Time Series Analysis Assignment -1

Name: PRATHAM RADHAKRISHNA

ID: s3997064

Introduction

This report presents an analysis of a dataset representing the return on investment (ROI) of a share market trader's portfolio. The dataset consists of 179 observations collected over a period of consecutive trading days within a single year, with each observation corresponding to the ROI on a specific trading day. Notably, the dataset captures trading activity over 179 out of a possible 252 trading days in the year, with weekends (Saturdays and Sundays) excluded from the analysis. It's important to highlight that for the purpose of this assignment, consecutive trading days are considered to be Fridays and Mondays, disregarding the market closure on weekends. The analysis aims to provide insights into the performance and trends of the trader's investment portfolio over the observed period, facilitating informed decision-making and strategic planning in investment management.

Objective

- To find the best fitting model among the linear, quadratic, cosine, cyclical or seasonal trend models by implementing the model-building strategy in Module 1.
- To give the predictions for the next 5 trading days using the best model you find.

```
library(TSA)
```

```
##
## Attaching package: 'TSA'
```

```
## The following objects are masked from 'package:stats':
##
##     acf, arima
```

```
## The following object is masked from 'package:utils':
##
##     tar
```

```
library(tseries)
```

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

```
# Loading  the dataset
setwd("/Users/sanjukthgowda/Desktop/time series")
dataset <- read.csv("assignment1Data2024.csv", header = FALSE)
#Naming the second column as returns
names(dataset)[2] <- "Returns"
dataset <- dataset[-1, ]
rownames(dataset) <- NULL
```

Renaming the column as ‘Returns’ and removing the unwanted space.This is my dataset below of the first few rows

```
head(dataset)
```

