# HW2

Derya Özşeker

5/7/2022

In this assignment, unloaded gasoline sales per quarters with respect to a number of independent variables that are;

RNUV: An index indicating the rate of new unleaded gasoline using vehicles being added to the traffic in a quarter, PU: Average price (adjusted with an index) of a liter of unleaded gasoline in a quarter, PG: Average price (adjusted with an index) of a liter of diesel gasoline in a quarter, NUGV: Number of unleaded gasoline using vehicles in the traffic, NDGV: Number of diesel gasoline using vehicles in the traffic (per 1000 people), GNPA: Agriculture component of Gross National Product (adjusted with an index), GNPC: Commerce component of Gross National Product (adjusted with an index), GNP: Grand total for GNP (agriculture, commerce and other components total);

is given. The aim is to forecast the sales of UGS for every quarter of 2007 after analyzing the previous data.

## 0. Installing Packages and Organizing the Data

```
install.packages("ggplot2", repos = "http://cran.us.r-project.org")
```

```
##
## The downloaded binary packages are in
##
/var/folders/zd/nqgnrbsj3x9882f5rt6ljggc0000gq/T//RtmpjO42SU/downloaded
_packages
```

```
library(ggplot2)
install.packages("forecast", repos = "http://cran.us.r-project.org")
```

```
##
## The downloaded binary packages are in
##
/var/folders/zd/nqgnrbsj3x9882f5rt6ljggc0000gq/T//RtmpjO42SU/downloaded
_packages
```

```
library(forecast)
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```
install.packages("data.table", repos = "http://cran.us.r-project.org")
```

```
##
## The downloaded binary packages are in
##
/var/folders/zd/nqgnrbsj3x9882f5rt6ljggc0000gq/T//RtmpjO42SU/downloaded
_packages
```

```
library(data.table)
install.packages("GGally", repos = "http://cran.us.r-project.org")
```

```
##
## The downloaded binary packages are in
##
/var/folders/zd/nqgnrbsj3x9882f5rt6ljggc0000gq/T//RtmpjO42SU/downloaded
_packages
```

```
library(GGally)
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```
install.packages("ggcorrplot", repos = "http://cran.us.r-project.org")
```

```
##
## The downloaded binary packages are in
##
/var/folders/zd/nqgnrbsj3x9882f5rt6ljggc0000gq/T//RtmpjO42SU/downloaded
_packages
```

```
library(ggcorrplot)
install.packages("zoo", repos = "http://cran.us.r-project.org")
```

```
##
## The downloaded binary packages are in
##
/var/folders/zd/nqgnrbsj3x9882f5rt6ljggc0000gq/T//RtmpjO42SU/downloaded
_packages
```

```
library(zoo)
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
SalesUGS=data.table(read.csv("~/Desktop/IE360 Spring22 HW2 Updated-
20220507/IE360_Spring22_HW2_data.csv",colClasses=c("character",
rep("numeric",10))))
colnames(SalesUGS) <- c("Quarter", "UGS",
"RNUV","NLPG","PU","PG","NUGV","NDGV","GNPA","GNPC","GNPT")
```

```
SalesUGS$Quarter <- as.yearqtr(SalesUGS$Quarter, format = "%Y_Q%q")
str(SalesUGS)
```

```
## Classes 'data.table' and 'data.frame':   32 obs. of  11 variables:
##  $ Quarter: 'yearqtr' num  2000 Q1 2000 Q2 2000 Q3 2000 Q4 ...
##  $ UGS    : num  1128971 1199569 1370167 1127548 1033918 ...
##  $ RNUV   : num  0.0146 0.0205 0.0207 0.0163 0.0071 0.0051 0.0041
0.0048 0.0012 0.0032 ...
##  $ NLPG   : num  940000 941000 943500 948000 950000 ...
##  $ PU     : num  469 459 440 402 412 ...
##  $ PG     : num  356 345 327 301 306 ...
##  $ NUGV   : num  4647500 4742876 4840931 4919685 4954754 ...
##  $ NDGV   : num  282 284 287 288 288 ...
##  $ GNPA   : num  1040173 1760460 6974808 3267125 1004528 ...
##  $ GNPC   : num  3483132 4525451 5915204 4929778 3418387 ...
##  $ GNPT   : num  18022686 21797130 30050207 24480153 15832648 ...
##  - attr(*, ".internal.selfref")=<externalptr>
```

```
head(SalesUGS)
```

