Time Series Analysis

Seasonal and non-seasonal

time series analysis and forecasting

Project Report
Under the Guidance of
Prof. Hadi Safari Katesari
By

Amit Singh [10465317]

Stevens Institute of Technology

Project 1— Motor Vehicle Retail Sales: Domestic Autos (Seasonal)

1. Introduction and Motivation

Sales forecasting is essential for any company to plan and execute according to a future sale. 'Time' is the most crucial factor which ensures success in a business. Sales prediction for revenue planning is essential so the resource can be managed according. It has been proven that forecasting sales can help efficient resource utilization and money management. Forecasting motor vehicles can give a company idea about the unit needed to produce during that month of the year.

Project Definition

This analysis aims to provide a forecast based on the latest available data to reflect the current conditions in the domestic auto sales by information management about the demand for motor vehicles that season for planning and development. The aim is to develop a model that can predict domestic auto sales.

Scope of this Project
I will use the ARIMA model using R. ARIMA models, which are popular and flexible forecasting models that utilize historical information to make predictions. This model is a basic forecasting technique that can be used as a foundation for more complex models. This project examines the time series for predicting domestic auto sales for 2019 in the U.S. by fitting an ARIMA model and creating a forecast.

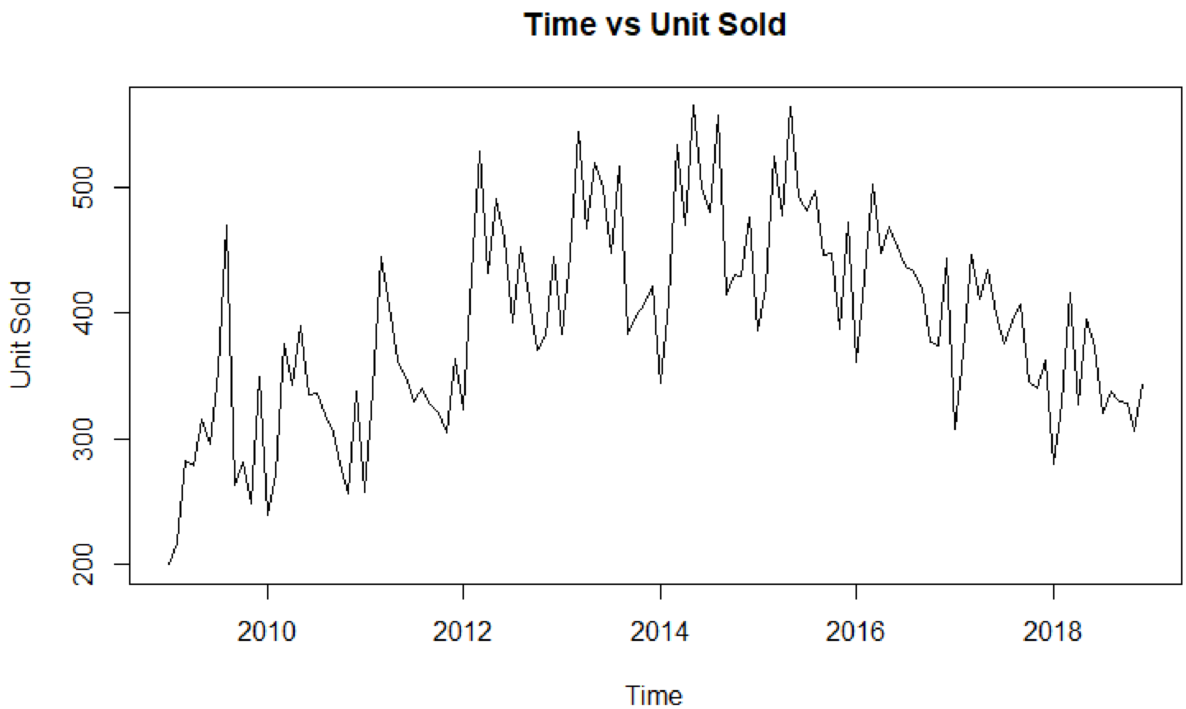
2. Data Description

Date Range: From 2009-01 to 2019-12
Frequency: Monthly
Data source Description: The dataset contains details of Date and the number of units sold monthly. Autos are all passenger cars, including station wagons. Domestic sales are all United States (U.S.) sales of vehicles assembled in the U.S., Canada, and Mexico. The data set consists of 132 rows, i.e., 132 months of data over ten years, and two columns (Date and Unit Sold):
Date: This contains the month of the year. (YYYY-mm).
DAUTONSA: This contains the number of units sold monthly.

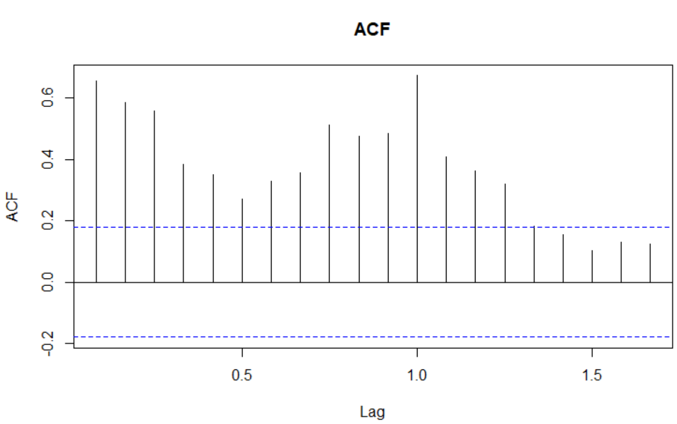
Data Source: https://fred.stlouisfed.org/series/DAUTONSA

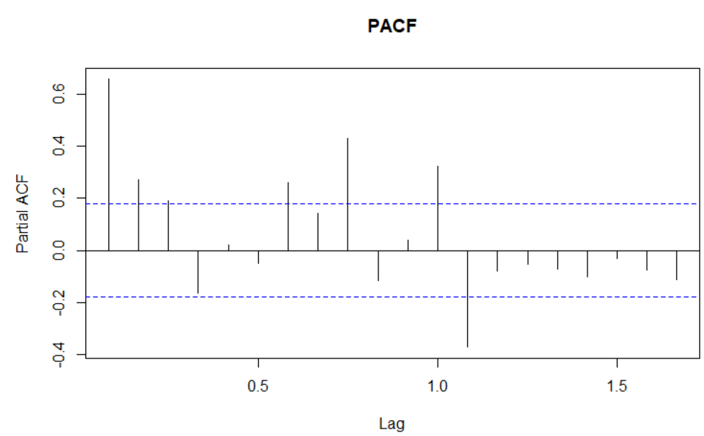
3. Time vs Unit Sold

First, let's understand the pattern unit of vehicles sold per month. We can see some monthly patterns with some upward and then downward trends in time series, as shown in Figure 1. As data is monthly, I used frequency =12. Contain no NA values.



ACF AND PACF of series.





ACF Showing pattern which suggest it is series is nonstationary.

4. Stationarity Test

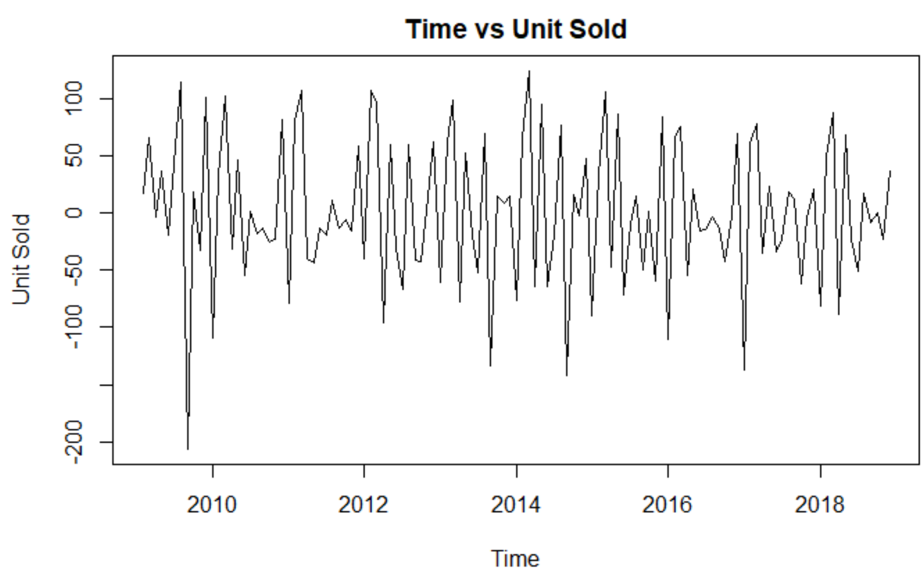
Next, we will perform the Augmented Dickey Fuller test to see if the trend is stationary or non-stationary.

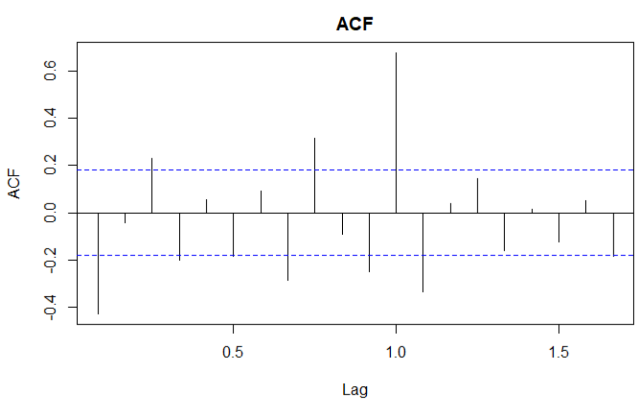
```
adf.test(train_ts)
```

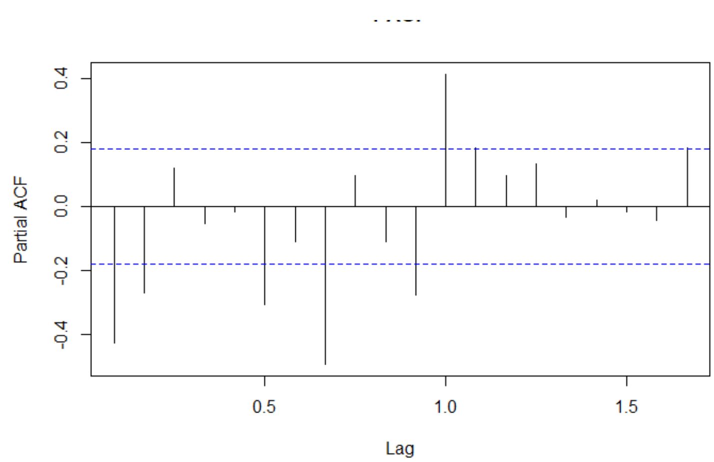
Augmented Dickey-Fuller Test
##
data: train_ts
Dickey-Fuller = -2.6963, Lag order = 4, p-value = 0.2874
alternative hypothesis: stationary

As we can observe Augmented Dickey fuller test has a P value of greater than 0.05 which seems to suggest that the time series is non-stationary. We need to take difference of lag to make it stationary. We need take seasonal difference also as these is seasonality in series.