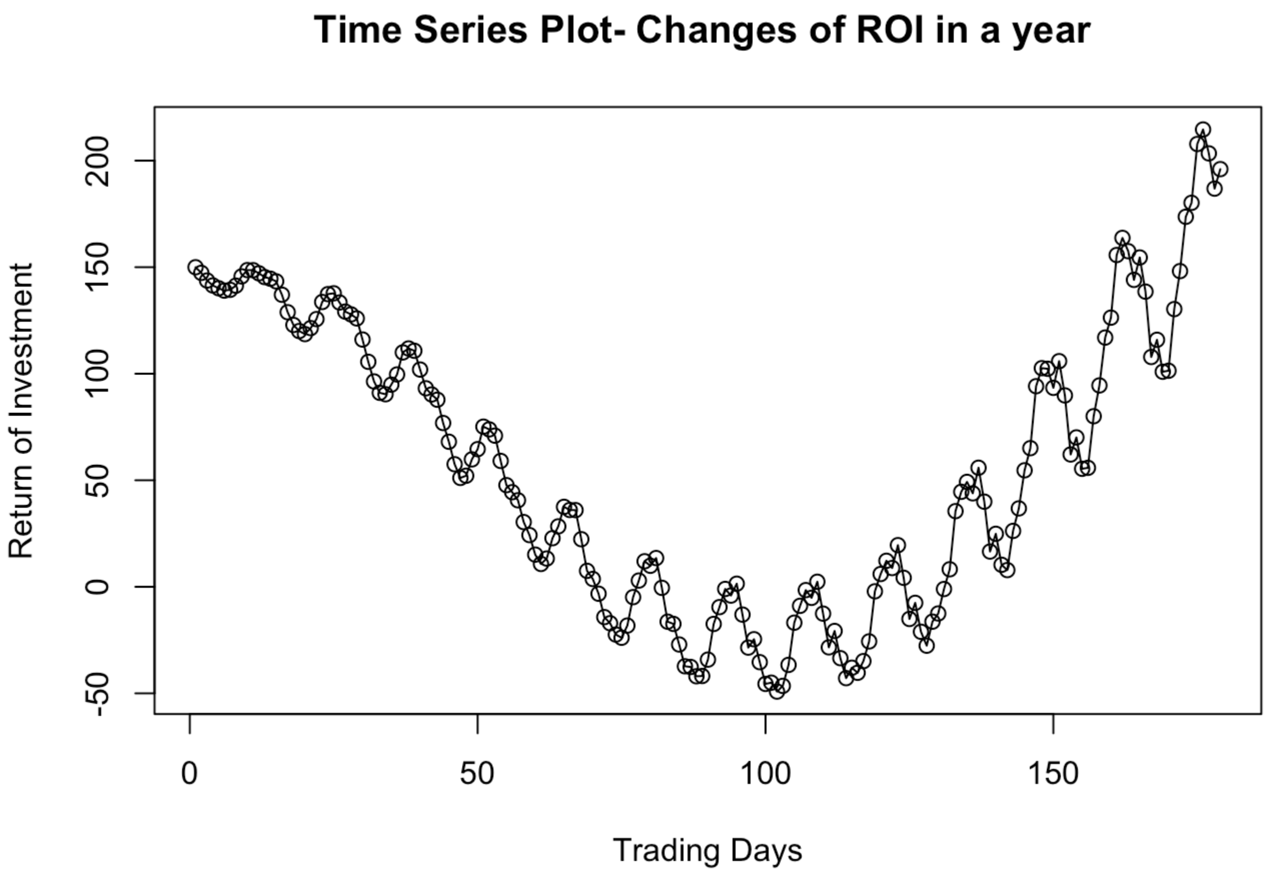
```
##   V1          Returns
## 1  1 149.976414104052
## 2  2 147.331236322272
## 3  3 143.742098742287
## 4  4 141.427000055757
## 5  5    140.113798047
## 6  6 139.005675801169
```

```
# Time Series of dataset
ReturnsTS <- ts(dataset$Returns,frequency =1)
ReturnsTS
```

```
## Time Series:
## Start = 1
## End = 179
## Frequency = 1
##   [1]   149.976414104052   147.331236322272   143.742098742287
##   [4]   141.427000055757      140.113798047   139.005675801169
##   [7]   139.381354679266   141.318554850487   145.670943553201
##  [10]   148.635903355871   148.597267866784   147.124844167853
##  [13]   145.357171597311    144.59142614917   143.353833560188
##  [16]   137.070643164612    128.86416171929   122.887505557005
##  [19]   120.024081778818   118.589227772159   121.415549617793
##  [22]   125.637189745714   133.662759726276   137.364137707533
##  [25]   137.742465911352   133.440363203005   129.155826556038
##  [28]   127.855799404128   125.957273871672   116.051386033735
##  [31]   105.598209887655   96.3822990763183   91.0377708071934
##  [34]   90.3617594233942   94.8646086928685   99.6581769183476
##  [37]   109.996677354026   111.857374078794   110.775465185494
##  [40]   101.963076550312   93.1641159877247   90.2672273006729
##  [43]    87.800739935508    76.867443248349   68.0830082412553
##  [46]   57.5297754656232   51.1049291737404   52.1373587267573
##  [49]   59.7974923631022   64.6062279375847   75.1565830797418
##  [52]   73.8825070180512   70.9435860556784   59.0720140026074
##  [55]   47.7063514939371   44.4034074264907   40.6121027043905
##  [58]   30.3832797762894   24.2679989050602   15.1092333224322
##  [61]   10.7206516725182   13.2665689322991   22.8130022402065
##  [64]   28.3532618577283   37.5251150307416   36.0476632859681
##  [67]   35.9861826481303   22.3339569430406    7.4891516405379
##  [70]   3.67410742642483  -3.24523525731857  -14.2196991214683
##  [73]  -17.0912632467583  -22.4107029576666  -23.9016263456746
##  [76]  -18.1923420489703  -4.85564312970439   2.93292953017365
##  [79]   11.9867518376814   9.87259873057437   13.4886954451295
##  [82] -0.482396347324055  -16.3714443481068  -17.4101838494503
##  [85]  -27.0785552540166  -37.2959158499908  -37.6258574587758
##  [88]  -41.9926588181003  -41.8538378943612  -34.1691177106952
##  [91]  -17.4319752271488  -9.53609522164319  -1.00838673518484
##  [94]  -4.11738225551431   1.45390124995333  -13.0478849308034
##  [97]  -28.5303281802403  -24.6028417356918  -35.3678916870082
## [100]     -45.5742870757  -44.9893083599443  -49.1673628955949
## [103]  -46.4989142046612  -36.6436061154577  -16.8369131809536
## [106]  -8.92759350051816  -1.54730694031932  -5.18259684887499
## [109]    2.3701171686169  -12.5857504509095  -28.4262709705163
## [112]  -20.7296777129063  -33.4946050082619  -42.8667694500957
## [115]  -37.9828798943016  -40.3481284589801  -34.8765161411299
## [118]  -25.6293115328057  -2.12452736502541   6.00131296087352
## [121]   12.2149068759398   8.80394478565799   19.5475042092371
## [124]   4.22519402996531  -15.0934417770393  -7.53330519487932
## [127]  -21.0962522521511  -27.6296070643687  -16.3524162349503
## [130]  -12.5228100565261  -1.05022398676422   8.22311490647405
## [133]   35.5015858139678   44.5841186859103   49.2380856568181
## [136]    43.860778458148   55.8762132879478   39.9568534816477
## [139]   16.5493029470778   24.8712715510026   10.3316686578495
## [142]   7.84105792785991   26.2363898578614   36.7919186856417
## [145]   54.6747601945916   65.0661158910161   94.1400067901697
## [148]   102.614046533448   102.325423451177   93.3756033785501
## [151]   105.900148196016   89.8233831795272   62.1062162961197
```

```
## [154]   70.1136166776001   55.3928602052406   55.8354773297659
## [157]   80.0485193620785    94.505128851083   116.973014237495
## [160]   126.323845105924   155.705622112767   163.743259494458
## [163]    157.49684927466   144.041793081926   154.549418941591
## [166]   138.529698792422   107.849252179142   115.927721545665
## [169]   100.928631466312   101.404439929157   130.324035701832
## [172]   148.163666468377   173.654176871585   180.208784332072
## [175]   207.806511921174   214.610521294681   203.409179185071
## [178]   186.867580151086   196.016410506297
```

```
### Time Series Plot
plot(ReturnsTS, ylab='Return of Investment', xlab='Trading Days', type='o',
     main ="Time Series Plot- Changes of ROI in a year ")
```



Observations from the above plot

Trend - An upward trend.

Seasonality - there are repeating patterns observed.

Changing Variance - change in variance is present.

Behavior - Both Moving Average and Auto-Regressive.

Change Point - There is no change point observed .

