortgage Payments' Influence

- › Mortgage Payments slightly influences Retail Sales inversely. The Retail Sales values resulting from Mortgage Payments range from \$405Bn on Mortgage Payments' low end and just under \$375MM on Mortgage Payments' high end. Also the T-Value is -17.1 which tells us that Mortgage Payments' impact is large enough in comparison to its standard error to keep in the model.
- › Additionally the slope of the relationship makes sense. As consumers spend more on their mortgages, retail sales are crowded out.

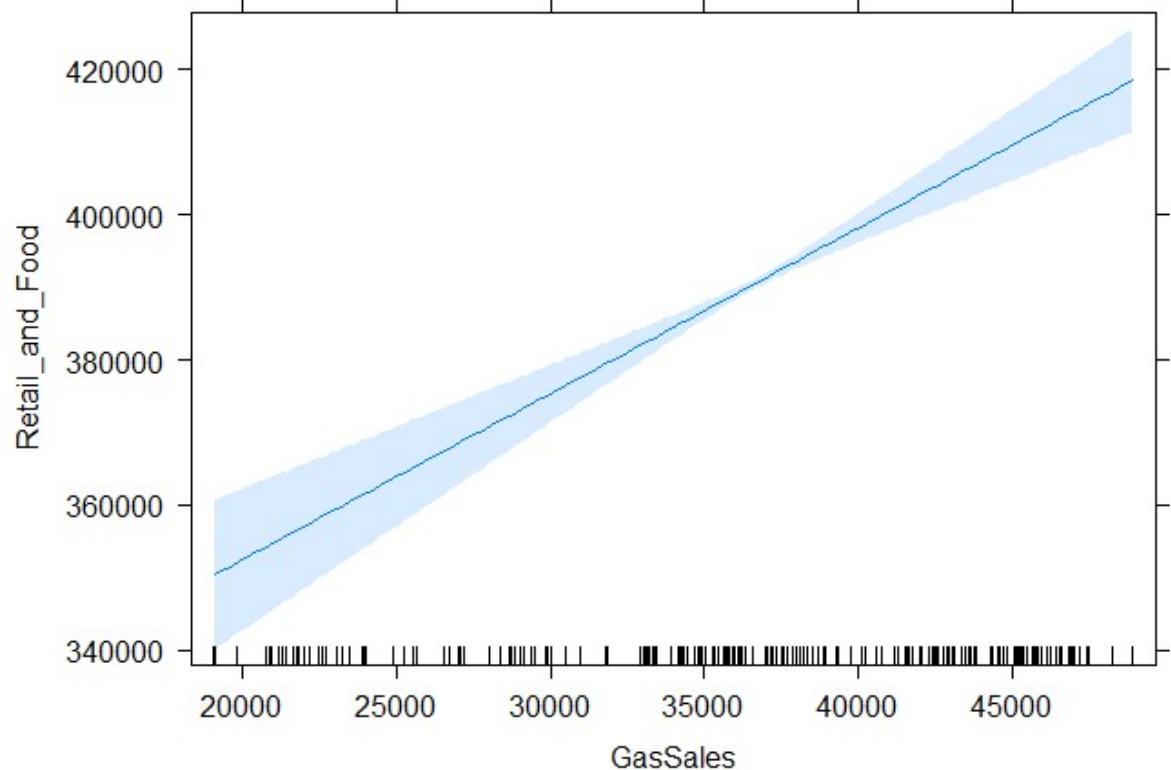
MortPayments effect plot



Gasoline Sales' Influence

- › Gasoline Sales will require more investigation before this project is submitted.
- › The statistics work, but I need to think through the real world implications. 😞