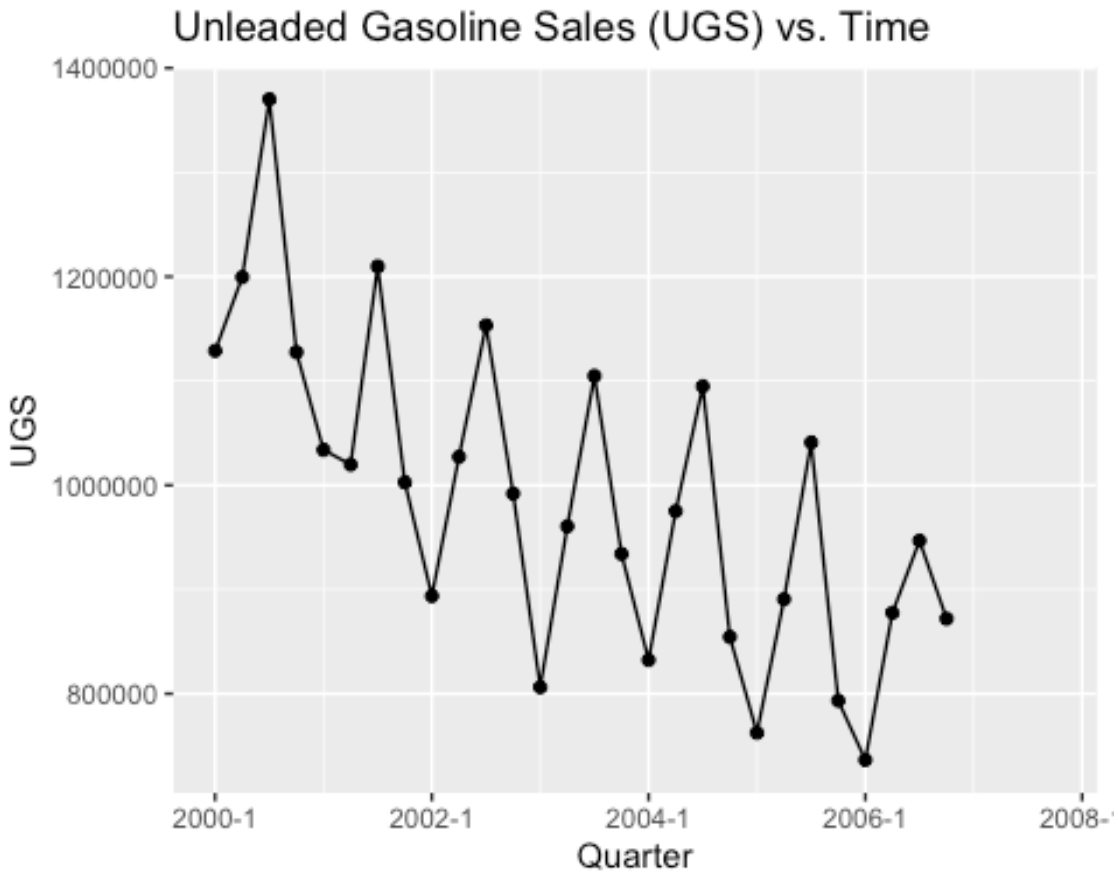
```
##    Quarter     UGS    RNUV   NLPG     PU     PG    NUGV     NDGV
GNPA    GNPC
## 1: 2000 Q1 1128971 0.0146 940000 469.03 355.69 4647500 281.9853
1040173 3483132
## 2: 2000 Q2 1199569 0.0205 941000 459.42 344.58 4742876 284.0813
1760460 4525451
## 3: 2000 Q3 1370167 0.0207 943500 439.98 327.21 4840931 286.7169
6974808 5915204
## 4: 2000 Q4 1127548 0.0163 948000 402.08 300.67 4919685 288.3137
3267125 4929778
## 5: 2001 Q1 1033918 0.0071 950000 411.58 305.75 4954754 287.6237
1004528 3418387
## 6: 2001 Q2 1019754 0.0051 955000 520.39 374.78 4980204 287.8814
1449357 4359831
##        GNPT
## 1: 18022686
## 2: 21797130
## 3: 30050207
## 4: 24480153
## 5: 15832648
## 6: 20296918
```

###1. Plotting the Time Series of UGS

```
ggplot(SalesUGS, aes(x=Quarter, y=UGS, group=1)) +
    geom_line()+
    geom_point()+
    ggtitle('Unleaded Gasoline Sales (UGS) vs. Time')
```

```
## Warning: Removed 4 row(s) containing missing values (geom_path).
```

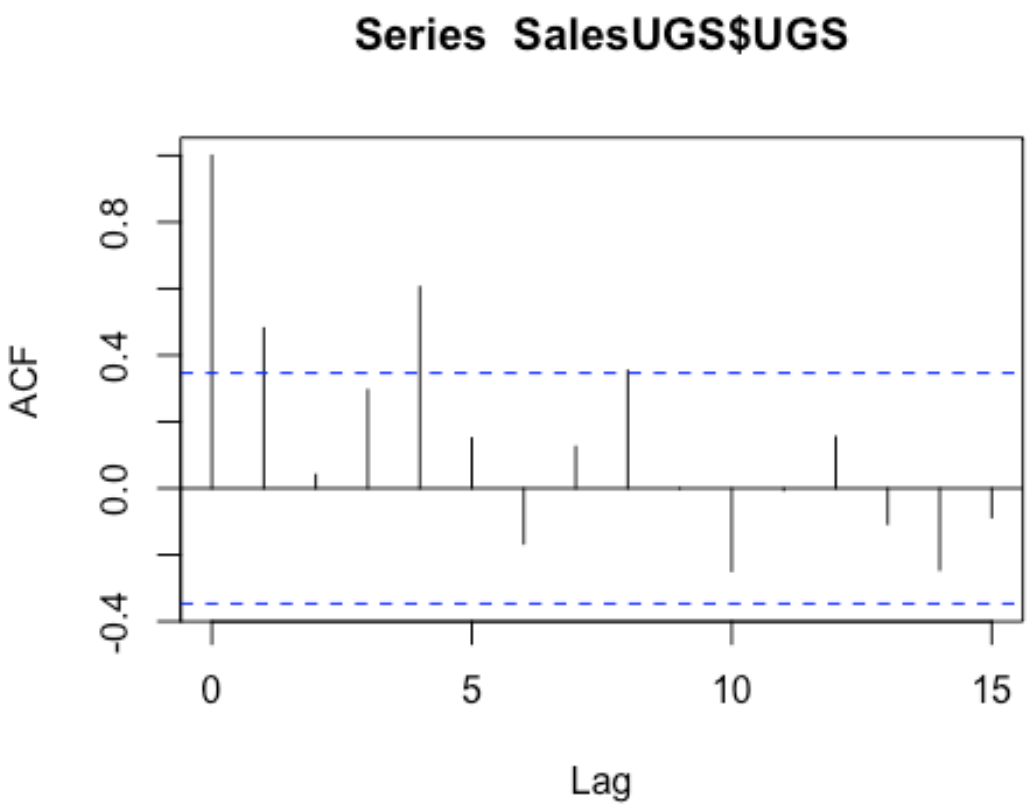
```
## Warning: Removed 4 rows containing missing values (geom_point).
```




