Miles S. Marimbire

Bachelor of Science in Economics University of Cincinnati

Last Modification: 4/27/2022

Project: Vehicle Forecasting

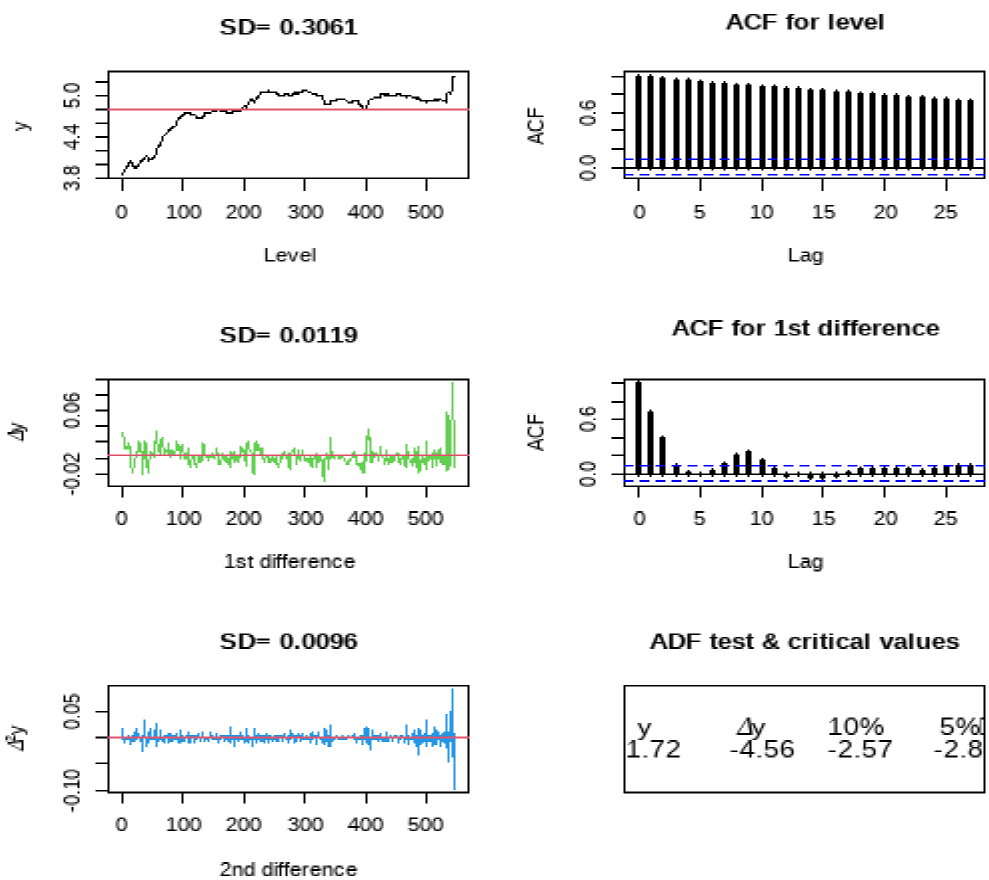
This project involved consumer price index data, vehicle total sales, and vehicle miles traveled from data ranging from the years 1976 to 2022. The data was collected from the website (https://fred.stlouisfed.org/). The kind of application I sought to implement was to forecast vehicle total sales based on data about vehicle miles traveled and consumer price index along with providing a combination of different visualizations.

After obtaining the data, I first created a sub-sample of the data for model building and estimation purposes. I dropped the last four observations of the time series. Below is a brief descriptive statistic of the differenced sub-sample data:

Consumer Price Index	Vehicle Miles Traveled	Vehicle Total Sales
Min. ~ -0.030466	Min. ~ -0.301672	Min. ~ -0.4365450
1st Qu. ~ -0.004200	1st Qu. ~ -0.053023	1st Qu. ~ -0.0223696
Median. ~ 0.001709	Median. ~ 0.003217	Median. ~ 0.0017871
Mean. ~ 0.002659	Mean. ~ 0.001821	Mean. ~ -0.0000175
3rd Qu. ~ 0.007537	3rd Qu. ~ 0.033584	3rd Qu. ~ 0.0300546
Max. ~ 0.096339	Max. ~ 0.276508	Max. ~ 0.3278745
SD. ~ 0.01192257	SD. ~ 0.06952254	SD. ~ 0.06894714
Observations. ~ 548	Observations. ~ 548	Observations. ~ 548

