BOGAZICI UNIVERSITY

IE360
Term Project

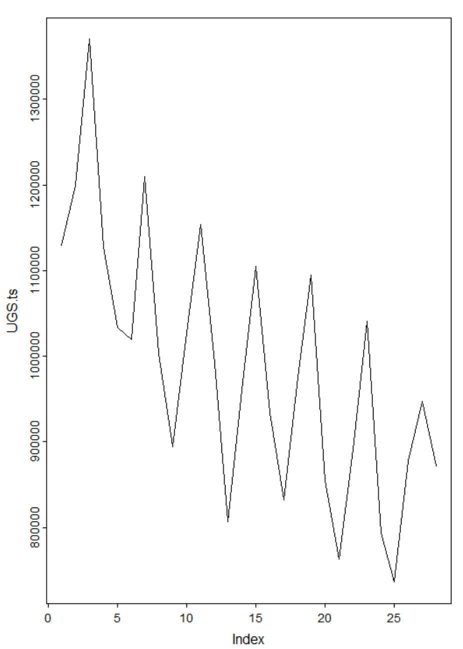
Instructor: Prof. Dr. Refik GÜLLÜ
TA: Emre KARA
SA: Halis OĞUZ

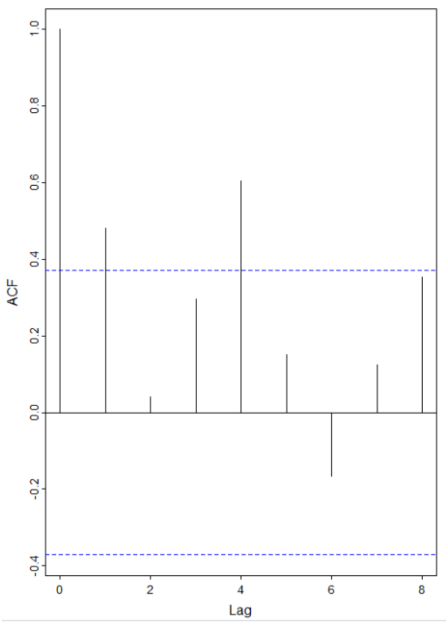
Arda Turan - 2019402075
Burak Berk Bulut - 2019402138
Batuhan Özkan - 2018402033
Mustafa Kutay Almak - 2019402144
Hasan Yağız Kılıç - 2019402147

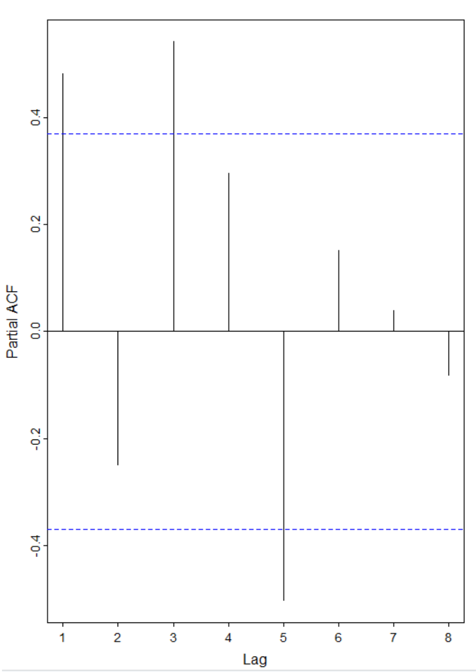
OBSERVATIONS OF THE DATA

UGS

```
UGS.ts <- ts(data$'Unleaded Gasoline Sale (UGS)', frequency=4)
UGS.ts <- UGS.ts[1:(length(UGS.ts)-4)]
plot(UGS.ts,type="l")
acf(UGS.ts,lag=8)
pacf(UGS.ts, lag=8)
plot(UGS.ts, type = "l", main = "Line Graph of UGS Time Series")
```

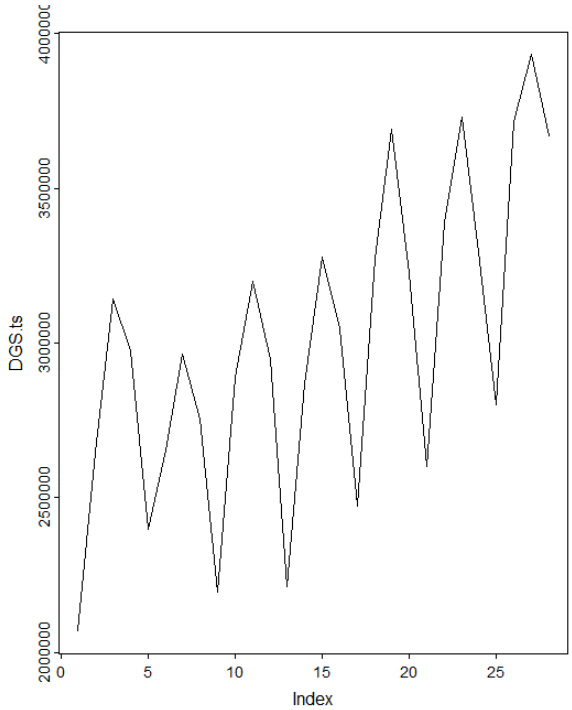


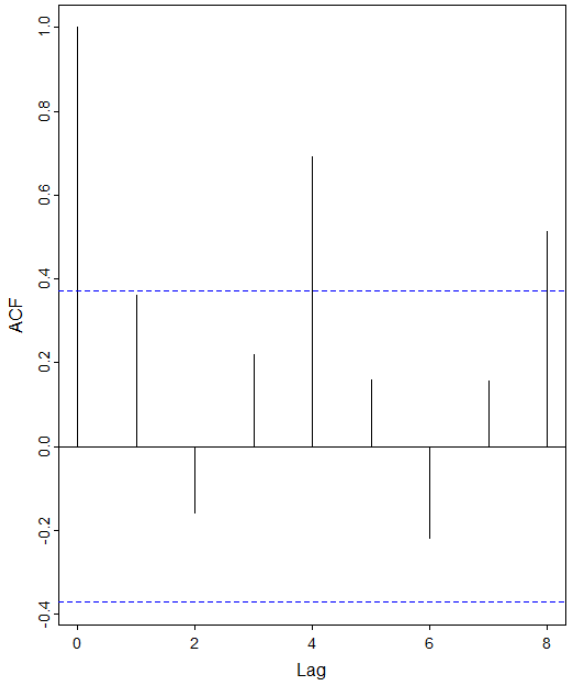


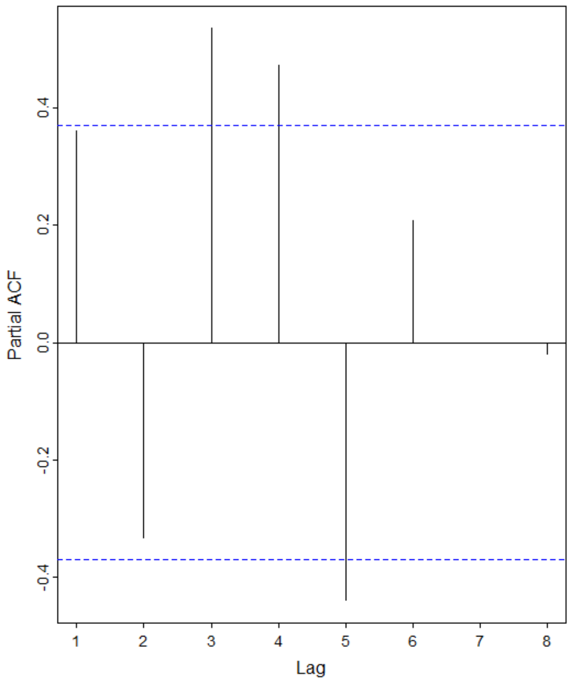


DGS

```
DGS.ts <- ts(data$'Diesel Gasoline Sale (DGS)',frequency=4)
DGS.ts <- DGS.ts[1:(length(DGS.ts)-4)]
plot(DGS.ts,type="l")
acf(DGS.ts,lag=8)
pacf(DGS.ts, lag=8)
plot(DGS.ts, type = "l", main = "Line Graph of UGS Time Series")
```







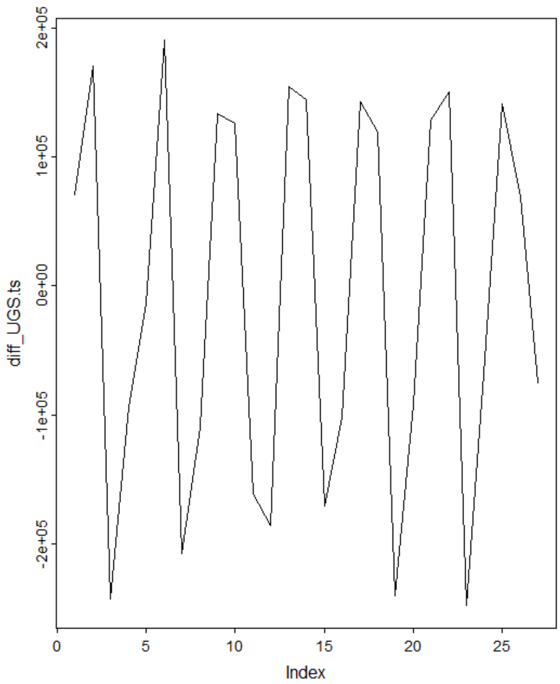
As it can be seen from the plots, both UGS and DGS data are not stationary and they need a preliminary transformation to induce their stationarity.

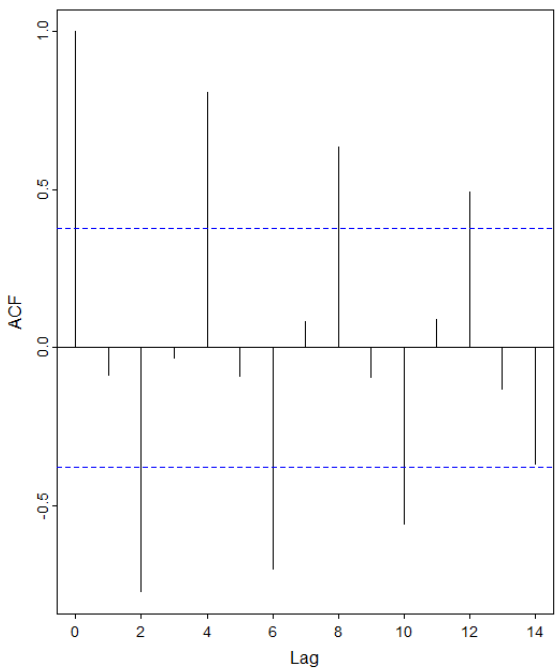
METHOD A

UGS

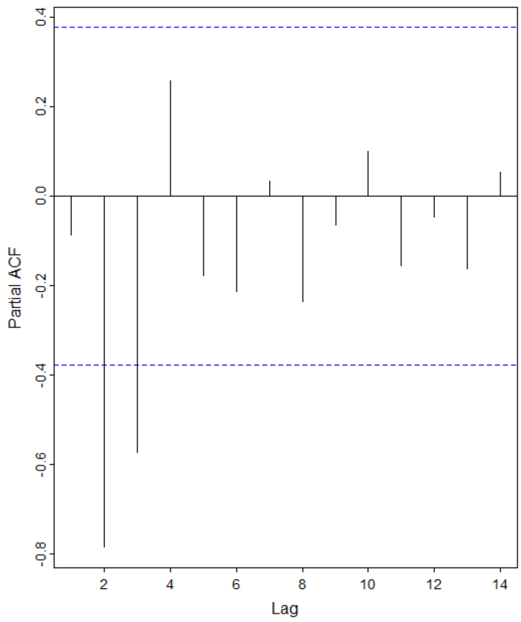
Part 1 and 2

```
diff_UGS.ts <- diff(UGS.ts)
plot(diff_UGS.ts, type = "l")
acf(diff_UGS.ts)
pacf(diff_UGS.ts)
# Perform the Augmented Dickey–Fuller test to check the stationarity.
adf_test <- adf.test(diff_UGS.ts)
print(adf_test)
# ADF test says that data is stationary.
```





To remove the trend, the difference of the data is taken. According to the ADF test, data is stationary now.
According to the ACF plot, data needs seasonal differencing.



Part 3

model <- auto.arima(UGS.ts, seasonal = TRUE)
model

```
Series: UGS.ts
ARIMA(0,1,0)

sigma^2 = 2.244e+10:  log likelihood = -360.07
AIC=722.15   AICc=722.31   BIC=723.44
```

Auto.arima function is used to find a reference ARIMA model. Note that the SARIMA model is not considered in this part.

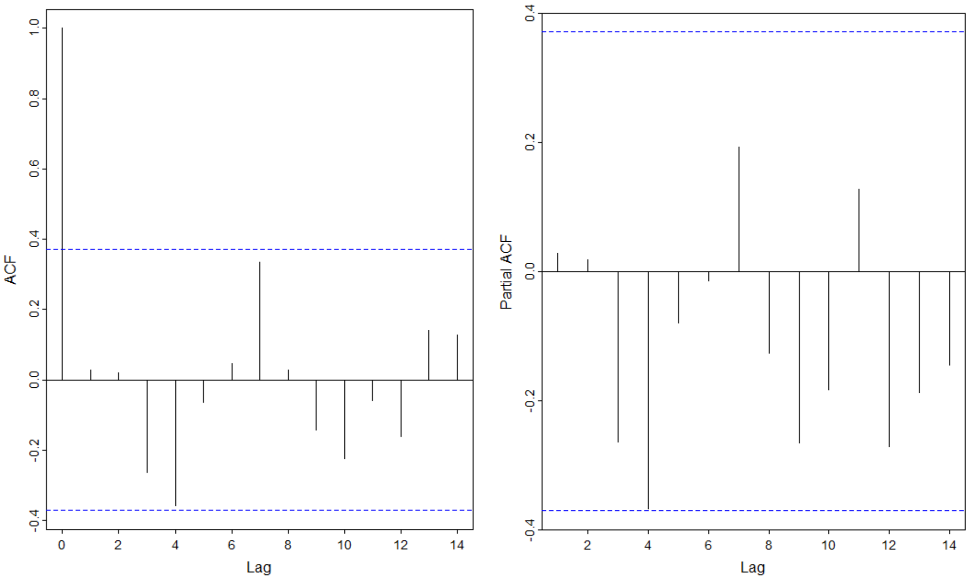
Part 4 and 5

model1<-sarima(UGS.ts,1,1,0,0,1,0,S=4,details=F)
model1

```
$fit

Call:
arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D, Q), 
period = S),
    include.mean = !no.constant, transform.pars = trans, fixed = fixed, 
optim.control = list(trace = trc,
        REPORT = 1, reltol = tol))

Coefficients:
          ar1
      -0.5877
s.e.   0.2145

sigma^2 estimated as 3.133e+09:  log likelihood = -284.3,  aic = 572.59

$degrees_of_freedom
[1] 22

$ttable
    Estimate     SE t.value p.value
ar1  -0.5877 0.2145 -2.7399   0.012
```

```
$AIC
[1] 24.89537

$AICc
[1] 24.90365

$BIC
[1] 24.99411
```

acf(resid(model1$fit), main = "ACF of residuals")
pacf(resid(model1$fit), main = "PACF of residuals")

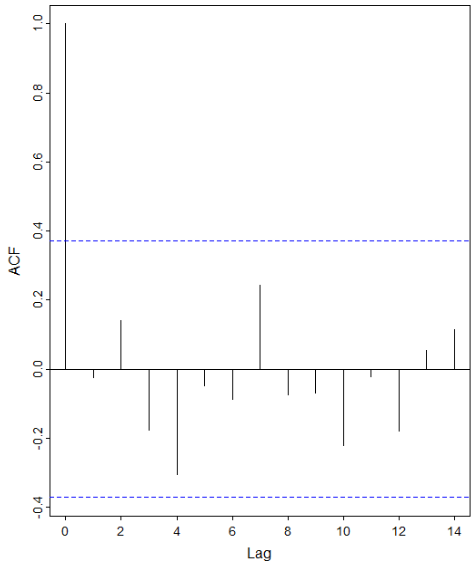


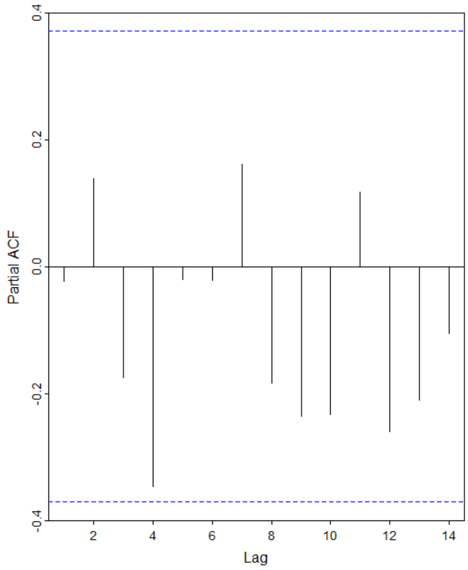
model2<-sarima(UGS.ts,0,1,1,0,1,0,S=4,details=F)
model2

```
$fit

Call:
arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D, Q),
period = S),
    include.mean = !no.constant, transform.pars = trans, fixed = fixed,
optim.control = list(trace = trc,
        REPORT = 1, reltol = tol))

Coefficients:
          ma1
      -0.5810
s.e.   0.2207

sigma^2 estimated as 3.239e+09:  log likelihood = -284.67,  aic = 573.35

$degrees_of_freedom
[1] 22

$ttable
    Estimate     SE t.value p.value
ma1   -0.581 0.2207 -2.6327  0.0152
$AIC
[1] 24.92816
```

```
$AICc
[1] 24.93644

$BIC
[1] 25.0269
```

acf(resid(model2$fit), main = "ACF of residuals")
pacf(resid(model2$fit), main = "PACF of residuals")

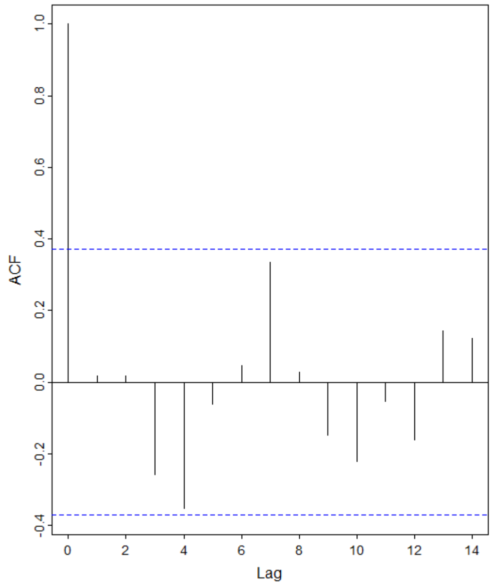


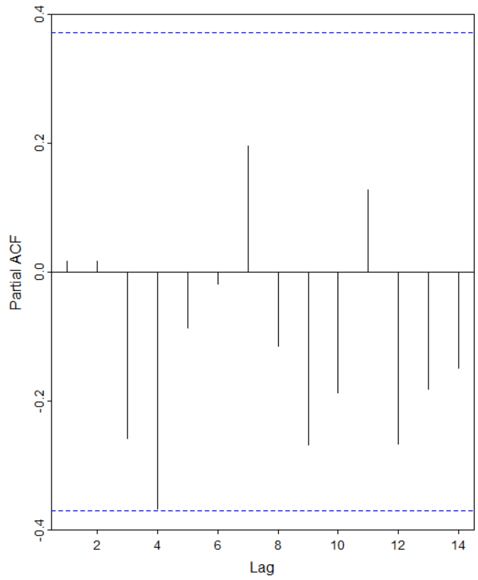


model3<-sarima(UGS.ts,1,1,1,0,1,0,S=4,details=F)
model3

```
$fit

Call:
arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D, Q), 
period = S), 
    include.mean = !no.constant, transform.pars = trans, fixed = fixed, 
optim.control = list(trace = trc, 
        REPORT = 1, reltol = tol))

Coefficients:
          ar1     ma1
      -0.6182  0.0412
s.e.   0.3302  0.3580

sigma^2 estimated as 3.13e+09:  log likelihood = -284.29,  aic = 574.58

$degrees_of_freedom
[1] 21

$ttable
    Estimate     SE t.value p.value
ar1  -0.6182 0.3302 -1.8720  0.0752
ma1   0.0412 0.3580  0.1151  0.9095

$AIC
[1] 24.98178

$AICc
```

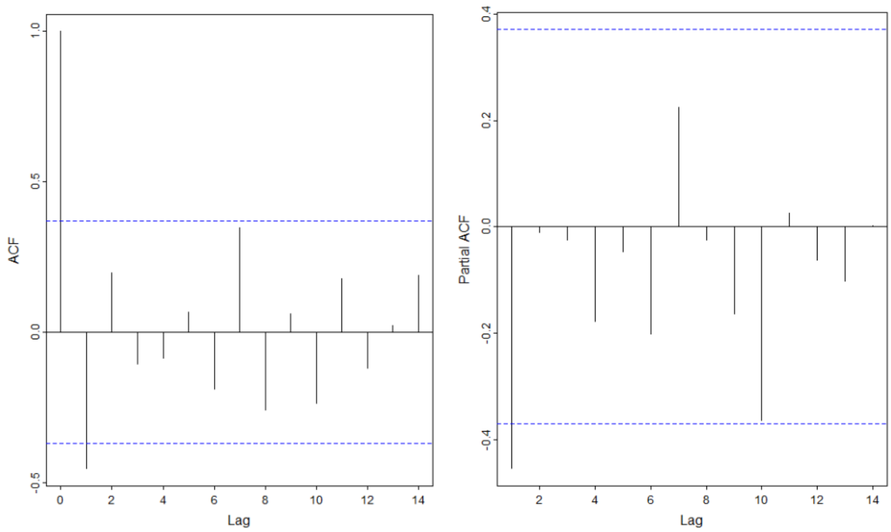
```
[1] 25.00787

$BIC
[1] 25.12989
```
acf(resid(model3$fit), main = "ACF of residuals")
pacf(resid(model3$fit), main = "PACF of residuals")





model4<-sarima(UGS.ts,0,1,0,1,1,0,S=4,details=F)
model4
```
$fit

Call:
arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D, Q),
period = S),
    include.mean = !no.constant, transform.pars = trans, fixed = fixed,
optim.control = list(trace = trc,
        REPORT = 1, reltol = tol))

Coefficients:
         sar1
      -0.2872
s.e.   0.2383

sigma^2 estimated as 3.86e+09:  log likelihood = -286.66,  aic = 577.32

$degrees_of_freedom
[1] 22

$ttable
     Estimate     SE t.value p.value
sar1  -0.2872 0.2383 -1.2052  0.2409

$AIC
[1] 25.1008

$AICc
[1] 25.10908
```

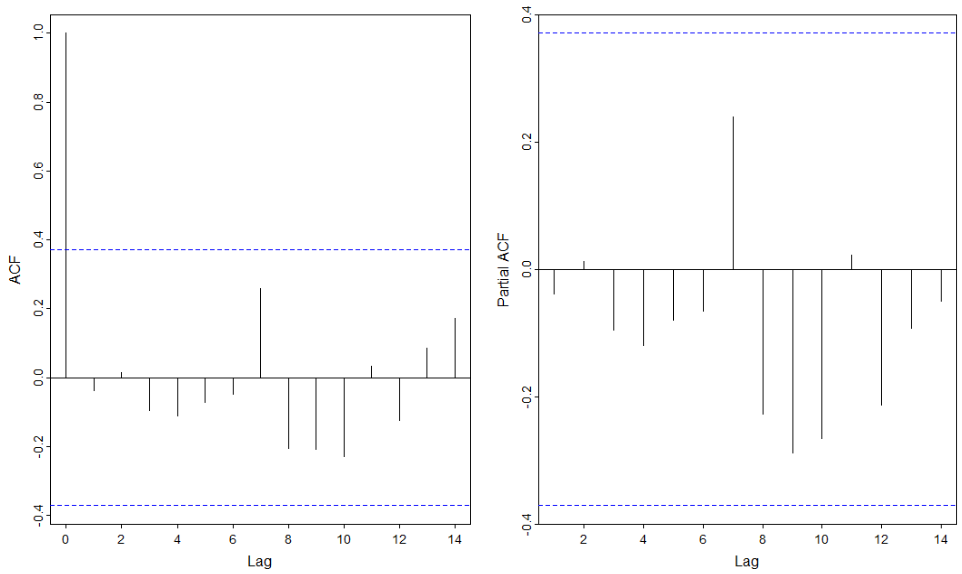
```
$BIC
[1] 25.19954
```
acf(resid(model4$fit), main = "ACF of residuals")
pacf(resid(model4$fit), main = "PACF of residuals")



model5<-sarima(UGS.ts,1,1,0,1,1,0,S=4,details=F)
model5

```
$fit

Call:
arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D, Q),
period = S),
    include.mean = !no.constant, transform.pars = trans, fixed = fixed,
optim.control = list(trace = trc,
        REPORT = 1, reltol = tol))

Coefficients:
          ar1      sar1
      -0.7037   -0.5098
s.e.   0.1940    0.2131

sigma^2 estimated as 2.457e+09:  log likelihood = -282.12,  aic = 570.24

$degrees_of_freedom
[1] 21

$ttable
     Estimate     SE t.value p.value
ar1   -0.7037 0.1940 -3.6272  0.0016
sar1  -0.5098 0.2131 -2.3919  0.0262

$AIC
[1] 24.79293

$AICc
[1] 24.81901

$BIC
[1] 24.94104
```
acf(resid(model5$fit), main = "ACF of residuals")

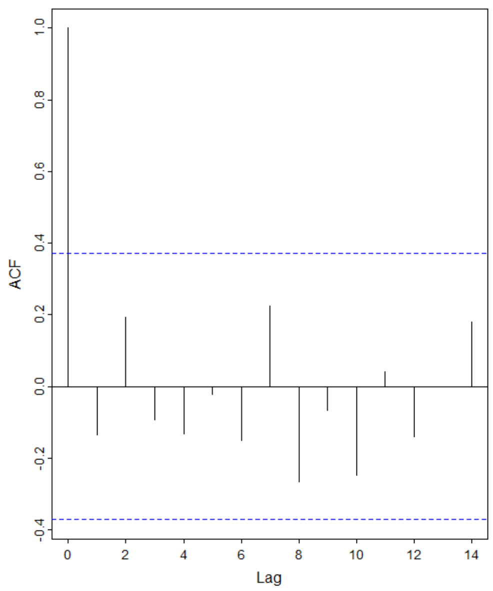
pacf(resid(model5$fit), main = "PACF of residuals")

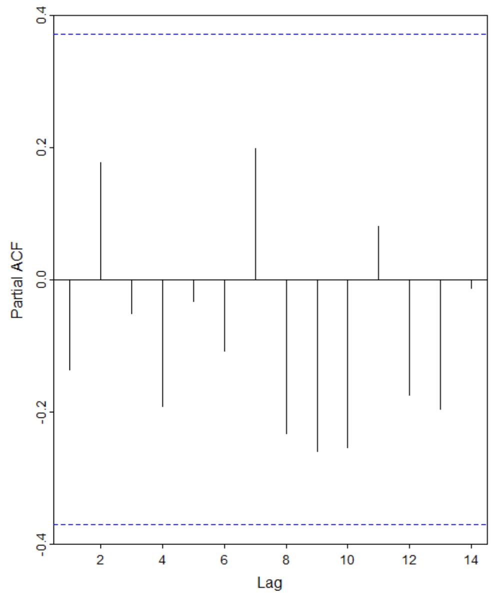


model6<-sarima(UGS.ts,0,1,1,1,1,0,S=4,details=F)
model6

```
$fit

Call:
arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D, Q), 
period = S),
    include.mean = !no.constant, transform.pars = trans, fixed = fixed, 
optim.control = list(trace = trc,
        REPORT = 1, reltol = tol))

Coefficients:
          ma1     sar1
      -0.5579  -0.3914
s.e.   0.1813   0.2279

sigma^2 estimated as 2.813e+09:  log likelihood = -283.4,  aic = 572.81

$degrees_of_freedom
[1] 21

$ttable
     Estimate     SE t.value p.value
ma1   -0.5579 0.1813 -3.0773  0.0057
sar1  -0.3914 0.2279 -1.7175  0.1006

$AIC
[1] 24.90459

$AICc
[1] 24.93068

$BIC
[1] 25.0527
```

acf(resid(model6$fit), main = "ACF of residuals")
pacf(resid(model6$fit), main = "PACF of residuals")

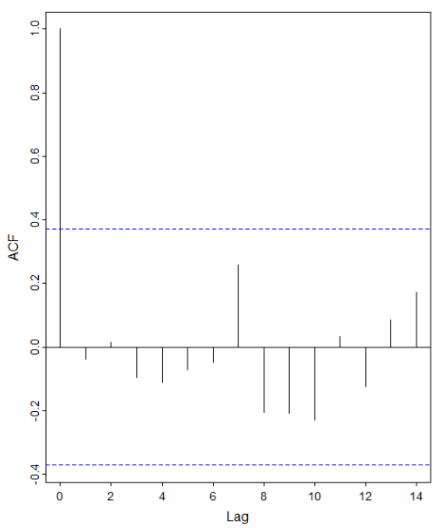


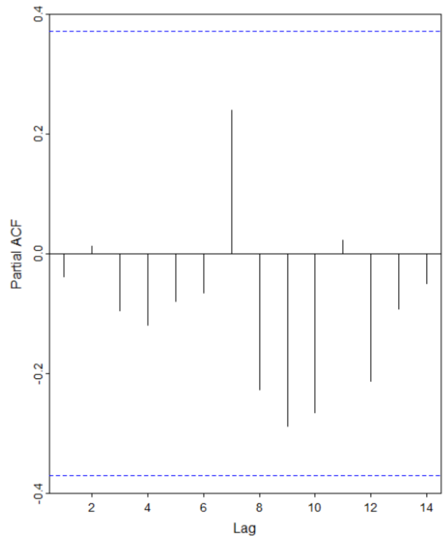


6 different SARIMA models have been tried with different (p,d,q)(P,D,Q) values and different AIC and AICc values have been obtained.