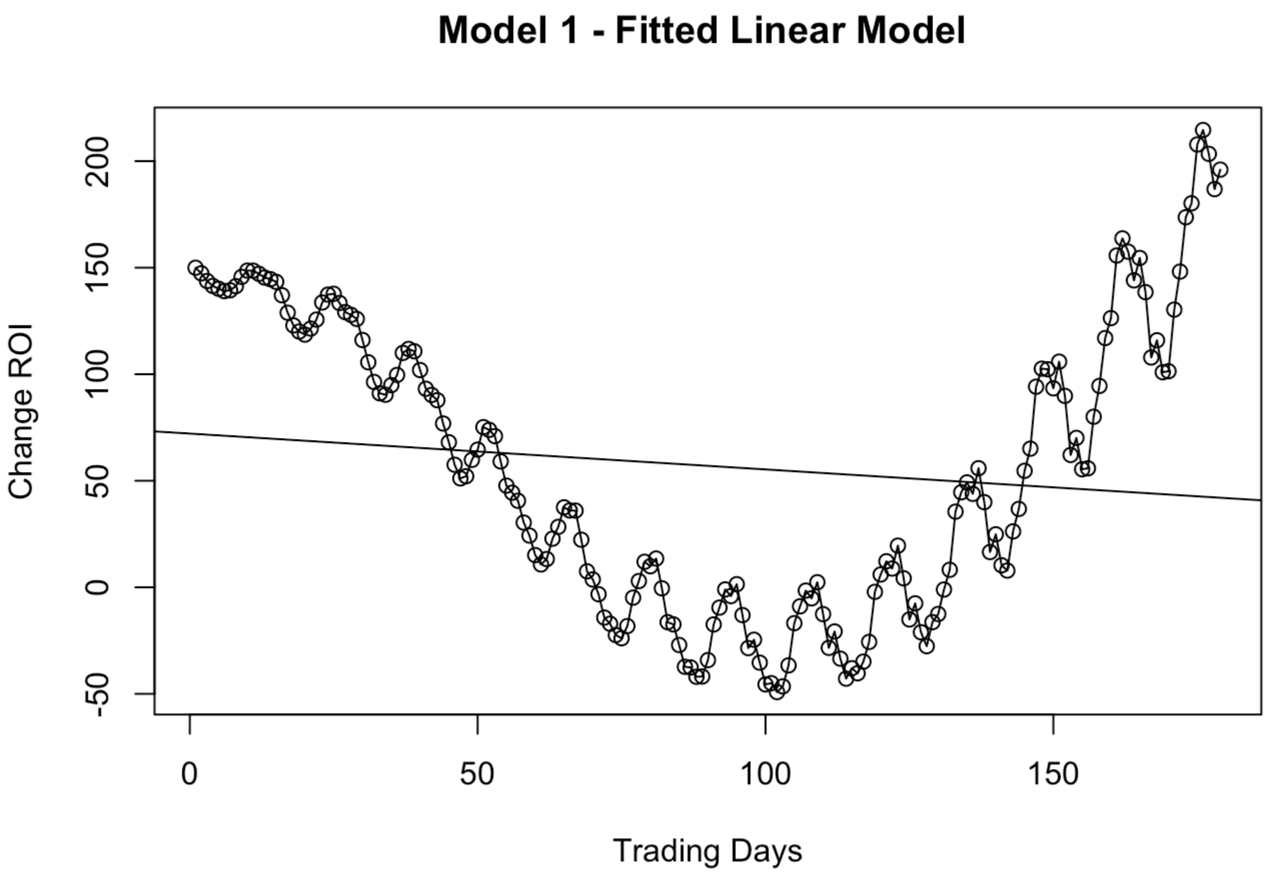
Task 1

Finding the best Fitting Model

Finding the best fitting model among the linear, quadratic, cosine, cyclical or seasonal trend models by implementing the model-building strategy.

Linear Model

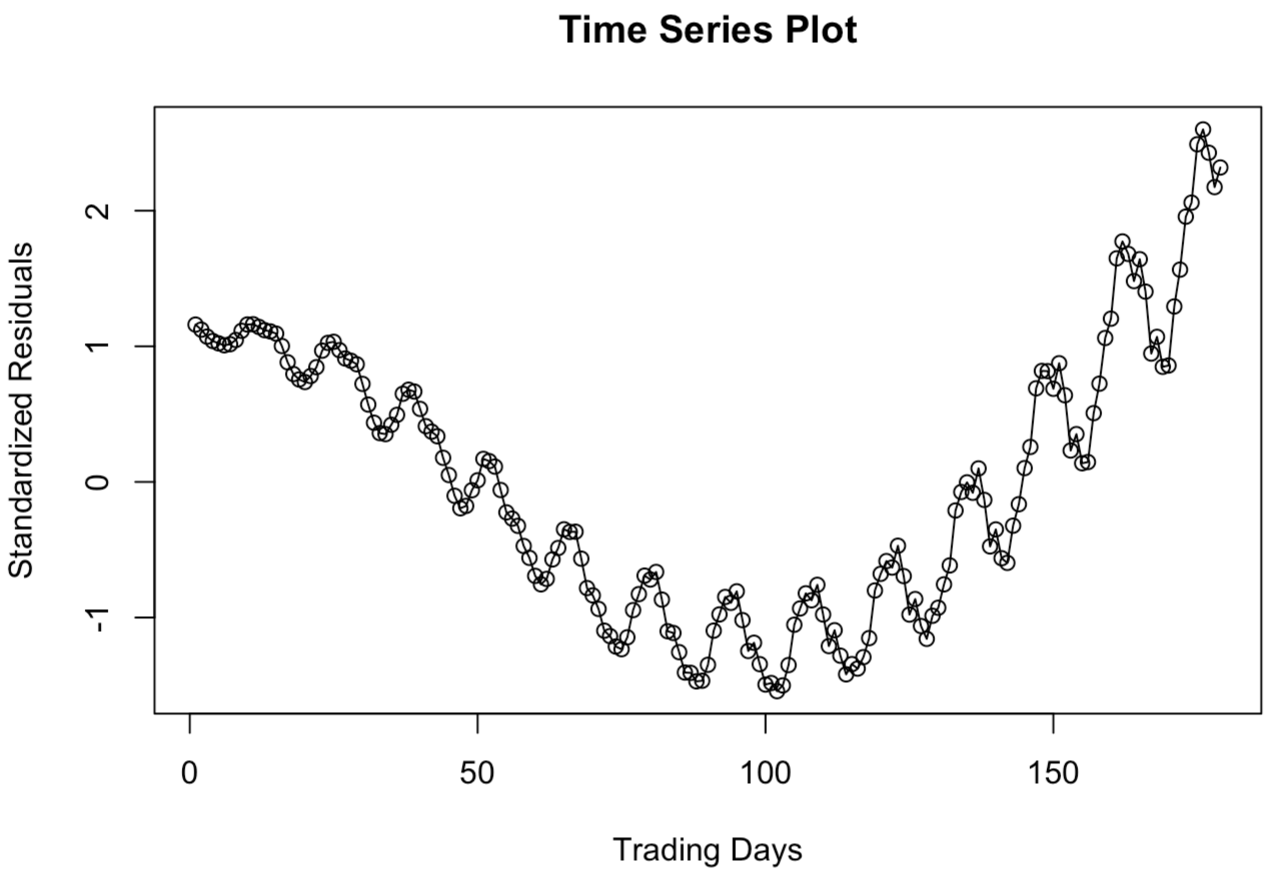
```
# extracting time from time series.
t <- time(ReturnsTS)

# Model 1 :Linear Regression Model

plot(ReturnsTS, ylab= 'Change ROI', xlab= 'Trading Days', type= 'o',
     main = "Model 1 - Fitted Linear Model ")
model1 <- lm(ReturnsTS ~ t)
abline(model1)
```