The mean of unleaded gas sales per quarter is not stationary. This can be seen from the fact that the UGS decreases over time meaning that it's dependent on time, trend is decreasing and the mean over four quarters each year decreases significantly as the years pass. However, the variance seems to be stationary and independent of time, as it doesn't seem to change much when the trend factor is removed and the years are evaluated separately. Because the first condition of stationarity, that is mean should be independent of time, doesn't hold, it can be concluded that the time series of UGS is not stationary and depends on time.

###2. Autocorrelation Functions of the Time Series UGS

```
acf(SalesUGS$UGS,na.action = na.pass)
```




The autocorrelation function shows a clear seasonality over the span of a year. At every four quarters, the UGS shows a similar repetitive pattern, which is a significant sign of yearly seasonality. This can be understood from the autocorrelation function via the lags that are relatively high. The relatively high autocorrelation at lag 1 shows that the data is correlated to the data one quarter before, which shows that data is correlated with the data of the previous quarter, which makes sense. In this case, 4th lag is very high relatively, which shows that the data is autocorrelated with lag 4 and is correlated to the data at the same quarter in the previous years. This proves the yearly seasonality.

###3. Defining Trend, Seasonality and Lagged Variables

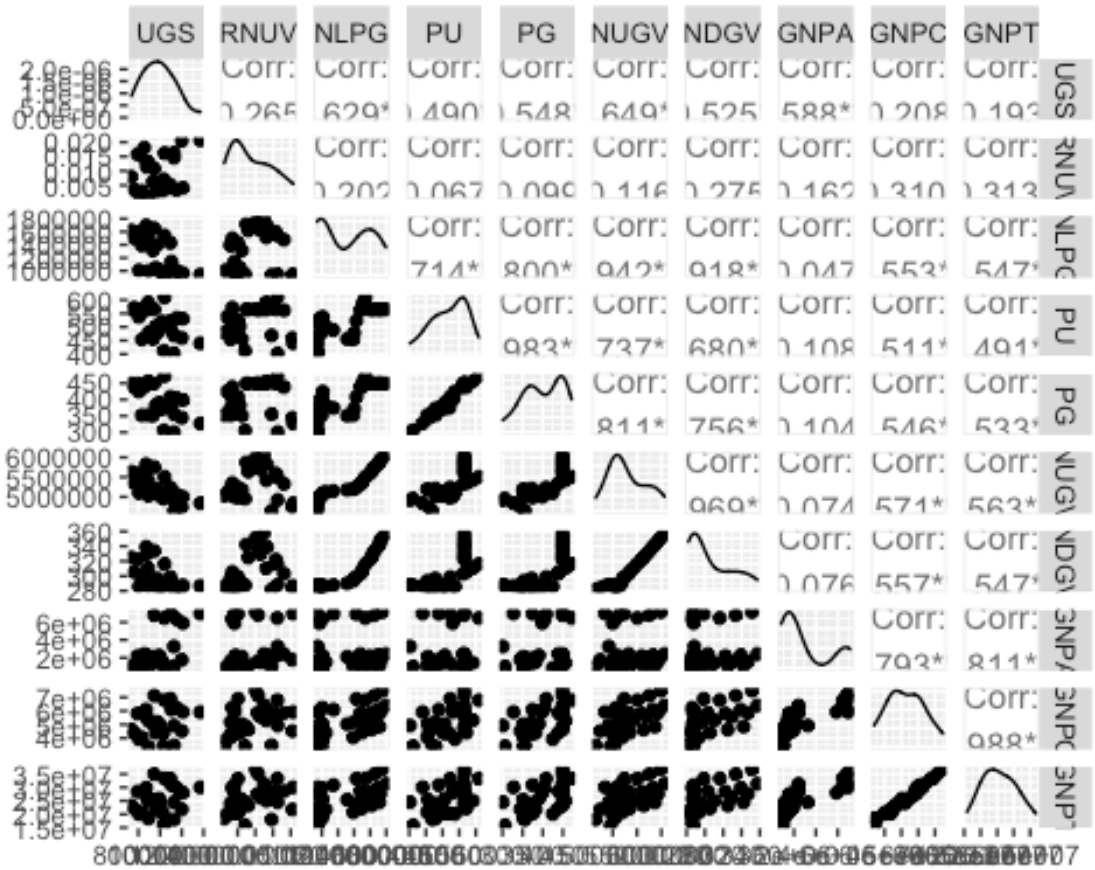
```
SalesUGS <- SalesUGS[, Trend:=(1:.N)]
SalesUGS <- SalesUGS[,Season:=(1:.N)%%4]
SalesUGS <- SalesUGS[Season==0,Season:=4]
SalesUGS <- SalesUGS[,Lagged1:=c(NA, SalesUGS$UGS[1:31])]
SalesUGS <- SalesUGS[,Lagged4:=c(rep(NA,4),SalesUGS$UGS[1:28])]
str(SalesUGS)

## Classes 'data.table' and 'data.frame':   32 obs. of  15 variables:
##  $ Quarter: 'yearqtr' num  2000 Q1 2000 Q2 2000 Q3 2000 Q4 ...
##  $ UGS    : num  1128971 1199569 1370167 1127548 1033918 ...
##  $ RNUV   : num  0.0146 0.0205 0.0207 0.0163 0.0071 0.0051 0.0041
0.0048 0.0012 0.0032 ...
```

```
##  $ NLPG   : num  940000 941000 943500 948000 950000 ...
##  $ PU     : num  469 459 440 402 412 ...
##  $ PG     : num  356 345 327 301 306 ...
##  $ NUGV   : num  4647500 4742876 4840931 4919685 4954754 ...
##  $ NDGV   : num  282 284 287 288 288 ...
##  $ GNPA   : num  1040173 1760460 6974808 3267125 1004528 ...
##  $ GNPC   : num  3483132 4525451 5915204 4929778 3418387 ...
##  $ GNPT   : num  18022686 21797130 30050207 24480153 15832648 ...
##  $ Trend  : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Season : num  1 2 3 4 1 2 3 4 1 2 ...
##  $ Lagged1: num  NA 1128971 1199569 1370167 1127548 ...
##  $ Lagged4: num  NA NA NA NA 1128971 ...
##  - attr(*, ".internal.selfref")=<externalptr>
##  - attr(*, "index")= int(0)

head(SalesUGS)

##    Quarter     UGS    RNUV   NLPG     PU     PG    NUGV     NDGV
GNPA    GNPC
## 1: 2000 Q1 1128971 0.0146 940000 469.03 355.69 4647500 281.9853
1040173 3483132
## 2: 2000 Q2 1199569 0.0205 941000 459.42 344.58 4742876 284.0813
1760460 4525451
## 3: 2000 Q3 1370167 0.0207 943500 439.98 327.21 4840931 286.7169
6974808 5915204
## 4: 2000 Q4 1127548 0.0163 948000 402.08 300.67 4919685 288.3137
3267125 4929778