The ACF/ PACF/EACF plots after taking difference of time series







Plot after taking lag difference looks more stationary. Some of the model suggested and we can check after interpreting ACF and PACF is ARIMA (2,1,3) X (0,1,1) [12], ARIMA (6,1,4) X (0,1,1) [12], ARIMA (0,1,2) X (2,1,1) [12].

```
eacf(train_diff_ts)
```

```
## AR/MA
##   0 1 2 3 4 5 6 7 8 9 10 11 12 13
## 0 x o x x o o o x x o x  x  x  o
## 1 x o x o o x o o x o o  x  x  o
## 2 x o o o o o o o o o o  x  x  o
## 3 x o o x o o o o o o o  x  o  o
## 4 x o o x o o o o o o o  x  o  o
## 5 o o x o o x o o o o o  x  x  o
## 6 x x x o o o o o o o o  x  o  x
## 7 x x o o o o o o o o o  x  o  x
```

The Stationarity Test after taking difference of time series

Augmented Dickey-Fuller Test
##
data: train_diff_ts
Dickey-Fuller = -5.486, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary

Now, we can observe data p-value is smaller than 0.05. Now series is stationary.

5. Model Building

Now as series is stationary, we can select and build model.

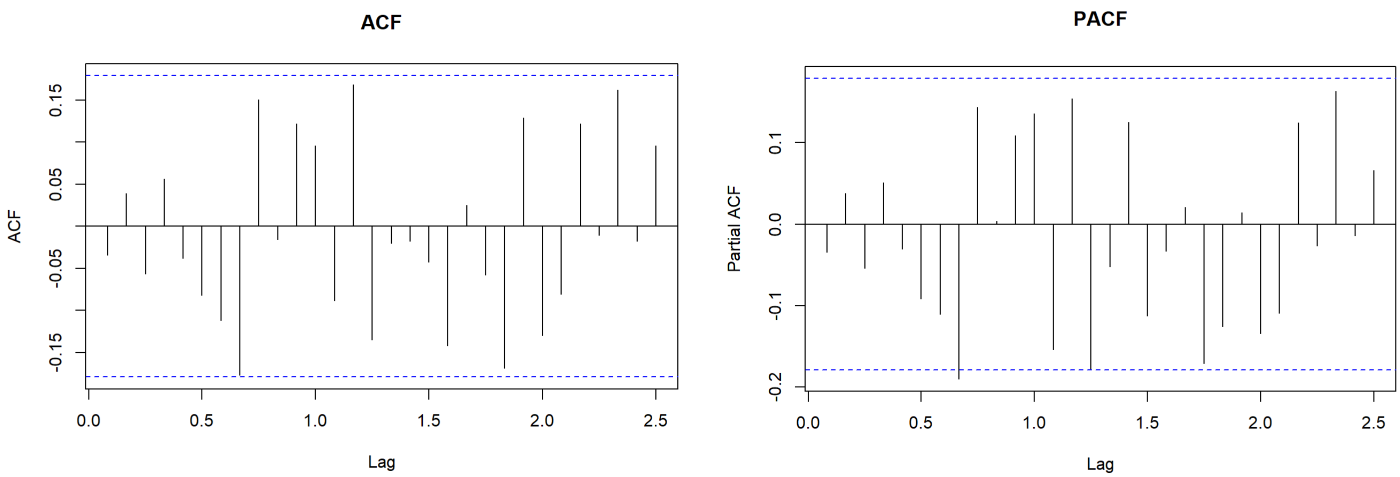
1) ARIMA (2,1,3) X (0,1,1) [12]

```
# ARIMA(2,1,3)X(0,1,1)[12]
model = Arima(train_ts, order=c(2,1,3), seasonal=c(0,1,1))
model
```

```
## Series: train_ts
## ARIMA(2,1,3)(0,1,1)[12]
##
## Coefficients:
##           ar1     ar2      ma1      ma2     ma3     sma1
##       -0.1662  0.6474  -0.3815  -0.8893  0.4000  -0.1902
## s.e.   0.3318  0.2840   0.3561   0.1048  0.2599   0.2390
##
## sigma^2 = 1269:  log likelihood = -531.87
## AIC=1077.74   AICc=1078.87   BIC=1096.45
```

AIC: 1077.74

Plots of ACF and PACF of residual of model ARIMA (2,1,3) X (0,1,1) [12]



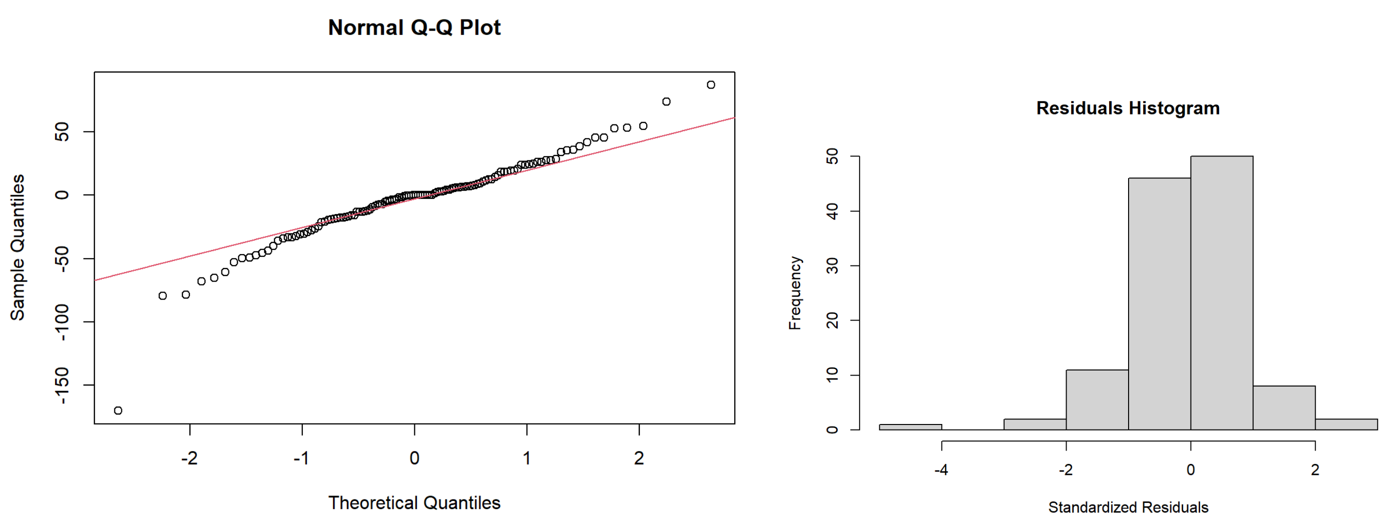
Residuals does not show any significant auto correlation which means that our model is adequately built. Let us further examine the residuals for test of significant autocorrelation by examining performing the Box test.

Box-Ljung test

```
## Box-Ljung test
##
## data:  model$resid
## X-squared = 10.922, df = 10, p-value = 0.3636
```

The P-value of the Box test is high suggesting that the residuals are not auto correlated.

QQ Plot and Histogram of Residual of model ARIMA (2,1,3) X (0,1,1) [12]



The standard assumption in linear regression is that the theoretical residuals are independent and normally distributed. We can see from the above histogram and the qq plot, that the residuals confirm to this assumption of normality.

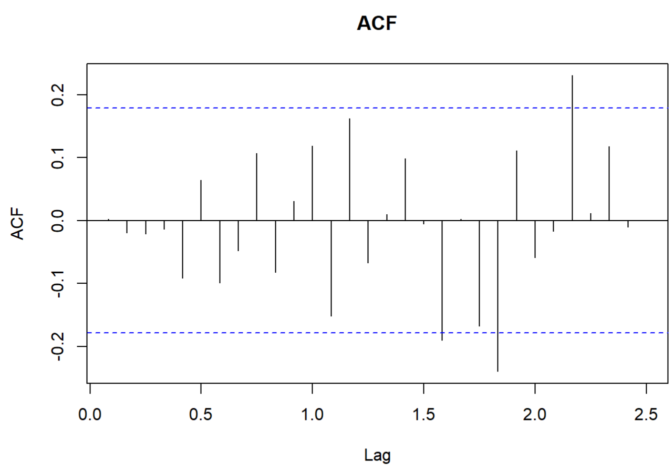
2) ARIMA (6,1,4) X (0,1,1) [12]

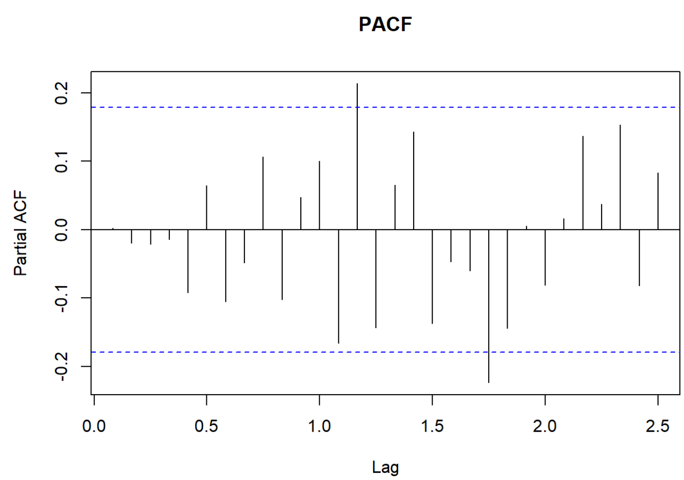
```r
# ARIMA(6,1,4)X(0,1,1)[12]
model2 = Arima(train_ts, order=c(6,1,4), seasonal=c(0,1,1))
model2
```

```
## Series: train_ts
## ARIMA(6,1,4)(0,1,1)[12]
##
## Coefficients:
##          ar1     ar2      ar3      ar4     ar5      ar6      ma1      ma2
##       0.5755  0.3888  -0.8082  -0.0004  0.0535  -0.2220  -1.2225  -0.0704
## s.e.  0.2316  0.1387   0.1571   0.1398  0.1374   0.1358   0.2238   0.2018
##          ma3      ma4     sma1
##       1.2286  -0.6049  -0.3633
## s.e.  0.1719   0.2171   0.2555
##
## sigma^2 = 1116:  log likelihood = -525.63
## AIC=1075.26   AICc=1078.58   BIC=1107.34
```