The next step in my analysis was to determine the order of integration of each variable, the first variable I looked at was the consumer price index, the visual inspection suggests an order of integration 1, and the standard deviation also suggests an order of integration 1 and the ADF test. Below are the results in Figure 1.1:

Figure 1.1

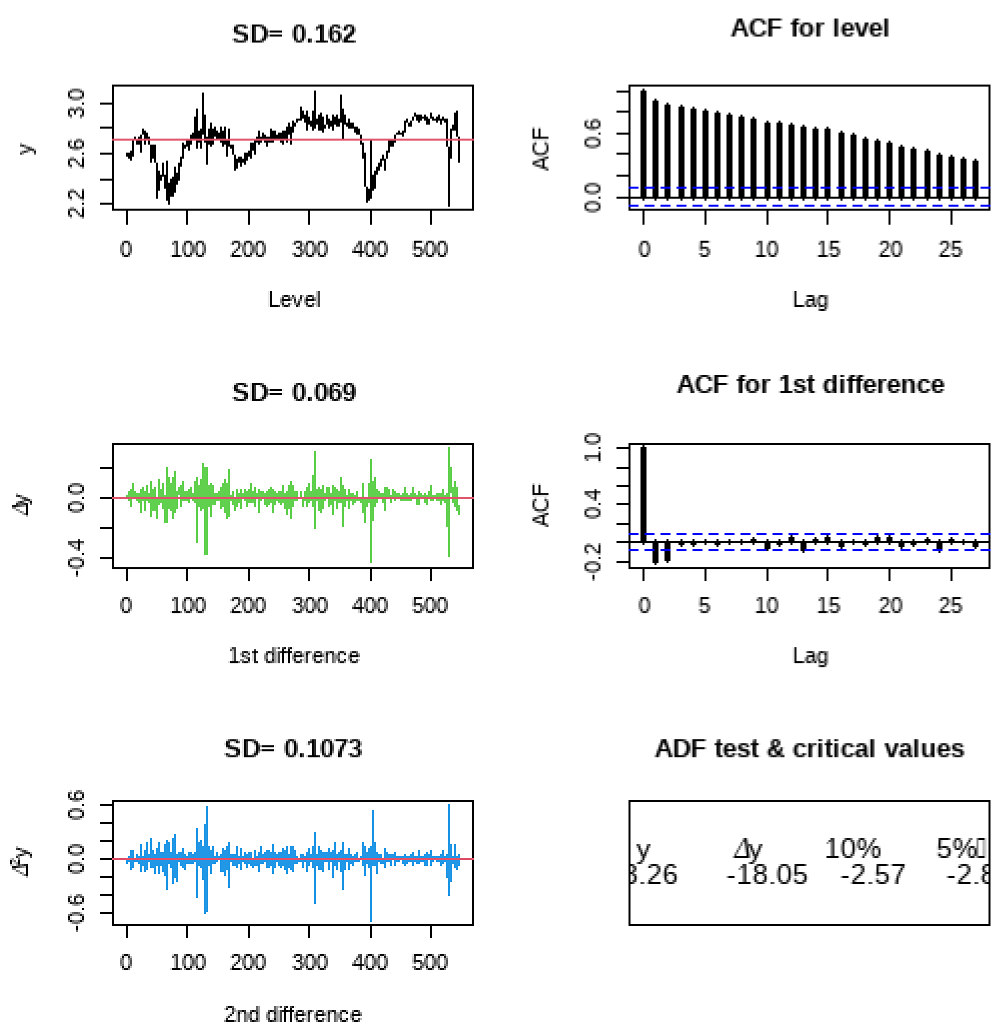



```
ADF t-value

     round1 round2

-1.72  -4.56
```

```
-2.57 -2.87 -3.44
```

My second variable is the vehicle total sales, Visual inspection, and Standard Deviation suggest an order of integration 1 below are the results Figure 1.2:

Figure 1.2

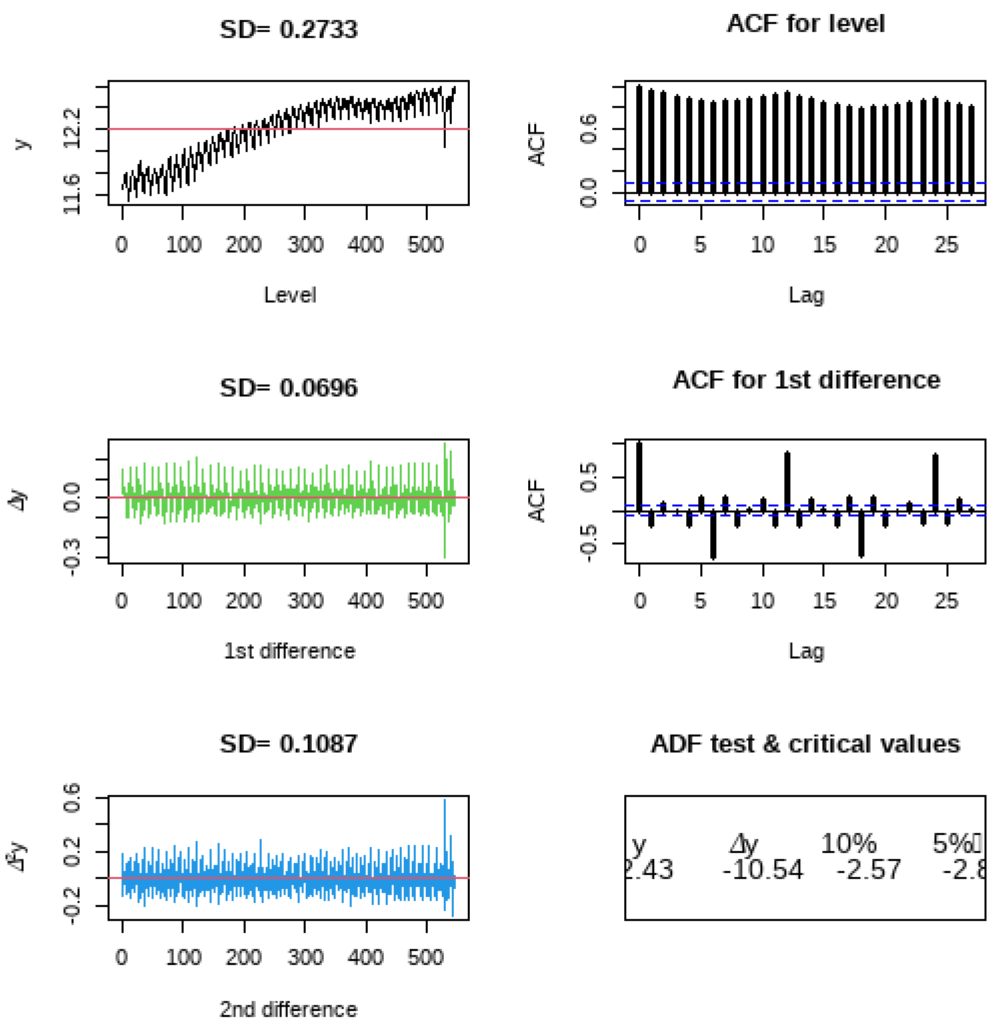



```
ADF t-value

     round1 round2

-3.26 -18.05
```

```
-2.57 -2.87 -3.44
```

My final variable was the miles traveled variable, all evidence suggests an order of integration 1; results are found below in Figure 1.3:

Figure 1.3




```
ADF t-value
     round1 round2
-2.43 -10.54
```

```
-2.57 -2.87 -3.44
```

The variables appear to be cointegrated, but it appears that around the 2008 financial crisis the variables dropped and spike back up again and drop again around the time of the pandemic. I want to see if either one of these influences the cointegration of the variables. You can see the visuals below; they are circled. Vehicle total sales are Figure 2.1, Consumer Price index in Figure 2.2 and Miles traveled in Figure 2.3