## 5: 2001 Q1 1033918 0.0071 950000 411.58 305.75 4954754 287.6237
1004528 3418387
## 6: 2001 Q2 1019754 0.0051 955000 520.39 374.78 4980204 287.8814
1449357 4359831
##        GNPT Trend Season Lagged1 Lagged4
## 1: 18022686     1      1      NA      NA
## 2: 21797130     2      2 1128971      NA
## 3: 30050207     3      3 1199569      NA
## 4: 24480153     4      4 1370167      NA
## 5: 15832648     5      1 1127548 1128971
## 6: 20296918     6      2 1033918 1199569

tail(SalesUGS)

##    Quarter    UGS    RNUV    NLPG     PU     PG    NUGV     NDGV
GNPA    GNPC
## 1: 2006 Q3 946783 0.0108 1653250 565.19 449.19 5754077 333.7144
6800688 7124204
## 2: 2006 Q4 872000 0.0133 1696000 565.19 449.19 5825866 339.2281
2303373 6093090
## 3: 2007 Q1     NA 0.0074 1715000 565.19 449.19 5869018 342.1729
1132973 4857305
## 4: 2007 Q2     NA 0.0106 1725300 565.19 449.19 5931348 346.9407
1570703 5852404
```

```
## 5: 2007 Q3     NA 0.0101 1751050 565.19 449.19 5991280 351.4449
7140722 7480414
## 6: 2007 Q4     NA 0.0124 1797400 565.19 449.19 6065597 357.2902
2418541 6397745
##        GNPT Trend Season Lagged1 Lagged4
## 1: 34992138    27      3  877614 1040946
## 2: 29867726    28      4  946783  793399
## 3: 24413807    29      1  872000  736580
## 4: 27597857    30      2      NA  877614
## 5: 36741745    31      3      NA  946783
## 6: 31361112    32      4      NA  872000
```

###4. Models

```
ggpairs(SalesUGS[,2:11])
```




From ggpairs table, it can be seen that UGS is most correlated with GNPA, NUGV and NLPG among all possible independent variables given. UGS is also correlated with PU, PG and NDGV, even though not as much. So, it makes sense to try different models via including these independent variables and observing the resulting model's significance.

```
model_with_GNPA <- lm(UGS~GNPA, SalesUGS)
summary(model_with_GNPA)
```

```
##
## Call:
## lm(formula = UGS ~ GNPA, data = SalesUGS)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -193499  -88313  -18851   90094  255627
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 8.754e+05  3.825e+04  22.885  < 2e-16 ***
## GNPA        3.895e-02  1.050e-02   3.708 0.000996 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 124500 on 26 degrees of freedom
##   (4 observations deleted due to missingness)
## Multiple R-squared:  0.3459, Adjusted R-squared:  0.3208
## F-statistic: 13.75 on 1 and 26 DF,  p-value: 0.0009958
```

When GNPA, which is one of the most correlated variables with UGS, is added to the model, the adjusted R-squared is 0.3208. And, GNPA is very significant. So, this variables explains a large portion of the data by itself.

```
model_with_GNPA_NUGV <- lm(UGS~GNPA+NUGV, SalesUGS)
summary(model_with_GNPA_NUGV)
```

```
##
## Call:
## lm(formula = UGS ~ GNPA + NUGV, data = SalesUGS)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -131950  -35309   -1196   39606  123585
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.657e+06  2.201e+05  12.075 6.28e-12 ***
## GNPA         4.197e-02  5.623e-03   7.464 8.12e-08 ***
## NUGV        -3.442e-01  4.233e-02  -8.132 1.74e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 66490 on 25 degrees of freedom
##   (4 observations deleted due to missingness)
## Multiple R-squared:  0.8206, Adjusted R-squared:  0.8062
## F-statistic: 57.17 on 2 and 25 DF,  p-value: 4.715e-10
```