AIC: 1075.26

Plots of ACF and PACF of residual of model ARIMA (6,1,4) X (0,1,1) [12]





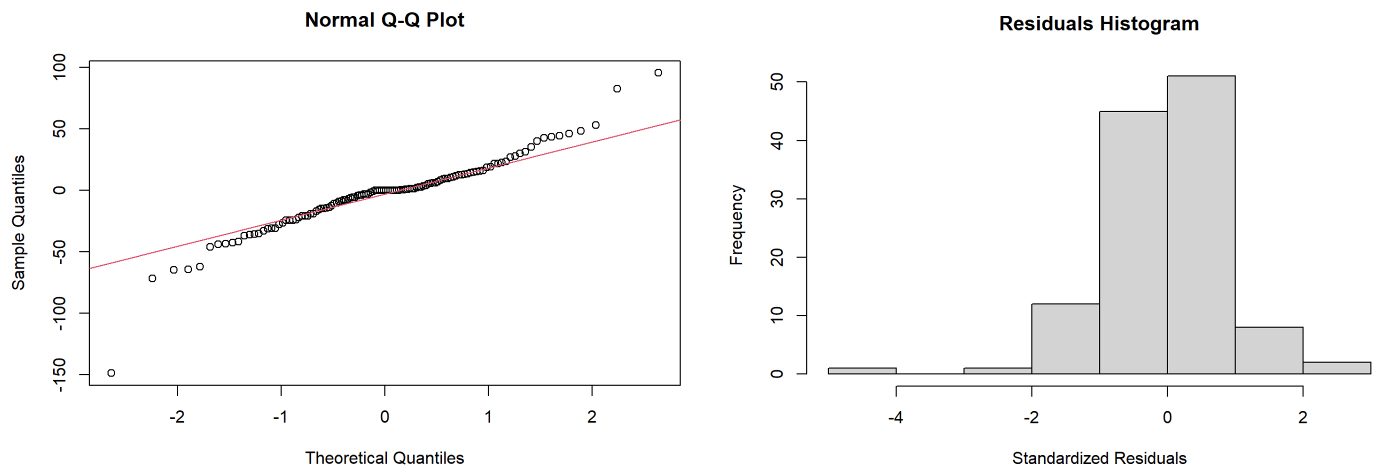
Residuals does not show any significant auto correlation which means that our model is adequately built. Looks better than ARIMA (2,1,3) X (0,1,1) [12]

Box-Ljung Test for ARIMA (6,1,4) X (0,1,1) [12]

Box-Ljung test
##
data: model2$resid
X-squared = 5.7126, df = 10, p-value = 0.8388

The P-value of the Box test is high suggesting that the residuals are not auto correlated.

QQ Plot and Histogram of Residual of model ARIMA (6,1,4) X (0,1,1) [12]



We can see from the above histogram and the qq plot, that the residuals confirm to this assumption of normality. Looks better than ARIMA (2,1,3) X (0,1,1) [12].

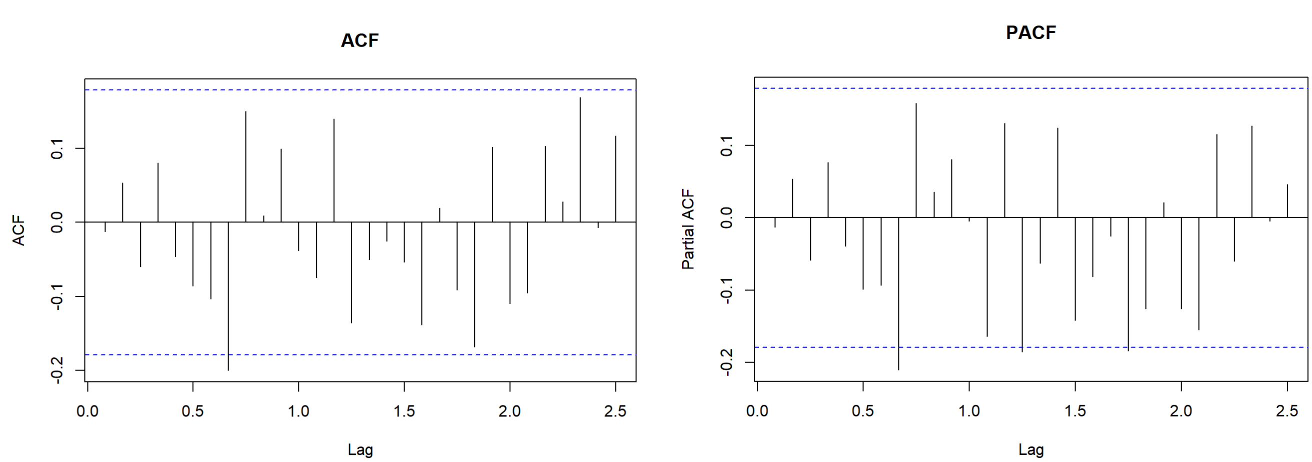
3) ARIMA (2,1,2) X (0,1,0) [12]

```
# ARIMA(2,1,2)X(0,1,0)[12]
model3 = Arima(train_ts, order=c(2,1,2), seasonal=c(0,1,0))
model3
```

```
## Series: train_ts
## ARIMA(2,1,2)(0,1,0)[12]
##
## Coefficients:
##           ar1     ar2     ma1      ma2
##       -0.8192  0.0747  0.2469  -0.7022
## s.e.   0.1780  0.1814  0.1473   0.1429
##
## sigma^2 = 1277:  log likelihood = -532.91
## AIC=1075.81   AICc=1076.41   BIC=1089.18
```

AIC: 1075.81

Plots of ACF and PACF of residual of model ARIMA (2,1,2) X (0,1,0) [12]

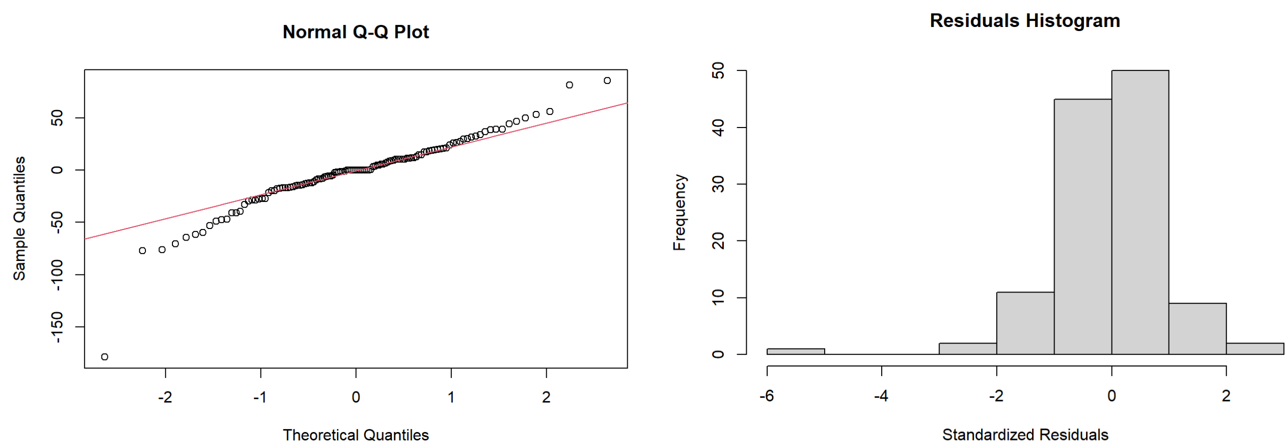


Box-Ljung Test for ARIMA (2,1,2) X (0,1,0) [12]

```
##
## Box-Ljung test
##
## data:  model3$resid
## X-squared = 12.481, df = 10, p-value = 0.2541
```

The P-value of the Box test is more than 0.05 suggesting that the residuals are not auto correlated.

QQ Plot and Histogram of Residual of model ARIMA (2,1,2) X (0,1,0) [12]



We can see from the above histogram and the qq plot, that the residuals confirm to this assumption of normality. Looks better than above two model.

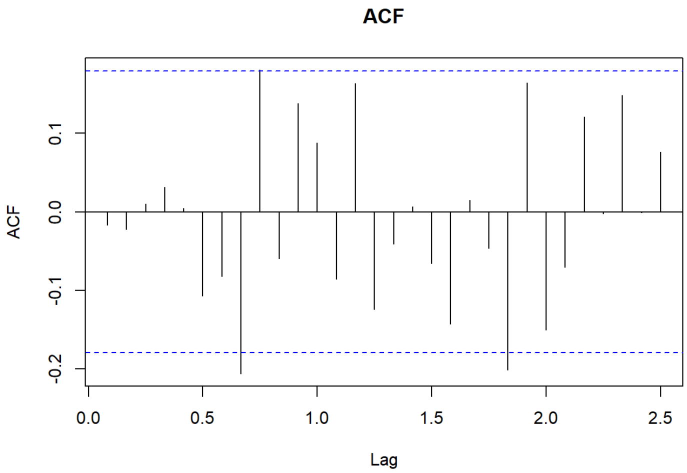
4) ARIMA (0,1,2) X (2,1,1) [12]

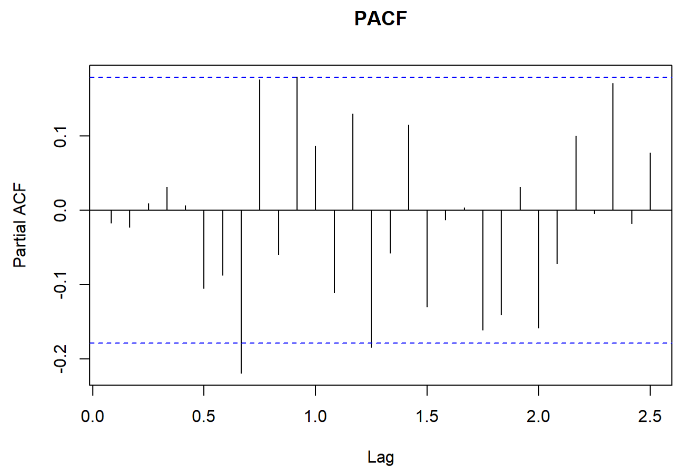
```
# ARIMA(2,1,1)X(0,1,1)[12]
model4 = Arima(train_ts, order=c(2,1,1), seasonal=c(0,1,1))
model4
```

```
## Series: train_ts
## ARIMA(2,1,1)(0,1,1)[12]
##
## Coefficients:
##          ar1     ar2      ma1     sma1
##       0.2460  0.0994  -0.8305  -0.2162
## s.e.  0.1442  0.1229   0.1045   0.2440
##
## sigma^2 = 1274:  log likelihood = -532.94
## AIC=1075.88   AICc=1076.47   BIC=1089.24
```