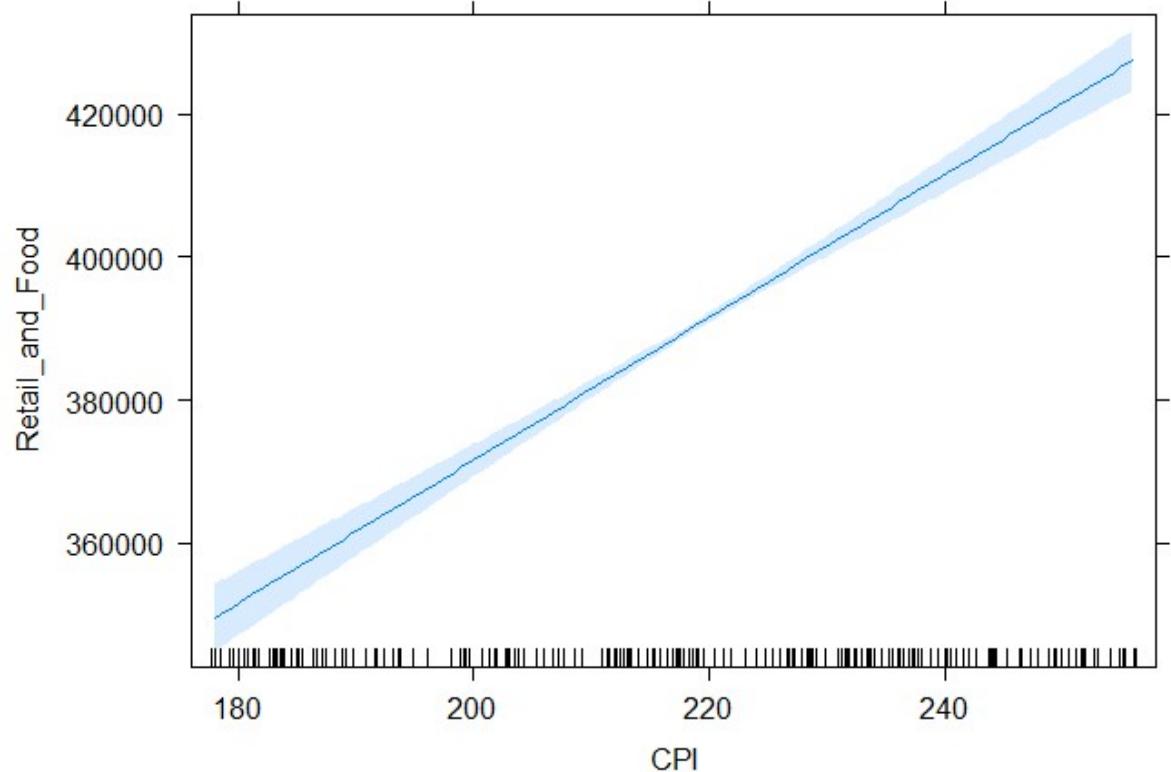
GasSales effect plot



CPI's Influence

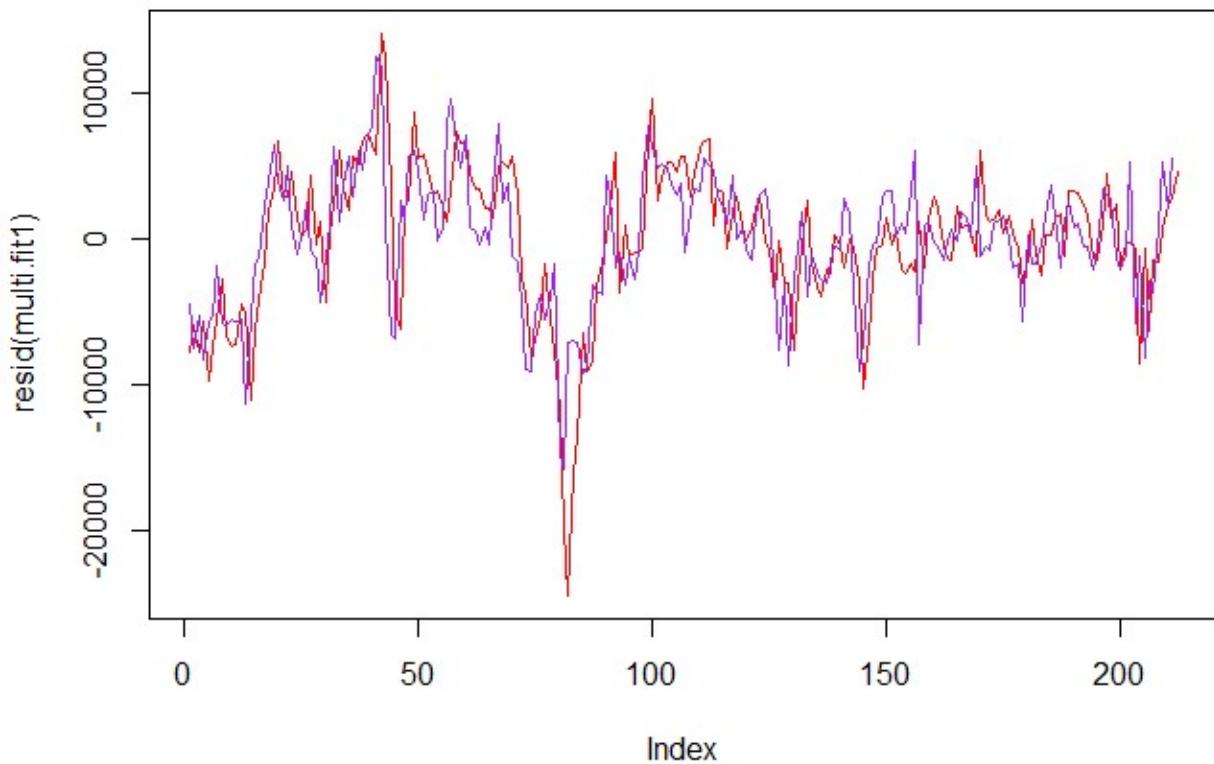
- › CPI influences Retail Sales positively. The Retail Sales values resulting from Mortgage Payments range from \$370Bn on CPI's low end and just over \$420MM on CPI's high end. Also the T-Value is 17.4 which tells us that CPI's impact is large enough in comparison to its standard error to keep in the model.
- › Additionally the slope of the relationship makes sense. As goods cost more, retail sales increase.