Adding NUGV increases adjusted R-squared significantly to 0.8092. So, this varivable also explains a significant portion of the data and should be added to the model.

```
model_with_GNPA_NUGV_NLPG <- lm(UGS~GNPA+NUGV+NLPG, SalesUGS)
summary(model_with_GNPA_NUGV_NLPG)

##
## Call:
## lm(formula = UGS ~ GNPA + NUGV + NLPG, data = SalesUGS)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -138411  -34379   -6271   38182  119783
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.508e+06  4.602e+05   5.449 1.34e-05 ***
## GNPA         4.183e-02  5.735e-03   7.294 1.55e-07 ***
## NUGV        -3.045e-01  1.151e-01  -2.645   0.0142 *
## NLPG        -4.578e-02  1.232e-01  -0.372   0.7134
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 67670 on 24 degrees of freedom
##   (4 observations deleted due to missingness)
## Multiple R-squared:  0.8216, Adjusted R-squared:  0.7993
## F-statistic: 36.84 on 3 and 24 DF,  p-value: 3.827e-09
```

When NLPG is added to the model, adjusted R-squared decreases slightly and NLPG is not as significant as other variables. So, it's not added to the model.

```
model_with_GNPA_NUGV_NDGV <- lm(UGS~GNPA+NUGV+NDGV, SalesUGS)
summary(model_with_GNPA_NUGV_NDGV)

##
## Call:
## lm(formula = UGS ~ GNPA + NUGV + NDGV, data = SalesUGS)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -89620 -47745   7755  34704  87358
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.522e+06  1.801e+05  14.005 4.79e-13 ***
## GNPA         4.172e-02  4.513e-03   9.244 2.23e-09 ***
## NUGV        -7.236e-01  1.042e-01  -6.942 3.53e-07 ***
## NDGV         7.066e+03  1.835e+03   3.850 0.000769 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 53360 on 24 degrees of freedom
##   (4 observations deleted due to missingness)
```

```
## Multiple R-squared:  0.8891, Adjusted R-squared:  0.8752
## F-statistic: 64.12 on 3 and 24 DF,  p-value: 1.325e-11
```

Adding NDGV increased adjusted R-squared significantly to 0.8752 and is very significant. So, this variable should be added to the model.

```
model_with_GNPA_NUGV_NDGV_PU <- lm(UGS~GNPA+NUGV+NDGV+PU, SalesUGS)
summary(model_with_GNPA_NUGV_NDGV_PU)

##
## Call:
## lm(formula = UGS ~ GNPA + NUGV + NDGV + PU, data = SalesUGS)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -79451 -47728   8834  37405  85321
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.481e+06  1.872e+05  13.256 2.96e-12 ***
## GNPA         4.209e-02  4.560e-03   9.231 3.38e-09 ***
## NUGV        -6.858e-01  1.138e-01  -6.027 3.81e-06 ***
## NDGV         6.958e+03  1.850e+03   3.760  0.00102 **
## PU          -2.435e+02  2.853e+02  -0.853  0.40231
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 53660 on 23 degrees of freedom
##   (4 observations deleted due to missingness)
## Multiple R-squared:  0.8925, Adjusted R-squared:  0.8738
## F-statistic: 47.73 on 4 and 23 DF,  p-value: 8.199e-11
```

Adding PU to the model decreased adjusted R-squared slightly and PU is not a significant variable by itself. So, it's not added to the model.

```
model_with_GNPA_NUGV_NDGV_PG <- lm(UGS~GNPA+NUGV+NDGV+PG, SalesUGS)
summary(model_with_GNPA_NUGV_NDGV_PG)

##
## Call:
## lm(formula = UGS ~ GNPA + NUGV + NDGV + PG, data = SalesUGS)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -77051 -45800   5524  40477  85004
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.398e+06  2.041e+05  11.750 3.36e-11 ***
## GNPA         4.225e-02  4.485e-03   9.421 2.33e-09 ***
## NUGV        -6.595e-01  1.154e-01  -5.713 8.10e-06 ***
```

```
## NDGV         6.947e+03  1.818e+03   3.821 0.000876 ***
## PG          -4.486e+02  3.627e+02  -1.237 0.228723
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 52780 on 23 degrees of freedom
##   (4 observations deleted due to missingness)
## Multiple R-squared:  0.896,  Adjusted R-squared:  0.8779
## F-statistic: 49.53 on 4 and 23 DF,  p-value: 5.617e-11
```