Part 6

model5 has the lowest AIC
AIC of model 5 is 24.79293
best_model <- model5
acf(resid(best_model$fit), main = "ACF of residuals (best model)")
pacf(resid(best_model$fit), main = "PACF of residuals (best model)")





Model 5 has the lowest AIC value and it has been chosen as the best model.
Consider that the dataset has few observations, looking the AICc values is more accurate. Since AICc is AIC(correction factor), it still gives the Model 5 as the best option.

Part 7

The ACF and the PACF of the residuals are analyzed. Addition to them, we constructed more statistical test to show the model’s validity.

# Apply the Ljung-Box test to the residuals
ljung_box_test <- Box.test(resid(best_model$fit), lag = 12, type = "Ljung-Box")
print(ljung_box_test)

```
Box-Ljung test

data:  resid(best_model$fit)
X-squared = 10.758, df = 12, p-value = 0.5497
```

#p-value is 0.5497. This shows that there is no significant evidence of autocorrelation in the residuals.
#Applying Box-Pierce
box_pierce_test <- Box.test(resid(best_model$fit), lag = 12, type = "Box-Pierce")
print(box_pierce_test)

```
Box-Pierce test

data:  resid(best_model$fit)
X-squared = 7.0673, df = 12, p-value = 0.8531
```

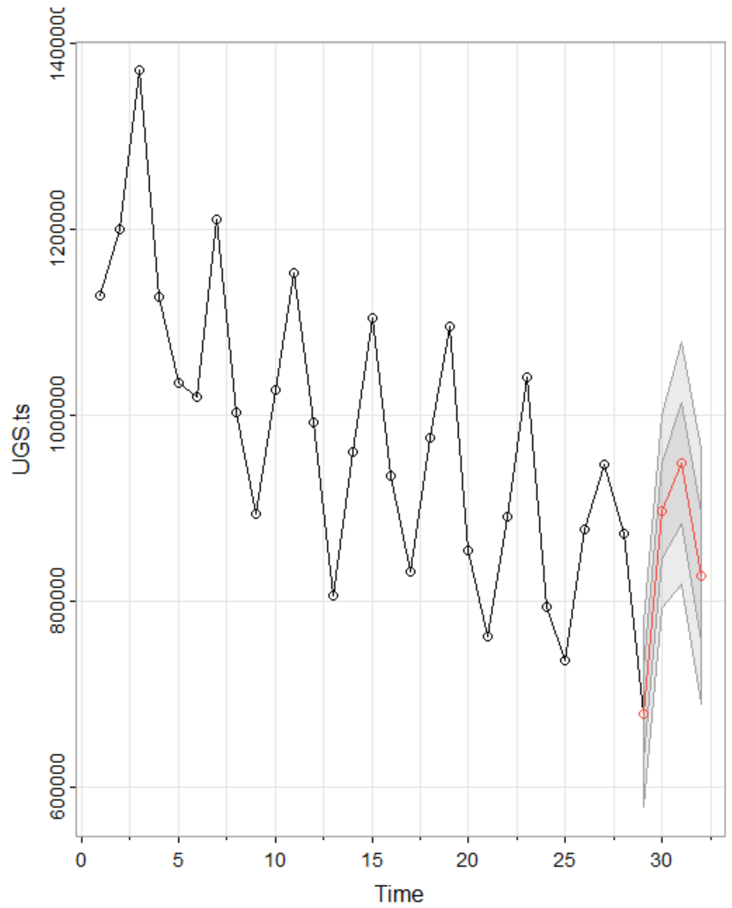
#p-value is 0.8531. This shows that there is no significant evidence of autocorrelation in the residuals.

To validate that the model that has been chosen is true, Ljung-Box and Box-Pierce tests have been applied and according to the p-values, the model has passed the tests.

Part 8

UGS_forecast <- sarima.for(UGS.ts,4,1,1,0,1,1,0,4)
UGS_forecast

```
$pred
Time Series:
Start = 29
End = 32
Frequency = 1
[1] 678288.8 896475.5 948056.1 826669.8

$se
Time Series:
Start = 29
End = 32
Frequency = 1
[1] 49572.18 51702.19 64904.67 68519.26
## Predicted values for 2007 sales:
## 2007_Q1 = 678288.8
## 2007_Q2 = 896475.5
## 2007_Q3 = 948056.1
## 2007_Q4 = 826669.8
```



Forecast of 4 quarters of 2007 is obtained by using the best SARIMA model.

DGS

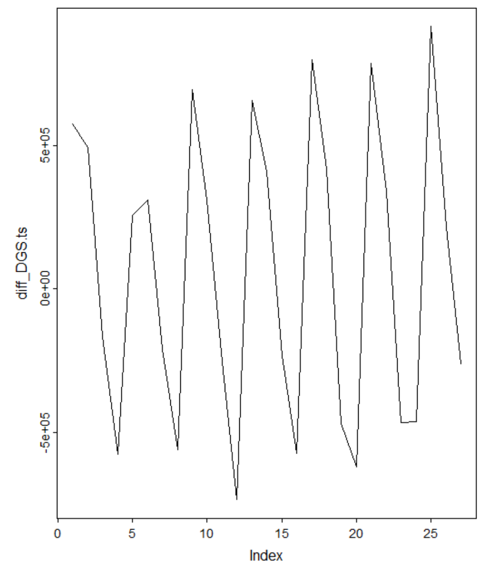
Part 1 and 2

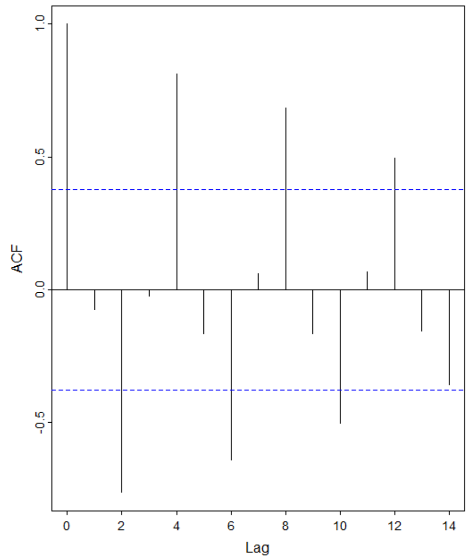
diff_DGS.ts <- diff(DGS.ts)
plot(diff_DGS.ts, type = "l")
acf(diff_DGS.ts)
pacf(diff_DGS.ts)
# Perform the ADF test
adf_test <- adf.test(diff_DGS.ts)
print(adf_test)

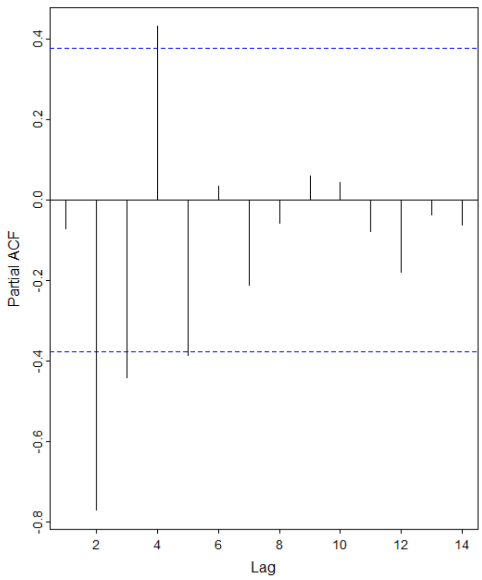
```
Augmented Dickey-Fuller Test

data:  diff_DGS.ts
Dickey-Fuller = -13.984, Lag order = 2, p-value = 0.01
alternative hypothesis: stationary
```

ADF test says that data is stationary.

To remove the trend, difference of the data is taken. According to the ADF test, data is stationary now.
According to the ACF plot, data needs seasonal differencing.







Part 3

model <- auto.arima(DGS.ts, seasonal = TRUE)
model

```
Series: DGS.ts
ARIMA(3,1,0) with drift

Coefficients:
          ar1      ar2      ar3      drift
      -0.8666  -0.9579  -0.8482  35175.579
s.e.   0.1101   0.0555   0.1016   9628.642

sigma^2 = 3.518e+10:  log likelihood = -366.89
AIC=743.77   AICc=746.63   BIC=750.25
```

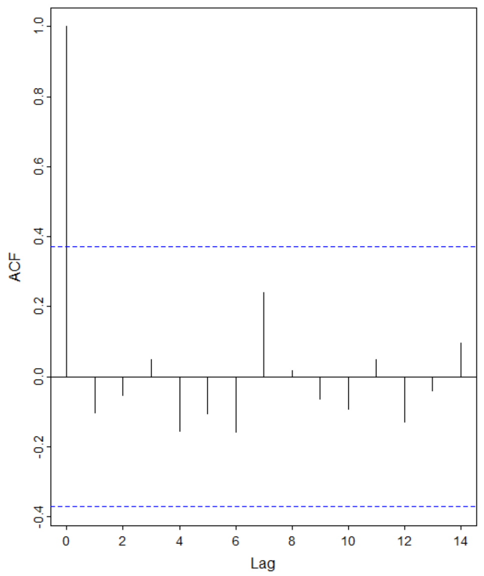
Auto.arima function is used to find a reference ARIMA model. Note that the SARIMA model is not considered in this part.

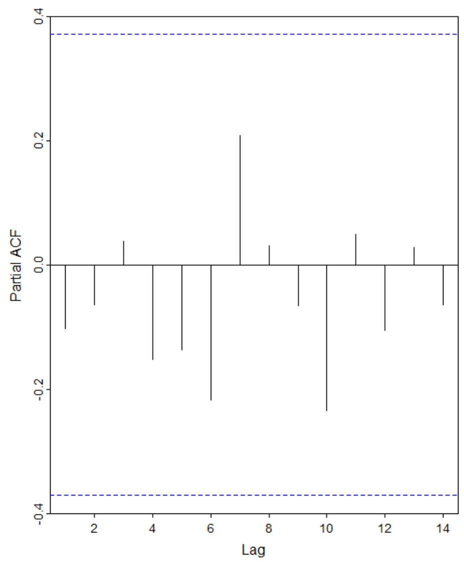
Part 4 and 5

model1<-sarima(DGS.ts,3,1,0,0,1,0,S=4,details=F)
model1

```
$fit

Call:
arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D, Q),
period = S),
    include.mean = !no.constant, transform.pars = trans, fixed = fixed,
optim.control = list(trace = trc,
        REPORT = 1, reltol = tol))

Coefficients:
         ar1      ar2      ar3
      0.1629  -0.0854  -0.3251
s.e.  0.2265   0.2293   0.2291

sigma^2 estimated as 2.171e+10:  log likelihood = -306.56,  aic = 621.11

$degrees_of_freedom
[1] 20

$ttable
    Estimate     SE t.value p.value
ar1   0.1629 0.2265  0.7193  0.4803
ar2  -0.0854 0.2293 -0.3725  0.7134
ar3  -0.3251 0.2291 -1.4190  0.1713

$AIC
[1] 27.00493

$AICc
[1] 27.05985

$BIC
[1] 27.2024
```

acf(resid(model1$fit), main = "ACF of residuals")
pacf(resid(model1$fit), main = "PACF of residuals")

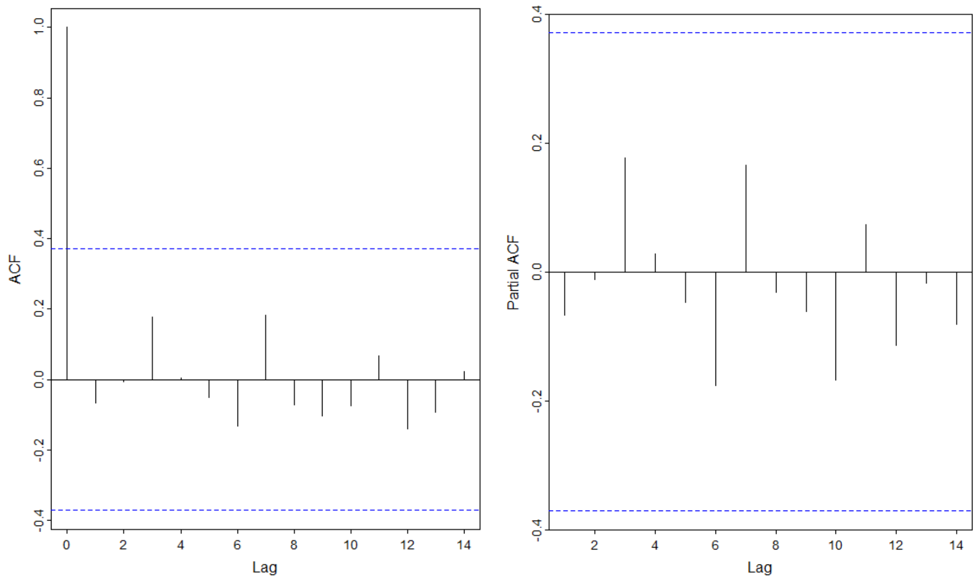




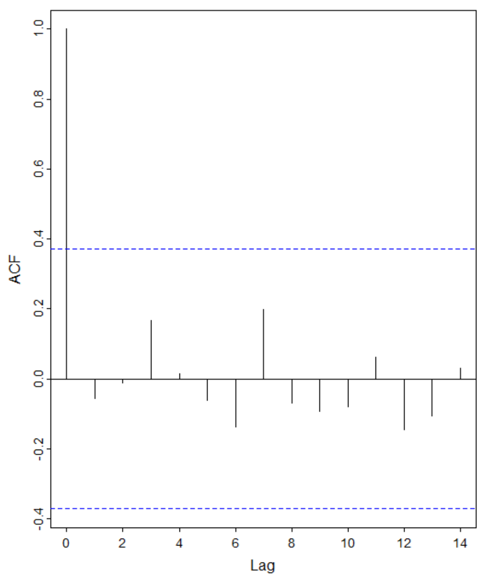
model2<-sarima(DGS.ts,3,1,1,0,1,0,S=4,details=F)
model2

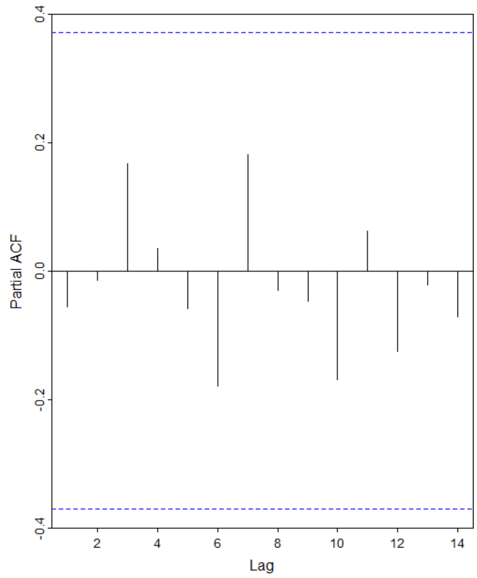
```
$fit

Call:
arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D, Q),
period = S),
    include.mean = !no.constant, transform.pars = trans, fixed = fixed,
optim.control = list(trace = trc,
        REPORT = 1, reltol = tol))

Coefficients:
         ar1      ar2      ar3      ma1
      0.6657  -0.1584  -0.3674  -0.8012
s.e.  0.2735   0.2758   0.2476   0.1835

sigma^2 estimated as 1.674e+10:  log likelihood = -304.25,  aic = 618.51

$degrees_of_freedom
[1] 19

$ttable
    Estimate     SE t.value p.value
ar1   0.6657 0.2735  2.4338  0.0250
ar2  -0.1584 0.2758 -0.5745  0.5724
ar3  -0.3674 0.2476 -1.4838  0.1543
ma1  -0.8012 0.1835 -4.3653  0.0003

$AIC
[1] 26.89153

$AICc
[1] 26.98815

$BIC
[1] 27.13838
```

acf(resid(model2$fit), main = "ACF of residuals")
pacf(resid(model2$fit), main = "PACF of residuals")



```
model3<-sarima(DGS.ts,3,1,1,0,1,1,S=4,details=F)
model3
$fit

Call:
arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D, Q),
period = S),
    include.mean = !no.constant, transform.pars = trans, fixed = fixed,
optim.control = list(trace = trc,
        REPORT = 1, reltol = tol))

Coefficients:
         ar1      ar2      ar3      ma1     sma1
      0.6529  -0.1635  -0.3373  -0.7911  -0.0597
s.e.  0.2961   0.2784   0.3487   0.2060   0.4519

sigma^2 estimated as 1.677e+10:  log likelihood = -304.24,  aic = 620.49

$degrees_of_freedom
[1] 18

$ttable
     Estimate     SE t.value p.value
ar1    0.6529 0.2961  2.2053  0.0407
ar2   -0.1635 0.2784 -0.5873  0.5643
ar3   -0.3373 0.3487 -0.9673  0.3462
ma1   -0.7911 0.2060 -3.8402  0.0012
sma1  -0.0597 0.4519 -0.1320  0.8964

$AIC
[1] 26.9777

$AICc
[1] 27.13115

$BIC
[1] 27.27392
acf(resid(model3$fit), main = "ACF of residuals")
pacf(resid(model3$fit), main = "PACF of residuals")
```

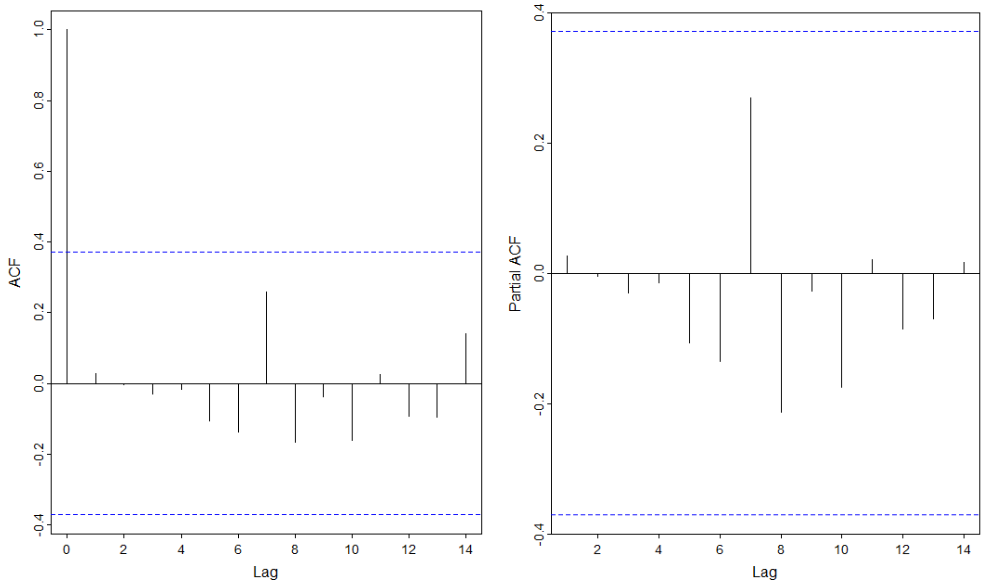




model4<-sarima(DGS.ts,3,1,0,1,1,0,S=4,details=F)
model4

```
$fit

Call:
arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D, Q),
period = S),
    include.mean = !no.constant, transform.pars = trans, fixed = fixed,
optim.control = list(trace = trc,
        REPORT = 1, reltol = tol))

Coefficients:
          ar1      ar2      ar3     sar1
      -0.1565  -0.2894  -0.2031  -0.5636
s.e.   0.2647   0.2463   0.2386   0.2447

sigma^2 estimated as 1.824e+10:  log likelihood = -305.1,  aic = 620.19

$degrees_of_freedom
[1] 19

$ttable
     Estimate     SE t.value p.value
ar1   -0.1565 0.2647 -0.5913  0.5613
ar2   -0.2894 0.2463 -1.1750  0.2545
ar3   -0.2031 0.2386 -0.8513  0.4052
sar1  -0.5636 0.2447 -2.3030  0.0327

$AIC
[1] 26.96483

$AICc
[1] 27.06145

$BIC
[1] 27.21167
```

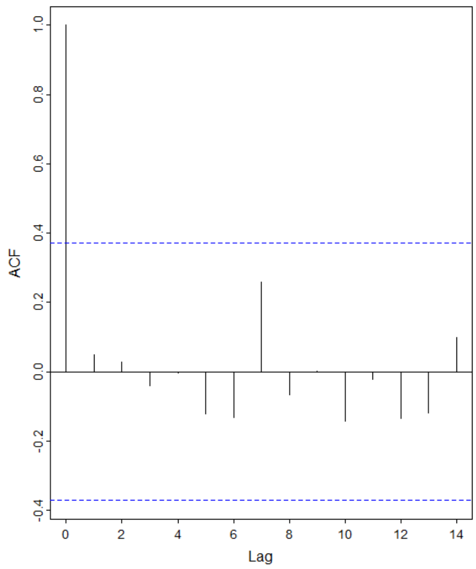
acf(resid(model4$fit), main = "ACF of residuals")
pacf(resid(model4$fit), main = "PACF of residuals")

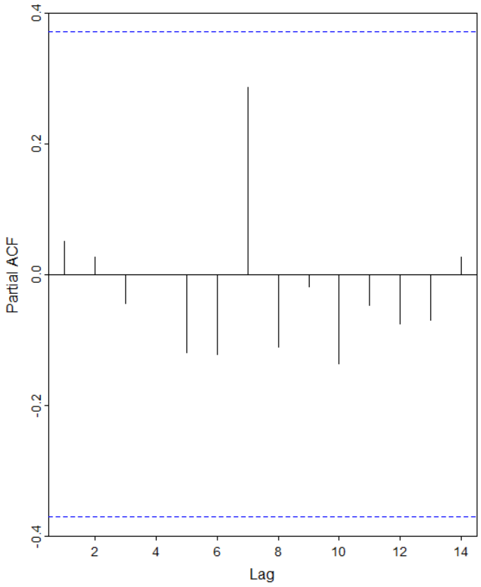


model5<-sarima(DGS.ts,3,1,0,1,1,1,S=4,details=F)
model5

```
$fit

Call:
arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D, Q),
period = S),
    include.mean = !no.constant, transform.pars = trans, fixed = fixed,
optim.control = list(trace = trc,
        REPORT = 1, reltol = tol))

Coefficients:
          ar1      ar2      ar3     sar1     sma1
      -0.1447  -0.4033  -0.1627  -0.3352  -0.3637
s.e.   0.2684   0.2903   0.2472   0.4729   0.4756

sigma^2 estimated as 1.768e+10:  log likelihood = -304.91,  aic = 621.81

$degrees_of_freedom
[1] 18

$ttable
     Estimate     SE t.value p.value
ar1   -0.1447 0.2684 -0.5390  0.5965
ar2   -0.4033 0.2903 -1.3893  0.1817
ar3   -0.1627 0.2472 -0.6581  0.5188
sar1  -0.3352 0.4729 -0.7089  0.4875
sma1  -0.3637 0.4756 -0.7648  0.4543

$AIC
[1] 27.03524

$AICc
[1] 27.18869

$BIC
[1] 27.33145
```

acf(resid(model5$fit), main = "ACF of residuals")
pacf(resid(model5$fit), main = "PACF of residuals")