CPI effect plot

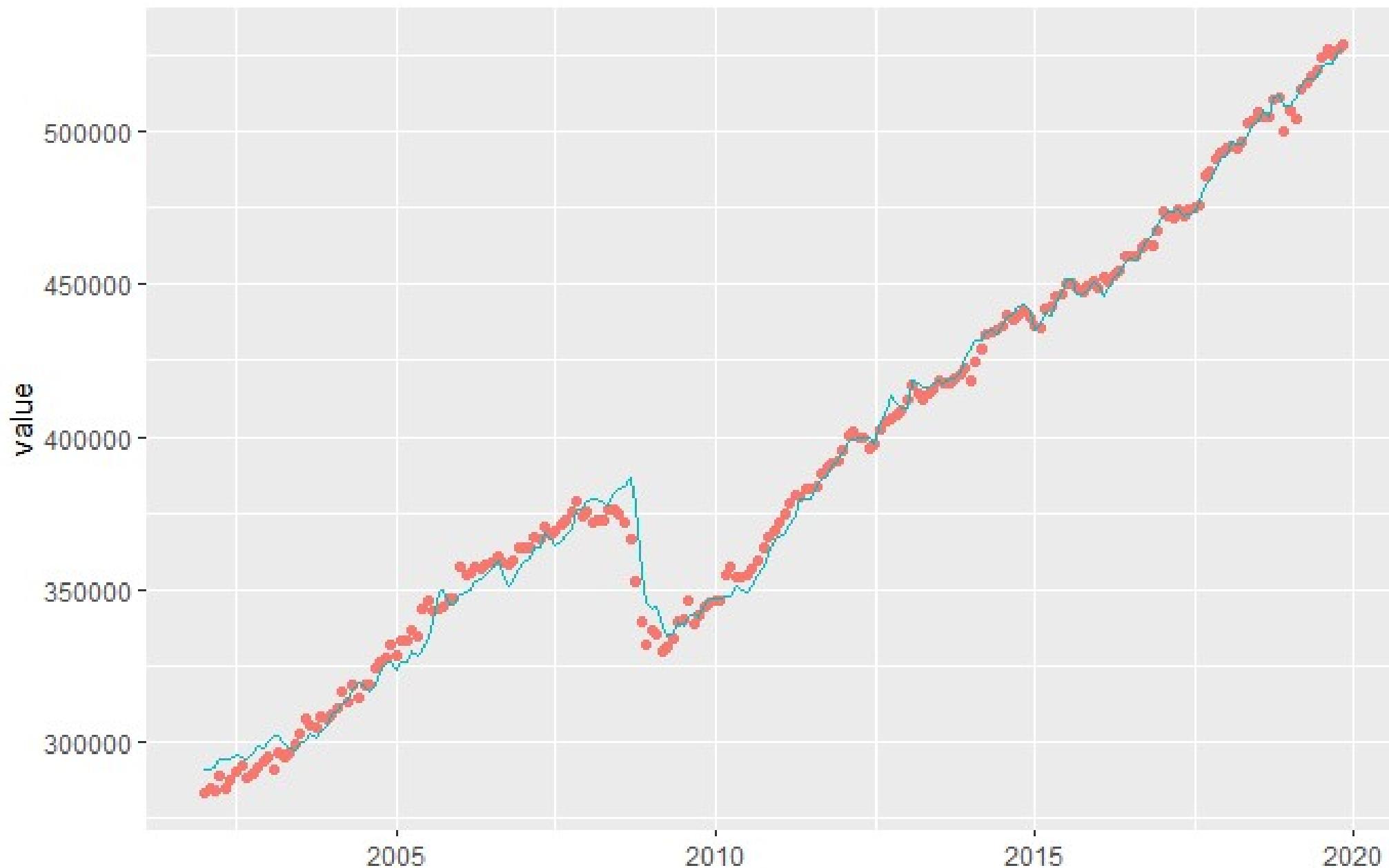


Lagging Variables

- › I used slide to lag the Price of Brent.
- › It helped get a slightly better R^2 and also reduced some of my residuals. The plot to the right show residuals without lagging in red and residuals with lagging in blue. By lagging Brent I was able to reduce the residuals.