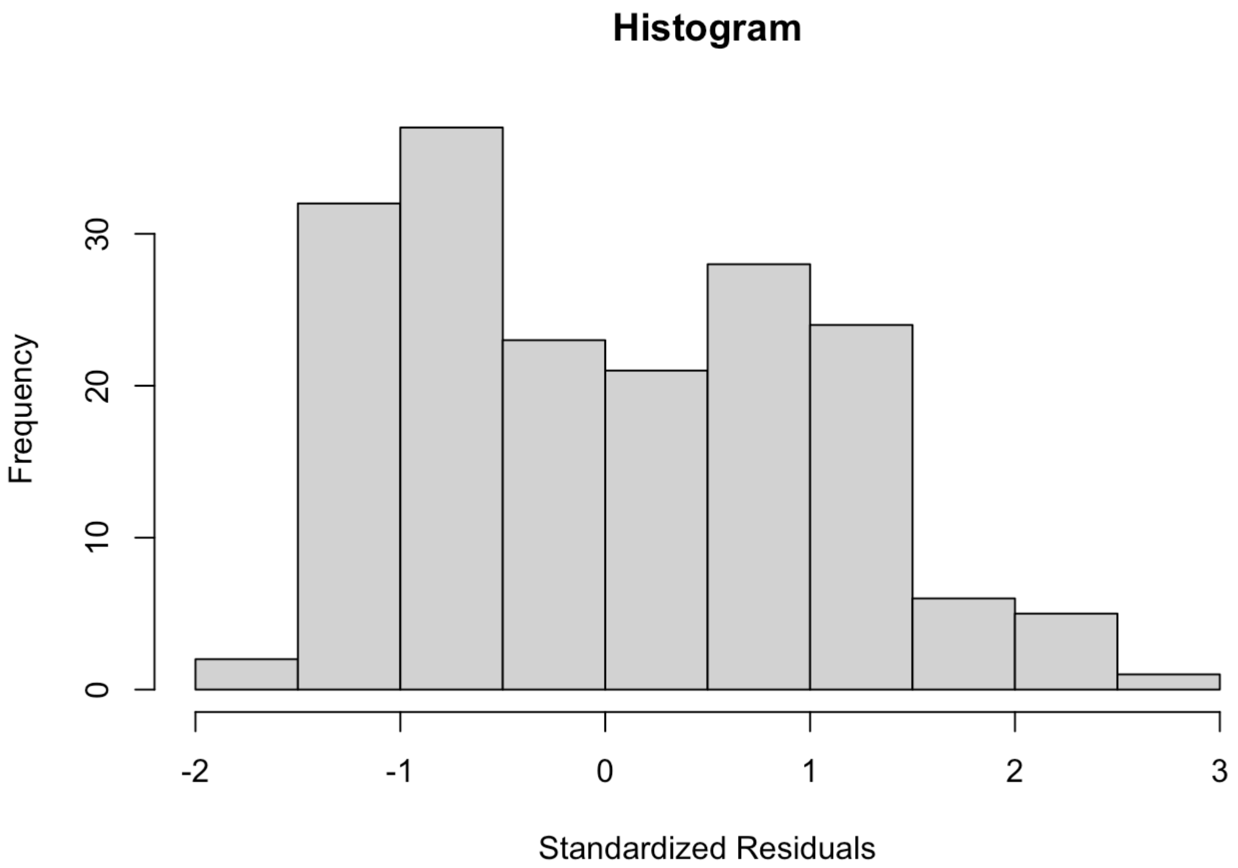

```
summary(model1)
```

```
##
## Call:
## lm(formula = ReturnsTS ~ t)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -104.186  -58.689   -5.424   57.532  172.072
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 72.22086   10.20622   7.076 3.33e-11 ***
## t           -0.16865    0.09835  -1.715   0.0881 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 67.99 on 177 degrees of freedom
## Multiple R-squared:  0.01634,    Adjusted R-squared:  0.01079
## F-statistic: 2.941 on 1 and 177 DF,  p-value: 0.08812
```