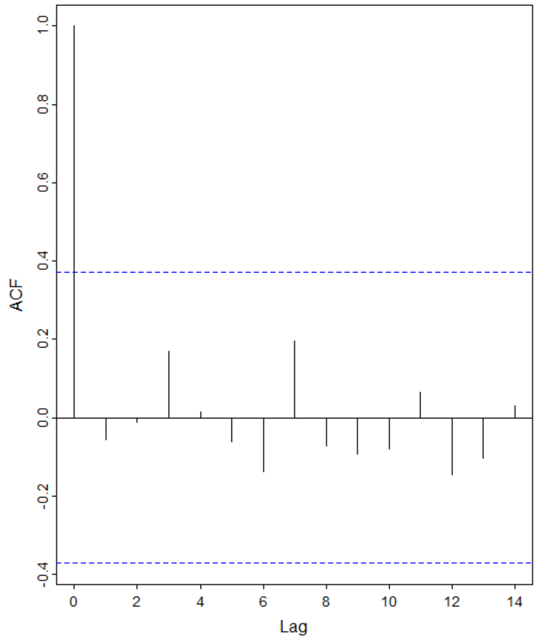


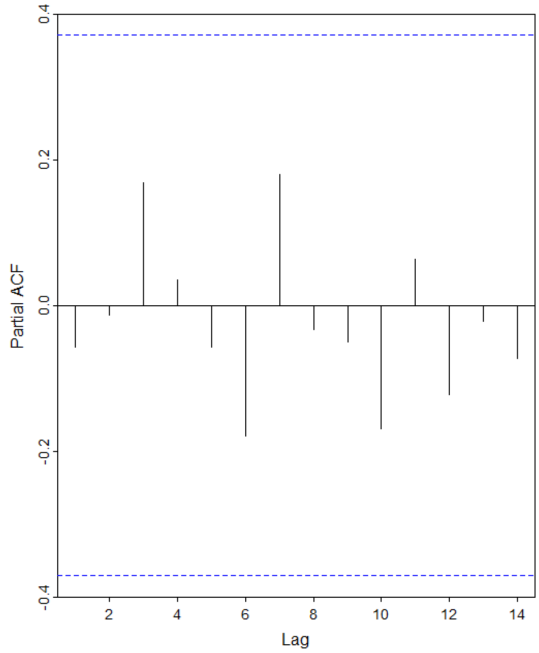


model6<-sarima(DGS.ts,3,1,1,1,1,0,S=4,details=F)
model6

```
$fit

Call:
arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D, Q),
period = S),
    include.mean = !no.constant, transform.pars = trans, fixed = fixed,
optim.control = list(trace = trc,
        REPORT = 1, reltol = tol))

Coefficients:
         ar1      ar2      ar3      ma1     sar1
      0.6546  -0.1619  -0.3418  -0.7930  -0.0511
s.e.  0.2954   0.2761   0.3378   0.2015   0.4308

sigma^2 estimated as 1.676e+10:  log likelihood = -304.24,  aic = 620.49

$degrees_of_freedom
[1] 18

$ttable
     Estimate     SE t.value p.value
ar1    0.6546 0.2954  2.2159  0.0398
ar2   -0.1619 0.2761 -0.5864  0.5649
ar3   -0.3418 0.3378 -1.0119  0.3250
ma1   -0.7930 0.2015 -3.9348  0.0010
sar1  -0.0511 0.4308 -0.1187  0.9069

$AIC
[1] 26.97782

$AICc
[1] 27.13127

$BIC
[1] 27.27403
```

```
acf(resid(model6$fit), main = "ACF of residuals")
pacf(resid(model6$fit), main = "PACF of residuals")
```

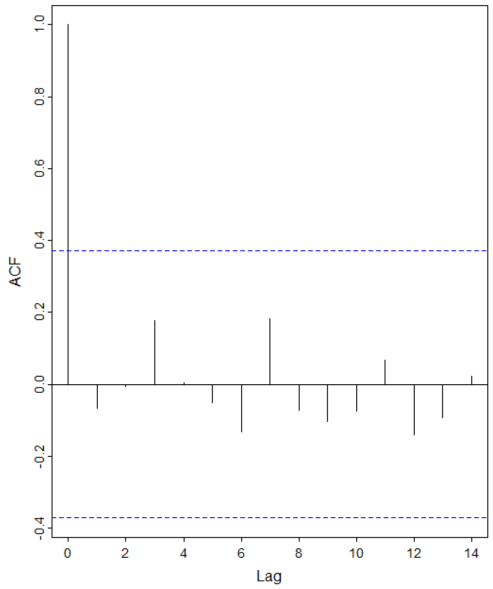


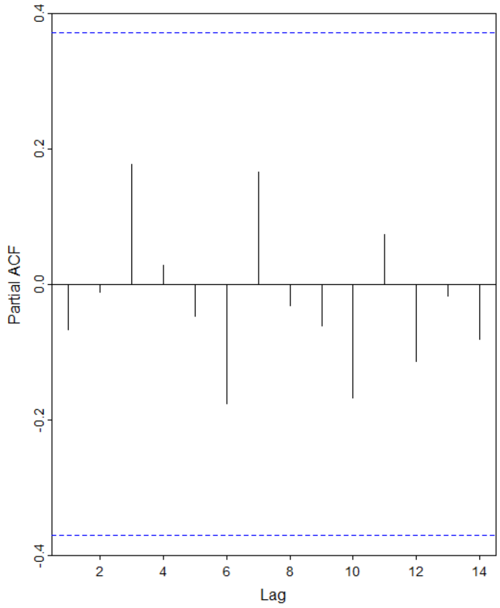


6 different SARIMA models have been tried with different (p,d,q)(P,D,Q) values and different AIC and AICc values have been obtained.

Part 6

```
## model2 has the lowest AIC
## AIC of model2 is 26.89153
best_model <- model2
acf(resid(best_model$fit), main = "ACF of residuals (best model)")
pacf(resid(best_model$fit), main = "PACF of residuals (best model)")
```





Model 5 has the lowest AIC value and it has been chosen as the best model.
Consider that the dataset has few observations, looking the AICc values is more accurate. Since AICc is AIC(correction factor), it still gives the Model 5 as the best option

Part 7

The ACF and the PACF of the residuals are analyzed. Addition to them, we constructed more statistical test to show the model's validity.

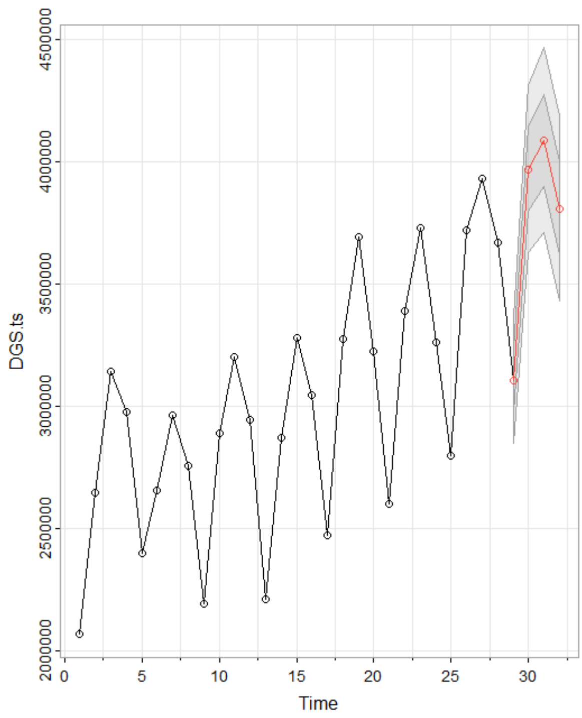
```
# Apply the Ljung-Box test to the residuals
ljung_box_test <- Box.test(resid(best_model$fit), lag = 12, type = "Ljung-Box")
print(ljung_box_test)
Box-Ljung test

data:  resid(best_model$fit)
X-squared = 5.4805, df = 12, p-value = 0.94
#p-value is 0.94. This shows that there is no significant evidence of autocorrelation in the residuals.
```

```
#Applying Box-Pierce
box_pierce_test <- Box.test(resid(best_model$fit), lag = 12, type = "Box-Pierce")
print(box_pierce_test)
Box-Pierce test

data:  resid(best_model$fit)
X-squared = 3.765, df = 12, p-value = 0.9873
#p-value is 0.9873. This shows that there is no significant evidence of autocorrelation in the residuals.
```

To validate that the model that has been chosen is true, Ljung-Box and Box-Pierce tests have been applied and according to the p-values, the model has passed the tests.

Part 8

```
DGS_forecast <- sarima.for(DGS.ts,4,3,1,1,0,1,0,4)
DGS_forecast
$pred
Time Series:
Start = 29
End = 32
Frequency = 1
[1] 3105950 3968121 4087278 3808176

$se
Time Series:
Start = 29
End = 32
Frequency = 1
[1] 129382.3 171026.2 188676.0 189157.4
## Predicted values for 2007 sales:
## 2007_Q1 = 3105950
## 2007_Q2 = 3968121
## 2007_Q3 = 4087278
## 2007_Q4 = 3808176
```



Forecast of 4 quarters of 2007 is obtained by using the best SARIMA model.

METHOD B

UGS

data<-read_excel("IE360-ProjectData.xlsx",col_names = TRUE)

cor(data[sapply(data, is.numeric)], use = "complete.obs")

	Unleaded Gasoline Sale (UGS)
Unleaded Gasoline Sale (UGS)	1.00000000
Diesel Gasoline Sale (DGS)	0.04670539
RNUV	0.26498942
# LPG Vehicles (NLPG)	-0.62890683
Price of Unleaded Gasoline (PU)	-0.48993252
Price of Diesel Gasoline (PG)	-0.54812079
# Unleaded Gasoline Vehicles (NUGV)	-0.64862885
# of Diesel Gasoline Vehicles (NDGV)	-0.52526929

GNP Agriculture	0.58815962
GNP Commerce	0.20843693
GNP Total	0.19348598

	Diesel Gasoline Sale (DGS)	RNUV
Unleaded Gasoline Sale (UGS)	0.04670539	0.26498942
Diesel Gasoline Sale (DGS)	1.00000000	0.30364955
RNUV	0.30364955	1.00000000
# LPG Vehicles (NLPG)	0.63895652	0.18522216
Price of Unleaded Gasoline (PU)	0.49630315	0.04177338
Price of Diesel Gasoline (PG)	0.54026568	0.07378405
# Unleaded Gasoline Vehicles (NUGV)	0.67288637	0.07413199
# of Diesel Gasoline Vehicles (NDGV)	0.65656807	0.30576056
GNP Agriculture	0.60365821	0.16530707
GNP Commerce	0.91741674	0.29698204
GNP Total	0.87962217	0.30260566

	# LPG Vehicles (NLPG)
Unleaded Gasoline Sale (UGS)	-0.62890683
Diesel Gasoline Sale (DGS)	0.63895652
RNUV	0.18522216
# LPG Vehicles (NLPG)	1.00000000
Price of Unleaded Gasoline (PU)	0.67602376
Price of Diesel Gasoline (PG)	0.76400431
# Unleaded Gasoline Vehicles (NUGV)	0.92714875
# of Diesel Gasoline Vehicles (NDGV)	0.90937746
GNP Agriculture	0.03664584
GNP Commerce	0.49970806
GNP Total	0.49188935

	Price of Unleaded Gasoline (PU)
Unleaded Gasoline Sale (UGS)	-0.48993252
Diesel Gasoline Sale (DGS)	0.49630315
RNUV	0.04177338
# LPG Vehicles (NLPG)	0.67602376
Price of Unleaded Gasoline (PU)	1.00000000
Price of Diesel Gasoline (PG)	0.98211135
# Unleaded Gasoline Vehicles (NUGV)	0.73923142

# of Diesel Gasoline Vehicles (NDGV)	0.68439875
GNP Agriculture	0.11288812
GNP Commerce	0.48075691
GNP Total	0.45775513
	Price of Diesel Gasoline (PG)
Unleaded Gasoline Sale (UGS)	-0.54812079
Diesel Gasoline Sale (DGS)	0.54026568
RNUV	0.07378405
# LPG Vehicles (NLPG)	0.76400431
Price of Unleaded Gasoline (PU)	0.98211135
Price of Diesel Gasoline (PG)	1.00000000
# Unleaded Gasoline Vehicles (NUGV)	0.81214044
# of Diesel Gasoline Vehicles (NDGV)	0.75830059
GNP Agriculture	0.10869816
GNP Commerce	0.51351013
GNP Total	0.49819657
	# Unleaded Gasoline Vehicles (NUGV)
Unleaded Gasoline Sale (UGS)	-0.64862885
Diesel Gasoline Sale (DGS)	0.67288637
RNUV	0.07413199
# LPG Vehicles (NLPG)	0.92714875
Price of Unleaded Gasoline (PU)	0.73923142
Price of Diesel Gasoline (PG)	0.81214044
# Unleaded Gasoline Vehicles (NUGV)	1.00000000
# of Diesel Gasoline Vehicles (NDGV)	0.94565611
GNP Agriculture	0.06599231
GNP Commerce	0.51971136
GNP Total	0.51050156
	# of Diesel Gasoline Vehicles (NDGV)
Unleaded Gasoline Sale (UGS)	-0.52526929
Diesel Gasoline Sale (DGS)	0.65656807
RNUV	0.30576056
# LPG Vehicles (NLPG)	0.90937746
Price of Unleaded Gasoline (PU)	0.68439875
Price of Diesel Gasoline (PG)	0.75830059

```
# Unleaded Gasoline Vehicles (NUGV)                              0.94565611
# of Diesel Gasoline Vehicles (NDGV)                             1.00000000
GNP Agriculture                                                  0.06716321
GNP Commerce                                                     0.50529894
GNP Total                                                        0.49132614
                                     GNP Agriculture GNP Commerce GNP Total
Unleaded Gasoline Sale (UGS)              0.58815962    0.2084369 0.1934860
Diesel Gasoline Sale (DGS)                0.60365821    0.9174167 0.8796222
RNUV                                      0.16530707    0.2969820 0.3026057
# LPG Vehicles (NLPG)                     0.03664584    0.4997081 0.4918894
Price of Unleaded Gasoline (PU)           0.11288812    0.4807569 0.4577551
Price of Diesel Gasoline (PG)             0.10869816    0.5135101 0.4981966
# Unleaded Gasoline Vehicles (NUGV)       0.06599231    0.5197114 0.5105016
# of Diesel Gasoline Vehicles (NDGV)      0.06716321    0.5052989 0.4913261
GNP Agriculture                           1.00000000    0.8170954 0.8343358
GNP Commerce                              0.81709538    1.0000000 0.9866848
GNP Total                                 0.83433576    0.9866848 1.0000000
```

x2<- 1:32 %% 4 == 2

x3<- 1:32 %% 4 == 3

x4<- 1:32 %% 4 == 0

#We created logical vectors x2, x3, and x4 that are used to create dummy variables to identify specific quarters in the dataset. Then we add them to the data.

data1<- data.frame(data,s2 = 1*x2,s3=1*x3,s4=1*x4)

data2<- data.frame(data,Quarters = 1:32,s2 = 1*x2,s3=1*x3,s4=1*x4)

lm1 <- lm(Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4, data = data1)

summary(lm1)

```
Call:
lm(formula = Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4, data = data1)

Residuals:
    Min      1Q  Median      3Q     Max
```

```
-184749  -86358  -22243   55218  244063

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   884908      47604  18.589 9.36e-16 ***
s2            108035      67323   1.605  0.12163    
s3            246624      67323   3.663  0.00123 ** 
s4             54557      67323   0.810  0.42568    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 125900 on 24 degrees of freedom
  (4 observations deleted due to missingness)
Multiple R-squared:  0.3819,	Adjusted R-squared:  0.3047 
F-statistic: 4.943 on 3 and 24 DF,  p-value: 0.008198
```

lm2<- lm(Unleaded.Gasoline.Sale..UGS.~s2+s3+s4+Quarters,data = data2)

summary(lm2)

```
Call:
lm(formula = Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters, 
    data = data2)

Residuals:
   Min     1Q Median     3Q    Max 
-81167 -31283  -3458  28640  94502 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1060372      23653  44.830  < 2e-16 ***
s2            121532      25987   4.677 0.000104 ***
s3            273618      26063  10.498 3.03e-10 ***
s4             95049      26189   3.629 0.001405 ** 
Quarters      -13497       1147 -11.764 3.28e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 48570 on 23 degrees of freedom
  (4 observations deleted due to missingness)
Multiple R-squared:  0.9119,	Adjusted R-squared:  0.8966 
F-statistic: 59.53 on 4 and 23 DF,  p-value: 8.446e-12
```

#Check the largest absolute correlation value with UGS from the table of correlation. NUGV is the largest one. First, add the NUGV to the model.

lm3<-lm(Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters + X..Unleaded.Gasoline.Vehicles..NUGV., data = data2)

summary(lm3)

```
Call:
```

```
lm(formula = Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters +
    X..Unleaded.Gasoline.Vehicles..NUGV., data = data2)

Residuals:
   Min     1Q Median     3Q    Max
-87257 -30446   1812  27878  86183

Coefficients:
                                       Estimate Std. Error t value
(Intercept)                           7.402e+05  5.981e+05   1.238
s2                                    1.209e+05  2.643e+04   4.575
s3                                    2.722e+05  2.661e+04  10.229
s4                                    9.203e+04  2.720e+04   3.384
Quarters                             -1.592e+04  4.674e+03  -3.406
X..Unleaded.Gasoline.Vehicles..NUGV.  6.856e-02  1.280e-01   0.536
                                     Pr(>|t|)
(Intercept)                          0.228896
s2                                   0.000148 ***
s3                                   7.98e-10 ***
s4                                   0.002673 **
Quarters                             0.002533 **
X..Unleaded.Gasoline.Vehicles..NUGV. 0.597494
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 49340 on 22 degrees of freedom
  (4 observations deleted due to missingness)
Multiple R-squared:  0.913,    Adjusted R-squared:  0.8933
F-statistic:  46.2 on 5 and 22 DF,  p-value: 6.113e-11
```

#Even though NUGV is the largest correlation value with UGS , When we look at the summmary NUGV seems unsignificant so try another one with the second largest correlation value which is NLPG

lm4<-lm(Unleaded.Gasoline.Sale..UGS.~s2+s3+s4+Quarters+X..LPG.Vehicles..NLPG.,data = data2)

summary(lm4)

```
Call:
lm(formula = Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters +
    X..LPG.Vehicles..NLPG., data = data2)

Residuals:
   Min     1Q Median     3Q    Max
-83022 -23287   3235  18555  90847

Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)             8.082e+05  7.363e+04  10.977 2.16e-10 ***
s2                      1.286e+05  2.128e+04   6.044 4.40e-06 ***
s3                      2.824e+05  2.140e+04  13.200 6.24e-12 ***
s4                      9.835e+04  2.137e+04   4.601 0.000139 ***
Quarters               -2.422e+04  3.163e+03  -7.657 1.21e-07 ***
```

```
X..LPG.Vehicles..NLPG.  3.259e-01  9.184e-02   3.549 0.001800 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 39600 on 22 degrees of freedom
  (4 observations deleted due to missingness)
Multiple R-squared:  0.944,    Adjusted R-squared:  0.9313
F-statistic: 74.15 on 5 and 22 DF,  p-value: 5.077e-13
```

#NLPG seems significant so we added NLPG to the model. We will decide what we should add to the model afterwards by doing anova

#We will go with the model which has the smallest p value

model<-lm(Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters + X..LPG.Vehicles..NLPG.,data = data2)

anova(model,lm(Unleaded.Gasoline.Sale..UGS.~s2+s3+s4+Quarters+X..LPG.Vehicles..NLPG.+RNUV,data = data2))

```
Analysis of Variance Table

Model 1: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters +
X..LPG.Vehicles..NLPG.
Model 2: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters +
X..LPG.Vehicles..NLPG. +
    RNUV
  Res.Df        RSS Df  Sum of Sq      F Pr(>F)
1     22 3.4505e+10
2     21 3.1850e+10  1 2654743072 1.7504 0.2001
```

anova(model,lm(Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters + X..LPG.Vehicles..NLPG.+Price.of.Unleaded.Gasoline..PU.,data = data2))

```
Analysis of Variance Table

Model 1: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters +
X..LPG.Vehicles..NLPG.
Model 2: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters +
X..LPG.Vehicles..NLPG. +
    Price.of.Unleaded.Gasoline..PU.
  Res.Df        RSS Df  Sum of Sq      F Pr(>F)
1     22 3.4505e+10
2     21 3.0810e+10  1 3695242120 2.5187 0.1274
```

anova(model,lm(Unleaded.Gasoline.Sale..UGS.~s2+s3+s4+Quarters+X..LPG.Vehicles..NLPG.+Price.of.Diesel.Gasoline..PG.,data = data2))

```
Analysis of Variance Table

Model 1: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters +
X..LPG.Vehicles..NLPG.
Model 2: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters +
X..LPG.Vehicles..NLPG. +
    Price.of.Diesel.Gasoline..PG.
  Res.Df        RSS Df Sum of Sq      F Pr(>F)
1     22 3.4505e+10
2     21 3.2381e+10  1 2.124e+09 1.3775 0.2537
```

anova(model,lm(Unleaded.Gasoline.Sale..UGS.~s2+s3+s4+Quarters+X..LPG.Vehicles..NLPG.+X..Unleaded.Gasoline.Vehicles..NUGV.,data = data2))

```
Analysis of Variance Table

Model 1: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters +
X..LPG.Vehicles..NLPG.
Model 2: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters +
X..LPG.Vehicles..NLPG. +
    X..Unleaded.Gasoline.Vehicles..NUGV.
  Res.Df        RSS Df Sum of Sq      F Pr(>F)
1     22 3.4505e+10
2     21 3.4071e+10  1 434605087 0.2679 0.6102
```

anova(model,lm(Unleaded.Gasoline.Sale..UGS.~s2+s3+s4+Quarters+X..LPG.Vehicles..NLPG.+X..of.Diesel.Gasoline.Vehicles..NDGV.,data = data2))

```
Analysis of Variance Table

Model 1: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters +
X..LPG.Vehicles..NLPG.
Model 2: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters +
X..LPG.Vehicles..NLPG. +
    X..of.Diesel.Gasoline.Vehicles..NDGV.
  Res.Df        RSS Df  Sum of Sq      F Pr(>F)
1     22 3.4505e+10
2     21 3.1873e+10  1 2632631294 1.7346  0.202
```

anova(model,lm(Unleaded.Gasoline.Sale..UGS.~s2+s3+s4+Quarters+NX..LPG.Vehicles..NLPG.LPG+GNP.Agriculture,data = data2))

```
Analysis of Variance Table
```

```
Model 1: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters +
X..LPG.Vehicles..NLPG.
Model 2: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters +
X..LPG.Vehicles..NLPG. +
    GNP.Agriculture
  Res.Df        RSS Df  Sum of Sq      F Pr(>F)
1     22 3.4505e+10
2     21 3.2293e+10  1 2212293139 1.4386 0.2437
```

anova(model,lm(Unleaded.Gasoline.Sale..UGS.~s2+s3+s4+Quarters+X..LPG.Vehicles..NLPG.+GNP.Commerce,data = data2))

```
Analysis of Variance Table

Model 1: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters +
X..LPG.Vehicles..NLPG.
Model 2: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters +
X..LPG.Vehicles..NLPG. +
    GNP.Commerce
  Res.Df        RSS Df Sum of Sq      F Pr(>F)
1     22 3.4505e+10
2     21 3.3559e+10  1 945922407 0.5919 0.4502
```

anova(model,lm(Unleaded.Gasoline.Sale..UGS.~s2+s3+s4+Quarters+X..LPG.Vehicles..NLPG.+GNP.Tota, data = data2))

```
Analysis of Variance Table

Model 1: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters +
X..LPG.Vehicles..NLPG.
Model 2: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters +
X..LPG.Vehicles..NLPG. +
    GNP.Total
  Res.Df        RSS Df  Sum of Sq      F Pr(>F)
1     22 3.4505e+10
2     21 3.2818e+10  1 1687207241 1.0796 0.3106
```

#PU has the lowest p value. So we update the model by adding PU to model.

model<-lm(Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters + X..LPG.Vehicles..NLPG.+Price.of.Unleaded.Gasoline..PU.,data = data2)

summary(model)

```
Call:
lm(formula = Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters +
```

```
    X..LPG.Vehicles..NLPG. + Price.of.Unleaded.Gasoline..PU.,
    data = data2)

Residuals:
   Min     1Q Median     3Q    Max
-68200 -23810  -3509  13864  90507

Coefficients:
                                  Estimate Std. Error t value Pr(>|t|)
(Intercept)                      9.780e+05  1.285e+05   7.610 1.82e-07
s2                               1.330e+05  2.077e+04   6.405 2.39e-06
s3                               2.883e+05  2.102e+04  13.715 5.97e-12
s4                               9.616e+04  2.072e+04   4.641  0.00014
Quarters                        -2.167e+04  3.456e+03  -6.271 3.21e-06
X..LPG.Vehicles..NLPG.           3.004e-01  9.027e-02   3.328  0.00319
Price.of.Unleaded.Gasoline..PU. -3.458e+02  2.179e+02  -1.587  0.12745

(Intercept)                     ***
s2                              ***
s3                              ***
s4                              ***
Quarters                        ***
X..LPG.Vehicles..NLPG.          **
Price.of.Unleaded.Gasoline..PU.
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 38300 on 21 degrees of freedom
  (4 observations deleted due to missingness)
Multiple R-squared:   0.95,    Adjusted R-squared:  0.9357
F-statistic: 66.47 on 6 and 21 DF,  p-value: 1.435e-12
```

PU looks like insignificant so we want to remove it from the model. But we should check it by removing NLPG from the model to see whether PU is significant or not.