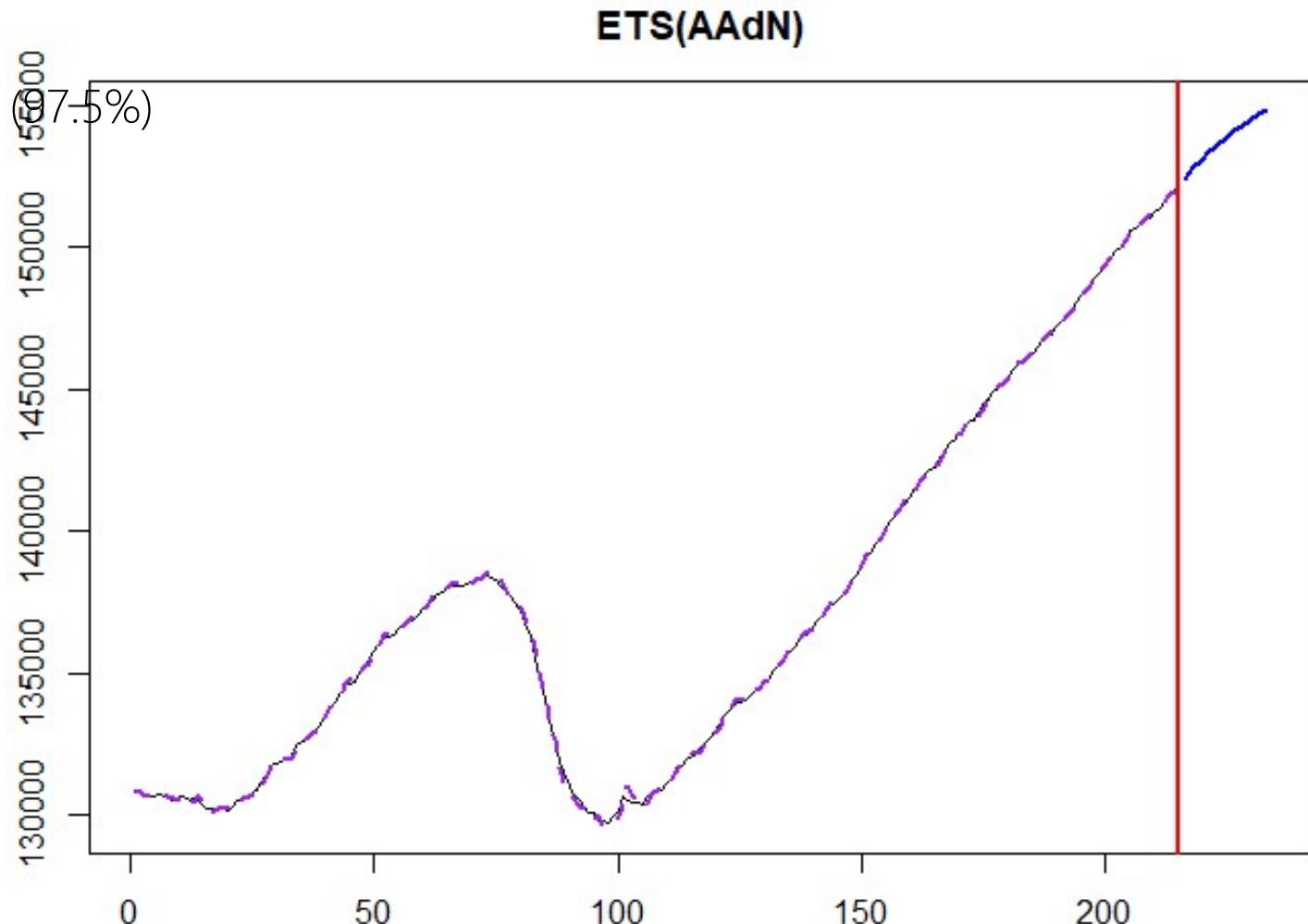


Fitted Values



Exponential Smoothing Payrolls

	Point forecast	Lower bound (2.5%)	Upper bound (97.5%)
216	152434.3	152204.3	152664.3
217	152628.0	152221.8	153034.1
218	152813.8	152209.1	153418.5
219	152992.3	152170.2	153814.3
220	153163.5	152108.5	154218.5
221	153327.8	152026.8	154628.9
222	153485.6	151927.3	155043.8
223	153637.0	151812.0	155461.9
224	153782.3	151682.5	155882.0
225	153921.7	151540.2	156303.2