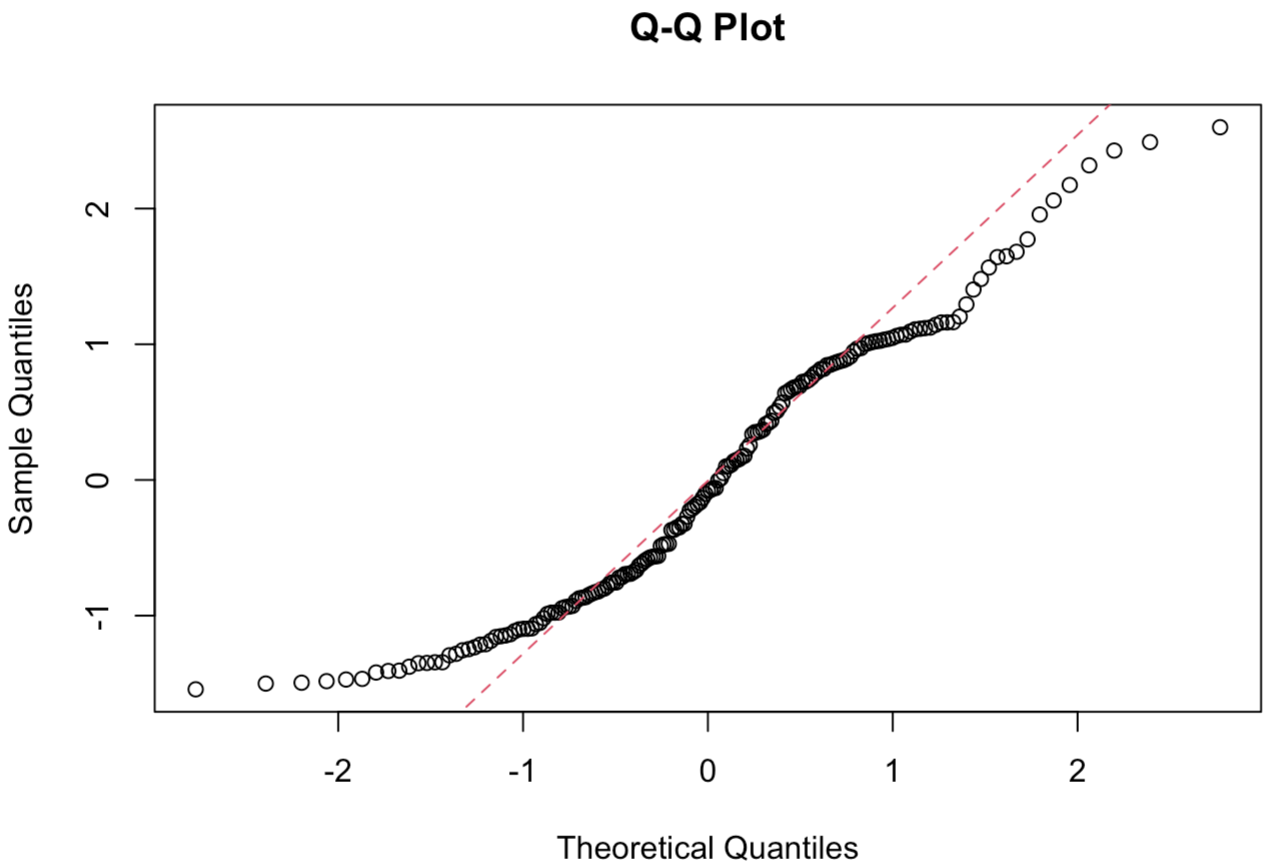
```
# Residual Analysis
plot(y= rstudent(model1), x= as.vector(time(ReturnsTS)), xlab= 'Trading Days',
     ylab= 'Standardized Residuals', type='o', main = "Time Series Plot")
```