anova(model,lm(Unleaded.Gasoline.Sale..UGS.~s2+s3+s4+Quarters+Price.of.Unleaded.Gasoline..PU.,data = data2))

```
Analysis of Variance Table

Model 1: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters +
X..LPG.Vehicles..NLPG. +
    Price.of.Unleaded.Gasoline..PU.
Model 2: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters +
Price.of.Unleaded.Gasoline..PU.
  Res.Df       RSS Df  Sum of Sq      F   Pr(>F)
1     21 3.081e+10
2     22 4.706e+10 -1 -1.625e+10 11.076 0.003193 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

#p value is so small so we can neither remove NLPG nor add PU to the model.

#Now we will add the variable with second smallest P value obtained from ANOVA which is RNUV

model<-lm(Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters + X..LPG.Vehicles..NLPG.+RNUV,data = data2)

summary(model)

```
Call:
lm(formula = Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters + 
    X..LPG.Vehicles..NLPG. + RNUV, data = data2)

Residuals:
   Min     1Q Median     3Q    Max 
-72932 -18011   1721  18130  86399 

Coefficients:
                         Estimate Std. Error t value Pr(>|t|)    
(Intercept)             8.799e+05  9.042e+04   9.731 3.12e-09 ***
s2                      1.184e+05  2.231e+04   5.308 2.91e-05 ***
s3                      2.708e+05  2.281e+04  11.870 8.90e-11 ***
s4                      8.355e+04  2.381e+04   3.509  0.00209 ** 
Quarters               -2.049e+04  4.197e+03  -4.882 7.93e-05 ***
X..LPG.Vehicles..NLPG.  2.148e-01  1.233e-01   1.742  0.09618 .  
RNUV                    2.377e+06  1.797e+06   1.323  0.20006    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 38940 on 21 degrees of freedom
  (4 observations deleted due to missingness)
Multiple R-squared:  0.9483,	Adjusted R-squared:  0.9335 
F-statistic: 64.19 on 6 and 21 DF,  p-value: 2.027e-12
```

#RNUV seems insignificant so we should check to remove NLPG with anova.

anova(model,lm(Unleaded.Gasoline.Sale..UGS.~s2+s3+s4+Quarters+RNUV,data = data2))

```
Analysis of Variance Table

Model 1: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters + 
X..LPG.Vehicles..NLPG. + 
    RNUV
Model 2: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters + RNUV
  Res.Df        RSS Df   Sum of Sq      F  Pr(>F)  
1     21 3.1850e+10                                
2     22 3.6452e+10 -1 -4601362376 3.0338 0.09618 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

#P value is large enough to remove NLPG.

model<-lm(Unleaded.Gasoline.Sale..UGS.~s2+s3+s4+Quarters+RNUV,data = data2)

summary(model)

```
Analysis of Variance Table

Model 1: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters + 
X..LPG.Vehicles..NLPG. + 
    RNUV
Model 2: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters + RNUV
  Res.Df        RSS Df   Sum of Sq      F  Pr(>F)
1     21 3.1850e+10
2     22 3.6452e+10 -1 -4601362376 3.0338 0.09618 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> model<-lm(Unleaded.Gasoline.Sale..UGS.~s2+s3+s4+Quarters+RNUV,data = data2)
> summary(model)

Call:
lm(formula = Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters + 
    RNUV, data = data2)

Residuals:
   Min     1Q Median     3Q    Max 
-63231 -15788  -4559  27431  83703 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 1033232.2    21482.5  48.096  < 2e-16 ***
s2           106743.1    22241.4   4.799 8.57e-05 ***
s3           257206.5    22409.0  11.478 9.30e-11 ***
s4            69115.6    23330.3   2.962  0.00719 ** 
Quarters     -13357.6      962.5 -13.878 2.32e-12 ***
RNUV        4507922.6  1375124.6   3.278  0.00344 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 40710 on 22 degrees of freedom
  (4 observations deleted due to missingness)
Multiple R-squared:  0.9408,	Adjusted R-squared:  0.9274 
F-statistic: 69.95 on 5 and 22 DF,  p-value: 9.243e-13
```

#RNUV is significant. We will look the ANOVA results to decide if we update the model with any variable.

anova(model,lm(Unleaded.Gasoline.Sale..UGS.~s2+s3+s4+Quarters+RNUV+Price.of.Diesel
.Gasoline..PG.,data = data2))

```
Analysis of Variance Table

Model 1: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters + RNUV
Model 2: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters + RNUV +
    Price.of.Diesel.Gasoline..PG.
  Res.Df        RSS Df  Sum of Sq      F  Pr(>F)
1     22 3.6452e+10
2     21 2.8679e+10  1 7772333539 5.6911 0.02655 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

anova(model,lm(Unleaded.Gasoline.Sale..UGS.~s2+s3+s4+Quarters+RNUV+X..Unleaded.Gasoline.Vehicles..NUGV. ,data = data2))

```
Analysis of Variance Table

Model 1: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters + RNUV
Model 2: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters + RNUV +
    X..Unleaded.Gasoline.Vehicles..NUGV.
  Res.Df        RSS Df Sum of Sq      F Pr(>F)
1     22 3.6452e+10
2     21 3.6431e+10  1  20673819 0.0119 0.9141
```

anova(model,lm(Unleaded.Gasoline.Sale..UGS.~s2+s3+s4+Quarters+RNUV+X..of.Diesel.Gasoline.Vehicles..NDGV.,data = data2))

```
Analysis of Variance Table

Model 1: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters + RNUV
Model 2: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters + RNUV +
    X..of.Diesel.Gasoline.Vehicles..NDGV.
  Res.Df        RSS Df  Sum of Sq      F Pr(>F)
1     22 3.6452e+10
2     21 3.4784e+10  1 1668140319 1.0071  0.327
```

anova(model,lm(Unleaded.Gasoline.Sale..UGS.~s2+s3+s4+Quarters+RNUV+GNP.Agriculture,data = data2))

```
Analysis of Variance Table
```

```
Model 1: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters + RNUV
Model 2: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters + RNUV +
    GNP.Agriculture
  Res.Df        RSS Df  Sum of Sq      F Pr(>F)
1     22 3.6452e+10
2     21 3.4987e+10  1 1465274695 0.8795  0.359
```

anova(model,lm(Unleaded.Gasoline.Sale..UGS.~s2+s3+s4+Quarters+RNUV+GNP.Commerce,data = data2))

```
Analysis of Variance Table

Model 1: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters + RNUV
Model 2: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters + RNUV +
    GNP.Commerce
  Res.Df        RSS Df  Sum of Sq      F Pr(>F)
1     22 3.6452e+10
2     21 3.4978e+10  1 1473618126 0.8847 0.3576
```

anova(model,lm(Unleaded.Gasoline.Sale..UGS.~s2+s3+s4+Quarters+RNUV+GNP.Total,data = data2))

```
Analysis of Variance Table

Model 1: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters + RNUV
Model 2: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters + RNUV +
    GNP.Total
  Res.Df        RSS Df Sum of Sq   F Pr(>F)
1     22 3.6452e+10
2     21 3.5938e+10  1 513355673 0.3 0.5897
```

#PG has the Smallest p value. So we should add the PG to the model

model<-lm(Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters + Price.of.Diesel.Gasoline..PG.+RNUV,data = data2)

summary(model)

```
Call:
lm(formula = Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters + 
    Price.of.Diesel.Gasoline..PG. + RNUV, data = data2)

Residuals:
   Min     1Q Median     3Q    Max 
-50232 -21999   -259  12073  81102 

Coefficients:
                               Estimate Std. Error t value Pr(>|t|)    
(Intercept)                   1240074.8    88870.8  13.954 4.30e-12 ***
s2                             109011.7    20214.8   5.393 2.39e-05 ***
s3                             263712.8    20526.6  12.847 2.05e-11 ***
s4                              63616.1    21306.2   2.986  0.00705 ** 
Quarters                       -10138.3     1607.7  -6.306 2.97e-06 ***
Price.of.Diesel.Gasoline..PG.    -656.3      275.1  -2.386  0.02655 *  
RNUV                          4952340.7  1262268.7   3.923  0.00078 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 36960 on 21 degrees of freedom
  (4 observations deleted due to missingness)
Multiple R-squared:  0.9534,	Adjusted R-squared:  0.9401 
F-statistic: 71.67 on 6 and 21 DF,  p-value: 6.808e-13
```

#PG is significant but we still want to try removing RNUV from the model to see if it is going better.

anova(model,lm(Unleaded.Gasoline.Sale..UGS.~s2+s3+s4+Quarters+Price.of.Diesel.Gasoline..PG.,data = data2))

```
Analysis of Variance Table

Model 1: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters + 
Price.of.Diesel.Gasoline..PG. + 
    RNUV
Model 2: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters + 
Price.of.Diesel.Gasoline..PG.
  Res.Df        RSS Df   Sum of Sq      F    Pr(>F)    
1     21 2.8679e+10                                    
2     22 4.9701e+10 -1 -2.1022e+10 15.393 0.0007802 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

#P value is so small so we can not remove RNUV. Now we do anova to see if we can add any more variables to the model.

anova(model,lm(Unleaded.Gasoline.Sale..UGS.~s2+s3+s4+Quarters+Price.of.Diesel.Gasoline..PG.+RNUV+X..Unleaded.Gasoline.Vehicles..NUGV. ,data = data2))

```
Analysis of Variance Table

Model 1: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters + 
Price.of.Diesel.Gasoline..PG. + 
    RNUV
Model 2: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters + 
Price.of.Diesel.Gasoline..PG. + 
    RNUV + X..Unleaded.Gasoline.Vehicles..NUGV.
  Res.Df        RSS Df Sum of Sq  F Pr(>F)
1     21 2.8679e+10                       
2     20 2.8679e+10  1    432.15  0 0.9996
```

anova(model,lm(Unleaded.Gasoline.Sale..UGS.~s2+s3+s4+Quarters+Price.of.Diesel.Gasoline..PG.+RNUV+X..of.Diesel.Gasoline.Vehicles..NDGV.,data = data2))

```
Analysis of Variance Table

Model 1: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters + 
Price.of.Diesel.Gasoline..PG. + 
    RNUV
Model 2: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters + 
Price.of.Diesel.Gasoline..PG. + 
    RNUV + X..of.Diesel.Gasoline.Vehicles..NDGV.
  Res.Df        RSS Df  Sum of Sq      F Pr(>F)
1     21 2.8679e+10                            
2     20 2.6662e+10  1 2017243108 1.5132 0.2329
```

anova(model,lm(Unleaded.Gasoline.Sale..UGS.~s2+s3+s4+Quarters+Price.of.Diesel.Gasoline..PG.+RNUV+GNP.Agriculture,data = data2))

```
Analysis of Variance Table

Model 1: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters + 
Price.of.Diesel.Gasoline..PG. + 
    RNUV
Model 2: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters + 
Price.of.Diesel.Gasoline..PG. + 
    RNUV + GNP.Agriculture
  Res.Df        RSS Df Sum of Sq      F Pr(>F)
1     21 2.8679e+10                           
2     20 2.8326e+10  1 353147552 0.2493  0.623
```

anova(model,lm(Unleaded.Gasoline.Sale..UGS.~s2+s3+s4+Quarters+Price.of.Diesel.Gasoline..PG.+RNUV+GNP.Commerce,data = data2))

```
Analysis of Variance Table

Model 1: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters +
Price.of.Diesel.Gasoline..PG. +
    RNUV
Model 2: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters +
Price.of.Diesel.Gasoline..PG. +
    RNUV + GNP.Commerce
  Res.Df        RSS Df  Sum of Sq      F Pr(>F)
1     21 2.8679e+10
2     20 2.6347e+10  1 2332493838 1.7706 0.1983
```

anova(model,lm(Unleaded.Gasoline.Sale..UGS.~s2+s3+s4+Quarters+Price.of.Diesel.Gasoline..PG.+RNUV+GNP.Total ,data = data2))

```
Analysis of Variance Table

Model 1: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters +
Price.of.Diesel.Gasoline..PG. +
    RNUV
Model 2: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters +
Price.of.Diesel.Gasoline..PG. +
    RNUV + GNP.Total
  Res.Df        RSS Df Sum of Sq      F Pr(>F)
1     21 2.8679e+10
2     20 2.8115e+10  1 564508508 0.4016 0.5335
```

#We added the variable with smallest p value, GNPC to the model.

model<-lm(Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters + Price.of.Diesel.Gasoline..PG.+RNUV+GNP.Commerce,data = data2)

summary(model)

```
Call:
lm(formula = Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters +
    Price.of.Diesel.Gasoline..PG. + RNUV + GNP.Commerce, data = data2)

Residuals:
   Min     1Q Median     3Q    Max
-55218 -18758   2045  14876  78354

Coefficients:
```

```
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                    1.074e+06  1.521e+05   7.062 7.56e-07 ***
s2                             6.306e+04  3.984e+04   1.583 0.129115
s3                             1.277e+05  1.042e+05   1.226 0.234388
s4                            -3.233e+03  5.442e+04  -0.059 0.953214
Quarters                      -1.321e+04  2.800e+03  -4.720 0.000131 ***
Price.of.Diesel.Gasoline..PG. -6.957e+02  2.718e+02  -2.560 0.018684 *
RNUV                           3.652e+06  1.578e+06   2.314 0.031424 *
GNP.Commerce                   5.851e-02  4.397e-02   1.331 0.198281
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 36300 on 20 degrees of freedom
  (4 observations deleted due to missingness)
Multiple R-squared:  0.9572,    Adjusted R-squared:  0.9423
F-statistic: 63.94 on 7 and 20 DF,  p-value: 2.633e-12
```

#GNPC is insignificant yet we still want to check if removing the others will make any difference

#Removing RNUV

anova(model,lm(Unleaded.Gasoline.Sale..UGS.~s2+s3+s4+Quarters+Price.of.Diesel.Gasoline..PG.+GNP.Commerce,data = data2))

```
Analysis of Variance Table

Model 1: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters + 
Price.of.Diesel.Gasoline..PG. + 
    RNUV + GNP.Commerce
Model 2: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters + 
Price.of.Diesel.Gasoline..PG. + 
    GNP.Commerce
  Res.Df        RSS Df   Sum of Sq      F  Pr(>F)  
1     20 2.6347e+10                                
2     21 3.3401e+10 -1 -7053624722 5.3544 0.03142 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

we can not remove RNUV from the model with this smallest p value

#Removing PG

```
Analysis of Variance Table
```

```
Model 1: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters +
Price.of.Diesel.Gasoline..PG. +
    RNUV + GNP.Commerce
Model 2: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters + RNUV +
    GNP.Commerce
  Res.Df        RSS Df   Sum of Sq      F  Pr(>F)
1     20 2.6347e+10
2     21 3.4978e+10 -1 -8631209251 6.5519 0.01868 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

we can not remove PG from the model with this smallest p value

#Removing both

anova(model,lm(Unleaded.Gasoline.Sale..UGS.~s2+s3+s4+Quarters+GNP.Commerce,data = data2))

```
Analysis of Variance Table

Model 1: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters +
Price.of.Diesel.Gasoline..PG. +
    RNUV + GNP.Commerce
Model 2: Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters +
GNP.Commerce
  Res.Df        RSS Df   Sum of Sq      F  Pr(>F)
1     20 2.6347e+10
2     22 4.1322e+10 -2 -1.4975e+10 5.6836 0.01111 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

we can not remove PG and RNUV from the model with this smallest p value.

#We did not add GNPC to the model. We should stop there because GNPC which has the best ANOVA result is not significant. So we don't have to try other ones.

model<-lm(Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters + RNUV+Price.of.Diesel.Gasoline..PG.,data = data2)

summary(model)

```
Call:
lm(formula = Unleaded.Gasoline.Sale..UGS. ~ s2 + s3 + s4 + Quarters +
    RNUV + Price.of.Diesel.Gasoline..PG., data = data2)
```

```
Residuals:
   Min     1Q Median     3Q    Max 
-50232 -21999   -259  12073  81102 

Coefficients:
                               Estimate Std. Error t value Pr(>|t|)    
(Intercept)                   1240074.8    88870.8  13.954 4.30e-12 ***
s2                             109011.7    20214.8   5.393 2.39e-05 ***
s3                             263712.8    20526.6  12.847 2.05e-11 ***
s4                              63616.1    21306.2   2.986  0.00705 ** 
Quarters                       -10138.3     1607.7  -6.306 2.97e-06 ***
RNUV                          4952340.7  1262268.7   3.923  0.00078 ***
Price.of.Diesel.Gasoline..PG.    -656.3      275.1  -2.386  0.02655 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 36960 on 21 degrees of freedom
  (4 observations deleted due to missingness)
Multiple R-squared:  0.9534,	Adjusted R-squared:  0.9401 
F-statistic: 71.67 on 6 and 21 DF,  p-value: 6.808e-13
```

#All variables are significant. We have a trend and seasonality variables. R-square values are 0.9534 and 0.9401 and p value is 6.808e-13 which are so satisfying.

#We will move on by checking model assumptions.

UGS_model<- model

library(car)

library(lmtest)

vif(UGS_model)

```
                           s2                            s3 
                     1.570896                      1.619730 
                           s4                      Quarters 
                     1.745089                      3.457702 
                         RNUV Price.of.Diesel.Gasoline..PG. 
                     1.158310                      3.486745 
```

#We performed Variance Inflation Factor test to see if there is multicollinearity in our model.

#since there is no value high enough to indicate there exist a multicollinearity in our model

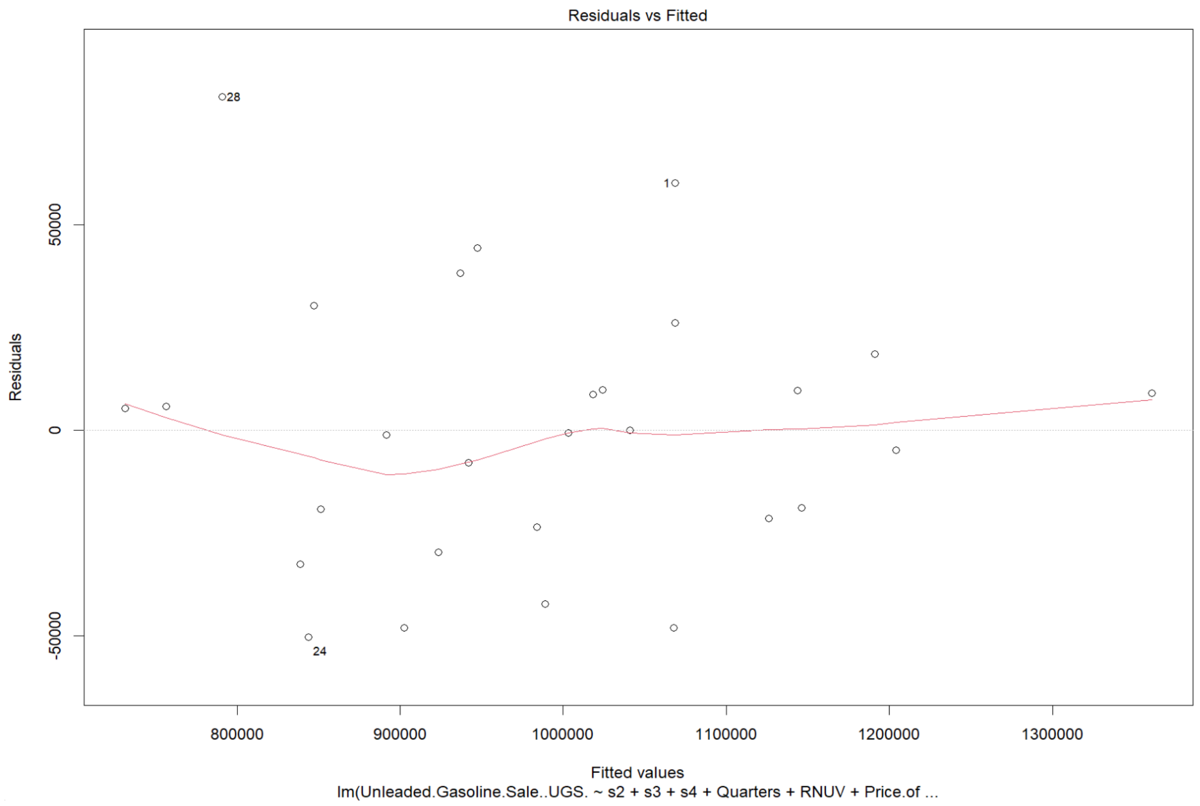
#Then we performed Dublin-Watson test to see if there is autocorrelation

dwtest(UGS_model)

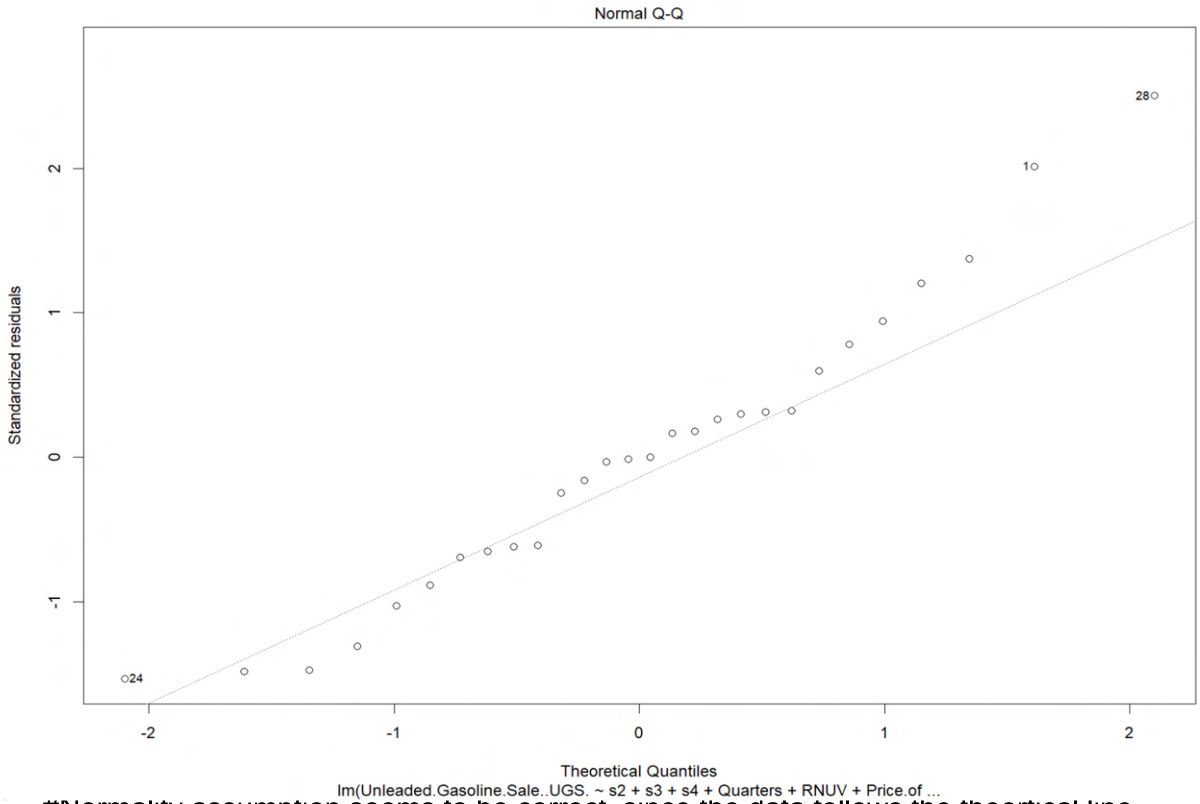
```
Durbin-Watson test