Exponential Smoothing Brent Price

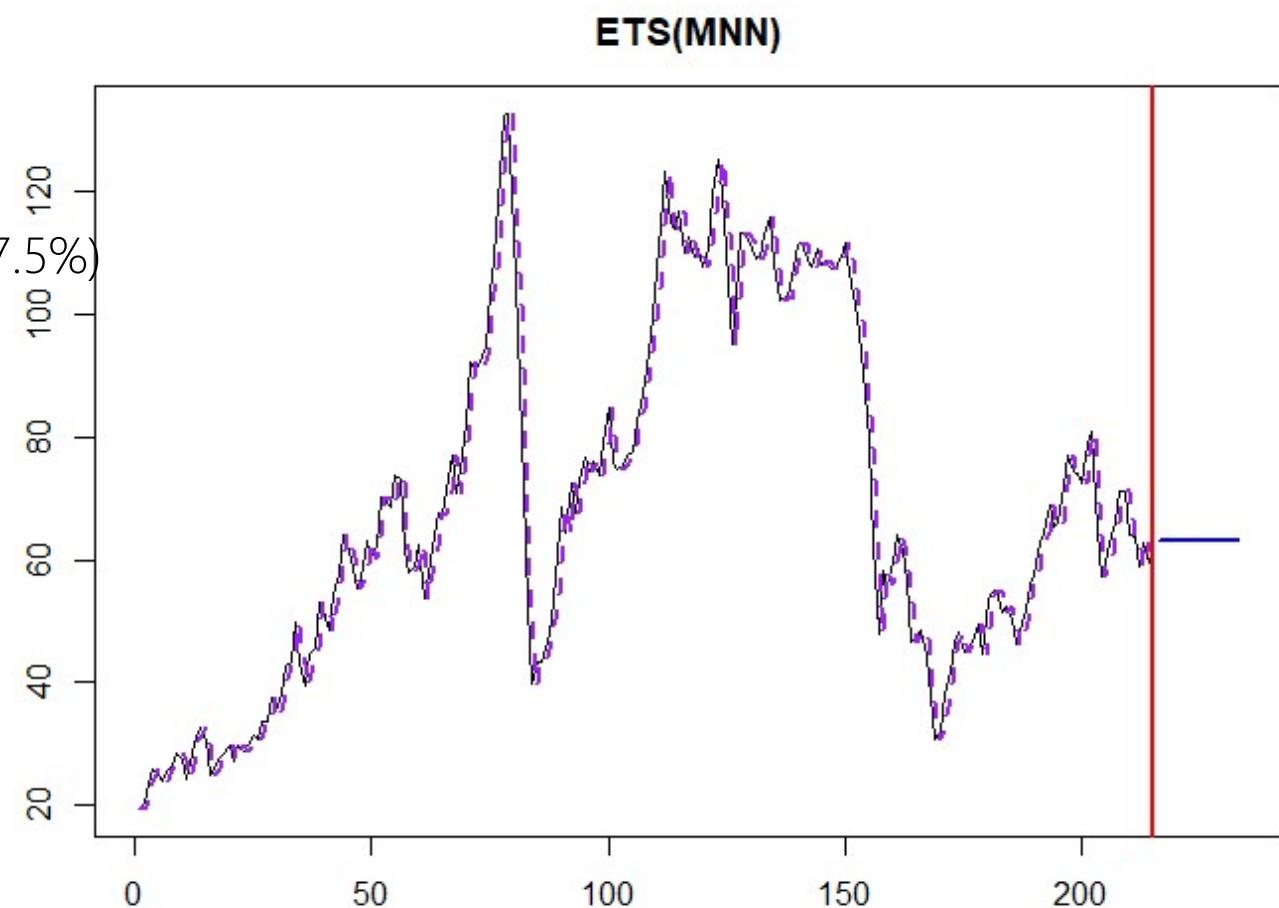
Time Series:

Start = 216

End = 225

Frequency = 1

	Point forecast	Lower bound (2.5%)	Upper bound (97.5%)
216	63.21	53.11813	75.21922
217	63.21	49.42558	80.83879
218	63.21	46.76720	85.43389
219	63.21	44.63748	89.51007
220	63.21	42.84167	93.26210
221	63.21	41.28040	96.78938
222	63.21	39.89493	100.15068
223	63.21	38.64718	103.38411
224	63.21	37.51083	106.51602
225	63.21	36.46680	109.56551