Adding PG to the model increased adjusted R-squared slightly. So, it's not added to the model.

```
model_with_GNPA_NUGV_NDGV_Trend<- lm(UGS~GNPA+NUGV+NDGV+Trend,
SalesUGS)
summary(model_with_GNPA_NUGV_NDGV_Trend)

##
## Call:
## lm(formula = UGS ~ GNPA + NUGV + NDGV + Trend, data = SalesUGS)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -83173 -36217  -3660  32733  83647
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.479e+06  6.814e+05   2.171   0.0405 *
## GNPA         4.145e-02  4.381e-03   9.461 2.15e-09 ***
## NUGV        -4.071e-01  2.241e-01  -1.816   0.0824 .
## NDGV         5.471e+03  2.046e+03   2.674   0.0136 *
## Trend       -8.780e+03  5.548e+03  -1.583   0.1272
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 51760 on 23 degrees of freedom
##   (4 observations deleted due to missingness)
## Multiple R-squared:    0.9,  Adjusted R-squared:  0.8826
## F-statistic: 51.73 on 4 and 23 DF,  p-value: 3.602e-11

model_with_GNPA_NUGV_NDGV_Trend_Seasonality<-
lm(UGS~GNPA+NUGV+NDGV+Trend+Season, SalesUGS)
summary(model_with_GNPA_NUGV_NDGV_Trend_Seasonality)

##
## Call:
## lm(formula = UGS ~ GNPA + NUGV + NDGV + Trend + Season, data =
SalesUGS)
##
## Residuals:
##    Min     1Q Median     3Q    Max
```

```
## -71054 -37897   1192  33312  86526
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.668e+06  7.007e+05   2.380  0.02642 *
## GNPA         3.898e-02  4.925e-03   7.915 7.04e-08 ***
## NUGV        -4.755e-01  2.321e-01  -2.049  0.05259 .
## NDGV         5.890e+03  2.075e+03   2.839  0.00955 **
## Trend       -7.286e+03  5.697e+03  -1.279  0.21421
## Season       1.116e+04  1.031e+04   1.082  0.29076
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 51570 on 22 degrees of freedom
##   (4 observations deleted due to missingness)
## Multiple R-squared:  0.905,  Adjusted R-squared:  0.8834
## F-statistic: 41.93 on 5 and 22 DF,  p-value: 1.594e-10
```

Adding Trend and Seasonality as variables to the model increased adjusted R-squared significantly. So, they're added to the model.

```
model_with_GNPA_NUGV_NDGV_PG_Trend_Seasonality_Lag1<-
lm(UGS~GNPA+NUGV+NDGV+PG+Trend+Season+Lagged1, SalesUGS)
summary(model_with_GNPA_NUGV_NDGV_PG_Trend_Seasonality_Lag1)

##
## Call:
## lm(formula = UGS ~ GNPA + NUGV + NDGV + PG + Trend + Season +
##     Lagged1, data = SalesUGS)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -40906 -16165  -1117  18770  42733
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.750e+06  4.438e+05   6.197 5.93e-06 ***
## GNPA         3.445e-02  2.769e-03  12.441 1.41e-10 ***
## NUGV        -7.538e-01  1.596e-01  -4.724 0.000148 ***
## NDGV         9.722e+03  1.467e+03   6.628 2.43e-06 ***
## PG          -6.302e+02  2.152e+02  -2.928 0.008624 **
## Trend       -1.008e+04  3.566e+03  -2.827 0.010770 *
## Season       5.753e+04  8.498e+03   6.769 1.82e-06 ***
## Lagged1     -5.938e-01  8.110e-02  -7.322 6.09e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 27620 on 19 degrees of freedom
##   (5 observations deleted due to missingness)
```

```
## Multiple R-squared:  0.9756, Adjusted R-squared:  0.9667
## F-statistic: 108.7 on 7 and 19 DF,  p-value: 5.672e-14
```

Adding a lagged variable of lag one increased adjusted R-squared significantly and increased significance of all variables. So, it's added.

```
model_with_GNPA_NUGV_NDGV_PG_Trend_Seasonality_Lag1_Lag4<-
lm(UGS~GNPA+NUGV+NDGV+PG+Trend+Season+Lagged4+Lagged1, SalesUGS)
summary(model_with_GNPA_NUGV_NDGV_PG_Trend_Seasonality_Lag1_Lag4)

##
## Call:
## lm(formula = UGS ~ GNPA + NUGV + NDGV + PG + Trend + Season +
##     Lagged4 + Lagged1, data = SalesUGS)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -38914 -16151  -5285  21708  42628
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.000e+06  8.187e+05   3.664 0.002303 **
## GNPA         3.748e-02  7.848e-03   4.775 0.000246 ***
## NUGV        -8.524e-01  3.473e-01  -2.454 0.026809 *
## NDGV         1.111e+04  3.932e+03   2.826 0.012771 *
## PG          -6.232e+02  2.518e+02  -2.475 0.025748 *
## Trend       -1.185e+04  5.975e+03  -1.984 0.065898 .
## Season       6.288e+04  1.311e+04   4.796 0.000236 ***
## Lagged4     -1.033e-01  1.962e-01  -0.526 0.606289
## Lagged1     -6.342e-01  1.318e-01  -4.814 0.000228 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 30050 on 15 degrees of freedom
##   (8 observations deleted due to missingness)
## Multiple R-squared:  0.9616, Adjusted R-squared:  0.9411
## F-statistic: 46.91 on 8 and 15 DF,  p-value: 3.096e-09
```