Figure 2.1

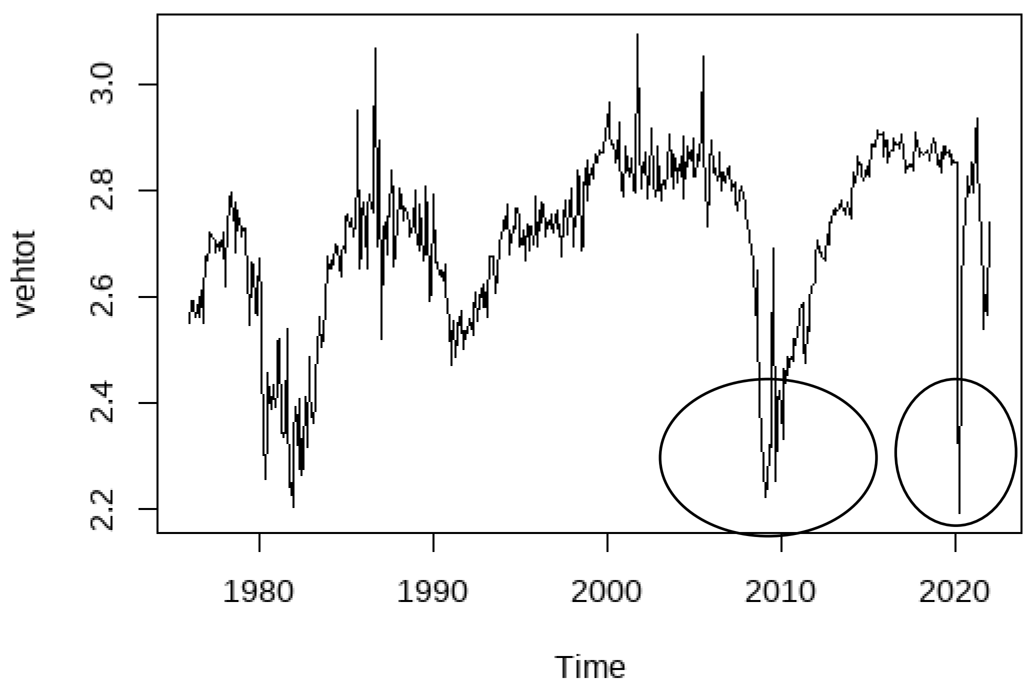
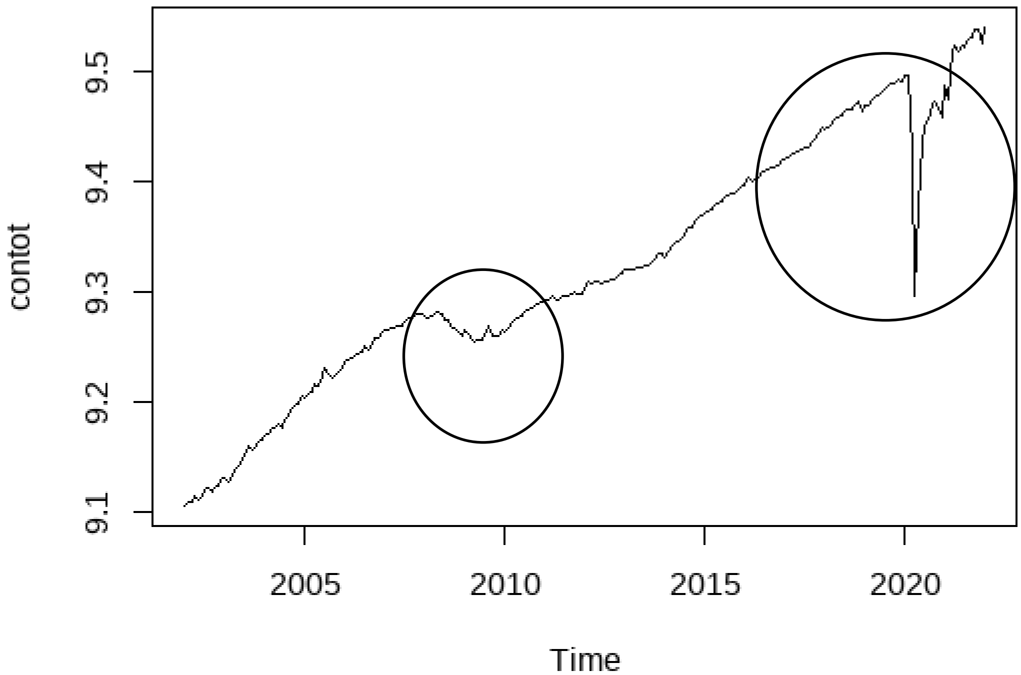