AIC: 1075.88

Plots of ACF and PACF of residual of model ARIMA (0,1,2) X (2,1,1) [12]



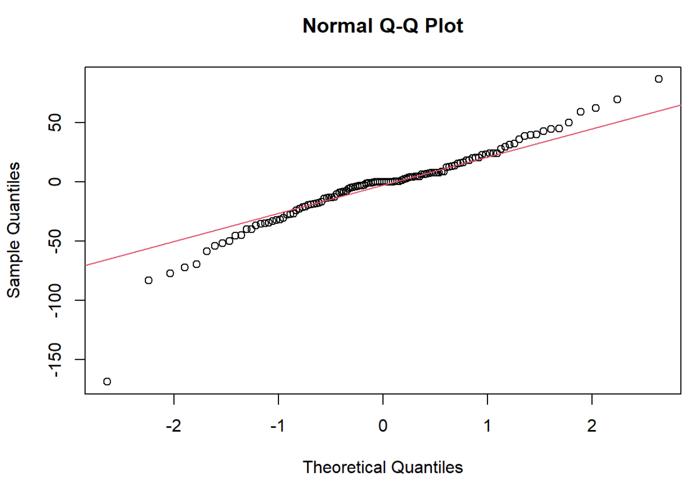


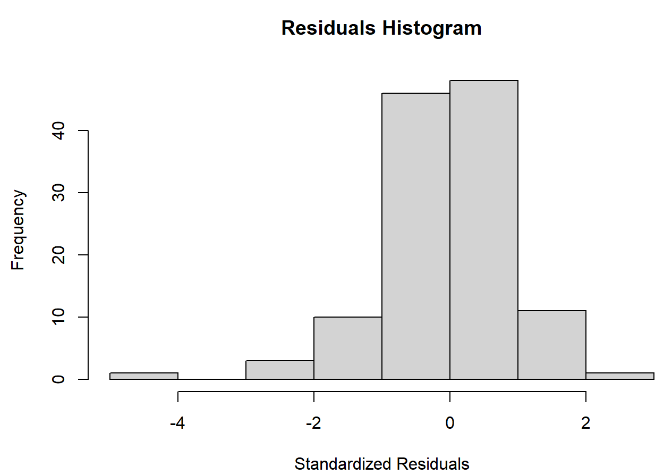
Box-Ljung Test for ARIMA (0,1,2) X (2,1,1) [12]

Box-Ljung test
##
data: model4$resid
X-squared = 12.876, df = 10, p-value = 0.2307

The P-value of the Box test is more than 0.05 suggesting that the residuals are not auto correlated.

QQ Plot and Histogram of Residual of model ARIMA (0,1,2) X (2,1,1) [12]

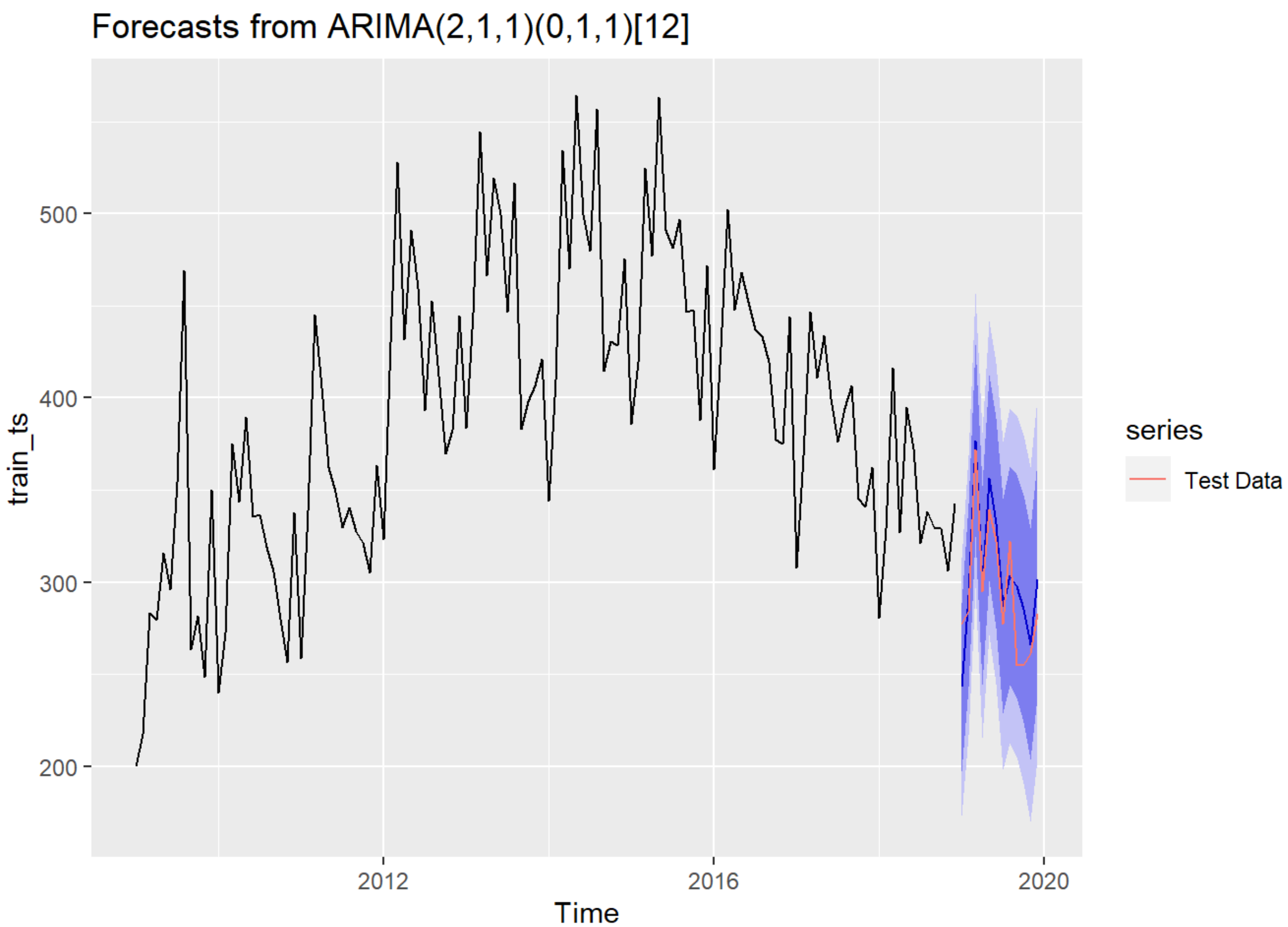




We can see from the above histogram and the qq plot, that the residuals confirm to this assumption of normality.

6. Forecasting

After comparing above models, based on AIC value and normality test, we got ARIMA (2,1,1) X (0,1,1) [12]as best model. Other model may overfit or under fit. We will forecast using ARIMA (2,1,1) X (0,1,1) [12] model.



As we can see forecasted blue line is close to red test or actual data. Therefore, we got best fit with ARIMA (2,1,1) X (0,1,1) [12] model on auto sales data.

Project 2— Reliance Industries Limited Stock Analysis

1. Introduction and Motivation

Stock market forecasting is helpful for traders and investors to know where stock can move in advance. Reliance Industries Limited is the most prominent firm in terms of market cap in India. Knowing forced value can help the trader to make their position in advance. This will help traders and investors to buy and sell stock in the coming days.

Project Definition

This analysis aims to provide a realistic forecast based on the latest available data to reflect the current conditions in the Reliance Industries by information traders about the demand for the stock. The aim is to develop a model that can predict the Reliance Industries stock price for next 50 days.

Scope of this Project

I will use the ARIMA model using R. ARIMA models, a popular and flexible forecasting model class that utilizes historical information to make predictions. This model is a basic forecasting technique that can be used as a foundation for more complex models. This project will examine the time series for predicting stock price.

2. Data Description

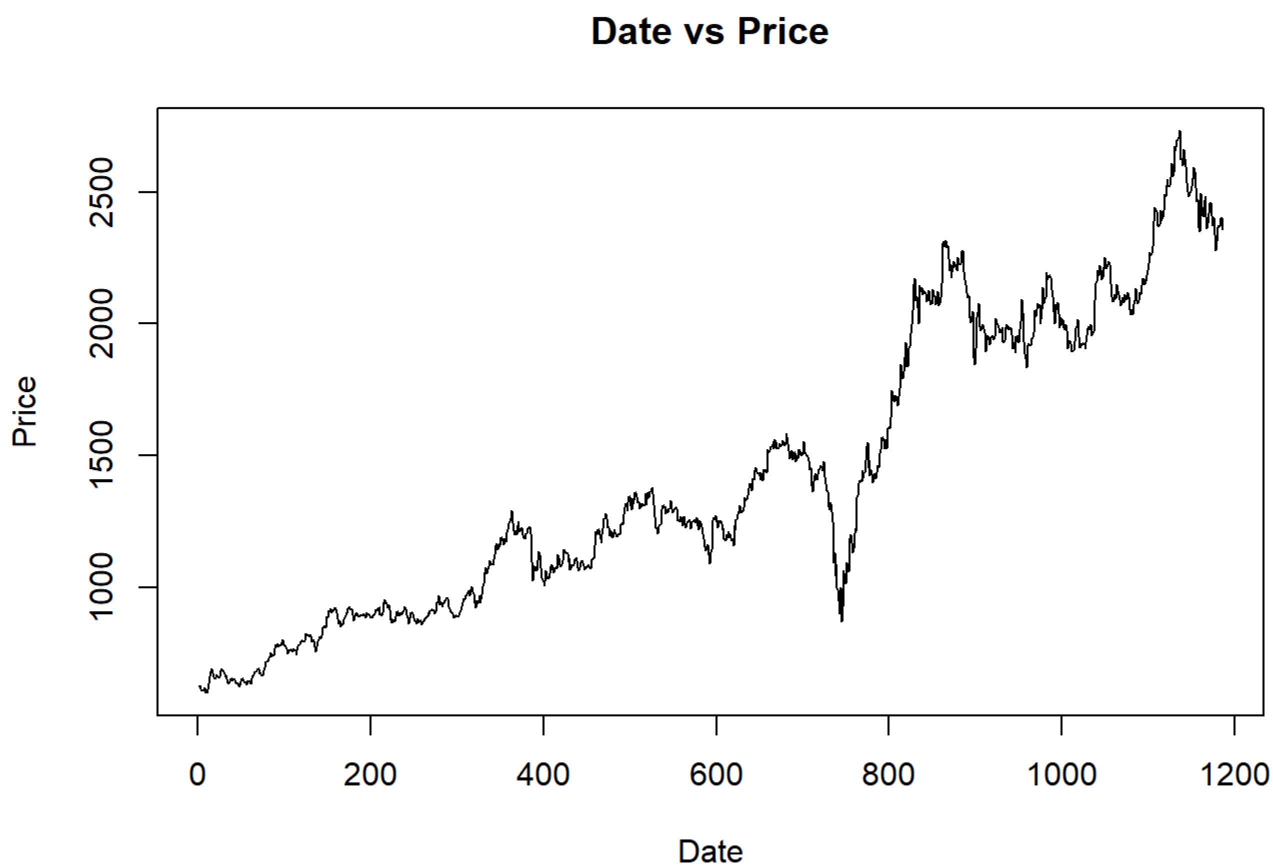
Date Range: From 2017-03-15 to 2022-03-14
Frequency: Daily
Data source Description: The dataset contains details of the date and price of a stock that day. The data set consists of 1236 rows, i.e., 1236 prices over the five years, and seven columns:
Date: Date of stock price
Adj. Close: It is the adjusted closing price of the stock.
It also contains five more columns: High, Low, Open, Close, and volume. But I will not be using this for this project.