data:  UGS_model
DW = 2.1835, p-value = 0.5329
alternative hypothesis: true autocorrelation is greater than 0
```

#dw value is so close to 2 and p value is 0.5329. Based on these results, we do not have sufficient evidence to reject the null hypothesis that the true autocorrelation is equal to 0.

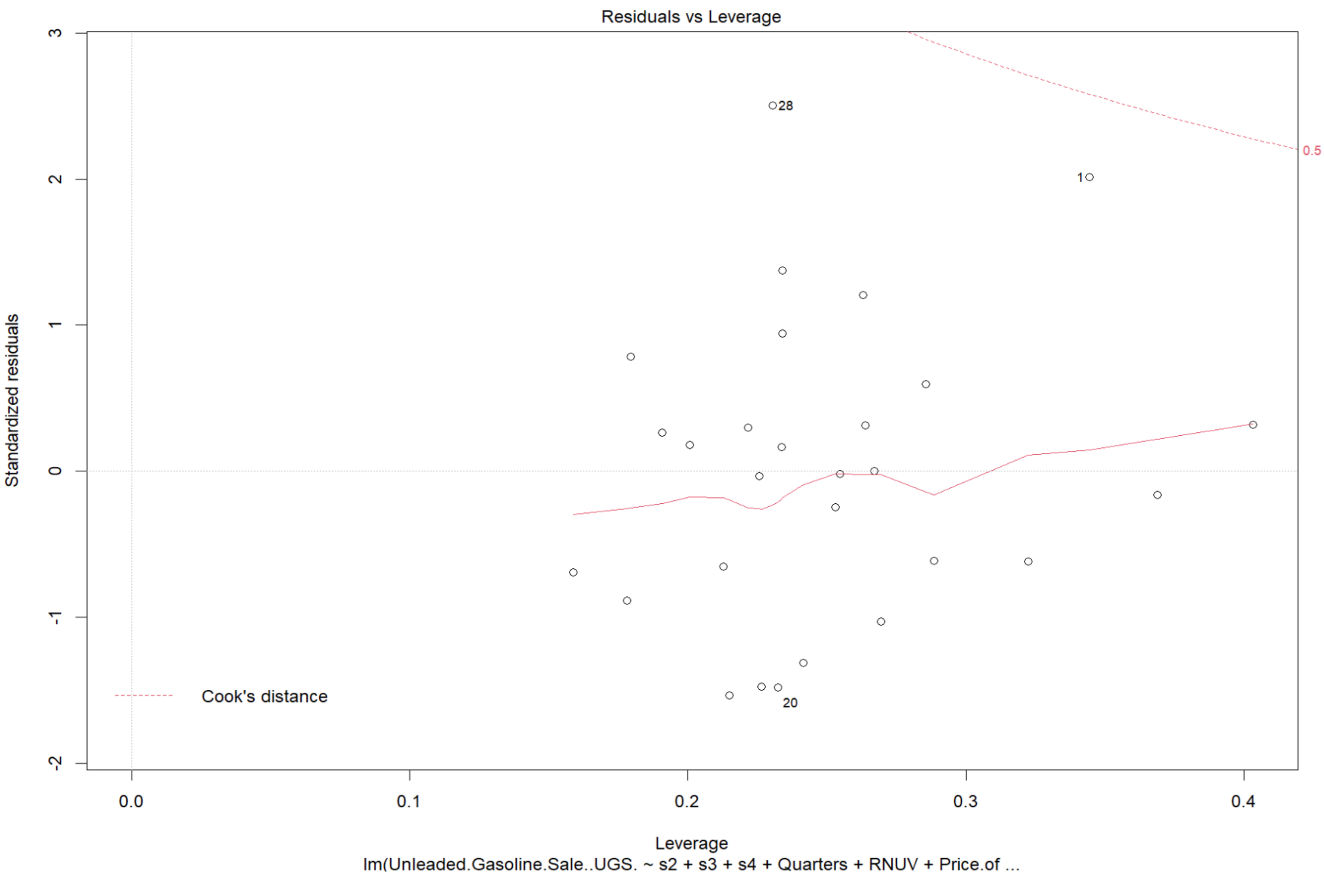
plot(UGS_model)

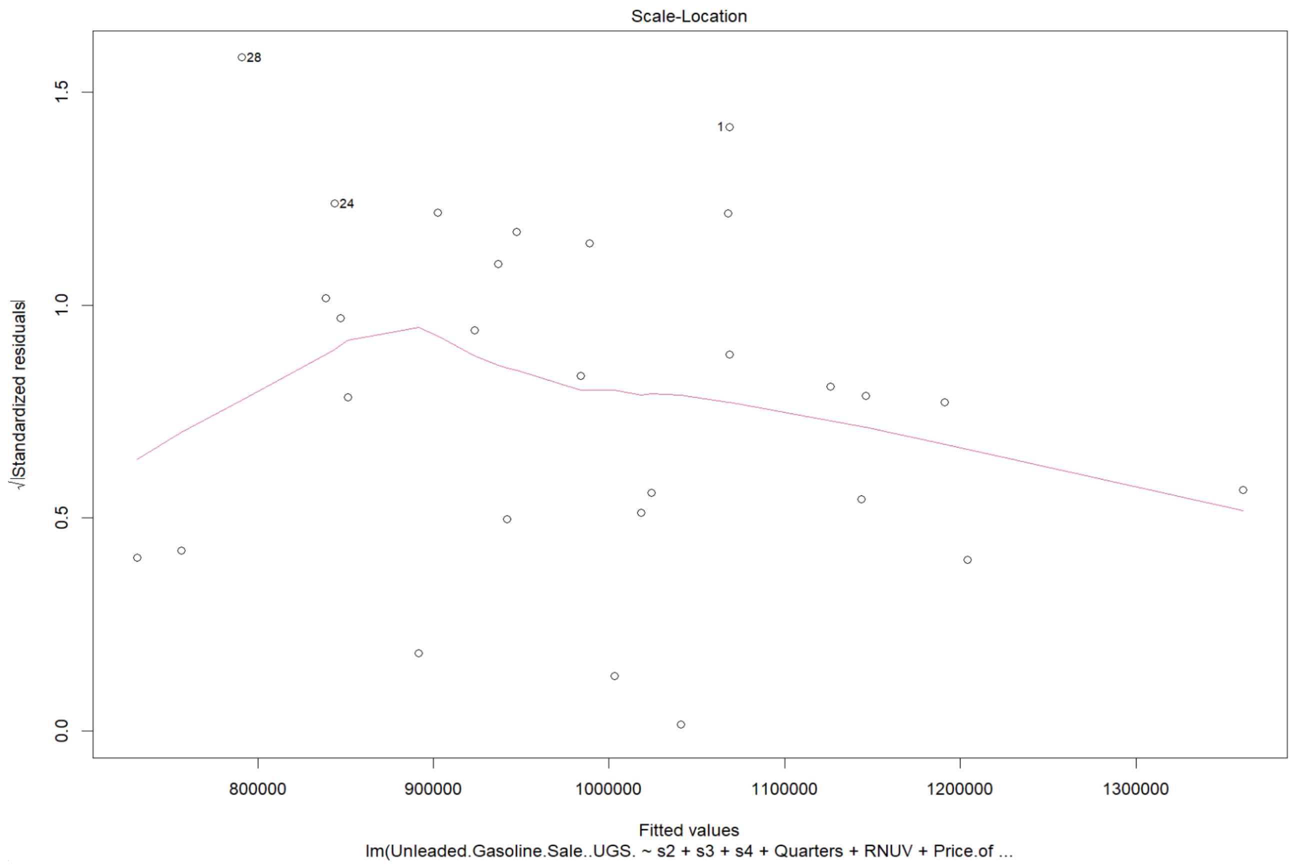


#When we look at the plot we can say that our linearity assumption holds.

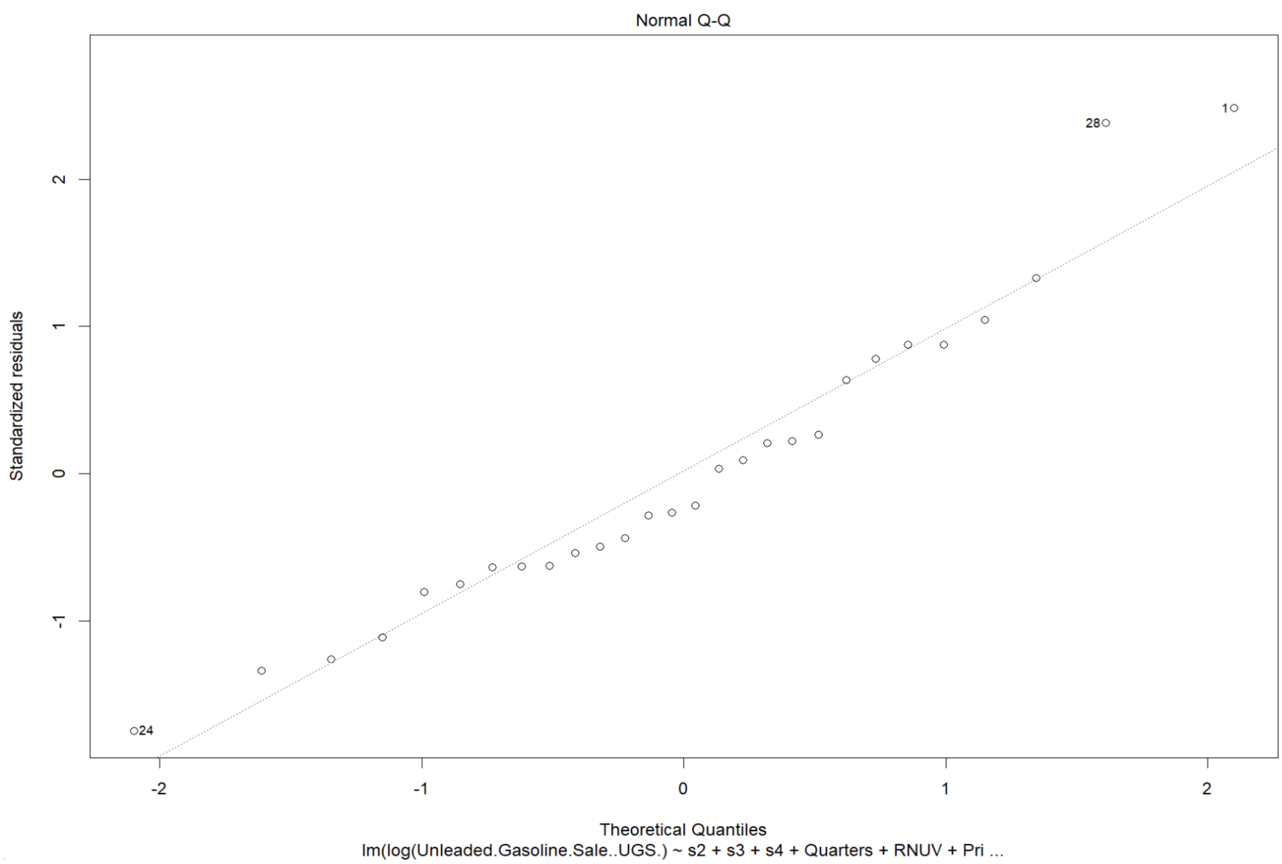


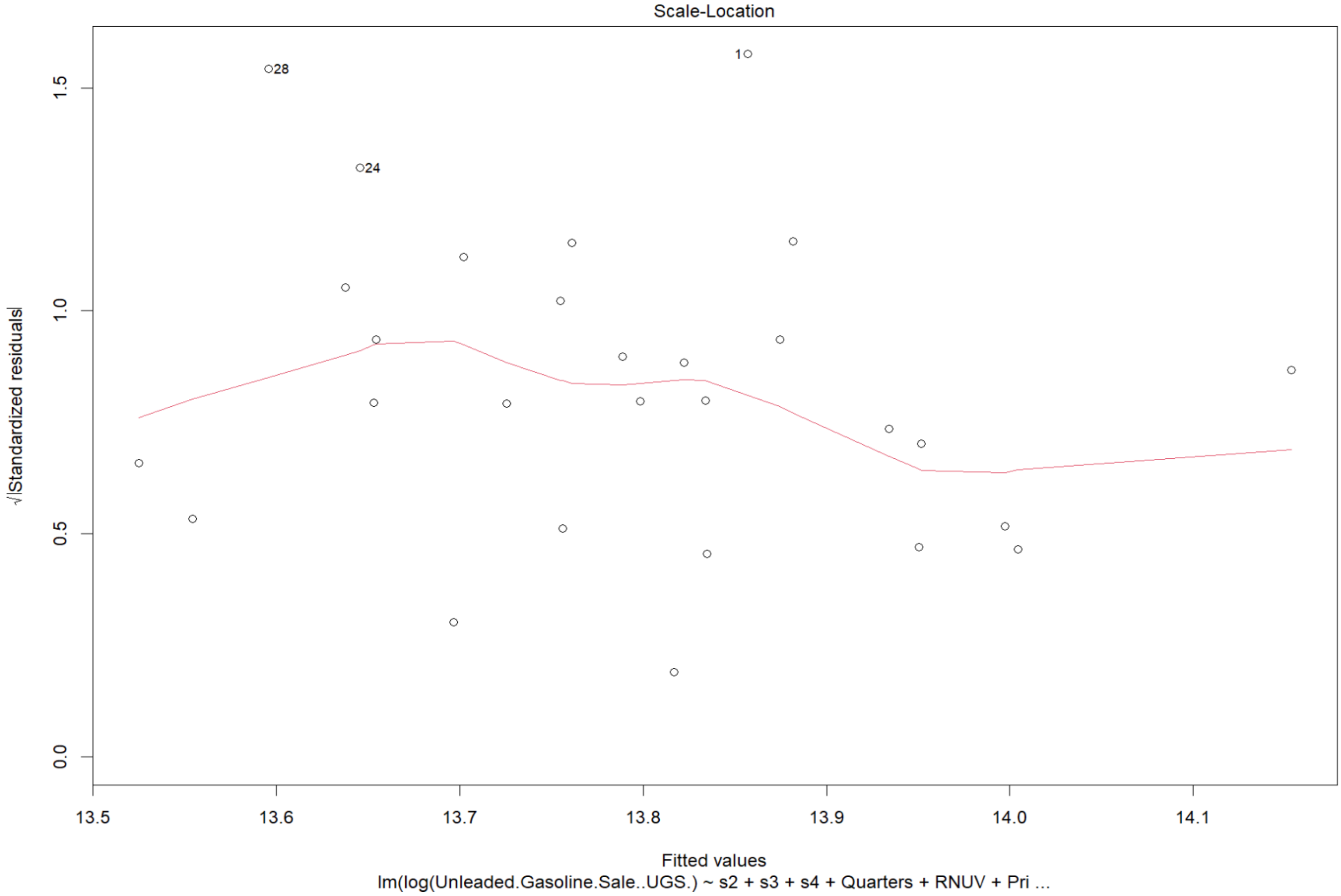
#Normality assumption seems to be correct, since the data follows the theortical line.

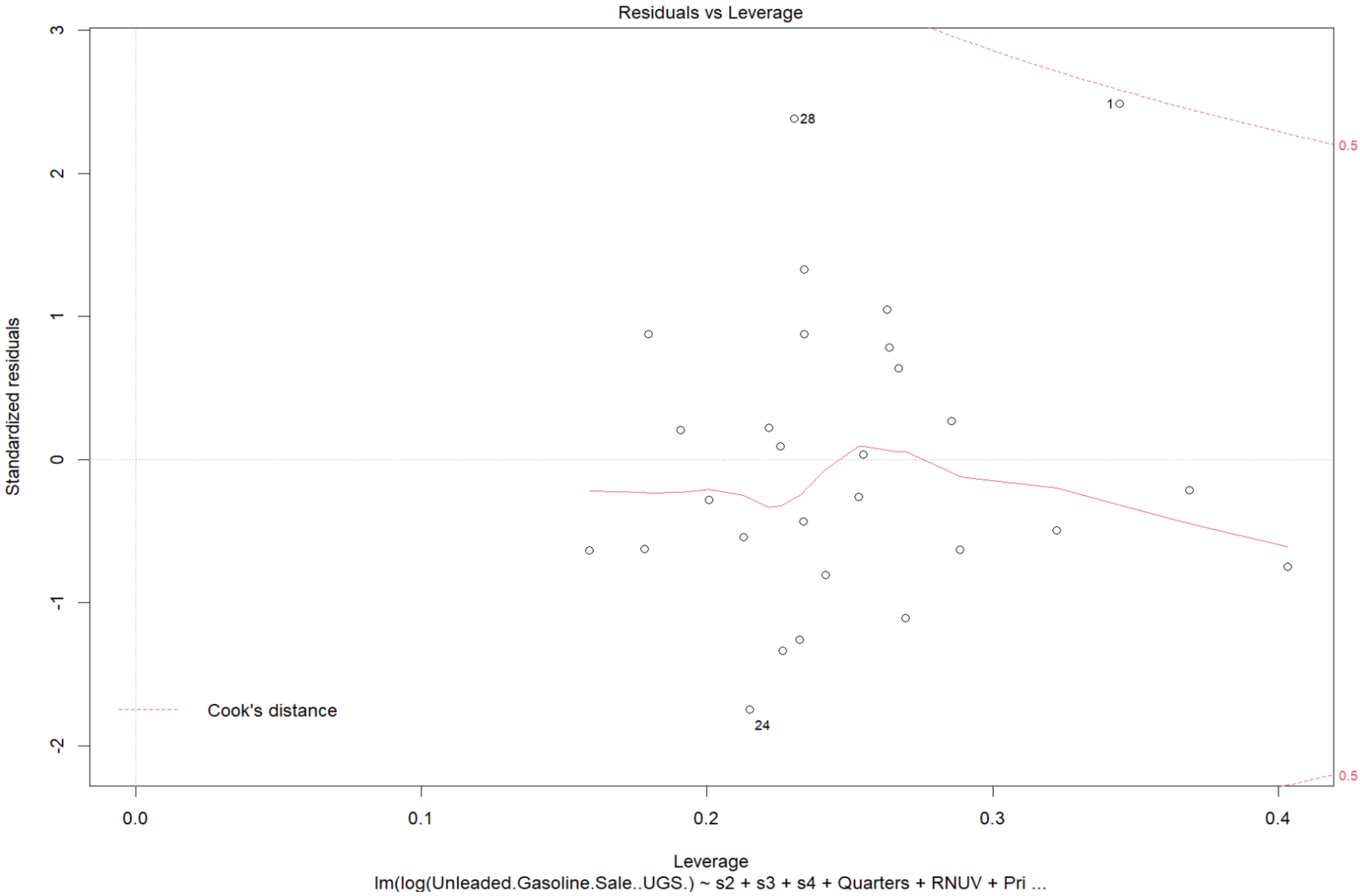


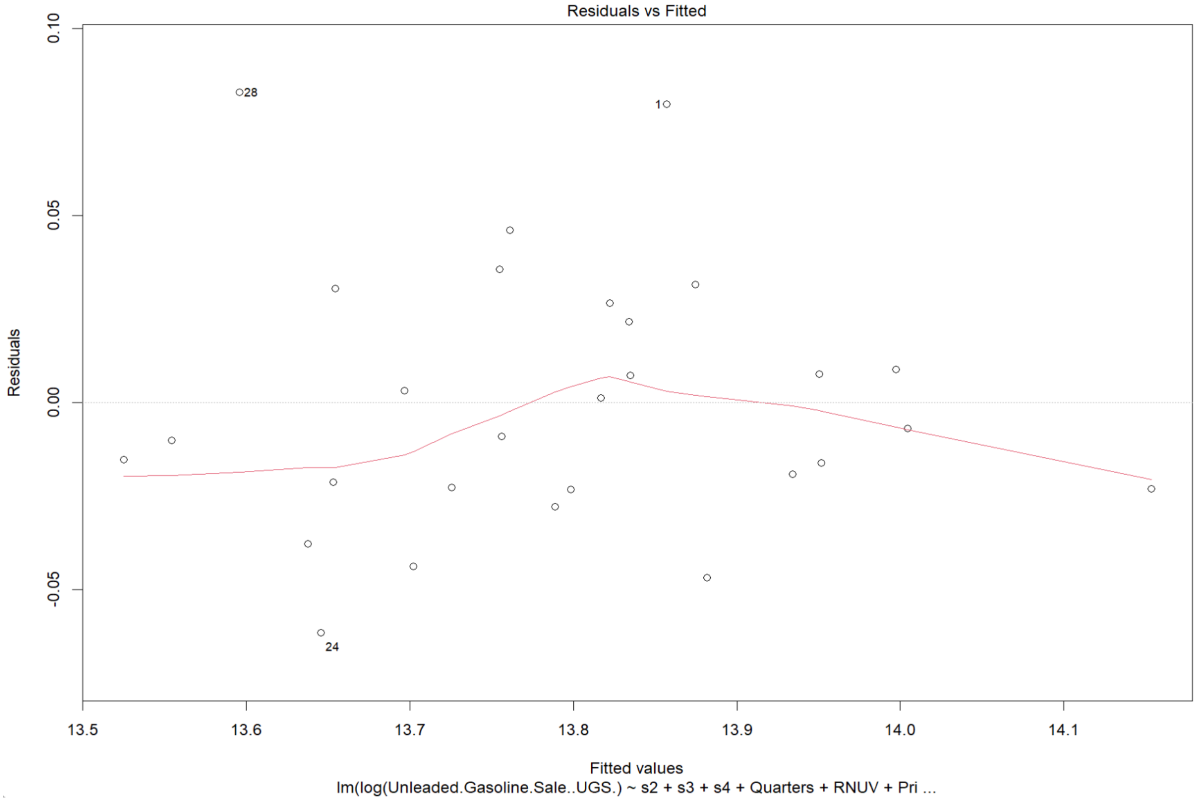



```
model<-lm(log(Unleaded.Gasoline.Sale..UGS.) ~ s2 + s3 + s4 + Quarters +
RNUV+Price.of.Diesel.Gasoline..PG.,data=data2)

plot(model)
```

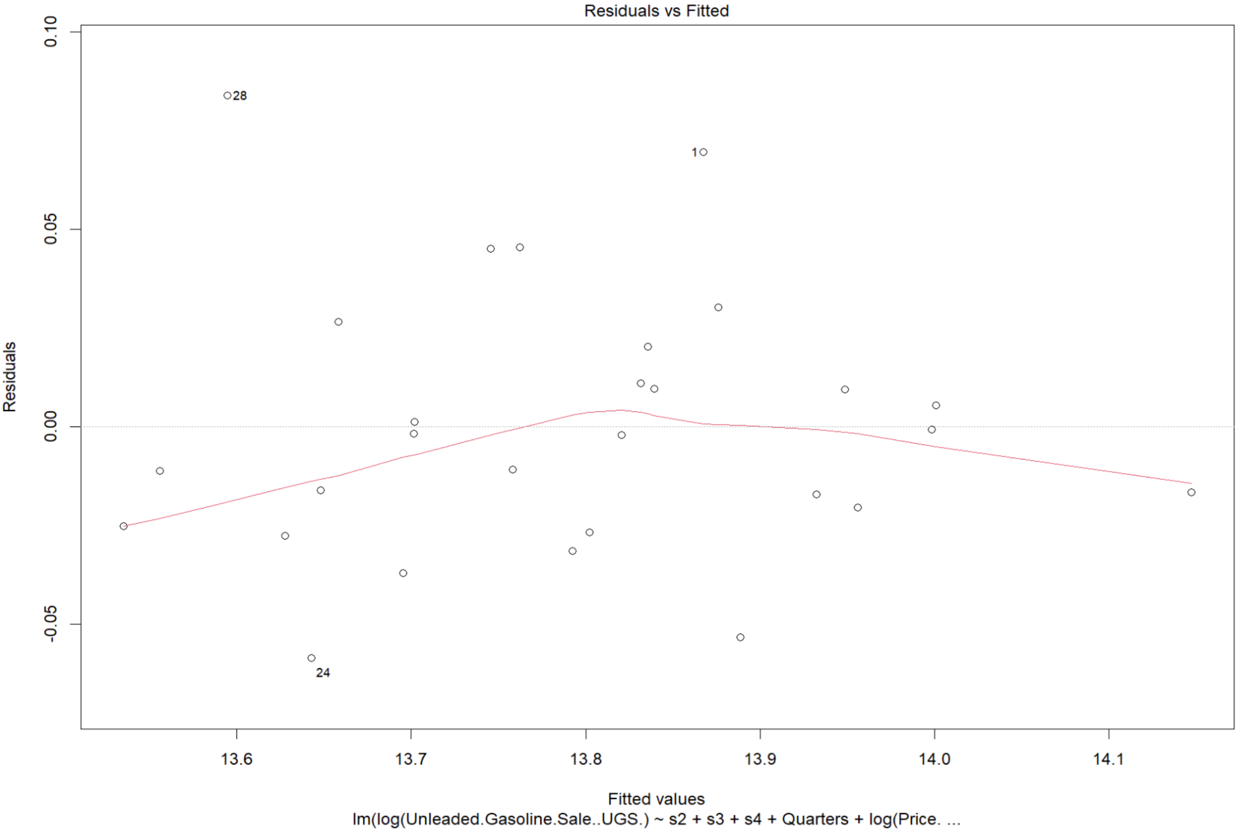



Scale-Location
√|Standardized residuals|
Fitted values
lm(log(Unleaded.Gasoline.Sale..UGS.) ~ s2 + s3 + s4 + Quarters + RNUV + Pri ...
28
24
1

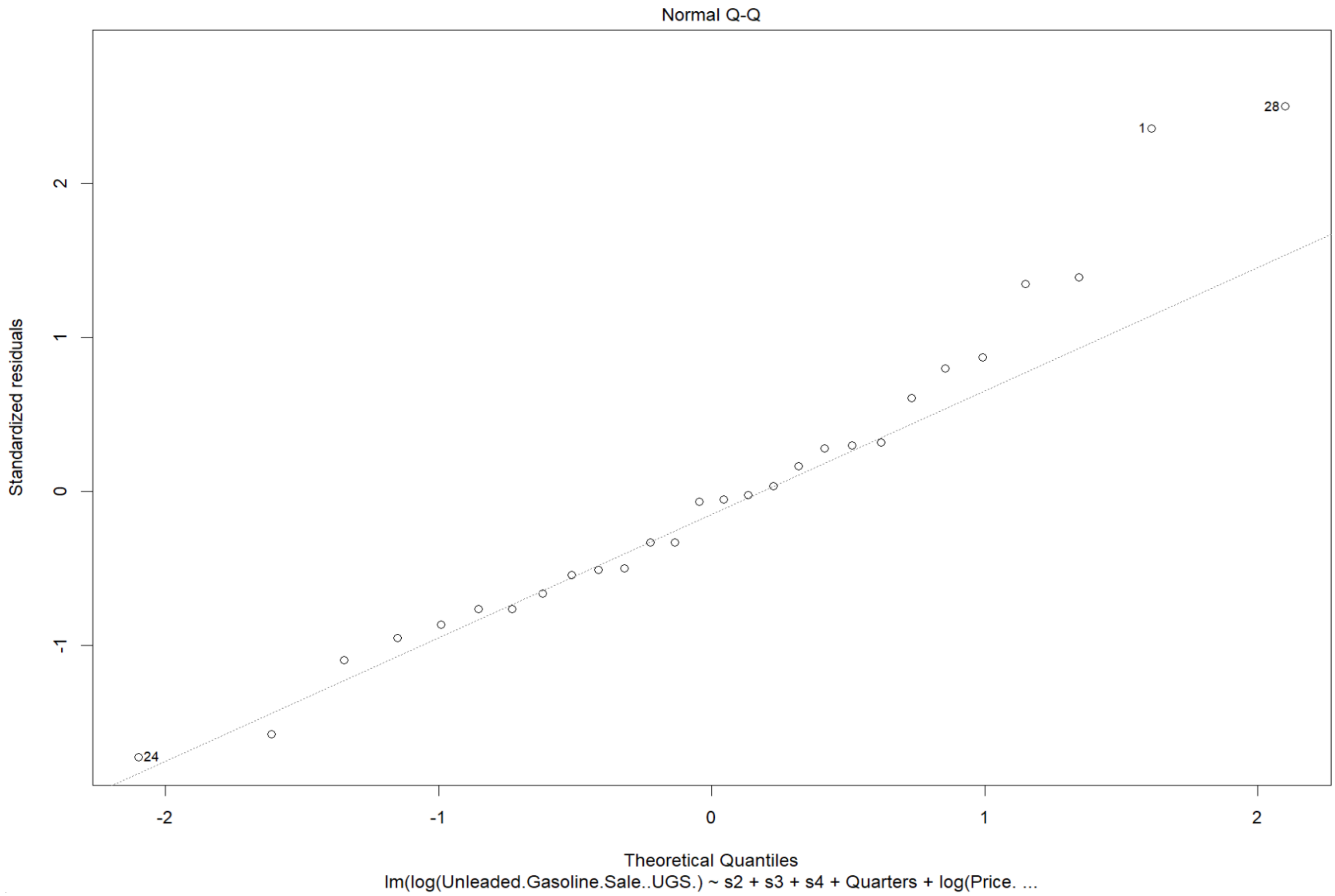

Residuals vs Leverage
Standardized residuals
Leverage
lm(log(Unleaded.Gasoline.Sale..UGS.) ~ s2 + s3 + s4 + Quarters + RNUV + Pri ...
Cook's distance
0.5
28
24
1



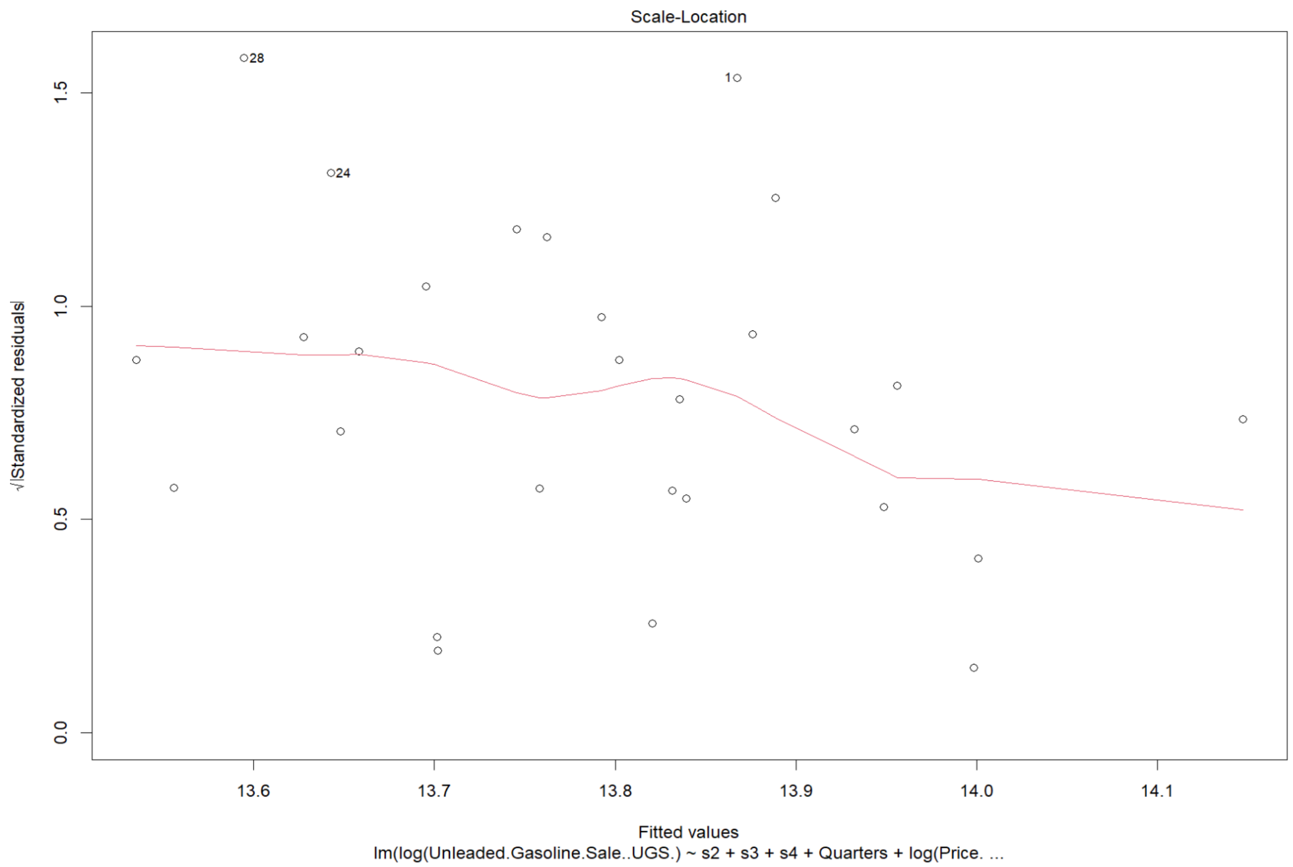
#The average value of standardized residuals appears to remain consistent across the range of fitted values, indicating that the assumption of constant variance is fulfilled. Linearity assumption does not maintain we should take the log of the explaining variables.



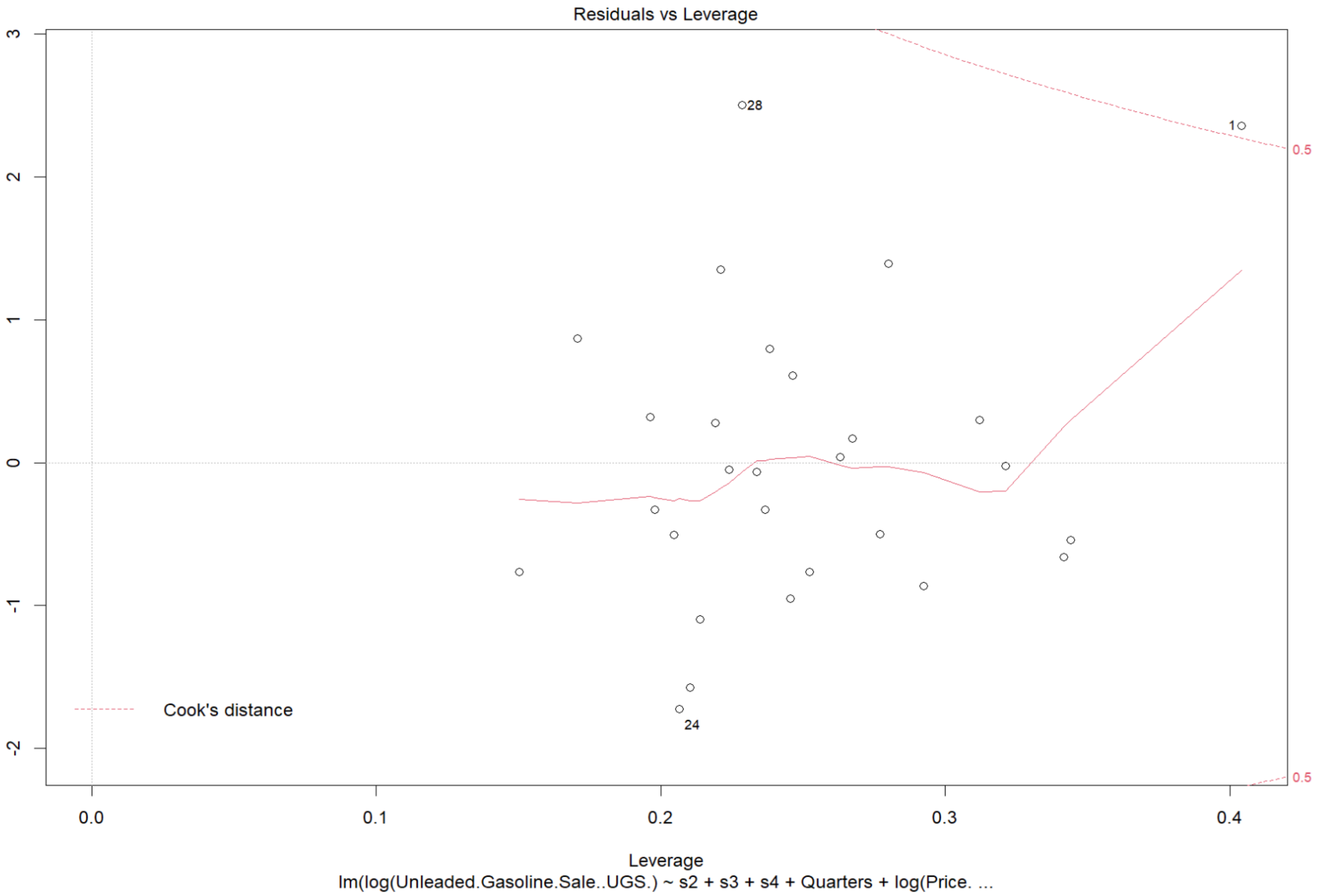
#Linearity assumption is maintained.



#Normality assumption is maintained.



#constant variance is maintained



#After that we should check the correlation between the explanatory variable and the residuals.