Exponential Smoothing Homeowner Mortgage Payments

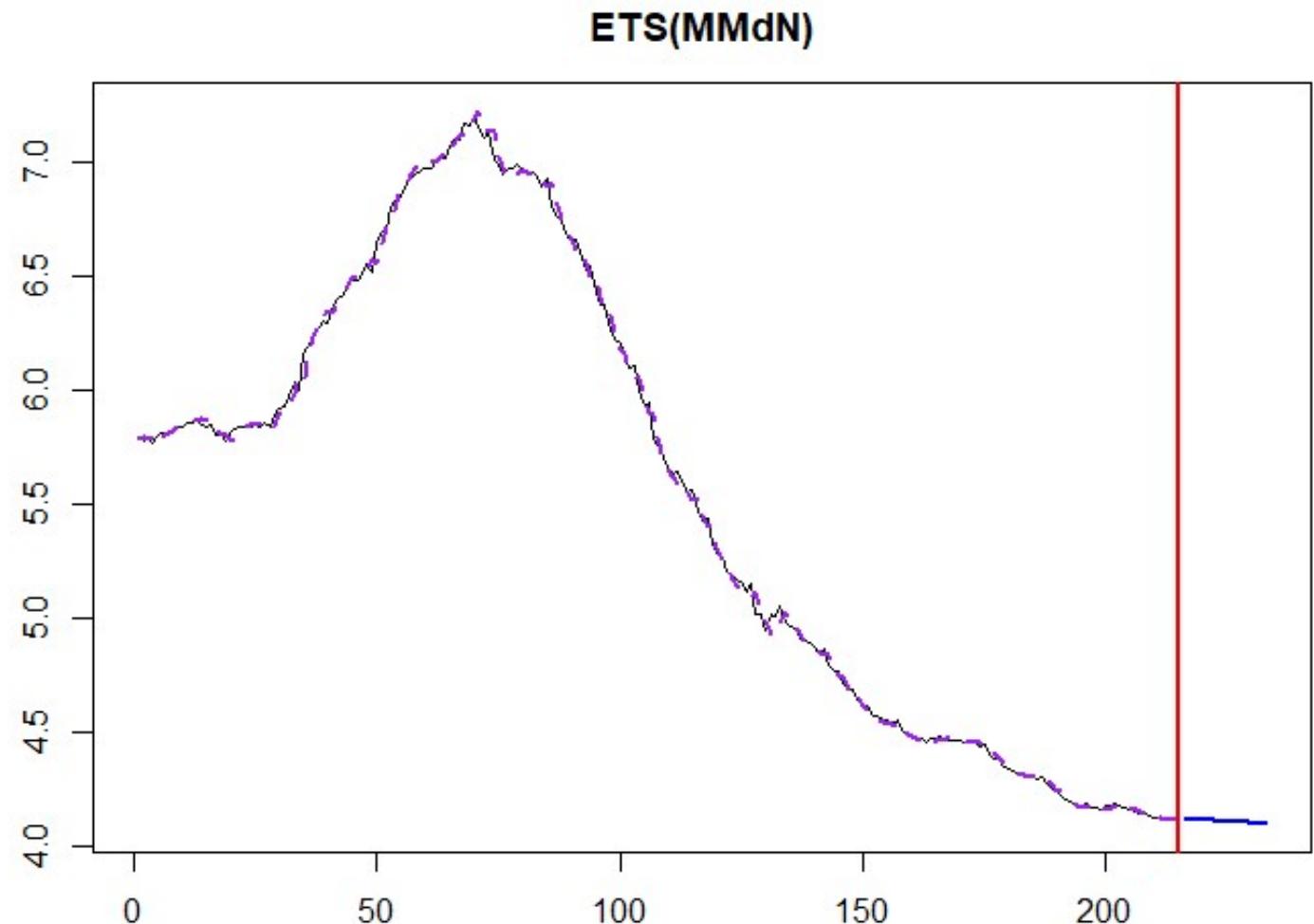
Time Series:

Start = 216

End = 225

Frequency = 1

	Point forecast	Lower bound (2.5%)	Upper bound (2.5%)
216	4.121307	4.076212	4.166902
217	4.119732	4.061027	4.179286
218	4.118195	4.044697	4.193028
219	4.116695	4.027435	4.20793
220	4.115231	4.009395	4.22386
221	4.113802	3.990695	4.24070
222	4.112408	3.971432	4.25838
223	4.111047	3.951685	4.276836
224	4.109719	3.931522	4.29599
225	4.108424	3.911002	4.315811