DataSource:https://finance.yahoo.com/quote/RELIANCE.NS/history?period1=1615766400&period2=1647302400&interval=1d&filter=history&frequency=1d&includeAdjustedClose=true

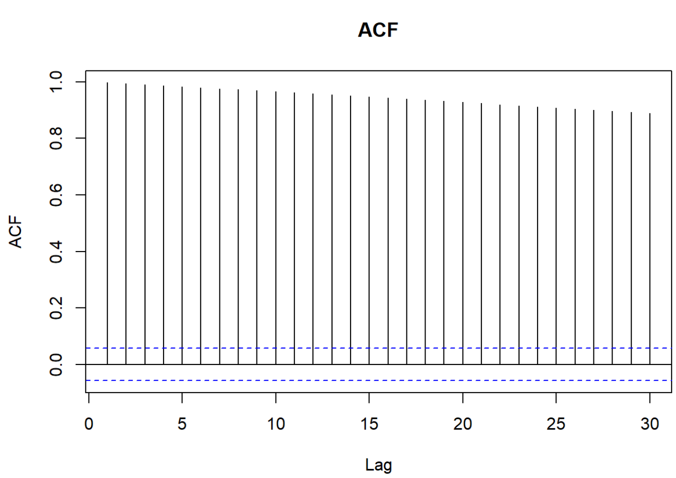
3. Time vs Unit Sold

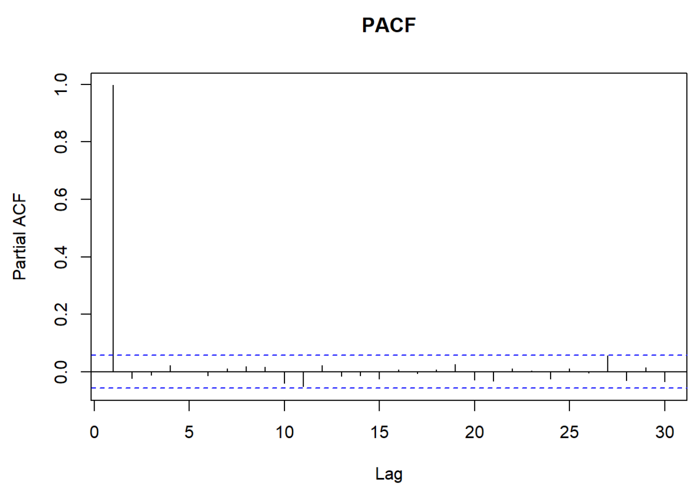
First, let's understand the pattern unit of date vs price. We can see some patterns with upward moment in time series, as shown in Figure 1. As data is daily, I used frequency = 365. Data contain no NA values.

Figure 1



ACF AND PACF of series.





ACF and PACF Showing pattern which suggest it is series is nonstationary.

4. Stationarity Test

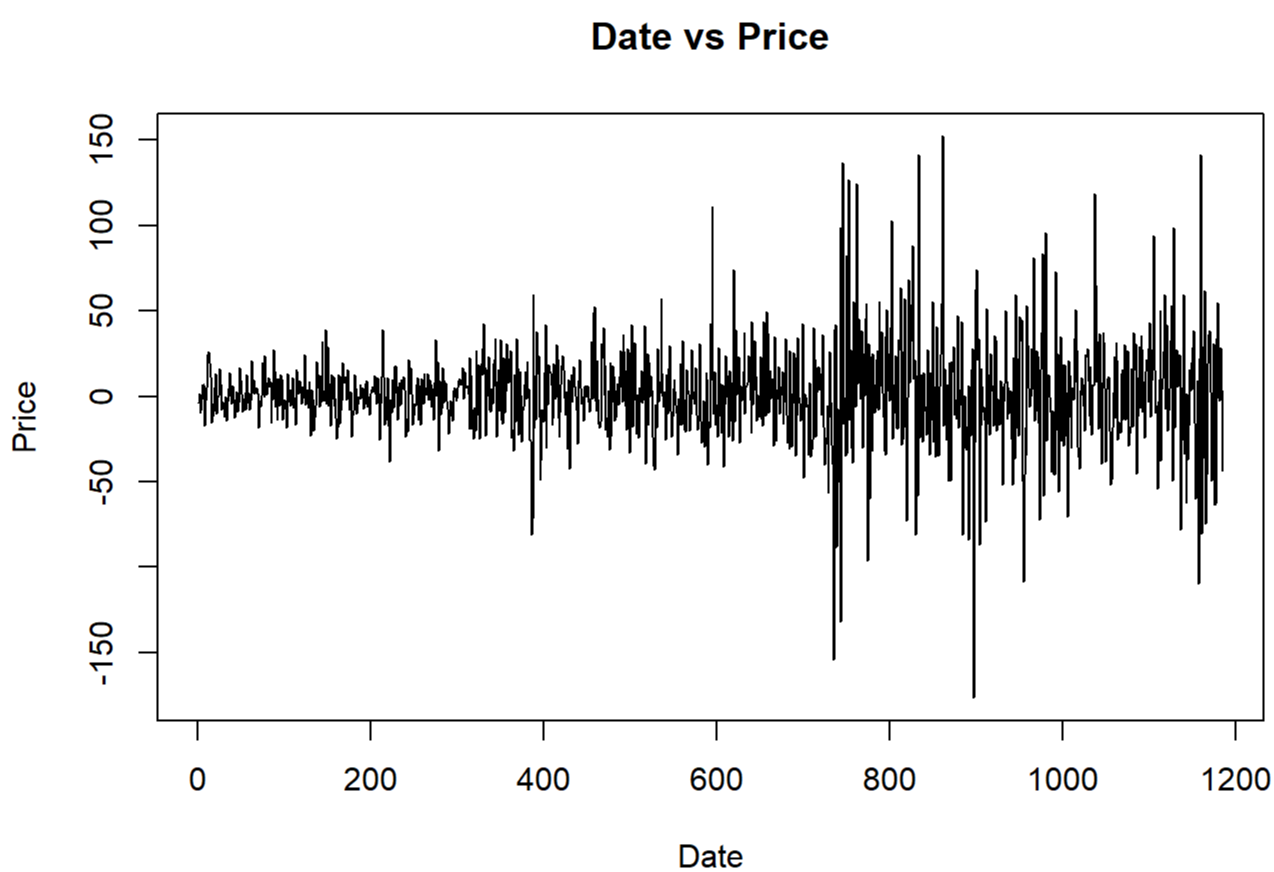
```
adf.test(train_ts)
```

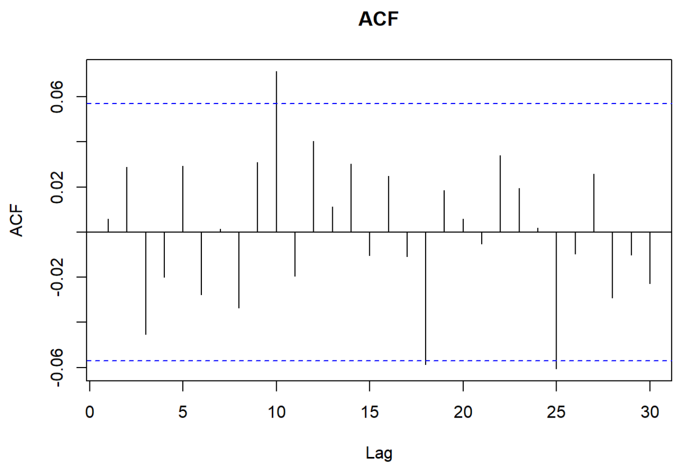
Next, we will perform the Augmented Dickey Fuller test to see if the trend is stationary or non-stationary.

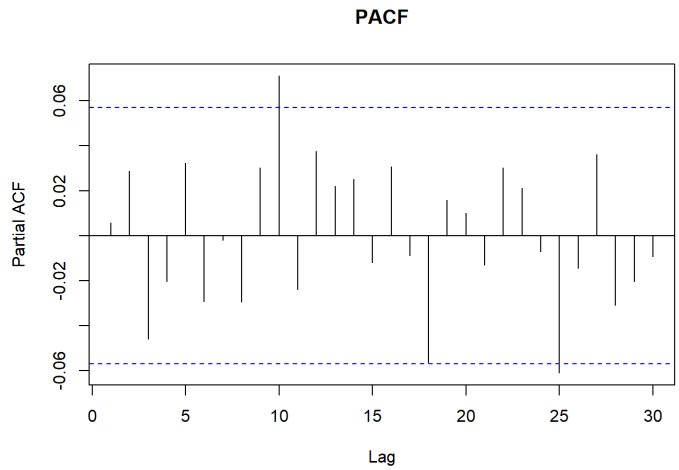
```
## Augmented Dickey-Fuller Test
##
## data:  train_ts
## Dickey-Fuller = -3.091, Lag order = 10, p-value = 0.1165
## alternative hypothesis: stationary
```

As we can observe Augmented Dickey fuller test has a P value of greater than 0.05 which seems to suggest that the time series is non-stationary. We need to take difference of lag to make it stationary.