```
y_predict<-predict(model,data2)

y_act<-log(data$'Unleaded Gasoline Sale (UGS)')

residuals<- y_act - y_predict

cor(data$'Unleaded Gasoline Sale (UGS)',residuals,  use = "complete.obs")
```

```
[1] 0.1966006
```

#Correlation between explanatory variables and residuals is 0.1966 which is insignificant. So our model satisfied the all model assumptions.

#We will make predicts for 2007 based on our model

```
predict2007Q1 <-
data.frame(s2=0,s3=0,s4=0,Quarters=29,Price.of.Diesel.Gasoline..PG.=449.1909,RNUV=0.007386855)
predict(model,predict2007Q1,level = 0.9,interval = "prediction")
```

```
       fit      lwr      upr
1 13.49022 13.41497 13.56547
```

```
exp(13.49022)
```

```
[1] 722317.4
```

```
predict2007Q2 <-
data.frame(s2=1,s3=0,s4=0,Quarters=30,Price.of.Diesel.Gasoline..PG.=449.1909,RNUV=0.010591663)
predict(model,predict2007Q2,level = 0.9,interval = "prediction")
```

```
      fit      lwr      upr
1 13.6137 13.53811 13.68929
```

```
exp(13.6137)
```

```
[1] 817249.7
```

```
predict2007Q3 <-
data.frame(s2=0,s3=1,s4=0,Quarters=31,Price.of.Diesel.Gasoline..PG.=449.1909,RNUV=0.010077553)
predict(model,predict2007Q3,level = 0.9,interval = "prediction")
```

```
       fit      lwr      upr
1 13.74766 13.67124 13.82409
```

```
exp(13.74766)
```

```
[1] 934400.1
```

```
predict2007Q4 <-
data.frame(s2=0,s3=0,s4=1,Quarters=32,Price.of.Diesel.Gasoline..PG.=449.1909,RNUV=0.012371491)
predict(model,predict2007Q4,level = 0.9,interval = "prediction")
```

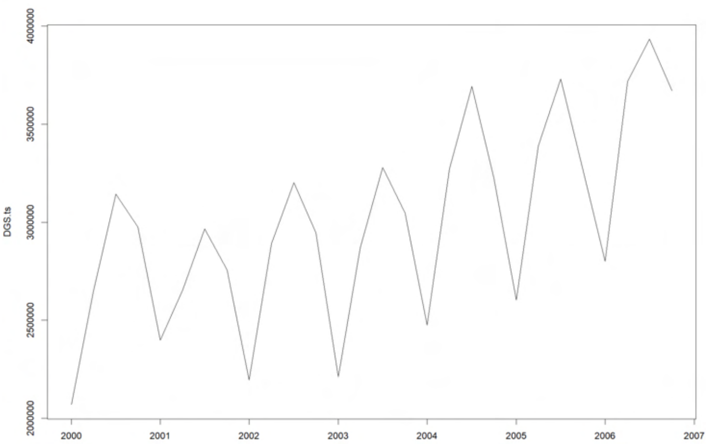
```
       fit      lwr      upr
1 13.55011 13.47487 13.62535
```

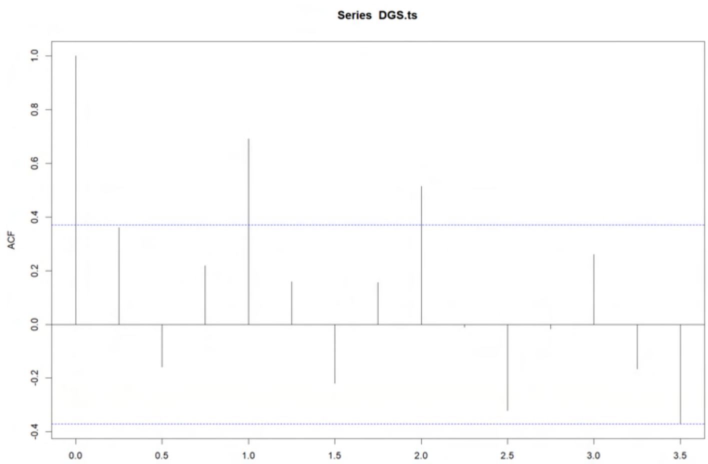
```
exp(13.55011)
```

```
[1] 766898.7
```

DGS

```
data<-read_excel("IE360-ProjectData.xlsx",col_names = TRUE)
DGS.ts<- ts(data$`Diesel Gasoline Sale (DGS)`,start=c(2000,1),end=c(2006,4),frequency=4)
plot(DGS.ts)
acf(DGS.ts)
```






```
#data is not stationary
#data seems seasonal
#We should add the trend and seasonality variable to the model
```

```
cor(data[sapply(data, is.numeric)],use = "complete.obs")
```

	Unleaded Gasoline Sale (UGS)
Unleaded Gasoline Sale (UGS)	1.00000000
Diesel Gasoline Sale (DGS)	0.04670539
RNUV	0.26498942
# LPG Vehicles (NLPG)	-0.62890683
Price of Unleaded Gasoline (PU)	-0.48993252
Price of Diesel Gasoline (PG)	-0.54812079
# Unleaded Gasoline Vehicles (NUGV)	-0.64862885
# of Diesel Gasoline Vehicles (NDGV)	-0.52526929
GNP Agriculture	0.58815962
GNP Commerce	0.20843693
GNP Total	0.19348598

	Diesel Gasoline Sale (DGS)	RNUV
Unleaded Gasoline Sale (UGS)	0.04670539	0.26498942
Diesel Gasoline Sale (DGS)	1.00000000	0.30364955
RNUV	0.30364955	1.00000000
# LPG Vehicles (NLPG)	0.63895652	0.18522216
Price of Unleaded Gasoline (PU)	0.49630315	0.04177338
Price of Diesel Gasoline (PG)	0.54026568	0.07378405
# Unleaded Gasoline Vehicles (NUGV)	0.67288637	0.07413199
# of Diesel Gasoline Vehicles (NDGV)	0.65656807	0.30576056
GNP Agriculture	0.60365821	0.16530707
GNP Commerce	0.91741674	0.29698204
GNP Total	0.87962217	0.30260566

	# LPG Vehicles (NLPG)
Unleaded Gasoline Sale (UGS)	-0.62890683
Diesel Gasoline Sale (DGS)	0.63895652
RNUV	0.18522216
# LPG Vehicles (NLPG)	1.00000000
Price of Unleaded Gasoline (PU)	0.67602376
Price of Diesel Gasoline (PG)	0.76400431
# Unleaded Gasoline Vehicles (NUGV)	0.92714875
# of Diesel Gasoline Vehicles (NDGV)	0.90937746

GNP Agriculture	0.03664584
GNP Commerce	0.49970806
GNP Total	0.49188935

	Price of Unleaded Gasoline (PU)
Unleaded Gasoline Sale (UGS)	-0.48993252
Diesel Gasoline Sale (DGS)	0.49630315
RNUV	0.04177338
# LPG Vehicles (NLPG)	0.67602376
Price of Unleaded Gasoline (PU)	1.00000000
Price of Diesel Gasoline (PG)	0.98211135
# Unleaded Gasoline Vehicles (NUGV)	0.73923142
# of Diesel Gasoline Vehicles (NDGV)	0.68439875
GNP Agriculture	0.11288812
GNP Commerce	0.48075691
GNP Total	0.45775513

	Price of Diesel Gasoline (PG)
Unleaded Gasoline Sale (UGS)	-0.54812079
Diesel Gasoline Sale (DGS)	0.54026568
RNUV	0.07378405
# LPG Vehicles (NLPG)	0.76400431
Price of Unleaded Gasoline (PU)	0.98211135
Price of Diesel Gasoline (PG)	1.00000000
# Unleaded Gasoline Vehicles (NUGV)	0.81214044
# of Diesel Gasoline Vehicles (NDGV)	0.75830059
GNP Agriculture	0.10869816
GNP Commerce	0.51351013
GNP Total	0.49819657

	# Unleaded Gasoline Vehicles (NUGV)
Unleaded Gasoline Sale (UGS)	-0.64862885
Diesel Gasoline Sale (DGS)	0.67288637
RNUV	0.07413199
# LPG Vehicles (NLPG)	0.92714875
Price of Unleaded Gasoline (PU)	0.73923142
Price of Diesel Gasoline (PG)	0.81214044
# Unleaded Gasoline Vehicles (NUGV)	1.00000000

```
# of Diesel Gasoline Vehicles (NDGV)                              0.94565611
GNP Agriculture                                                   0.06599231
GNP Commerce                                                      0.51971136
GNP Total                                                         0.51050156
                                     # of Diesel Gasoline Vehicles (NDGV)
Unleaded Gasoline Sale (UGS)                                     -0.52526929
Diesel Gasoline Sale (DGS)                                        0.65656807
RNUV                                                              0.30576056
# LPG Vehicles (NLPG)                                             0.90937746
Price of Unleaded Gasoline (PU)                                   0.68439875
Price of Diesel Gasoline (PG)                                     0.75830059
# Unleaded Gasoline Vehicles (NUGV)                               0.94565611
# of Diesel Gasoline Vehicles (NDGV)                              1.00000000
GNP Agriculture                                                   0.06716321
GNP Commerce                                                      0.50529894
GNP Total                                                         0.49132614
                                     GNP Agriculture GNP Commerce GNP Total
Unleaded Gasoline Sale (UGS)              0.58815962    0.2084369 0.1934860
Diesel Gasoline Sale (DGS)                0.60365821    0.9174167 0.8796222
RNUV                                      0.16530707    0.2969820 0.3026057
# LPG Vehicles (NLPG)                     0.03664584    0.4997081 0.4918894
Price of Unleaded Gasoline (PU)           0.11288812    0.4807569 0.4577551
Price of Diesel Gasoline (PG)             0.10869816    0.5135101 0.4981966
# Unleaded Gasoline Vehicles (NUGV)       0.06599231    0.5197114 0.5105016
# of Diesel Gasoline Vehicles (NDGV)      0.06716321    0.5052989 0.4913261
GNP Agriculture                           1.00000000    0.8170954 0.8343358
GNP Commerce                              0.81709538    1.0000000 0.9866848
GNP Total                                 0.83433576    0.9866848 1.0000000
```

#We added dummy quarter variables

x2<- 1:32 %% 4==2

```
x3<- 1:32 %% 4==3
x4<- 1:32 %% 4==0

data1<-data.frame(data, s2=1*x2,s3=1*x3,s4=1*x4)
data2<-data.frame(data,quarters=1:32,s2=1*x2,s3=1*x3,s4=1*x4)

lm1<- lm(Diesel.Gasoline.Sale..DGS.~s2+s3+s4,data=data1)
summary(lm1)
```

```
Call:
lm(formula = Diesel.Gasoline.Sale..DGS. ~ s2 + s3 + s4, data = data1)

Residuals:
    Min      1Q  Median      3Q     Max 
-454639 -203224 -110736  225573  653772 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  2393141     126376  18.937 6.16e-16 ***
s2            670435     178722   3.751 0.000985 ***
s3           1027346     178722   5.748 6.35e-06 ***
s4            732856     178722   4.101 0.000409 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 334400 on 24 degrees of freedom
  (4 observations deleted due to missingness)
Multiple R-squared:  0.5958,	Adjusted R-squared:  0.5453 
F-statistic: 11.79 on 3 and 24 DF,  p-value: 6.065e-05
```

```
lm2<- lm(Diesel.Gasoline.Sale..DGS.~s2+s3+s4+quarters,data=data2)
summary(lm2)
```

```
Call:
lm(formula = Diesel.Gasoline.Sale..DGS. ~ s2 + s3 + s4 + quarters,
    data = data2)

Residuals:
    Min      1Q  Median      3Q     Max
-193300  -81875  -35990   93354  285163

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1937690      70613  27.441  < 2e-16 ***
s2            635400      77580   8.190 2.86e-08 ***
s3            957277      77806  12.303 1.34e-11 ***
s4            627752      78182   8.029 4.03e-08 ***
quarters       35035       3425  10.228 4.98e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 145000 on 23 degrees of freedom
  (4 observations deleted due to missingness)
Multiple R-squared:  0.9272,   Adjusted R-squared:  0.9145
F-statistic: 73.18 on 4 and 23 DF,  p-value: 9.646e-13
```

#Check the largest absolute correlation value with DGS from the table of correlation. GNPC is the largest one.First, add the GNPC to the model

lm3<- lm(Diesel.Gasoline.Sale..DGS.~s2+s3+s4+quarters+GNP.Commerce,data=data2)

summary(lm3)

```
Call:
lm(formula = Diesel.Gasoline.Sale..DGS. ~ s2 + s3 + s4 + quarters +
    GNP.Commerce, data = data2)
```

```
Residuals:
    Min      1Q  Median      3Q     Max 
-192864  -89343   17474   51635  298413 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)  
(Intercept)  8.735e+05  3.842e+05   2.274   0.0331 *
s2           3.471e+05  1.232e+05   2.817   0.0100 *
s3           1.502e+05  2.956e+05   0.508   0.6163  
s4           2.040e+05  1.659e+05   1.230   0.2318  
quarters     1.658e+04  7.230e+03   2.293   0.0318 *
GNP.Commerce 3.344e-01  1.192e-01   2.806   0.0103 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 127200 on 22 degrees of freedom
  (4 observations deleted due to missingness)
Multiple R-squared:  0.9464,   Adjusted R-squared:  0.9342 
F-statistic: 77.63 on 5 and 22 DF,  p-value: 3.162e-13
```

#s3 and s4 have large p value so remove them.

lm3<-lm(Diesel.Gasoline.Sale..DGS.~s2+quarters+GNP.Total,data=data2)
summary(lm3)

```
Call:
lm(formula = Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + GNP.Total, 
    data = data2)

Residuals:
    Min      1Q  Median      3Q     Max 
-258030  -91264  -37451   58326  395684 
```

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 6.090e+05  1.606e+05   3.792 0.000891 ***
s2          3.528e+05  6.814e+04   5.177 2.66e-05 ***
quarters    1.397e+04  4.081e+03   3.424 0.002221 **
GNP.Total   8.359e-02  6.938e-03  12.049 1.15e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 149100 on 24 degrees of freedom
  (4 observations deleted due to missingness)
Multiple R-squared:  0.9197,   Adjusted R-squared:  0.9096
F-statistic: 91.59 on 3 and 24 DF,  p-value: 2.801e-13
```

model<-lm(Diesel.Gasoline.Sale..DGS.~s2+quarters+GNP.Total,data=data2)

#Now we do anova to see if we can add any more variables to the model

anova(model,lm(Diesel.Gasoline.Sale..DGS.~s2+quarters+GNP.Total+RNUV,data=data2))

```
Analysis of Variance Table

Model 1: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + GNP.Total
Model 2: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + GNP.Total + RNUV
  Res.Df        RSS Df  Sum of Sq      F Pr(>F)
1     24 5.3326e+11
2     23 5.1938e+11  1 1.3879e+10 0.6146 0.4411
```

anova(model,lm(Diesel.Gasoline.Sale..DGS.~s2+quarters+GNP.Total+X..LPG.Vehicles..NLPG.,data=data2))

```
Analysis of Variance Table
```

```
Model 1: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + GNP.Total

Model 2: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + GNP.Total +
X..LPG.Vehicles..NLPG.

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

1     24 5.3326e+11

2     23 5.1363e+11  1 1.9629e+10 0.879 0.3582
```

anova(model,lm(Diesel.Gasoline.Sale..DGS.~s2+quarters+GNP.Total+Price.of.Unleaded.Gasoline..PU.,data=data2))

```
Analysis of Variance Table


Model 1: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + GNP.Total

Model 2: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + GNP.Total +
Price.of.Unleaded.Gasoline..PU.

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

1     24 5.3326e+11

2     23 4.2717e+11  1 1.0609e+11 5.7125 0.02543 *

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

anova(model,lm(Diesel.Gasoline.Sale..DGS.~s2+quarters+GNP.Total+Price.of.Diesel.Gasoline..PG.,data=data2))

```
Analysis of Variance Table


Model 1: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + GNP.Total

Model 2: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + GNP.Total +
Price.of.Diesel.Gasoline..PG.

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

```
1     24 5.3326e+11
2     23 4.1300e+11  1 1.2026e+11 6.6971 0.01645 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

anova(model,lm(Diesel.Gasoline.Sale..DGS.~s2+quarters+GNP.Total+X..Unleaded.Gasoline.Vehicles..NUGV. ,data=data2))

```
Analysis of Variance Table

Model 1: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + GNP.Total
Model 2: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + GNP.Total +
X..Unleaded.Gasoline.Vehicles..NUGV.
  Res.Df        RSS Df  Sum of Sq     F Pr(>F)
1     24 5.3326e+11
2     23 4.2182e+11  1 1.1144e+11 6.076 0.0216 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

anova(model,lm(Diesel.Gasoline.Sale..DGS.~s2+quarters+GNP.Total+X..of.Diesel.Gasoline.Vehicles..NDGV.,data=data2))

```
Analysis of Variance Table

Model 1: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + GNP.Total
Model 2: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + GNP.Total +
X..of.Diesel.Gasoline.Vehicles..NDGV.
  Res.Df        RSS Df  Sum of Sq     F  Pr(>F)
1     24 5.3326e+11
2     23 4.5828e+11  1 7.4978e+10 3.763 0.06476 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

anova(model,lm(Diesel.Gasoline.Sale..DGS.~s2+quarters+GNP.Total+GNP.Agriculture,data=data2))

```
Analysis of Variance Table

Model 1: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + GNP.Total
Model 2: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + GNP.Total +
GNP.Agriculture
  Res.Df        RSS Df  Sum of Sq      F Pr(>F)
1     24 5.3326e+11
2     23 5.3039e+11  1 2871620039 0.1245 0.7274
```

anova(model,lm(Diesel.Gasoline.Sale..DGS.~s2+quarters+GNP.Total+GNP.Commerce,data=data2))

```
Analysis of Variance Table

Model 1: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + GNP.Total
Model 2: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + GNP.Total +
GNP.Commerce
  Res.Df        RSS Df  Sum of Sq      F Pr(>F)
1     24 5.3326e+11
2     23 4.2975e+11  1 1.0351e+11 5.5401 0.0275 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

#We selected to PG which has the smallest p-value.

model<-lm(Diesel.Gasoline.Sale..DGS.~s2+quarters+GNP.Total+Price.of.Diesel.Gasoline..PG.,data=data2)

summary(model)

```
Call:
lm(formula = Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + GNP.Total +
    Price.of.Diesel.Gasoline..PG., data = data2)

Residuals:
    Min      1Q  Median      3Q     Max
-217662  -76646  -10177   88420  314403

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                    1.358e+06  3.234e+05   4.199 0.000343 ***
s2                             3.691e+05  6.158e+04   5.994 4.12e-06 ***
quarters                       2.538e+04  5.733e+03   4.426 0.000195 ***
GNP.Total                      8.658e-02  6.343e-03  13.650 1.62e-12 ***
Price.of.Diesel.Gasoline..PG. -2.524e+03  9.755e+02  -2.588 0.016450 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 134000 on 23 degrees of freedom
  (4 observations deleted due to missingness)
Multiple R-squared:  0.9378,   Adjusted R-squared:  0.927
F-statistic: 86.67 on 4 and 23 DF,  p-value: 1.589e-13
```

#PG is indeed significant. But we still want to check if removing GNPT increases the significance of PG and therefore improves the model.

```
Analysis of Variance Table

Model 1: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + GNP.Total +
Price.of.Diesel.Gasoline..PG.
Model 2: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters +
Price.of.Diesel.Gasoline..PG.
  Res.Df        RSS Df   Sum of Sq      F    Pr(>F)
1     23 4.1300e+11
2     24 3.7585e+12 -1 -3.3455e+12 186.31 1.623e-12 ***
```

```
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

#We can not remove the GNPT from the model.

#So check anova results to determine variable which will be add

anova(model,lm(Diesel.Gasoline.Sale..DGS.~s2+quarters+GNP.Total+Price.of.Diesel.Gasoline..PG.+RNUV,data=data2))

```
Analysis of Variance Table

Model 1: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + GNP.Total +
Price.of.Diesel.Gasoline..PG.
Model 2: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + GNP.Total +
Price.of.Diesel.Gasoline..PG. +
    RNUV
  Res.Df        RSS Df  Sum of Sq      F Pr(>F)
1     23 4.1300e+11
2     22 3.9326e+11  1 1.9745e+10 1.1046 0.3047
```

anova(model,lm(Diesel.Gasoline.Sale..DGS.~s2+quarters+GNP.Total+Price.of.Diesel.Gasoline..PG.+X..LPG.Vehicles..NLPG.,data=data2))

```
Analysis of Variance Table

Model 1: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + GNP.Total +
Price.of.Diesel.Gasoline..PG.
Model 2: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + GNP.Total +
Price.of.Diesel.Gasoline..PG. +
    X..LPG.Vehicles..NLPG.
  Res.Df        RSS Df  Sum of Sq     F Pr(>F)
1     23 4.1300e+11
2     22 4.0666e+11  1 6340148042 0.343 0.5641
```

anova(model,lm(Diesel.Gasoline.Sale..DGS.~s2+quarters+GNP.Total+Price.of.Diesel.Gasoline..PG.+Price.of.Unleaded.Gasoline..PU.,data=data2))

```
Analysis of Variance Table

Model 1: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + GNP.Total +
Price.of.Diesel.Gasoline..PG.