Figure 2.2



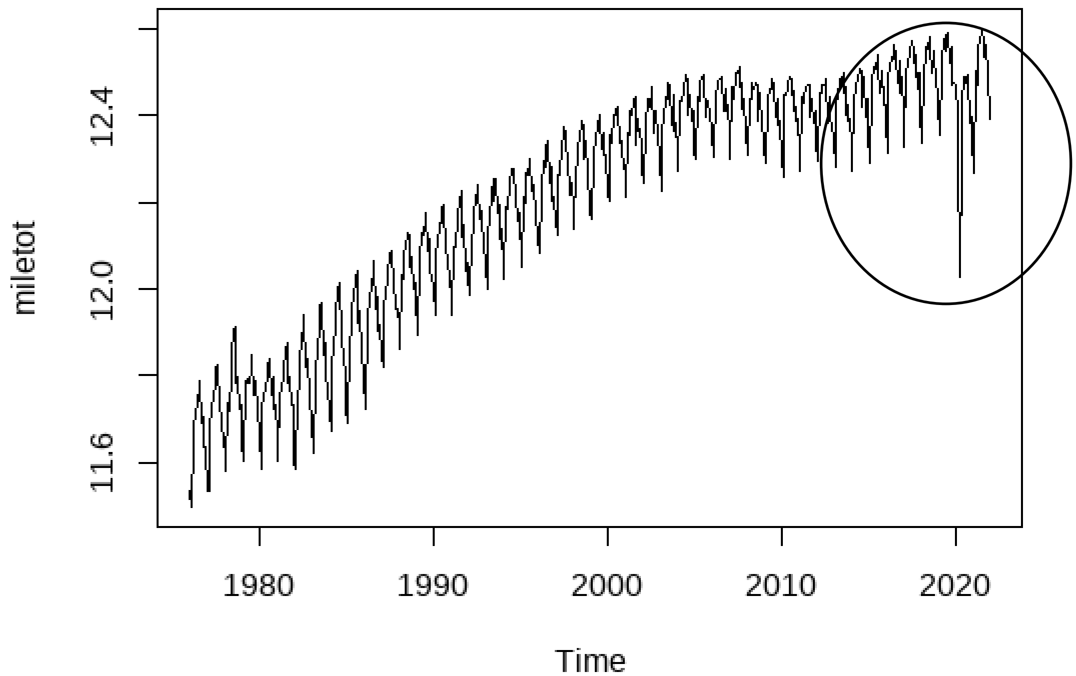


Fig 2.3

I used a combination of several regression models to test for cointegration. The first being before the pandemic, the second being before the 2008 financial crisis, and the third was all time from 1/1/1976 to 1/1/2022. Using the Engle-Granger cointegration method the results are as follows; Before the pandemic, there is no evidence of cointegration and before the 2008 financial crisis there is no evidence of cointegration. I concluded that neither one of these affect the cointegration of the variables. The first graph listed below is for the pandemic and the second one is a crisis; you can find the graphs below in Figures 3.1 and 3.2.