The ACF/ PACF/EACF plots after taking difference of time series







Plot after taking lag difference looks more stationary. Some of the model suggested and we can check after interpreting ACF and PACF is ARIMA (0,1,0), ARIMA (1,1,0), ARIMA (0,1,1).

```
eacf(train_diff_ts)
```

```
## AR/MA
##   0 1 2 3 4 5 6 7 8 9 10 11 12 13
## 0 o o o o o o o o o x o  o  o  o
## 1 x o o o o o o o o x o  o  o  o
## 2 x x o o o o o o o x o  o  o  o
## 3 x x x o o o o o o x o  o  o  o
## 4 x x x x o o o o o x o  o  o  o
## 5 x x x x x o o o o x o  o  o  o
## 6 x x x o x x o o o x o  o  o  o
## 7 x x x x o x x o o x o  o  o  o
```

The Stationarity Test after taking difference of time series

```
adf.test(train_diff_ts)
```

Augmented Dickey-Fuller Test
##
data: train_diff_ts
Dickey-Fuller = -9.9751, Lag order = 10, p-value = 0.01
alternative hypothesis: stationary

Now, we can observe data p-value is smaller than 0.05. Now series is stationary.

5. Model Building

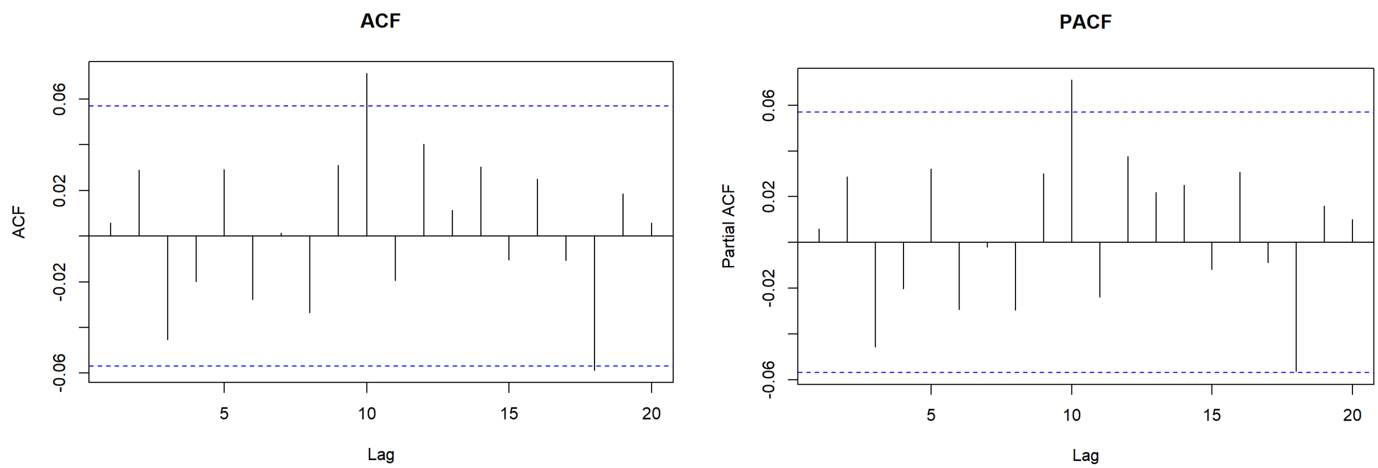
1) Model 1: ARIMA (0,1,0)

```
# ARIMA(0,1,0)
model <- Arima(train_ts,order=c(0,1,0))
model
```

```
## Series: train_ts
## ARIMA(0,1,0)
##
## sigma^2 = 820.8:  log likelihood = -5657.25
## AIC=11316.5   AICc=11316.5   BIC=11321.57
```

AIC: 11316.5

Plots of ACF and PACF of residual of model ARIMA (0,1,0)



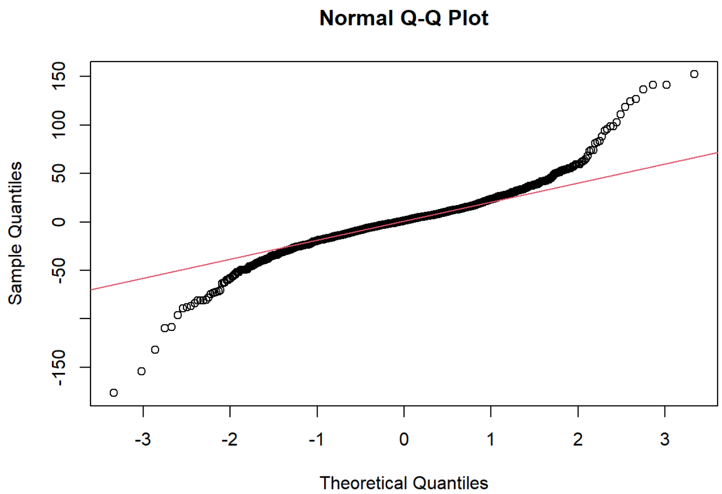
Residuals does not show any significant auto correlation which means that our model is adequately built. These are few one or two lag show significance. Let us further examine the residuals for test of significant autocorrelation by examining performing the Box test.

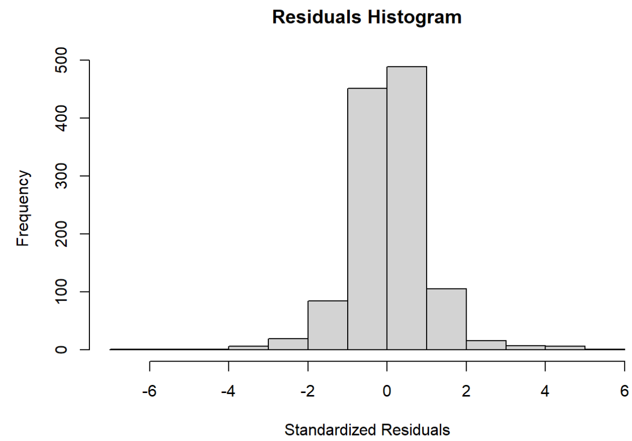
Box-Ljung test for ARIMA (0,1,0)

Box-Ljung test
##
data: model$resid
X-squared = 14.46, df = 10, p-value = 0.153

The P-value of the Box test is high than 0.05 suggesting that the residuals are not auto correlated.

QQ Plot and Histogram of Residual of model ARIMA (0,1,0)





The standard assumption in linear regression is that the theoretical residuals are independent and normally distributed. We can see from the above histogram and the qq plot, that the residuals confirm to this assumption of normality.

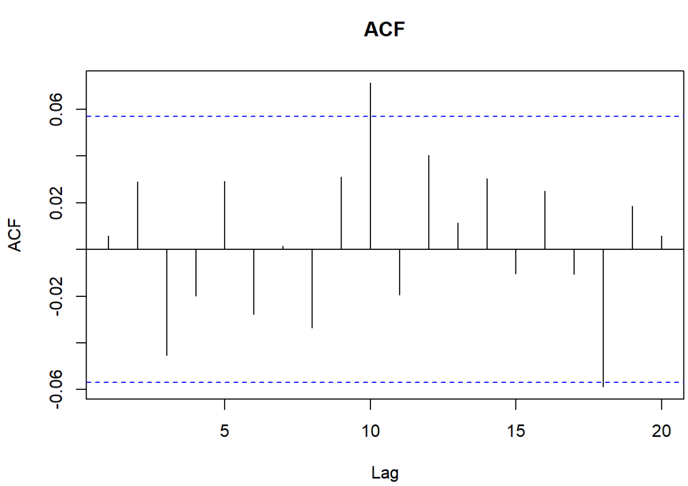
2) Model 2: ARIMA (1,1,0)

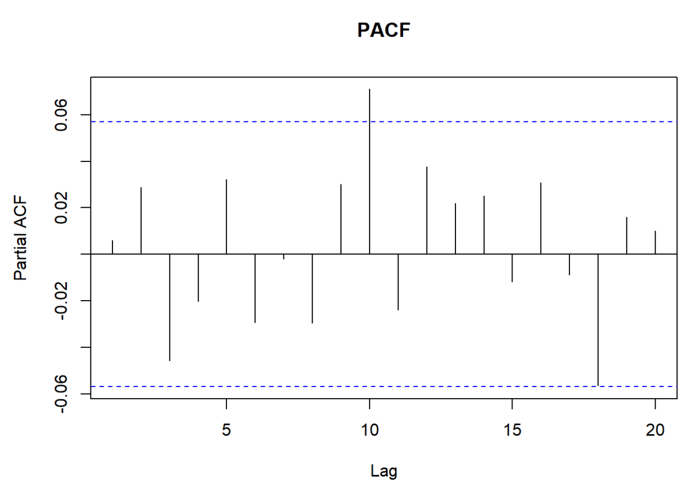
```
# ARIMA(1,1,0)
model2 <- Arima(train_ts,order=c(1,1,0))
model2
```

```
## Series: train_ts
## ARIMA(1,1,0)
##
## Coefficients:
##          ar1
##       0.0085
## s.e.  0.0291
##
## sigma^2 = 821.4:  log likelihood = -5657.2
## AIC=11318.41   AICc=11318.42   BIC=11328.56
```

AIC : 11318.41, not better than ARIMA(0,1,0)

Plots of ACF and PACF of residual of model ARIMA (1,1,0)



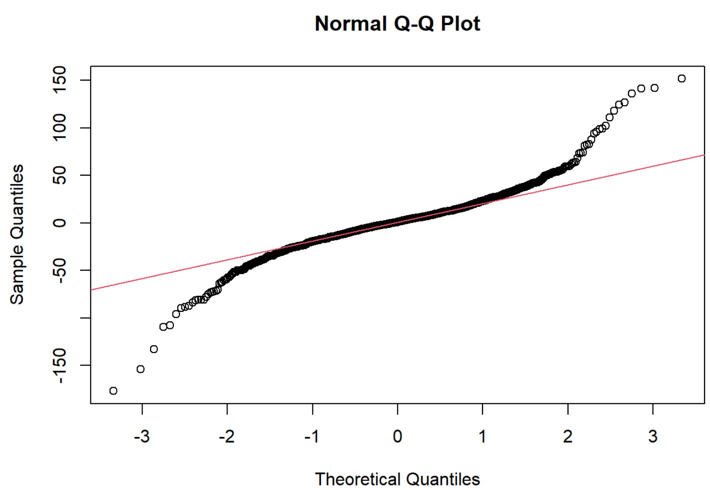


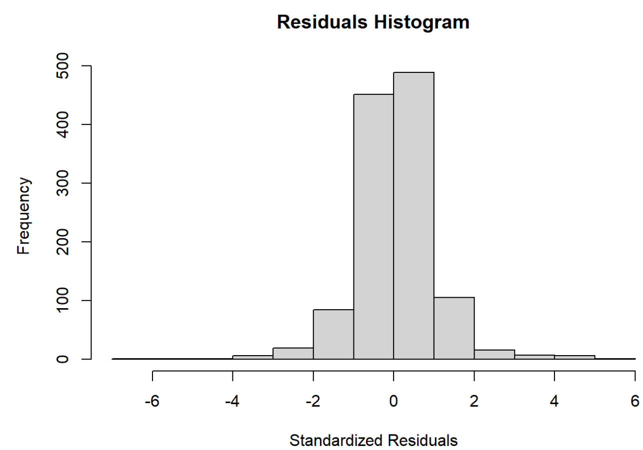
Box-Ljung Test ARIMA (1,1,0)

```
##
##  Box-Ljung test
##
## data:  model2$resid
## X-squared = 14.494, df = 10, p-value = 0.1516
```

The P-value of the Box test is high than 0.05 suggesting that the residuals are not auto correlated.