Model 2: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + GNP.Total +
Price.of.Diesel.Gasoline..PG. +

    Price.of.Unleaded.Gasoline..PU.

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

1     23 4.1300e+11

2     22 4.0922e+11  1 3783654094 0.2034 0.6564
```

anova(model,lm(Diesel.Gasoline.Sale..DGS.~s2+quarters+GNP.Total+Price.of.Diesel.Gasoline..PG.+X..Unleaded.Gasoline.Vehicles..NUGV.

,data=data2))

```
Analysis of Variance Table

Model 1: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + GNP.Total +
Price.of.Diesel.Gasoline..PG.

Model 2: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + GNP.Total +
Price.of.Diesel.Gasoline..PG. +

    X..Unleaded.Gasoline.Vehicles..NUGV.

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

1     23 4.1300e+11

2     22 2.8833e+11  1 1.2467e+11 9.5124 0.005421 **

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

anova(model,lm(Diesel.Gasoline.Sale..DGS.~s2+quarters+GNP.Total+Price.of.Diesel.Gasoline..PG.+X..of.Diesel.Gasoline.Vehicles..NDGV.,data=data2))

```
Analysis of Variance Table

Model 1: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + GNP.Total +
Price.of.Diesel.Gasoline..PG.
Model 2: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + GNP.Total +
Price.of.Diesel.Gasoline..PG. +
    X..of.Diesel.Gasoline.Vehicles..NDGV.
  Res.Df        RSS Df  Sum of Sq      F  Pr(>F)
1     23 4.1300e+11
2     22 3.1641e+11  1 9.6596e+10 6.7164 0.01665 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

anova(model,lm(Diesel.Gasoline.Sale..DGS.~s2+quarters+GNP.Total+Price.of.Diesel.Gasoline..PG.+GNP.Agriculture,data=data2))

```
Analysis of Variance Table

Model 1: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + GNP.Total +
Price.of.Diesel.Gasoline..PG.
Model 2: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + GNP.Total +
Price.of.Diesel.Gasoline..PG. +
    GNP.Agriculture
  Res.Df        RSS Df  Sum of Sq     F Pr(>F)
1     23 4.1300e+11
2     22 4.0633e+11  1 6668217786 0.361 0.5541
```

anova(model,lm(Diesel.Gasoline.Sale..DGS.~s2+quarters+GNP.Total+Price.of.Diesel.Gasoline..PG.+GNP.Commerce,data=data2))

```
Analysis of Variance Table

Model 1: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + GNP.Total + 
Price.of.Diesel.Gasoline..PG.

Model 2: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + GNP.Total + 
Price.of.Diesel.Gasoline..PG. + 

    GNP.Commerce

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

1     23 4.130e+11                                

2     22 2.794e+11  1 1.3361e+11 10.52 0.003729 **

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

#GNPC is selected because the smallest p-value is in GNPC

model<-lm(Diesel.Gasoline.Sale..DGS.~s2+quarters+GNP.Total+Price.of.Diesel.Gasoline..PG.+GNP.Commerce,data=data2)

summary(model)

```
Call:

lm(formula = Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + GNP.Total + 

    Price.of.Diesel.Gasoline..PG. + GNP.Commerce, data = data2)

Residuals:

    Min      1Q  Median      3Q     Max 

-204125  -85488  -14629   92439  178383 

Coefficients:

                                Estimate Std. Error t value Pr(>|t|)    

(Intercept)                    1.611e+06  2.830e+05   5.694 9.99e-06 ***
```

```
s2                           2.307e+05  6.711e+04   3.438  0.00235 **
quarters                     2.690e+04  4.845e+03   5.553 1.40e-05 ***
GNP.Total                   -2.817e-02  3.578e-02  -0.787  0.43946
Price.of.Diesel.Gasoline..PG. -2.843e+03  8.262e+02  -3.441  0.00233 **
GNP.Commerce                 5.409e-01  1.668e-01   3.243  0.00373 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 112700 on 22 degrees of freedom
  (4 observations deleted due to missingness)
Multiple R-squared:  0.9579,   Adjusted R-squared:  0.9483
F-statistic: 100.1 on 5 and 22 DF,  p-value: 2.232e-14
```

#GNPT seems insignificant we should do ANOVA to determine remove or not

#We still want to check for removing both GNPT or PG

#For PG:

anova(model,lm(Diesel.Gasoline.Sale..DGS.~s2+quarters+GNP.Total+GNP.Commerce,data=data2))

```
Analysis of Variance Table

Model 1: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + GNP.Total +
Price.of.Diesel.Gasoline..PG. +
    GNP.Commerce
Model 2: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + GNP.Total +
GNP.Commerce
  Res.Df        RSS Df   Sum of Sq      F   Pr(>F)
1     22 2.7940e+11
2     23 4.2975e+11 -1 -1.5035e+11 11.838 0.002333 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Since p value is quite small, we can not remove PG

#For GNPT:

anova(model,lm(Diesel.Gasoline.Sale..DGS.~s2+quarters+Price.of.Diesel.Gasoline..PG.+GNP.Commerce,data=data2))

```
Analysis of Variance Table

Model 1: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + GNP.Total +
Price.of.Diesel.Gasoline..PG. +
    GNP.Commerce
Model 2: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters +
Price.of.Diesel.Gasoline..PG. +
    GNP.Commerce
  Res.Df        RSS Df   Sum of Sq    F Pr(>F)
1     22 2.7940e+11
2     23 2.8727e+11 -1 -7873681465 0.62 0.4395
```

Due to the high p value, we can remove GNPT

model<-lm(Diesel.Gasoline.Sale..DGS.~s2+quarters+Price.of.Diesel.Gasoline..PG.+GNP.Commerce,data=data2)

summary(model)

```
Call:
lm(formula = Diesel.Gasoline.Sale..DGS. ~ s2 + quarters +
Price.of.Diesel.Gasoline..PG. +
    GNP.Commerce, data = data2)

Residuals:
    Min      1Q  Median      3Q     Max
-204583  -71473  -20736   89535  180997

Coefficients:
                               Estimate Std. Error t value Pr(>|t|)
(Intercept)                   1.543e+06  2.670e+05   5.778 6.92e-06 ***
s2                            2.658e+05  4.975e+04   5.343 2.00e-05 ***
```

```
quarters                         2.643e+04  4.768e+03   5.544 1.22e-05 ***
Price.of.Diesel.Gasoline..PG. -2.784e+03  8.160e+02  -3.412  0.00239 **
GNP.Commerce                     4.111e-01  2.466e-02  16.671 2.45e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 111800 on 23 degrees of freedom
  (4 observations deleted due to missingness)
Multiple R-squared:  0.9567,   Adjusted R-squared:  0.9492
F-statistic: 127.1 on 4 and 23 DF,  p-value: 2.489e-15
```

All variables are significant, so check anova results to determine variable which will be add.

anova(model,lm(Diesel.Gasoline.Sale..DGS.~s2+quarters+Price.of.Diesel.Gasoline..PG.+GNP.Commerce+RNUV,data=data2))

```
Analysis of Variance Table

Model 1: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + Price.of.Diesel.Gasoline..PG. +
    GNP.Commerce
Model 2: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + Price.of.Diesel.Gasoline..PG. +
    GNP.Commerce + RNUV
  Res.Df        RSS Df Sum of Sq      F Pr(>F)
1     23 2.8727e+11
2     22 2.5981e+11  1 2.746e+10 2.3252 0.1415
```

anova(model,lm(Diesel.Gasoline.Sale..DGS.~s2+quarters+Price.of.Diesel.Gasoline..PG.+GNP.Commerce+X..LPG.Vehicles..NLPG.,data=data2))

```
Analysis of Variance Table


Model 1: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters +
Price.of.Diesel.Gasoline..PG. +

    GNP.Commerce

Model 2: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters +
Price.of.Diesel.Gasoline..PG. +

    GNP.Commerce + X..LPG.Vehicles..NLPG.

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

1     23 2.8727e+11

2     22 2.8655e+11  1 717964400 0.0551 0.8166
```

anova(model,lm(Diesel.Gasoline.Sale..DGS.~s2+quarters+Price.of.Diesel.Gasoline..PG.+GNP.Commerce+Price.of.Unleaded.Gasoline..PU.,data=data2))

```
Analysis of Variance Table


Model 1: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters +
Price.of.Diesel.Gasoline..PG. +

    GNP.Commerce

Model 2: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters +
Price.of.Diesel.Gasoline..PG. +

    GNP.Commerce + Price.of.Unleaded.Gasoline..PU.

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

1     23 2.8727e+11

2     22 2.8529e+11  1 1979616563 0.1527 0.6998
```

anova(model,lm(Diesel.Gasoline.Sale..DGS.~s2+quarters+Price.of.Diesel.Gasoline..PG.+GNP.Commerce+X..Unleaded.Gasoline.Vehicles..NUGV. ,data=data2))

```
Analysis of Variance Table

Model 1: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters +
Price.of.Diesel.Gasoline..PG. +
    GNP.Commerce
Model 2: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters +
Price.of.Diesel.Gasoline..PG. +
    GNP.Commerce + X..Unleaded.Gasoline.Vehicles..NUGV.
  Res.Df        RSS Df  Sum of Sq      F   Pr(>F)
1     23 2.8727e+11
2     22 2.0453e+11  1 8.2744e+10 8.9004 0.006854 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

anova(model,lm(Diesel.Gasoline.Sale..DGS.~s2+quarters+Price.of.Diesel.Gasoline..PG.+GNP.Commerce+X..of.Diesel.Gasoline.Vehicles..NDGV.,data=data2))

```
Analysis of Variance Table

Model 1: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters +
Price.of.Diesel.Gasoline..PG. +
    GNP.Commerce
Model 2: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters +
Price.of.Diesel.Gasoline..PG. +
    GNP.Commerce + X..of.Diesel.Gasoline.Vehicles..NDGV.
  Res.Df        RSS Df  Sum of Sq      F Pr(>F)
1     23 2.8727e+11
2     22 2.1980e+11  1 6.7466e+10 6.7526 0.0164 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

anova(model,lm(Diesel.Gasoline.Sale..DGS.~s2+quarters+Price.of.Diesel.Gasoline..PG.+GNP.Commerce+GNP.Agriculture,data=data2))

```
Analysis of Variance Table

Model 1: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + Price.of.Diesel.Gasoline..PG. +
    GNP.Commerce
Model 2: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + Price.of.Diesel.Gasoline..PG. +
    GNP.Commerce + GNP.Agriculture
  Res.Df        RSS Df  Sum of Sq      F Pr(>F)
1     23 2.8727e+11
2     22 2.6199e+11  1 2.5278e+10 2.1226 0.1593
```

#select the variable which has a smallest p-value. NUGV is the smallest one.

model<-lm(Diesel.Gasoline.Sale..DGS.~s2+quarters+Price.of.Diesel.Gasoline..PG. + GNP.Commerce+X..Unleaded.Gasoline.Vehicles..NUGV. ,data=data2)

summary(model)

```
Call:
lm(formula = Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + Price.of.Diesel.Gasoline..PG. +
    GNP.Commerce + X..Unleaded.Gasoline.Vehicles..NUGV., data = data2)

Residuals:
    Min      1Q  Median      3Q     Max
-144167  -71433  -17206   76505  121482

Coefficients:
                                       Estimate Std. Error t value Pr(>|t|)
(Intercept)                          -1.844e+06  1.158e+06  -1.592 0.125676
s2                                    2.730e+05  4.299e+04   6.351 2.16e-06
quarters                              1.118e+03  9.429e+03   0.119 0.906685
Price.of.Diesel.Gasoline..PG.        -2.881e+03  7.048e+02  -4.088 0.000486
```

```
GNP.Commerce                          4.052e-01  2.136e-02  18.970 4.01e-15
X..Unleaded.Gasoline.Vehicles..NUGV.  7.344e-01  2.462e-01   2.983 0.006854

(Intercept)
s2                                   ***
quarters
Price.of.Diesel.Gasoline..PG.        ***
GNP.Commerce                         ***
X..Unleaded.Gasoline.Vehicles..NUGV. **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 96420 on 22 degrees of freedom
  (4 observations deleted due to missingness)
Multiple R-squared:  0.9692,   Adjusted R-squared:  0.9622
F-statistic: 138.4 on 5 and 22 DF,  p-value: 7.336e-16
```

#All variables seem significant but we do ANOVA to remove any variable.

anova(model,lm(Diesel.Gasoline.Sale..DGS.~s2+quarters+Price.of.Diesel.Gasoline..PG.+X..Unleaded.Gasoline.Vehicles..NUGV. ,data=data2))

```
Analysis of Variance Table

Model 1: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + Price.of.Diesel.Gasoline..PG. +
    GNP.Commerce + X..Unleaded.Gasoline.Vehicles..NUGV.
Model 2: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + Price.of.Diesel.Gasoline..PG. +
    X..Unleaded.Gasoline.Vehicles..NUGV.
  Res.Df        RSS Df   Sum of Sq      F   Pr(>F)
1     22 2.0453e+11
2     23 3.5502e+12 -1 -3.3456e+12 359.87 4.01e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

#We can not remove GNPC.

anova(model,lm(Diesel.Gasoline.Sale..DGS.~s2+quarters+GNP.Commerce+X..Unleaded.Gasoline.Vehicles..NUGV. ,data=data2))

```
Analysis of Variance Table

Model 1: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + Price.of.Diesel.Gasoline..PG. +
    GNP.Commerce + X..Unleaded.Gasoline.Vehicles..NUGV.
Model 2: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + GNP.Commerce + X..Unleaded.Gasoline.Vehicles..NUGV.
  Res.Df        RSS Df  Sum of Sq      F    Pr(>F)
1     22 2.0453e+11
2     23 3.5993e+11 -1 -1.554e+11 16.715 0.0004861 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

#We can not remove PG.

#Now determine the next variable to add the model.

anova(model,lm(Diesel.Gasoline.Sale..DGS.~s2+quarters+Price.of.Diesel.Gasoline..PG.+GNP.Commerce+X..Unleaded.Gasoline.Vehicles..NUGV.
+RNUV,data=data2))

```
Analysis of Variance Table

Model 1: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + Price.of.Diesel.Gasoline..PG. +
    GNP.Commerce + X..Unleaded.Gasoline.Vehicles..NUGV.
Model 2: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + Price.of.Diesel.Gasoline..PG. +
    GNP.Commerce + X..Unleaded.Gasoline.Vehicles..NUGV. + RNUV
  Res.Df        RSS Df  Sum of Sq      F Pr(>F)
1     22 2.0453e+11
2     21 1.9579e+11  1 8734607531 0.9368 0.3441
```

```
anova(model,lm(Diesel.Gasoline.Sale..DGS.~s2+quarters+Price.of.Diesel.Gasoline..PG.+GNP.Commerce+X..Unleaded.Gasoline.Vehicles..NUGV.
+X..LPG.Vehicles..NLPG.,data=data2))
```

```
Analysis of Variance Table

Model 1: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters +
Price.of.Diesel.Gasoline..PG. +
    GNP.Commerce + X..Unleaded.Gasoline.Vehicles..NUGV.
Model 2: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters +
Price.of.Diesel.Gasoline..PG. +
    GNP.Commerce + X..Unleaded.Gasoline.Vehicles..NUGV. +
X..LPG.Vehicles..NLPG.
  Res.Df        RSS Df Sum of Sq      F Pr(>F)
1     22 2.0453e+11
2     21 2.0426e+11  1 266700287 0.0274 0.8701
```

```
anova(model,lm(Diesel.Gasoline.Sale..DGS.~s2+quarters+Price.of.Diesel.Gasoline..PG.+GNP.Commerce+X..Unleaded.Gasoline.Vehicles..NUGV. +
Price.of.Unleaded.Gasoline..PU.,data=data2))
```

```
Analysis of Variance Table

Model 1: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters +
Price.of.Diesel.Gasoline..PG. +
    GNP.Commerce + X..Unleaded.Gasoline.Vehicles..NUGV.
Model 2: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters +
Price.of.Diesel.Gasoline..PG. +
    GNP.Commerce + X..Unleaded.Gasoline.Vehicles..NUGV. +
Price.of.Unleaded.Gasoline..PU.
  Res.Df        RSS Df Sum of Sq     F Pr(>F)
1     22 2.0453e+11
2     21 2.0453e+11  1    986634 1e-04 0.9921
```

anova(model,lm(Diesel.Gasoline.Sale..DGS.~s2+quarters+Price.of.Diesel.Gasoline..PG.+GNP.Commerce+X..Unleaded.Gasoline.Vehicles..NUGV. + X..of.Diesel.Gasoline.Vehicles..NDGV.,data=data2))

```
Analysis of Variance Table


Model 1: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters +
Price.of.Diesel.Gasoline..PG. +
    GNP.Commerce + X..Unleaded.Gasoline.Vehicles..NUGV.
Model 2: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters +
Price.of.Diesel.Gasoline..PG. +
    GNP.Commerce + X..Unleaded.Gasoline.Vehicles..NUGV. +
X..of.Diesel.Gasoline.Vehicles..NDGV.
  Res.Df        RSS Df Sum of Sq      F Pr(>F)
1     22 2.0453e+11
2     21 2.0243e+11  1 2.093e+09 0.2171  0.646
```

anova(model,lm(Diesel.Gasoline.Sale..DGS.~s2+quarters+Price.of.Diesel.Gasoline..PG.+GNP.Commerce+X..Unleaded.Gasoline.Vehicles..NUGV. +GNP.Agriculture,data=data2))

```
Analysis of Variance Table


Model 1: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters +
Price.of.Diesel.Gasoline..PG. +
    GNP.Commerce + X..Unleaded.Gasoline.Vehicles..NUGV.
Model 2: Diesel.Gasoline.Sale..DGS. ~ s2 + quarters +
Price.of.Diesel.Gasoline..PG. +
    GNP.Commerce + X..Unleaded.Gasoline.Vehicles..NUGV. + GNP.Agriculture
  Res.Df        RSS Df  Sum of Sq      F Pr(>F)
1     22 2.0453e+11
2     21 1.9172e+11  1 1.2809e+10 1.4031 0.2494
```

#We checked all the variables and they cannot be added to the model because of their ANOVA results, no variable seems significant due to high p values.

model<-lm(Diesel.Gasoline.Sale..DGS.~s2+quarters+Price.of.Diesel.Gasoline..PG.+GNP.Commerce+X..Unleaded.Gasoline.Vehicles..NUGV.

,data=data2)

summary(model)

```
Call:

lm(formula = Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + 
Price.of.Diesel.Gasoline..PG. + 

    GNP.Commerce + X..Unleaded.Gasoline.Vehicles..NUGV., data = data2)


Residuals:

    Min      1Q  Median      3Q     Max 

-144167  -71433  -17206   76505  121482 


Coefficients:

                                       Estimate Std. Error t value Pr(>|t|)

(Intercept)                          -1.844e+06  1.158e+06  -1.592 0.125676

s2                                    2.730e+05  4.299e+04   6.351 2.16e-06

quarters                              1.118e+03  9.429e+03   0.119 0.906685

Price.of.Diesel.Gasoline..PG.        -2.881e+03  7.048e+02  -4.088 0.000486

GNP.Commerce                          4.052e-01  2.136e-02  18.970 4.01e-15

X..Unleaded.Gasoline.Vehicles..NUGV.  7.344e-01  2.462e-01   2.983 0.006854


(Intercept)                             

s2                                   ***

quarters                                

Price.of.Diesel.Gasoline..PG.        ***

GNP.Commerce                         ***

X..Unleaded.Gasoline.Vehicles..NUGV. ** 

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Residual standard error: 96420 on 22 degrees of freedom
```

```
  (4 observations deleted due to missingness)
Multiple R-squared:  0.9692,   Adjusted R-squared:  0.9622
F-statistic: 138.4 on 5 and 22 DF,  p-value: 7.336e-16
```

#All variables are significant but NUGV and PG are correlated. So, we should remove one of them.

model1<-lm(Diesel.Gasoline.Sale..DGS.~s2+quarters+Price.of.Diesel.Gasoline..PG.+GNP.Commerce,data=data2)

summary(model1)

```
Call:
lm(formula = Diesel.Gasoline.Sale..DGS. ~ s2 + quarters +
Price.of.Diesel.Gasoline..PG. +
    GNP.Commerce, data = data2)

Residuals:
    Min      1Q  Median      3Q     Max
-204583  -71473  -20736   89535  180997

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                    1.543e+06  2.670e+05   5.778 6.92e-06 ***
s2                             2.658e+05  4.975e+04   5.343 2.00e-05 ***
quarters                       2.643e+04  4.768e+03   5.544 1.22e-05 ***
Price.of.Diesel.Gasoline..PG. -2.784e+03  8.160e+02  -3.412  0.00239 **
GNP.Commerce                   4.111e-01  2.466e-02  16.671 2.45e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 111800 on 23 degrees of freedom
  (4 observations deleted due to missingness)
Multiple R-squared:  0.9567,   Adjusted R-squared:  0.9492
F-statistic: 127.1 on 4 and 23 DF,  p-value: 2.489e-15
```

model2<-lm(Diesel.Gasoline.Sale..DGS.~s2+quarters+GNP.Commerce+X..Unleaded.Gasoline.Vehicles..NUGV. ,data=data2)

summary(model2)

```
Call:
lm(formula = Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + GNP.Commerce +
    X..Unleaded.Gasoline.Vehicles..NUGV., data = data2)

Residuals:
    Min      1Q  Median      3Q     Max
-222507  -96935    9087   82912  199487

Coefficients:
                                       Estimate Std. Error t value Pr(>|t|)
(Intercept)                          -2.481e+06  1.489e+06  -1.666 0.109203
s2                                    2.577e+05  5.556e+04   4.638 0.000115
quarters                             -1.019e+04  1.170e+04  -0.871 0.392684
GNP.Commerce                          3.885e-01  2.720e-02  14.282 6.36e-13
X..Unleaded.Gasoline.Vehicles..NUGV.  6.877e-01  3.190e-01   2.156 0.041809

(Intercept)
s2                                   ***
quarters
GNP.Commerce                         ***
X..Unleaded.Gasoline.Vehicles..NUGV. *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 125100 on 23 degrees of freedom
  (4 observations deleted due to missingness)
Multiple R-squared:  0.9458,   Adjusted R-squared:  0.9363
F-statistic: 100.3 on 4 and 23 DF,  p-value: 3.293e-14
```

#We should remove the NUGV

model<-model1

summary(model)

```
Call:

lm(formula = Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + Price.of.Diesel.Gasoline..PG. +

    GNP.Commerce, data = data2)


Residuals:

    Min      1Q  Median      3Q     Max

-204583  -71473  -20736   89535  180997



Coefficients:

                                Estimate Std. Error t value Pr(>|t|)

(Intercept)                    1.543e+06  2.670e+05   5.778 6.92e-06 ***

s2                             2.658e+05  4.975e+04   5.343 2.00e-05 ***

quarters                       2.643e+04  4.768e+03   5.544 1.22e-05 ***

Price.of.Diesel.Gasoline..PG. -2.784e+03  8.160e+02  -3.412  0.00239 **

GNP.Commerce                   4.111e-01  2.466e-02  16.671 2.45e-14 ***

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1



Residual standard error: 111800 on 23 degrees of freedom

  (4 observations deleted due to missingness)

Multiple R-squared:  0.9567,   Adjusted R-squared:  0.9492

F-statistic: 127.1 on 4 and 23 DF,  p-value: 2.489e-15
```

library(lmtest)

library(car)

Now we do Model Analysis with Durbin-Watson test.

dwtest(model)