Figure 3.1

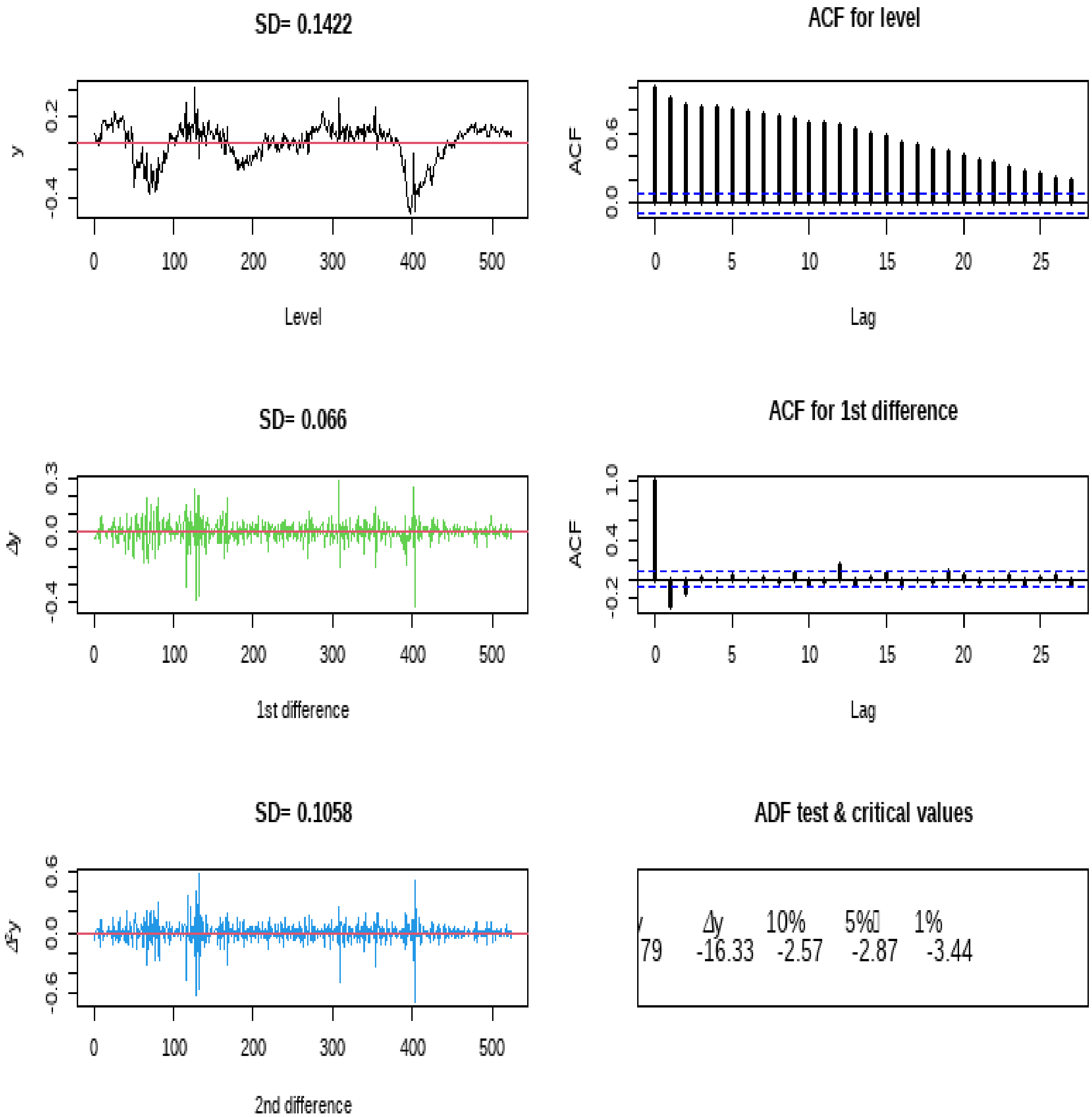
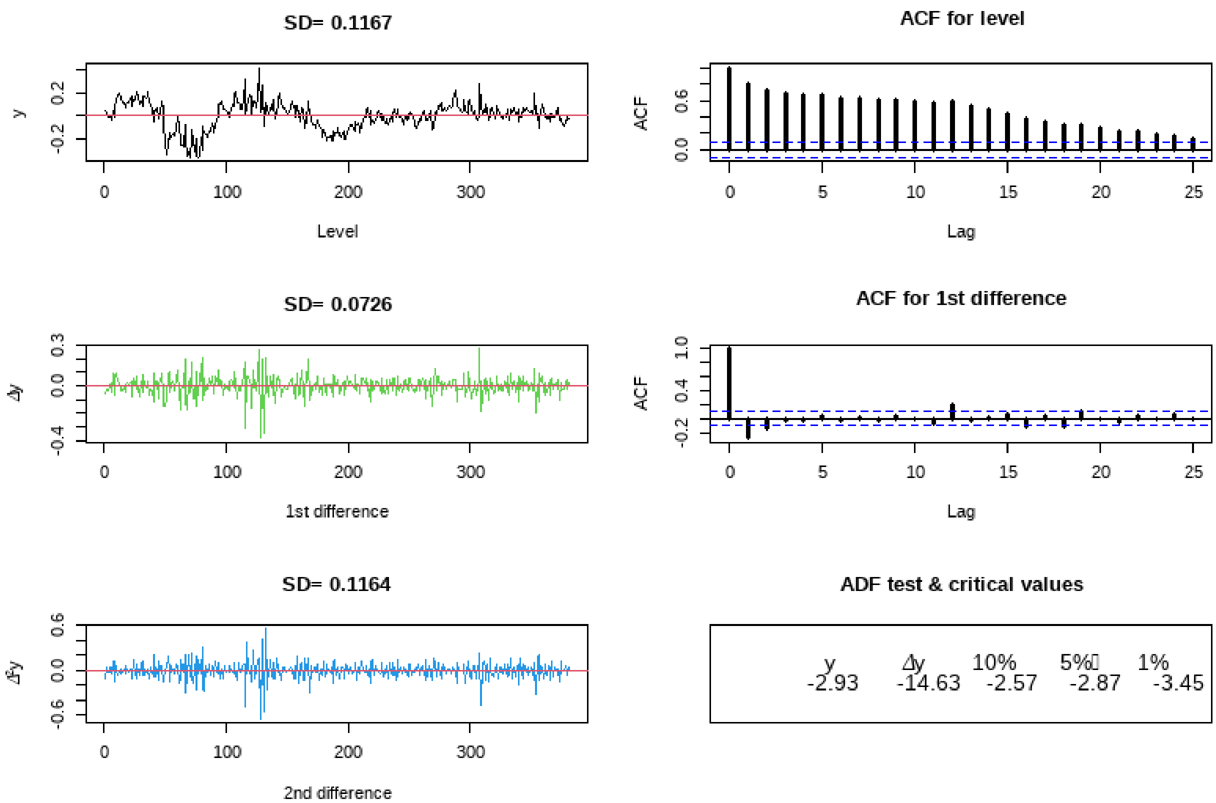