Exponential Smoothing Gasoline Sales

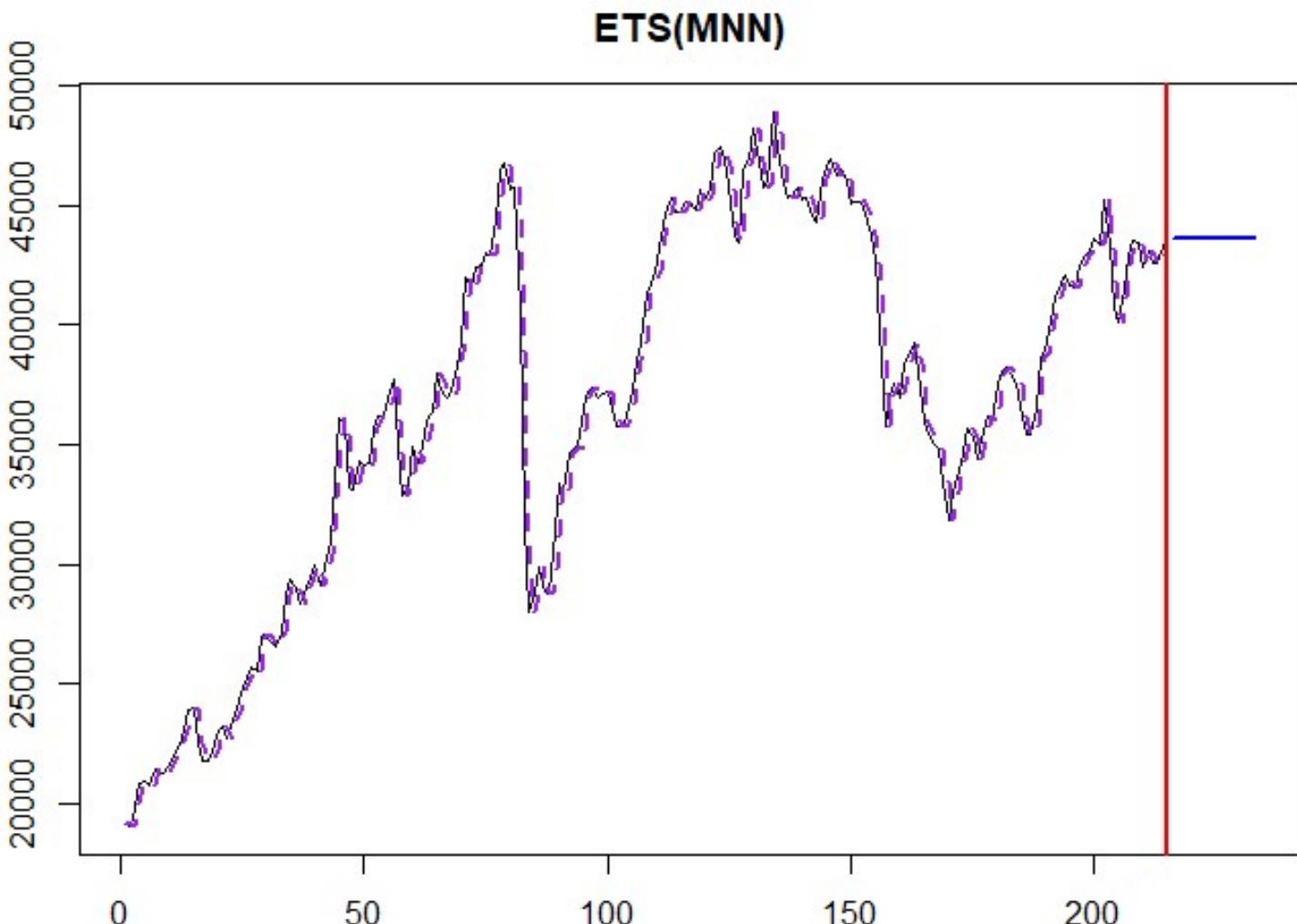
Time Series:

Start = 216

End = 225

Frequency = 1

	Point forecast	Lower bound (2.5%)	Upper bound (97.5%)
216	43598	40639.59	46771.77
217	43598	39473.78	48153.12
218	43598	38601.95	49240.67
219	43598	37881.93	50176.57
220	43598	37258.73	51015.85
221	43598	36704.13	51786.69
222	43598	36201.42	52505.83
223	43598	35739.70	53184.15
224	43598	35311.40	53829.23
225	43598	34911.03	54446.56