QQ Plot and Histogram of Residual of model ARIMA (1,1,0)





We can see from the above histogram and the qq plot, that the residuals confirm to this assumption of normality.

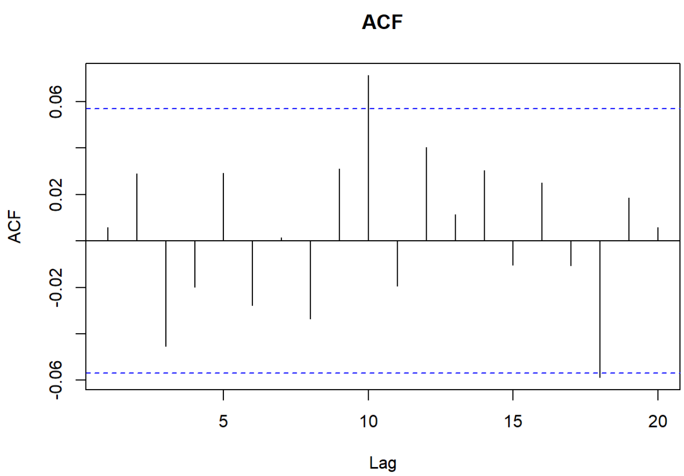
3) Model 3 : ARIMA(0,1,1)

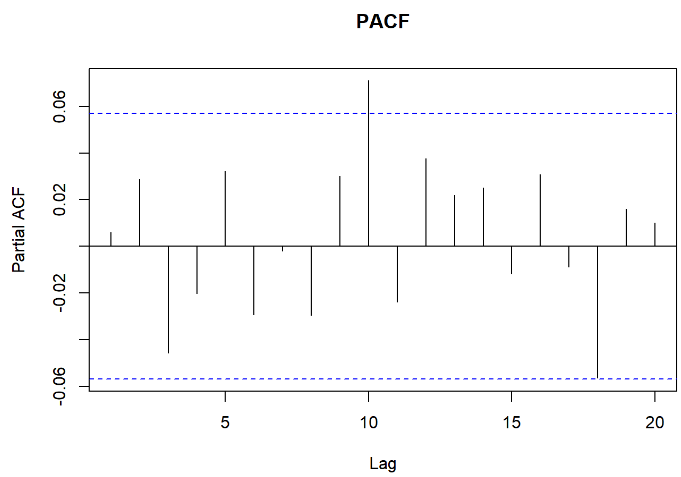
```
# ARIMA(0,1,1)
model3 <- Arima(train_ts,order=c(0,1,1))
model3
```

```
## Series: train_ts
## ARIMA(0,1,1)
##
## Coefficients:
##          ma1
##       0.0080
## s.e.  0.0282
##
## sigma^2 = 821.4:  log likelihood = -5657.21
## AIC=11318.41   AICc=11318.42   BIC=11328.57
```

AIC : 11318.41

Plots of ACF and PACF of residual of model ARIMA (0,1,1)

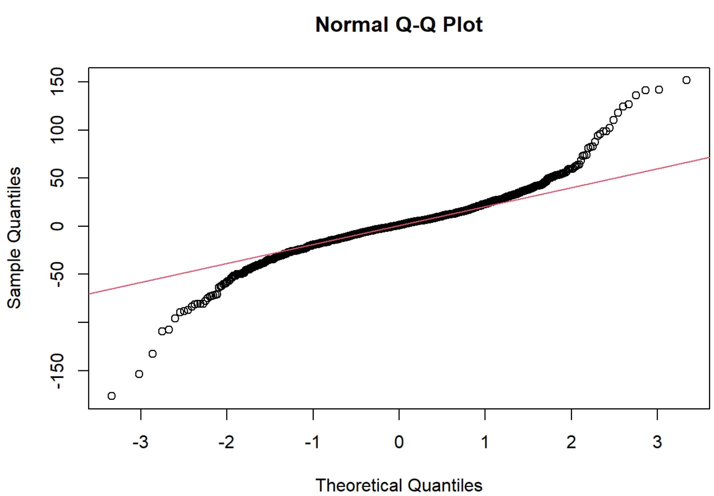


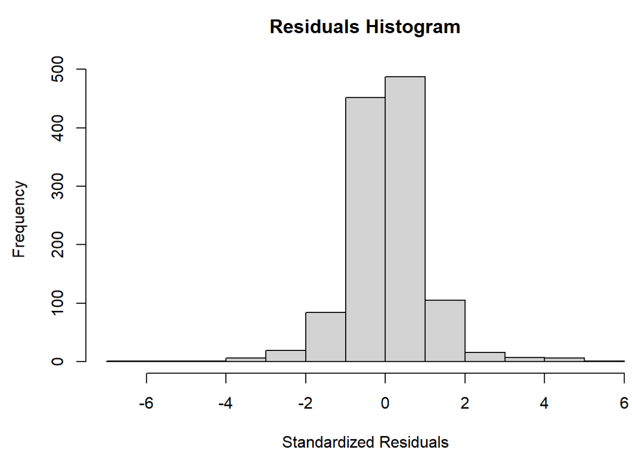


Box-Ljung test for ARIMA (0,1,1)

```
##
##  Box-Ljung test
##
## data:  model3$resid
## X-squared = 14.491, df = 10, p-value = 0.1518
```

QQ Plot and Histogram of Residual of model ARIMA (0,1,1)

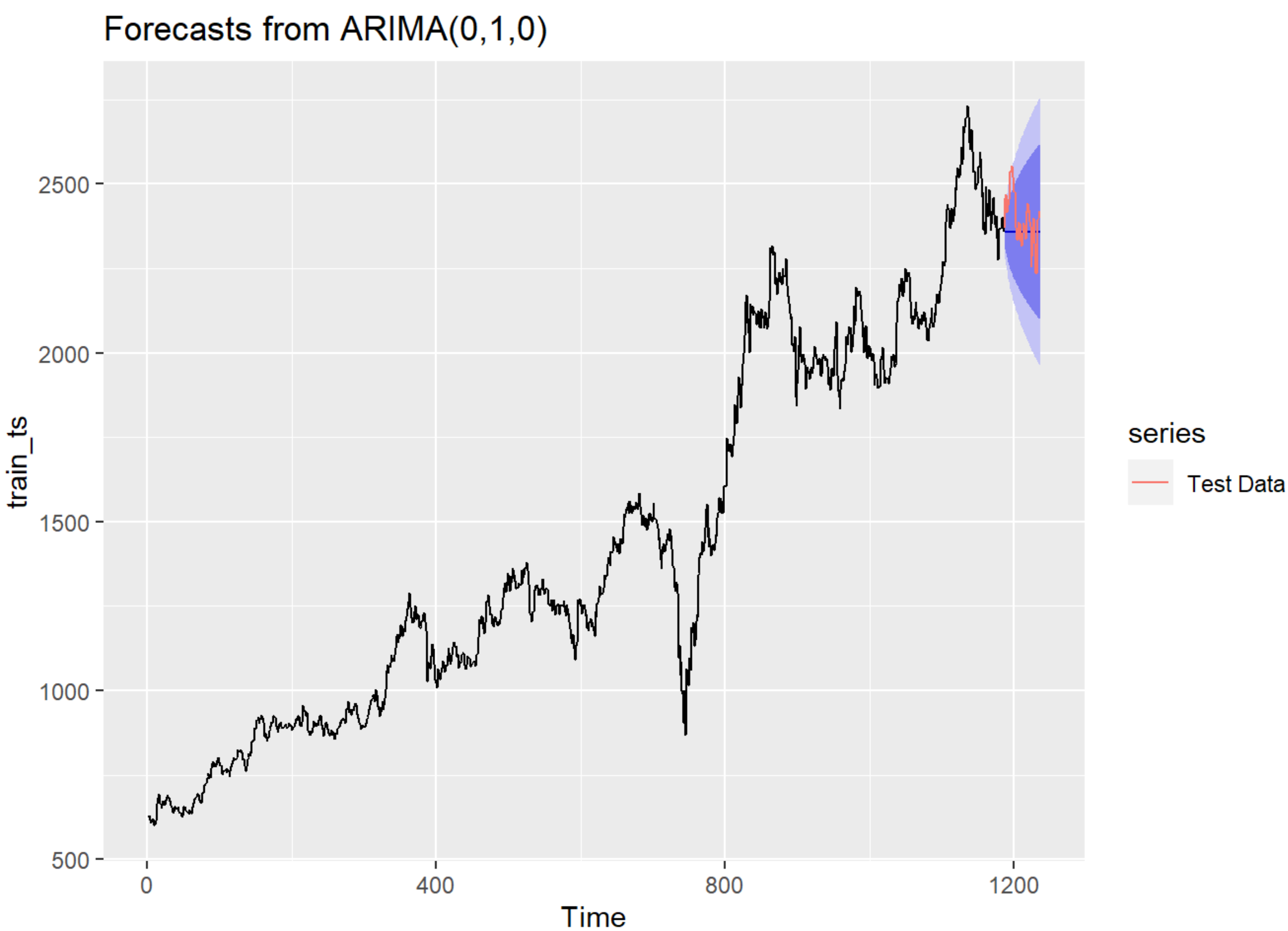




Residuals does not show any significant auto correlation which means that our model is adequately built. These are few one or two lag show significance. Let us further examine the residuals for test of significant autocorrelation by examining performing the Box test.