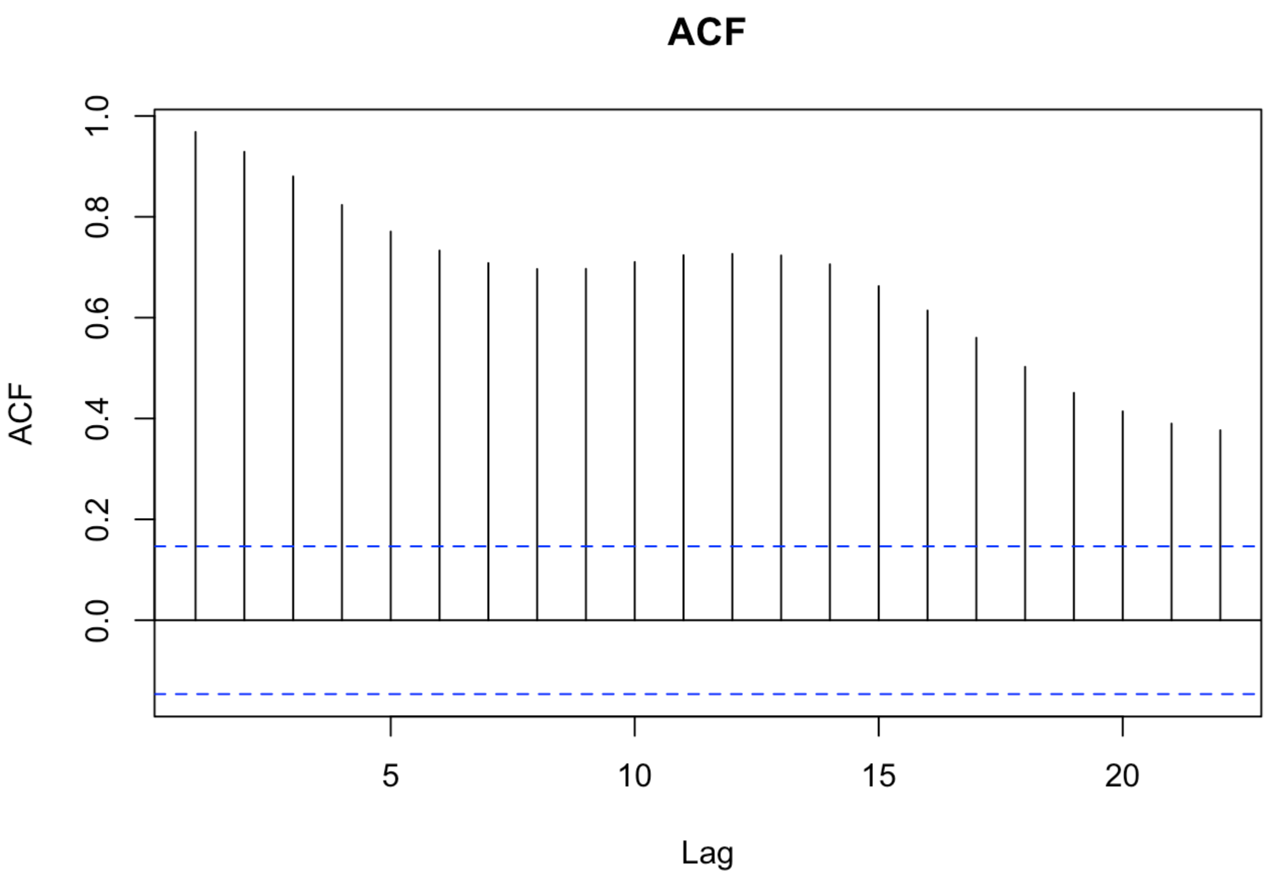

```
# Histogram Distribution
hist(rstudent(model1), xlab= 'Standardized Residuals', main = "Histogram")
```




```
# Q-Q plots
y = rstudent(model1)
qqnorm(y, main= "Q-Q Plot")
qqline(y, col= 2, lwd= 1, lty= 2)
```




```
# ACF of linear model.
acf(rstudent(model1), main = "ACF")
```



```
# Shapiro-Wilk test of linear model.
y = rstudent(model1)
shapiro.test(y)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  y
## W = 0.9525, p-value = 1.022e-05
```

Linear Model analysis

The p value of the slope is 0.08812 which is more than 0.05 , we can say that the effect of t on ReturnsTs may not be statistically significant. The R- Squared value is 0.01634,indicating that only approximately 1.634% of the variance in ReturnsTS is explained by the predictor variable t. The R-Squared value is slightly lower, suggesting the model's explanatory power is limited when adjusted for the number of predictors. The histogram plot does not captures a symmetrical distribution of data.The Q-Q plot appears to be nearly like a straight line with slight deviations at the ends which indicates that the data points are normally distributed.

The Shapiro-Wilk normality test is a statistical test used to assess whether a given sample comes from a normally distributed population. In this case, the test was applied to the variable y.The test statistic (W) is a measure of how closely the sample data resembles a normal distribution. In this output, the test statistic is 0.9525.The p-value associated with the test statistic indicates the probability of observing a test statistic as extreme as, or more extreme than, the one calculated from the sample data, assuming that the null hypothesis is true (i.e., assuming the data comes from a normally distributed population).Here, the p-value is reported as 1.022e-05, which is a very small value (1.022 x 10^-5), much smaller than the commonly used significance level of 0.05.Since the p-value is less than 0.05, we reject the null hypothesis that the data comes from a normally distributed population.Therefore, based on the Shapiro-Wilk test, we have evidence to suggest that the variable y is not normally distributed.

The ACF plots shows that there are all significant lags present which indicates a strong autocorrelation in the residuals.