```
	Durbin-Watson test

data:  model
DW = 1.8296, p-value = 0.2319
alternative hypothesis: true autocorrelation is greater than 0
```

Residuals are independent and there is no autocorrelation

#Now we perform vif test to check collinearity

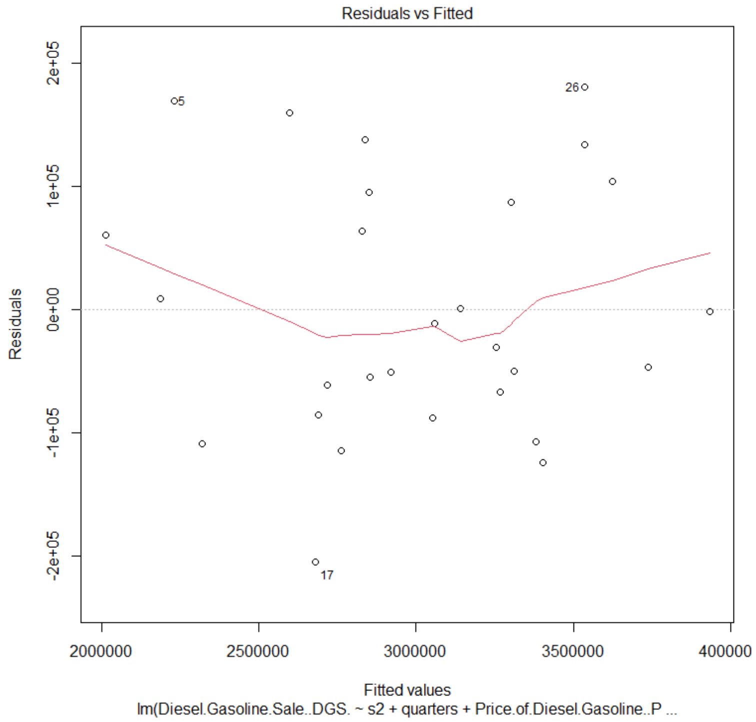
vif(model)

```
                           s2                        quarters
                     1.040329                        3.324759
Price.of.Diesel.Gasoline..PG.                    GNP.Commerce
                     3.354455                        1.454041
```

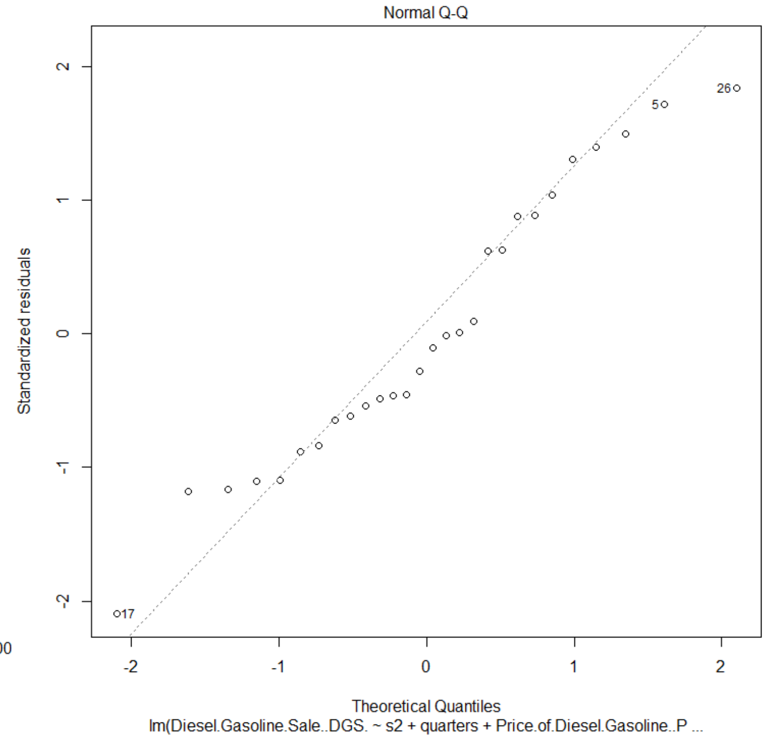
#No collinearity problem

#Now we should check model assumptions.

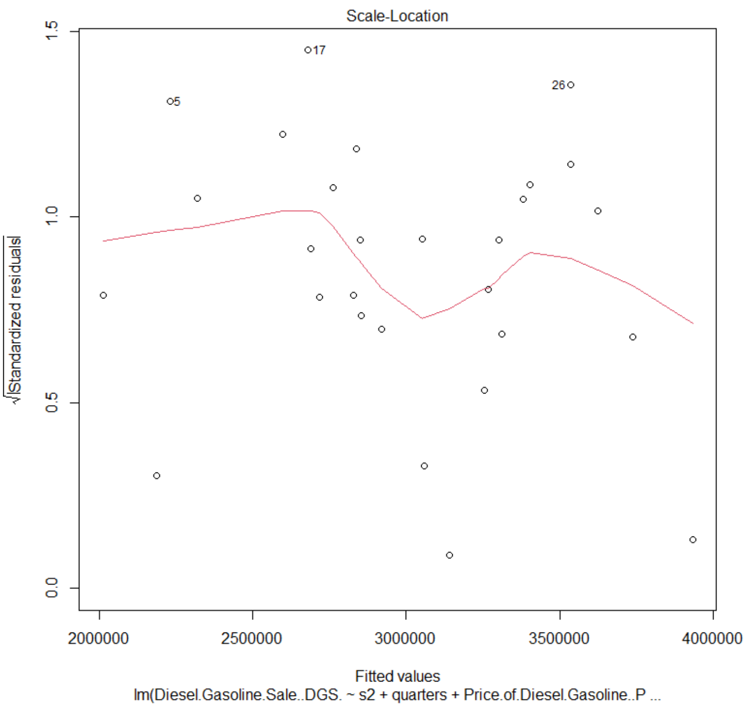
plot(model)



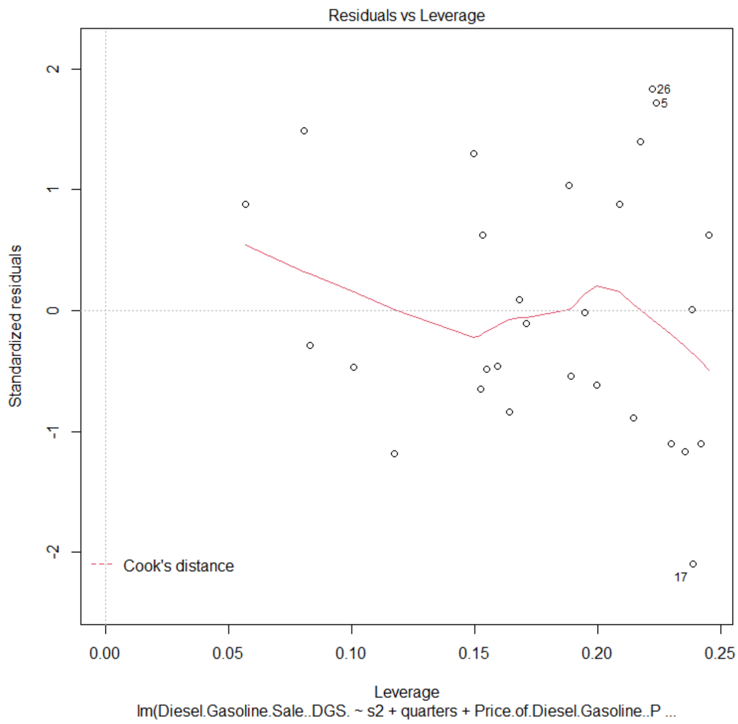
Residuals vs Fitted
Residuals
Fitted values
lm(Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + Price.of.Diesel.Gasoline..P ...
2e+05
1e+05
0e+00
-1e+05
-2e+05
2000000
2500000
3000000
3500000
400000
5
26
17



Normal Q-Q
Standardized residuals
Theoretical Quantiles
lm(Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + Price.of.Diesel.Gasoline..P ...
2
1
0
-1
-2
26
5
17



Scale-Location
√|Standardized residuals|
Fitted values
lm(Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + Price.of.Diesel.Gasoline..P ...
1.5
1.0
0.5
0.0
2000000
2500000
3000000
3500000
4000000
17
5
26



Residuals vs Leverage
Standardized residuals
Leverage
lm(Diesel.Gasoline.Sale..DGS. ~ s2 + quarters + Price.of.Diesel.Gasoline..P ...
Cook's distance
2
1
0
-1
-2
0.00
0.05
0.10
0.15
0.20
0.25
26
5
17

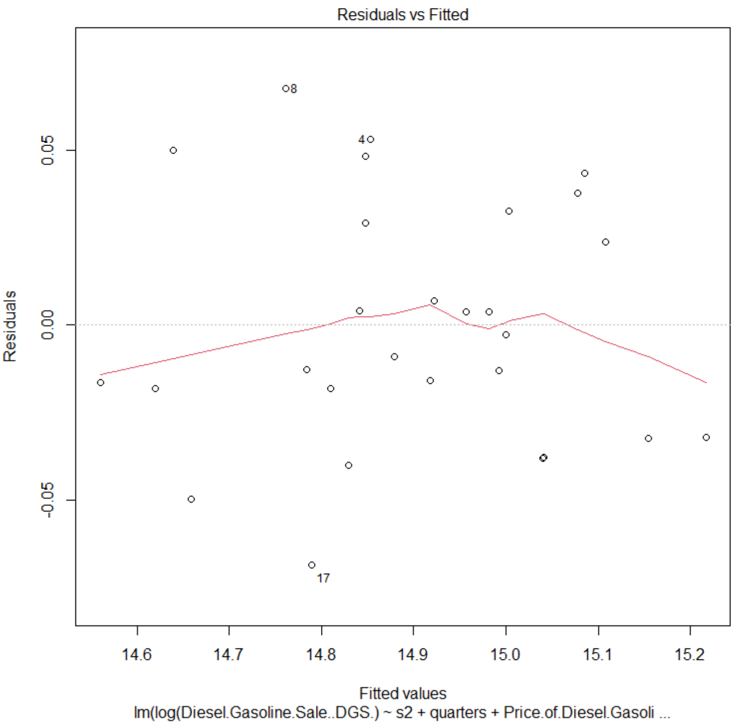
#We should use log transformation to solve linearity problem

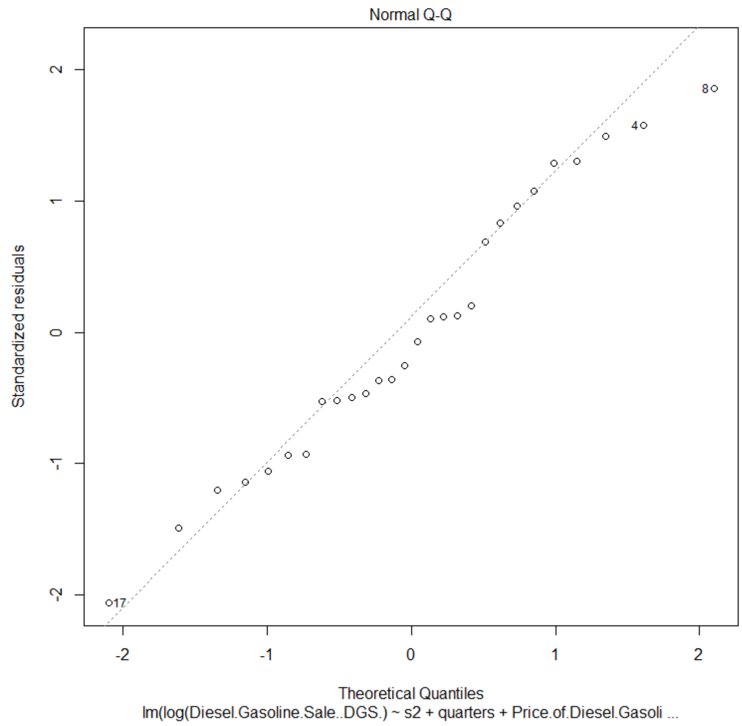
model<-lm(log(Diesel.Gasoline.Sale..DGS.)~s2+quarters+Price.of.Diesel.Gasoline..PG.+GNP.Commerce,data=data2)

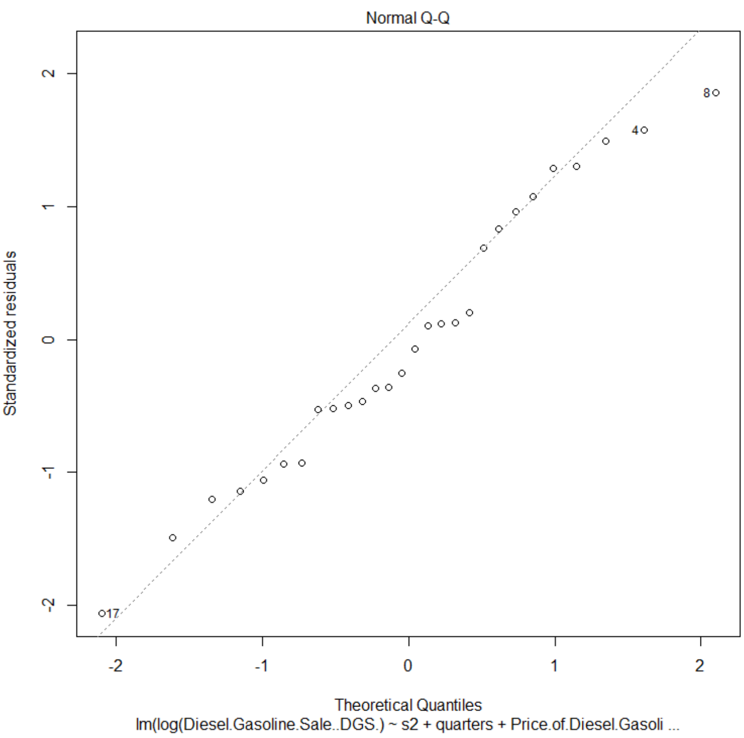
summary(model)

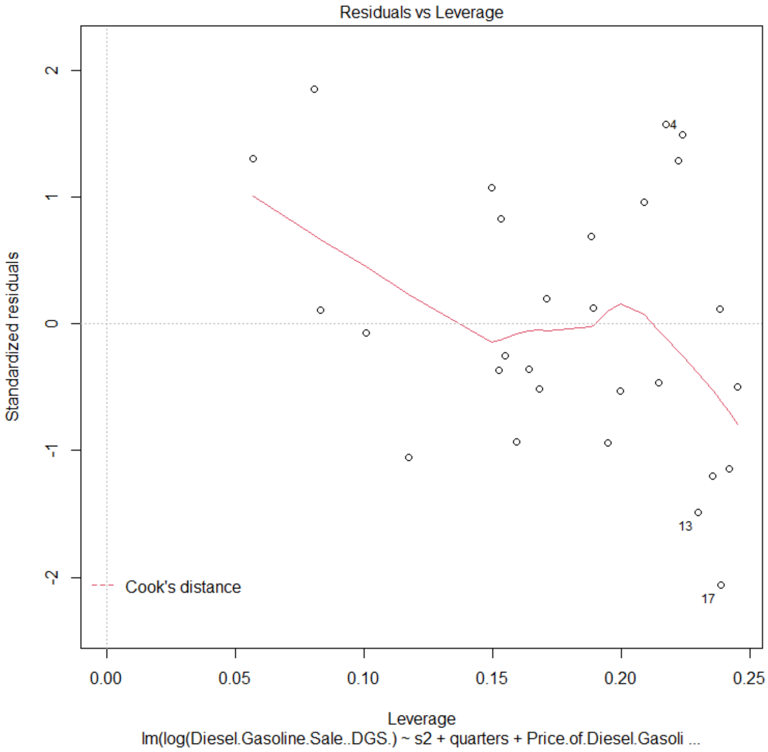
```
Call:
lm(formula = log(Diesel.Gasoline.Sale..DGS.) ~ s2 + quarters +
    Price.of.Diesel.Gasoline..PG. + GNP.Commerce, data = data2)

Residuals:
     Min       1Q   Median       3Q      Max
-0.06873 -0.02161 -0.00580  0.03005  0.06788

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                    1.443e+01  9.130e-02 158.047  < 2e-16 ***
s2                             9.932e-02  1.701e-02   5.838 5.99e-06 ***
quarters                       9.015e-03  1.630e-03   5.530 1.27e-05 ***
Price.of.Diesel.Gasoline..PG. -1.062e-03  2.791e-04  -3.806 0.000909 ***
GNP.Commerce                   1.433e-07  8.432e-09  16.996 1.62e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.03822 on 23 degrees of freedom
  (4 observations deleted due to missingness)
Multiple R-squared:  0.9567,   Adjusted R-squared:  0.9492
F-statistic: 127.2 on 4 and 23 DF,  p-value: 2.482e-15
```

plot(model)


Residuals vs Fitted
Residuals
Fitted values
lm(log(Diesel.Gasoline.Sale..DGS.) ~ s2 + quarters + Price.of.Diesel.Gasoli ...
14.6 14.7 14.8 14.9 15.0 15.1 15.2
-0.05 0.00 0.05
8
4
17


Normal Q-Q
Standardized residuals
Theoretical Quantiles
lm(log(Diesel.Gasoline.Sale..DGS.) ~ s2 + quarters + Price.of.Diesel.Gasoli ...
-2 -1 0 1 2
8
4
17

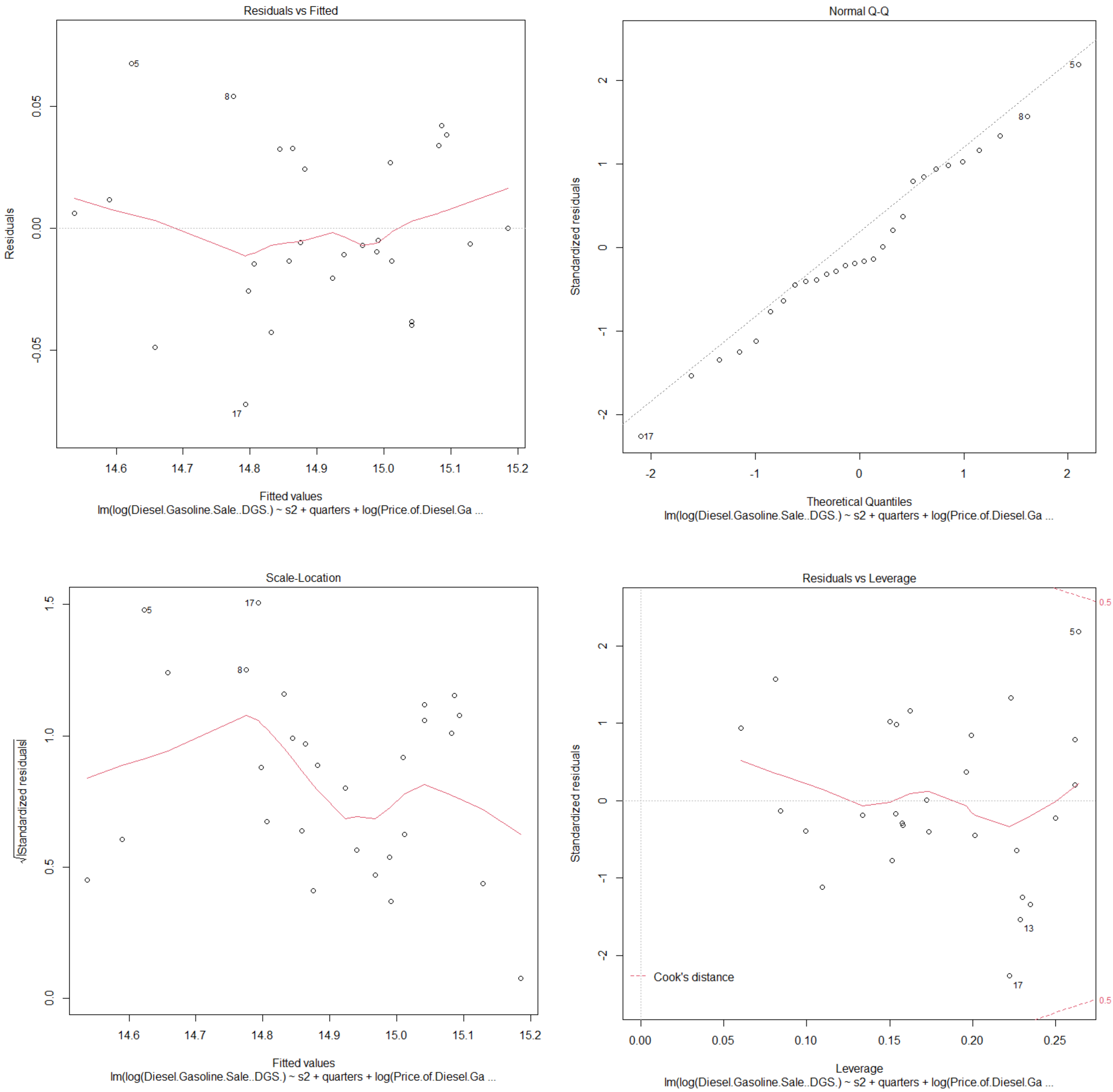

Normal Q-Q
Standardized residuals
Theoretical Quantiles
lm(log(Diesel.Gasoline.Sale..DGS.) ~ s2 + quarters + Price.of.Diesel.Gasoli ...
-2 -1 0 1 2
8
4
17


Residuals vs Leverage
Standardized residuals
Leverage
lm(log(Diesel.Gasoline.Sale..DGS.) ~ s2 + quarters + Price.of.Diesel.Gasoli ...
0.00 0.05 0.10 0.15 0.20 0.25
-2 -1 0 1 2
Cook's distance
4
13
17

#We should do log transformations to other variables because assumptions are not satisfied.

```
Call:
lm(formula = log(Diesel.Gasoline.Sale..DGS.) ~ s2 + quarters +
    log(Price.of.Diesel.Gasoline..PG.) + log(GNP.Commerce), data = data2)

Residuals:
      Min        1Q    Median        3Q       Max
-0.072111 -0.016057 -0.006014  0.028559  0.067776

Coefficients:
                                    Estimate Std. Error t value Pr(>|t|)
(Intercept)                         6.208686   0.750343   8.274 2.39e-08 ***
s2                                  0.085690   0.015971   5.365 1.89e-05 ***
quarters                            0.008971   0.001509   5.945 4.63e-06 ***
log(Price.of.Diesel.Gasoline..PG.) -0.413934   0.099120  -4.176 0.000363 ***
log(GNP.Commerce)                   0.713719   0.039551  18.045 4.49e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.03613 on 23 degrees of freedom
  (4 observations deleted due to missingness)
Multiple R-squared:  0.9613,   Adjusted R-squared:  0.9546
F-statistic:   143 on 4 and 23 DF,  p-value: 6.83e-16
```

plot(model)



Residuals vs Fitted
Residuals
Fitted values
lm(log(Diesel.Gasoline.Sale..DGS.) ~ s2 + quarters + log(Price.of.Diesel.Ga ...
Normal Q-Q
Standardized residuals
Theoretical Quantiles
lm(log(Diesel.Gasoline.Sale..DGS.) ~ s2 + quarters + log(Price.of.Diesel.Ga ...
Scale-Location
√|Standardized residuals|
Fitted values
lm(log(Diesel.Gasoline.Sale..DGS.) ~ s2 + quarters + log(Price.of.Diesel.Ga ...
Residuals vs Leverage
Standardized residuals
Cook's distance
Leverage
lm(log(Diesel.Gasoline.Sale..DGS.) ~ s2 + quarters + log(Price.of.Diesel.Ga ...

#Now constant variance and normality assumptions seem better in this model.

```
model<-lm(Diesel.Gasoline.Sale..DGS.~s2+quarters+Price.of.Diesel.Gasoline..PG.+GNP.Commerce,data=data2)

modelDGS<-model
```

#Correlation between residuals and explanatory variable

```
y_predict<-predict(model,data2)

y_act<-data$`Diesel Gasoline Sale (DGS)`

residuals<- y_act - y_predict

cor(data$`Diesel Gasoline Sale (DGS)`,residuals,  use = "complete.obs")

[1] 0.2080289
```

#There is no correlation between residuals and explanatory variables.

```
model<-lm(log(Diesel.Gasoline.Sale..DGS.)~s2+quarters+log(Price.of.Diesel.Gasoline..PG.)+log(GNP.Commerce),data=data2)

predict2007Q1 <-
data.frame(s2=0,quarters=29,Price.of.Diesel.Gasoline..PG.=449.1909,GNP.Commerce=4857305,12539562)

predict(model,predict2007Q1,level = 0.9,interval = "prediction")

      fit      lwr      upr

1 14.9292 14.86063 14.99777

exp(14.9292)

[1] 3045574

predict2007Q2 <-
data.frame(s2=1,quarters=30,Price.of.Diesel.Gasoline..PG.=449.1909,GNP.Commerce=5852403,78328151)

predict(model,predict2007Q2,level = 0.9,interval = "prediction")
```

```
      fit      lwr      upr
1 15.15688 15.08596 15.22779
```

exp(15.15688)

```
[1] 3824277
```

predict2007Q3 <- data.frame(s2=0,quarters=31,Price.of.Diesel.Gasoline..PG.=449.1909,GNP.Commerce=7480414,37162258)

predict(model,predict2007Q3,level = 0.9,interval = "prediction")

```
      fit      lwr      upr
1 15.25533 15.18549 15.32516
```

exp(15.25533)

```
[1] 4219934
```

predict2007Q4 <- data.frame(s2=0,quarters=32,Price.of.Diesel.Gasoline..PG.=449.1909,GNP.Commerce=6397744,76104164)

predict(model,predict2007Q4,level = 0.9,interval = "prediction")

```
      fit      lwr      upr
1 15.15272 15.08332 15.22211
```

exp(15.15272)

```
[1] 3808401
```

FORECASTS

UGS	Method A	Method B
2007_Q1	678288	722317
2007_Q2	896475	817249
2007_Q3	948056	934400
2007_Q4	826669	766898

DGS	Method A	Method B
2007_Q1	3105950	3045574
2007_Q2	3968121	3824277
2007_Q3	4087278	4219934
2007_Q4	3808176	3808401

COMPARISON & CONCLUSION

SARIMA is preferred when the focus is on modeling and capturing the inherent temporal patterns and autocorrelation in the time series. Time series regression is more suitable when the goal is to understand the influence of external factors and covariates on the time series behavior.

UGS has a decreasing trend, that's why Method B is overestimating UGS. Similarly, DGS has an increasing trend and Method B is underestimating DGS. It is observed that both UGS and DGS have seasonality. Method A includes seasonal and regular differencing factors, this is the reason that Method A gives better results than Method B for this data. That's why Method A is used for forecasting this data.

In summary, SARIMA is commonly used for forecasting time series with seasonality and autocorrelation, while time series regression is useful when there are external factors that can explain the variation in the time series. The choice depends on the specific characteristics and objectives of the analysis.