6. Forcasting

After comparing above models, based on AIC value and normality test, we got ARIMA (0,1,0) as best model. Other model may overfit or under fit. We will forecast using ARIMA (0,1,0) model.



As we can see forecasted blue line is close to red test or actual data. Therefore, we got best fit with ARIMA (0,1,0) model on auto sales data.

Reference

1. Seasonal Dataset : Motor Vehicle Retail Sales: Domestic Autos
   Data Source: https://fred.stlouisfed.org/series/DAUTONSA
2. Non-Seasonal Dataset: Reliance Industries Limited Stock Analysis
   DataSource: https://finance.yahoo.com/quote/RELIANCE.NS/history?period1=1615766400&period2=1647302400&interval=1d&filter=history&frequency=1d&includeAdjustedClose=true

Appendix

R Code - Motor Vehicle Retail Sales: Domestic Autos

```
library(tseries)
library(forecast)
library(dplyr)
library(TSA)

# Load data
monthly_sales <- read.csv("C:\\Users\\amitk\\Desktop\\MA 641\\Project\\DAUTONSA.csv",header = TRUE
, stringsAsFactors = FALSE)
head(monthly_sales)

summary(monthly_sales)

# Check if na
is.na(monthly_sales)

# Split into training and testing data
n = 12
monthly_sales$DATE <- as.Date(monthly_sales$DATE, format = "%Y-%m-%d")
train <- monthly_sales %>% dplyr::filter(DATE<='2018-12-01')
train_ts <- ts(train$DAUTONSA,start=c(2009,1), end=c(2018,12), frequency=n)
test <- monthly_sales %>% dplyr::filter(DATE>='2019-01-01')
test_ts <- ts(test$DAUTONSA,start=c(2019,1), end=c(2019,12), frequency=n)

# Plotting time series
ts.plot(train_ts, xlab="Time", ylab = "Unit Sold", main="Time vs Unit Sold")

# ACF and PACF of series
acf(train_ts, main = "ACF")

pacf(train_ts, main = "PACF")

# Stationary Test
# Augmented Dickey Fuller
adf.test(train_ts)

# Taking diff on time series.
train_diff_ts <- diff(train_ts, differences = 1)
# Plotting time series after taking difference of time series
ts.plot(train_diff_ts, xlab="Time", ylab = "Unit Sold", main="Time vs Unit Sold")
```

```r
# The ACF/ PACF / EACF plots after taking difference of time series
acf(train_diff_ts, main = "ACF")

pacf(train_diff_ts, main = "PACF")

eacf(train_diff_ts)

# Stationary Test
# Augmented Dickey Fuller Test after taking difference of time series
adf.test(train_diff_ts)

#####################################################################################

# ARIMA(2,1,3)X(0,1,1)[12]

model = Arima(train_ts, order=c(2,1,3), seasonal=c(0,1,1))

model

# ACF/PACF of residual
resid = model$resid
acf(resid,lag.max=30, main = "ACF")

pacf(resid,lag.max=30, main = "PACF")

# Ljung-Box Test
Box.test(model$resid, lag = 10, type ="Ljung-Box")

# QQ plot for normality
qqnorm(residuals(model))
qqline(residuals(model), col=2)

# Histogram plot for normality
hist(window(rstandard(model)),xlab='Standardized Residuals',  main="Residuals Histogram")

#####################################################################################\

# ARIMA(6,1,4)X(0,1,1)[12]
model2 = Arima(train_ts, order=c(6,1,4), seasonal=c(0,1,1))

# ACF/PACF of residual
resid2 = model2$resid
acf(resid2,lag.max=30, main = "ACF")

pacf(resid2,lag.max=30, main = "PACF")

# Ljung-Box Test
Box.test(model2$resid, lag = 10, type ="Ljung-Box")

# QQ plot for normality
qqnorm(residuals(model2))
qqline(residuals(model2), col=2)

# Histogram plot for normality
hist(window(rstandard(model2)),xlab='Standardized Residuals',  main="Residuals Histogram")

#####################################################################################

# ARIMA(2,1,2)X(0,1,0)[12]

# ACF/PACF of residual
resid3 = model3$resid
acf(resid3,lag.max=30, main = "ACF")
```

```r
pacf(resid3,lag.max=30, main = "PACF")

# Ljung-Box Test
Box.test(model3$resid, lag = 10, type ="Ljung-Box")

# QQ plot for normality
qqnorm(residuals(model3))
qqline(residuals(model3), col=2)

# Histogram plot for normality
hist(window(rstandard(model3)),xlab='Standardized Residuals',  main="Residuals Histogram")

#######################################################################################

# ARIMA(2,1,1)X(0,1,1)[12]
model4 = Arima(train_ts, order=c(2,1,1), seasonal=c(0,1,1))
model4

## Series: train_ts
## ARIMA(2,1,1)(0,1,1)[12]
##
## Coefficients:
##          ar1     ar2      ma1     sma1
##       0.2460  0.0994  -0.8305  -0.2162
## s.e.  0.1442  0.1229   0.1045   0.2440
##
## sigma^2 = 1274:  log likelihood = -532.94
## AIC=1075.88   AICc=1076.47   BIC=1089.24

# ACF/PACF of residual
resid4 = model4$resid
acf(resid4,lag.max=30, main = "ACF")

pacf(resid4,lag.max=30, main = "PACF")

# Ljung-Box Test
Box.test(model4$resid, lag = 10, type ="Ljung-Box")

##
##  Box-Ljung test
##
## data:  model4$resid
## X-squared = 12.876, df = 10, p-value = 0.2307

# QQ plot for normality
qqnorm(residuals(model4))
qqline(residuals(model4), col=2)

# Histogram plot for normality
hist(window(rstandard(model4)),xlab='Standardized Residuals',  main="Residuals Histogram")

#######################################################################################

# Forecasting using ARIMA (2,1,3) X (0,1,1) [12]

forecast <- forecast(model4, h=n)
autoplot(forecast) +
  autolayer(test_ts, series = "Test Data")
```

R Code - Reliance Industries Limited

```
library(tseries)
library(forecast)
library(dplyr)
library(TSA)
library(quantmod)

# Loading Data
reliance = read.csv("C:\\Users\\amitk\\Desktop\\MA 641\\Project\\RELIANCE.NS.csv", sep=',', header
= T)
summary(reliance)

head(reliance)

# Check if na
is.na(reliance)

# Change to Date
reliance$Date <- as.Date(reliance$Date, format = "%Y-%m-%d")

# Sort by Date
reliance <- reliance %>%
  select(Date, Adj.Close) %>%
  arrange(Date)

# Split into test and train data
n=50
train_ts <-  head(Cl(reliance), length(Cl(reliance))-n)
test_ts <- tail(Cl(reliance), n)

# Plotting time series
ts.plot(train_ts, xlab="Date", ylab = "Price", main="Date vs Price")

# Loading Data
# ACF and PACF of series
acf(train_ts, lag.max=30, main="ACF")

pacf(train_ts, lag.max=30, main="PACF")

# Stationary Test
# Augmented Dickey Fuller
adf.test(train_ts)

# Taking diff on time series.
train_diff_ts <- diff(train_ts, differences = 1)
# Plotting time series after taking difference of time series
ts.plot(train_diff_ts, xlab="Date", ylab = "Price", main="Date vs Price")

# Stationary Test
# Augmented Dickey Fuller Test after taking difference of time series
adf.test(train_diff_ts)

# The ACF/ PACF / EACF plots after taking difference of time series
acf(train_diff_ts, main="ACF")

pacf(train_diff_ts, main="PACF")

eacf(train_diff_ts)

##########################################################################
```

```r
# ARIMA(0,1,0)
model <- Arima(train_ts,order=c(0,1,0))
model

# ACF/PACF of residual
resid = model$resid
acf(resid,lag.max=20, main = "ACF")

pacf(resid,lag.max=20, main = "PACF")

# Ljung-Box Test
Box.test(model$resid, lag = 10, type ="Ljung-Box")

# QQ plot for normality
qqnorm(residuals(model))
qqline(residuals(model), col=2)

# Histogram plot for normality
hist(window(rstandard(model)),xlab='Standardized Residuals',  main="Residuals Histogram")

##########################################################################

# ARIMA(1,1,0)
model2 <- Arima(train_ts,order=c(1,1,0))
model2

# ACF/PACF of residual
resid2 = model$resid
acf(resid2,lag.max=20, main = "ACF")

pacf(resid2,lag.max=20, main = "PACF")

# Ljung-Box Test
Box.test(model2$resid, lag = 10, type ="Ljung-Box")

# QQ plot for normality
qqnorm(residuals(model2))
qqline(residuals(model2), col=2)

# Histogram plot for normality
hist(window(rstandard(model2)),xlab='Standardized Residuals',  main="Residuals Histogram")

##########################################################################

# ARIMA(0,1,1)
model3 <- Arima(train_ts,order=c(0,1,1))
model3

# ACF/PACF of residual
resid3 = model$resid
acf(resid3,lag.max=20, main = "ACF")

pacf(resid3,lag.max=20, main = "PACF")

# Ljung-Box Test
Box.test(model3$resid, lag = 10, type ="Ljung-Box")

# QQ plot for normality
qqnorm(residuals(model3))
qqline(residuals(model3), col=2)

# Histogram plot for normality
hist(window(rstandard(model3)),xlab='Standardized Residuals',  main="Residuals Histogram")
```

```r
##########################################################################

# ARIMA(2,1,3)
model4 <- Arima(train_ts,order=c(2,1,3))
model4

# ACF/PACF of residual
resid4 = model4$resid
acf(resid4,lag.max=20, main = "ACF")

pacf(resid4,lag.max=20, main = "PACF")

# Ljung-Box Test
Box.test(model4$resid, lag = 10, type ="Ljung-Box")

# QQ plot for normality
qqnorm(residuals(model4))
qqline(residuals(model4), col=2)

# Histogram plot for normality
hist(window(rstandard(model4)),xlab='Standardized Residuals',  main="Residuals Histogram")

##########################################################################

# Forecasting using ARIMA(0,1,0)

forecast <- forecast(model, h=n)
autoplot(forecast) +
  autolayer(ts(test_ts, start=length(train_ts)), series = "Test Data")
```