Exponential Smoothing CPI

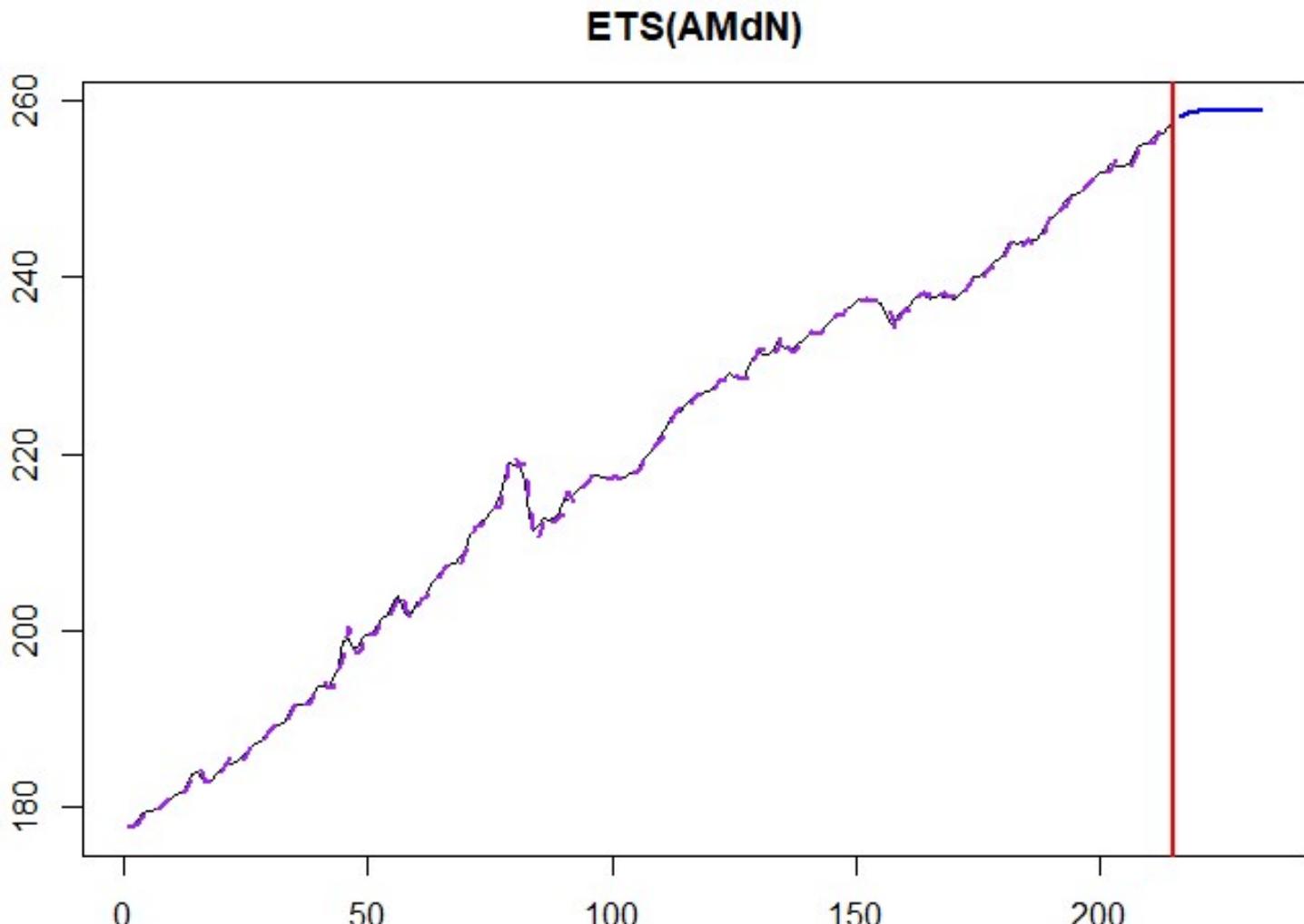
Time Series:

Start = 216

End = 225

Frequency = 1

	Point forecast	Lower bound (2.5%)	Upper bound (2.5%)
216	258.3282	257.1726	259.4814
217	258.5594	256.3750	260.7314
218	258.6956	255.5758	261.8238
219	258.7758	254.7984	262.7812
220	258.8230	254.0597	263.624
221	258.8508	253.3769	264.3601
222	258.8672	252.7556	265.046
223	258.8768	252.1554	265.6740
224	258.8825	251.6095	266.2492
225	258.8858	251.1223	266.7778



Plot Forecast



Variance

Month	Forecast	Actual	Variance (\$)	Variance (%)
Dec. 2019	529,404	528,367	-1,037	-0.19%
Jan. 2020	530,924	529,766	-1,158	-0.22%
Feb. 2020	532,443			
Mar. 2020	533,962			
Apr. 2020	535,481			
May 2020	537,000			
Jun. 2020	538,519			
Jul. 2020	540,039			
Aug. 2020	541,558			

Potential Improvements

1. Think though gasoline sales and maybe replace with gas prices.
2. Smarten up predictor forecasts.
3. Build scenarios based on flexing input variables.
4. Clean up forecast generation.