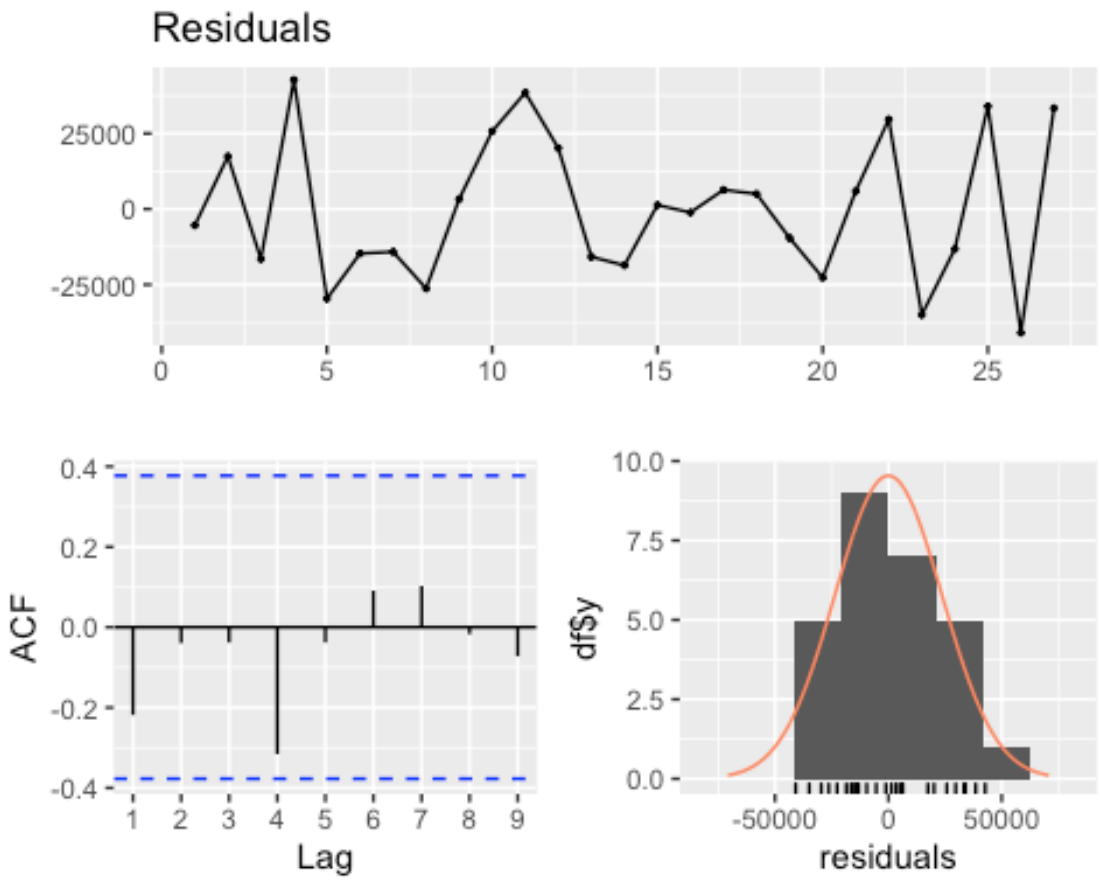
Adding a lagged four variable decreased the adjusted R-squared and the significance of variables. So, it's not added.

###5. Final Model Analysis

```
model<- lm(UGS~GNPA+NUGV+NDGV+PG+Trend+Season+Lagged1, SalesUGS)
summary(model)

##
## Call:
## lm(formula = UGS ~ GNPA + NUGV + NDGV + PG + Trend + Season +
##     Lagged1, data = SalesUGS)
##
```

```
## Residuals:
##    Min     1Q Median     3Q    Max
## -40906 -16165  -1117  18770  42733
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.750e+06  4.438e+05   6.197 5.93e-06 ***
## GNPA         3.445e-02  2.769e-03  12.441 1.41e-10 ***
## NUGV        -7.538e-01  1.596e-01  -4.724 0.000148 ***
## NDGV         9.722e+03  1.467e+03   6.628 2.43e-06 ***
## PG          -6.302e+02  2.152e+02  -2.928 0.008624 **
## Trend       -1.008e+04  3.566e+03  -2.827 0.010770 *
## Season       5.753e+04  8.498e+03   6.769 1.82e-06 ***
## Lagged1     -5.938e-01  8.110e-02  -7.322 6.09e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 27620 on 19 degrees of freedom
##   (5 observations deleted due to missingness)
## Multiple R-squared:  0.9756, Adjusted R-squared:  0.9667
## F-statistic: 108.7 on 7 and 19 DF,  p-value: 5.672e-14
```

```
checkresiduals(model)
```



```
##
##  Breusch-Godfrey test for serial correlation of order up to 11
##
## data:  Residuals
## LM test = 13.277, df = 11, p-value = 0.2756
```

The final model has variables GNPA, NUGV, NDGV, PG, Trend, Seasonality and lagged variable with lag one. This model's adjusted R-squared value is 0.9667, which is a really good result. The model explains 97.56% of the data and all of the variables in the model show high significance values with low p values of the t test. These show that the model is generally very good at explaining the data and all the variables in the model are significant. In the residual analysis, autocorrelation function shows that there're no lags that have relatively high correlation. And the autocorrelation function shows that the autocorrelation in different lags show no pattern and seems to be random and within the limits. This proves that the errors are independent from each other and show no seasonality. The normality assumption of the error term also holds as it can be seen in the histogram of the residuals that is close the normal and is not lagged to one side. The residuals are distributed around mean zero at all times and the variance seem to be constant and doesn't show a pattern. So, the residual analysis show that the assumptions of the regression model hold in this model.