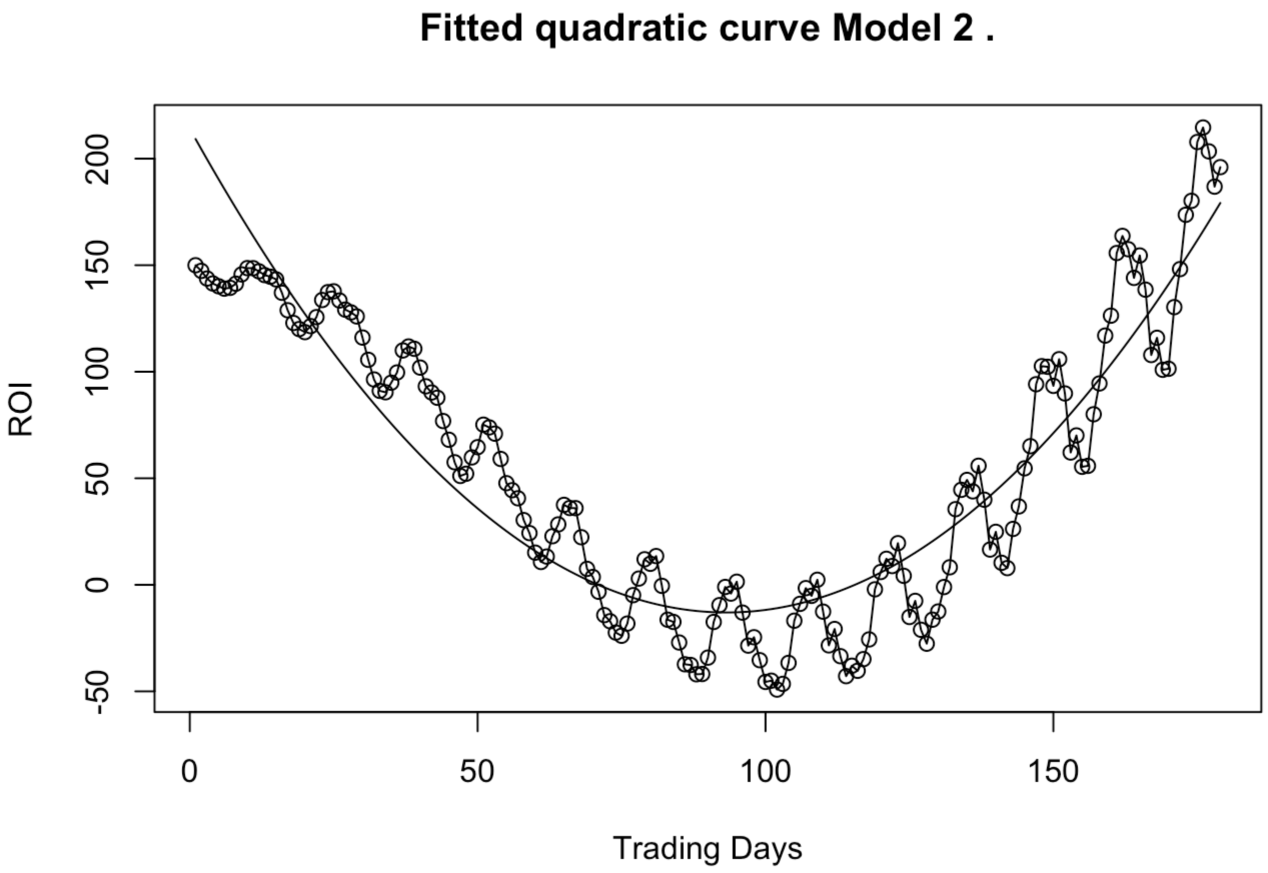
Overall I conclude that linear model cannot be considered as an ideal model for this data.

Quadratic Model

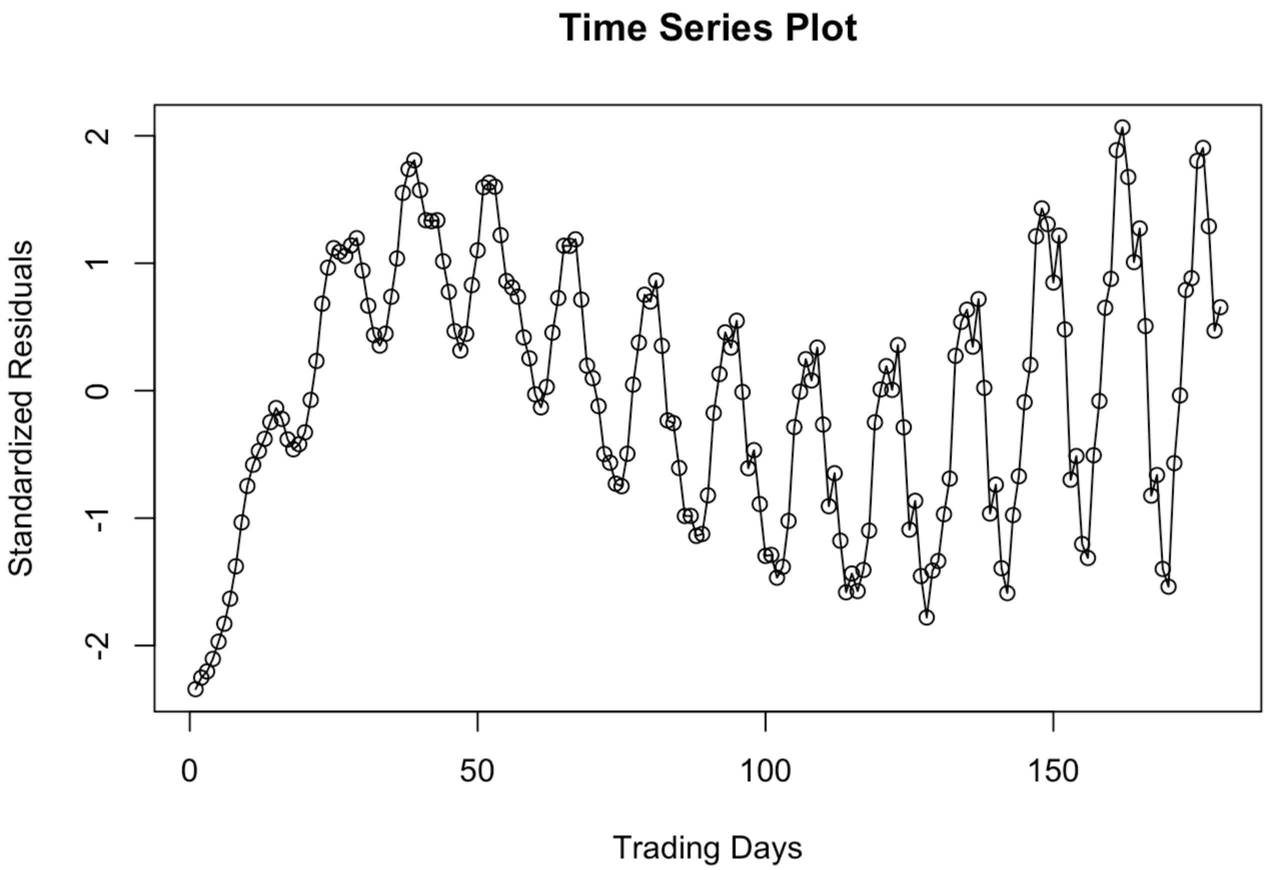
```
# Model 2 - Quadratic Model.
t = time(ReturnsTS)
t2 = t**2
model2 = lm(ReturnsTS ~ t + t2)
summary(model2)
```

```
##
## Call:
## lm(formula = ReturnsTS ~ t + t2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -59.258 -19.321   0.571  19.992  53.165
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.141e+02  5.958e+00   35.93   <2e-16 ***
## t           -4.871e+00  1.528e-01  -31.87   <2e-16 ***
## t2           2.612e-02  8.224e-04   31.77   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 26.27 on 176 degrees of freedom
## Multiple R-squared:  0.8539, Adjusted R-squared:  0.8523
## F-statistic: 514.4 on 2 and 176 DF,  p-value: < 2.2e-16
```