SD= 0.1422
ACF for level
y
ACF
Level
Lag
SD= 0.066
ACF for 1st difference
Δy
ACF
1st difference
Lag
SD= 0.1058
ADF test & critical values
2nd difference
Δy 10% 5% 1%
-16.33 -2.57 -2.87 -3.44

Figure 3.2


SD= 0.1167
y
Level
ACF for level
ACF
Lag
SD= 0.0726
Δy
1st difference
ACF for 1st difference
ACF
Lag
SD= 0.1164
2nd difference
ADF test & critical values
y Δy 10% 5% 1%
-2.93 -14.63 -2.57 -2.87 -3.45

I then tested for cointegration using a bunch of different combinations ranging from vehicle sales, miles traveled, and consumer price index. Another combination I used was miles traveled and consumer price index, etc. There was no evidence of cointegration of any of the variables. I then tested for cointegration using the Johansen approach and for the Max eigenvalue test the results show no evidence of cointegration past the 10% level, the results are below:

	test	10pct	5pct	1pct
r = 0 |	21.70	19.77	22.00	26.81

Strong evidence (@10%) of cointegration

Now the second test was the Trace eigen value test also shows there is no evidence of cointegration, the results are below:

	test	10pct	5pct	1pct
r = 0 |	40.73	32.00	34.91	41.07

Strong evidence (@10% and 5%) of cointegration

Other levels suggest that there is no level of cointegration

Now that there is no cointegration in our data, I built an autoregressive model with vehicle total sales as my dependent variable, and I built a dynamic regression model and tested for seasonality and serial correlation and found that there was no seasonality or serial correlation in the results are below:

```
Analysis of Variance Table

 Res. Df    RSS Df Sum of Sq      F Pr(>F)

1    511 2.2384

2    500 2.1915 11  0.046917 0.9731 0.4699
```

The first dynamic model consisted of 18 Lags and only six came back statistically significant, so I dropped the total Lags to four total lags and compared the AIC and BIC to see which of the two were the better models. The results were that the model preferred fewer lags thus leading me to believe that the fewer Lags is the better model. The results are below:

```
     Regression with more Lags
AIC
-1353.488
BIC
-1268.031
     Regression with fewer Lags
AIC
-1417.262
BIC
-1391.469
```

Since there is no cointegration evidence, I will build a VAR model. With our other variables to see if more Lags are appropriate to assist our model. The better model is the one with the fewer Lags as seen by the results below:

The model with 20 Lags

AIC

-7231.256

BIC

-6437.204

AIC

-1326.866

BIC

-1057.913

The model with 13 Lags

AIC

-7301.676

BIC

-6774.957

AIC

-1366.776

BIC

-1186.921

Now testing for serial correlation, I found there is no serial correlation in the model; results are found below:

```
      0.8284884 0.8395488 0.8599361 0.9433553 0.9788978
0.9811123 0.9558258
       0.9785543 0.9901634 0.9946020 0.9976281 0.9895667
```

When testing the Var model for Granger causality I noticed that there was evidence of Granger causality of consumer price index Lags. When testing for the Vehicle total sales I noticed there was evidence of Granger Causality as well. When performing the variance decomposition tests the results of each equation shows that each variance of the lags is explained by itself, results are found below:

Differenced Consumer Price Index Variance

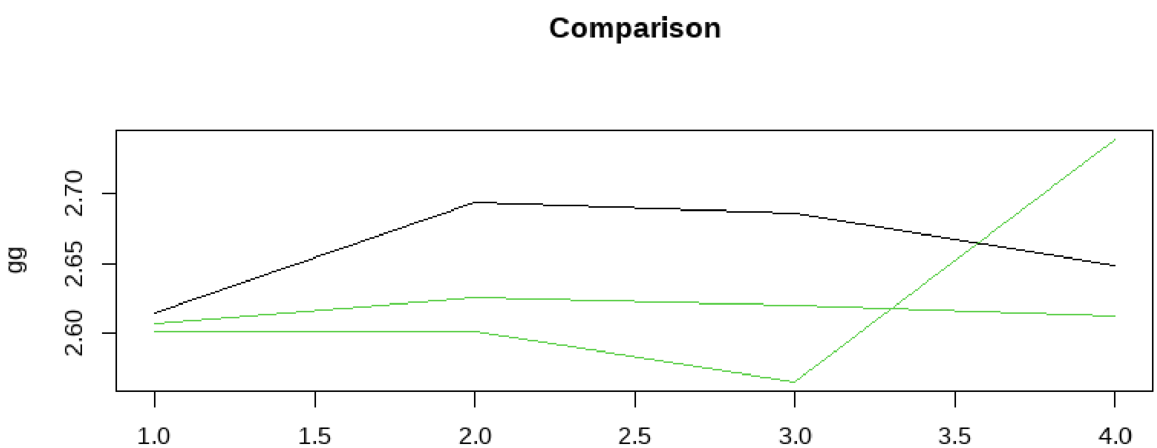
	dvehicle	dconsumer	dmiles
[1,]	0.01336316	0.9866368	0.000000000
[2,]	0.02015714	0.9732921	0.006550775
[3,]	0.02311439	0.9650889	0.011796755
[4,]	0.02479690	0.9546150	0.020588119
[5,]	0.02705929	0.9474951	0.025445647
[6,]	0.03139012	0.9131675	0.055442358
[7,]	0.03274497	0.9065027	0.060752343
[8,]	0.03318927	0.9016387	0.065172028
[9,]	0.03675627	0.8885210	0.074722720

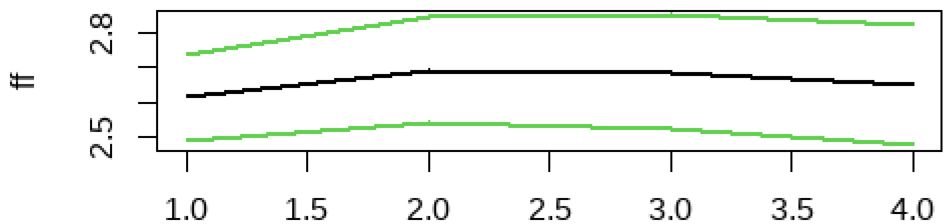
```
[10,] 0.03673542 0.8770356 0.086229021

[11,] 0.03707209 0.8699629 0.092964968

[12,] 0.03806235 0.8536566 0.108281070
```

7.

```
$dc_var

         fcst    lower    upper         CI

[1,] 5.268615 5.253503 5.283726 0.01511171

[2,] 5.276199 5.246564 5.305833 0.02963472

[3,] 5.286779 5.241307 5.332251 0.04547193

[4,] 5.292539 5.233897 5.351181 0.05864227


$dm_var

         fcst    lower    upper         CI

[1,] 12.56203 12.51308 12.61099 0.04895206

[2,] 12.50611 12.43853 12.57368 0.06757243

[3,] 12.51332 12.43801 12.58862 0.07530700

[4,] 12.44339 12.36361 12.52316 0.07977645
```

```
$dv_var

          fcst     lower     upper         CI

[1,] 2.614257 2.490502 2.738013 0.1237554

[2,] 2.693146 2.541719 2.844574 0.1514277

[3,] 2.685535 2.522002 2.849069 0.1635334

[4,] 2.648425 2.474227 2.822623 0.1741981
```




ff
2.5
2.8
1.0
1.5
2.0
2.5
3.0
3.5
4.0

Data Sources

https://fred.stlouisfed.org/series/CUSR0000SETA02

https://fred.stlouisfed.org/series/TOTALSA

https://fred.stlouisfed.org/series/TRFVOLUSM227NFWA