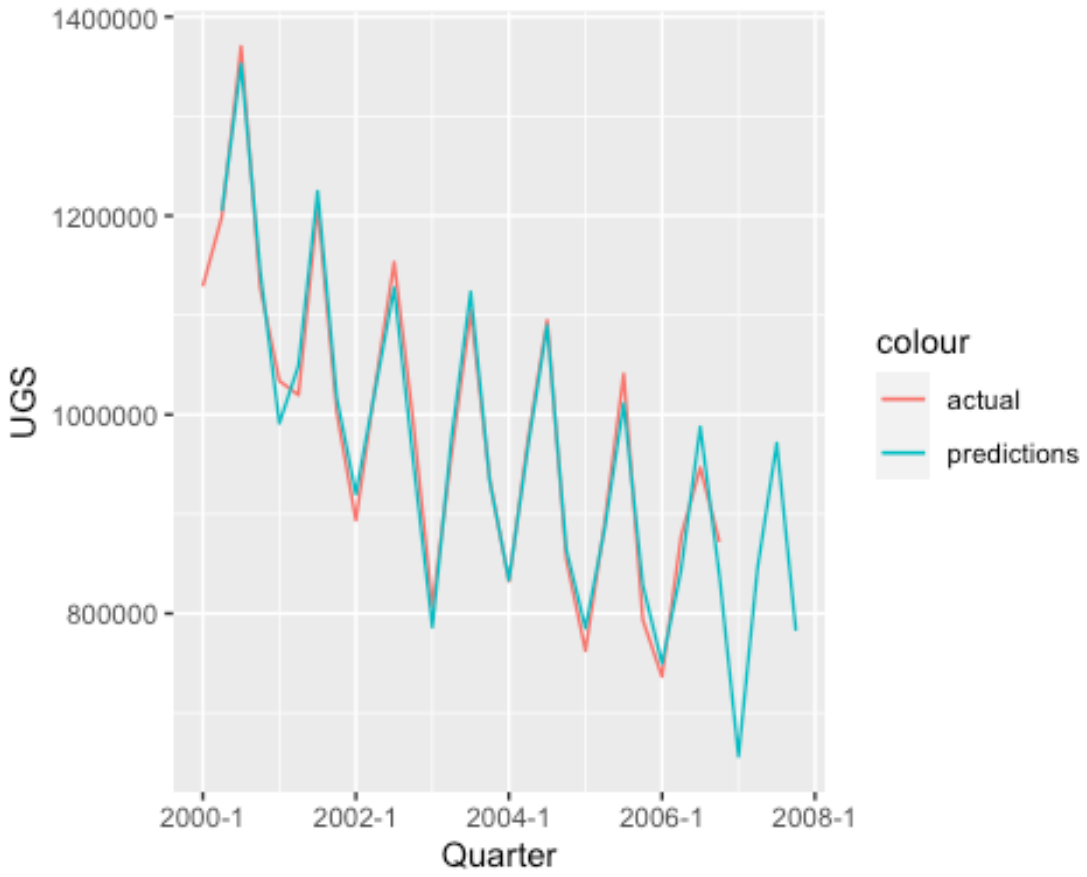
###6. Forecasting

```
tmp=
SalesUGS[29:32,c("GNPA","NUGV","NDGV","PG","Trend","Season","Lagged1")]
predictions = rep(0,4)

predictions[1] = predict(model,tmp[1,])
tmp[2,"Lagged1"] = predictions[1]

predictions[2] = predict(model,tmp[2,])
tmp[3,"Lagged1"] = predictions[2]

predictions[3] = predict(model,tmp[3,])
tmp[4,"Lagged1"] = predictions[3]

predictions[4] = predict(model,tmp[4,])

Predicted <-
data.table(cbind(1:32,c(NA,model$fitted.values,predictions)))
SalesUGS[,Prediction:=Predicted[,2]]
SalesUGS$Prediction[29:32]

## [1] 656158.4 846216.0 971297.7 782610.3

tail(SalesUGS)
```

```
##    Quarter    UGS   RNUV    NLPG     PU     PG    NUGV     NDGV
GNPA    GNPC
## 1: 2006 Q3 946783 0.0108 1653250 565.19 449.19 5754077 333.7144
6800688 7124204
## 2: 2006 Q4 872000 0.0133 1696000 565.19 449.19 5825866 339.2281
2303373 6093090
## 3: 2007 Q1     NA 0.0074 1715000 565.19 449.19 5869018 342.1729
1132973 4857305
## 4: 2007 Q2     NA 0.0106 1725300 565.19 449.19 5931348 346.9407
1570703 5852404
## 5: 2007 Q3     NA 0.0101 1751050 565.19 449.19 5991280 351.4449
7140722 7480414
## 6: 2007 Q4     NA 0.0124 1797400 565.19 449.19 6065597 357.2902
2418541 6397745
##        GNPT Trend Season Lagged1 Lagged4 Prediction
## 1: 34992138    27      3  877614 1040946   987689.2
## 2: 29867726    28      4  946783  793399   838630.0
## 3: 24413807    29      1  872000  736580   656158.4
## 4: 27597857    30      2      NA  877614   846216.0
## 5: 36741745    31      3      NA  946783   971297.7
## 6: 31361112    32      4      NA  872000   782610.3
```

For 2007 quarter 1, it's predicted that UGS will be 656158.4. For 2007 quarter 2, it's predicted that UGS will be 846216.0.2.4.For 2007 quarter 3, it's predicted that UGS will be 971297.7.For 2007 quarter 4, it's predicted that UGS will be 782610.3.

```
ggplot(SalesUGS ,aes(x=Quarter)) +
  geom_line(aes(y=UGS,color='actual', group = 1)) +
  geom_line(aes(y=Prediction, color = 'predictions', group = 1))

## Warning: Removed 4 row(s) containing missing values (geom_path).

## Warning: Removed 1 row(s) containing missing values (geom_path).
```




The predictions for the year 2007 seems to be in line with the previous data.