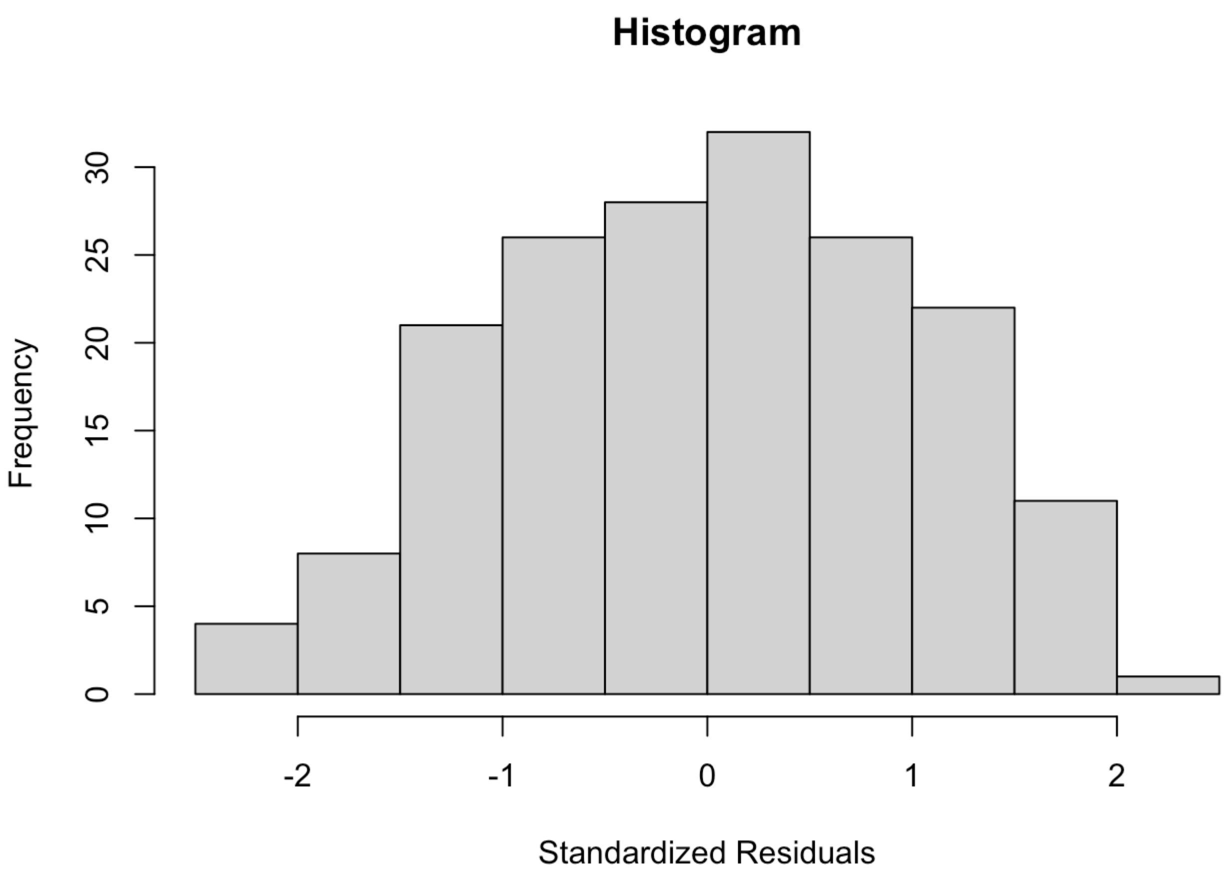
```
# Convert fitted values to numeric
fitted_values <- as.numeric(fitted(model2))    #converting  model2 and
ReturnTS from character to integer type

# Convert ReturnsTS to numeric
returns_values <- as.numeric(as.vector(ReturnsTS))

# Calculate the limits for the y-axis
y_min <- min(c(fitted_values, returns_values), na.rm = TRUE)
y_max <- max(c(fitted_values, returns_values), na.rm = TRUE)

# Plot the fitted values
plot(ts(fitted_values), ylab = 'ROI',xlab='Trading Days', main = "Fitted quadratic cu
rve Model 2 .",
     ylim = c(y_min, y_max))

# Add lines for the actual values
lines(returns_values, type = "o")
```



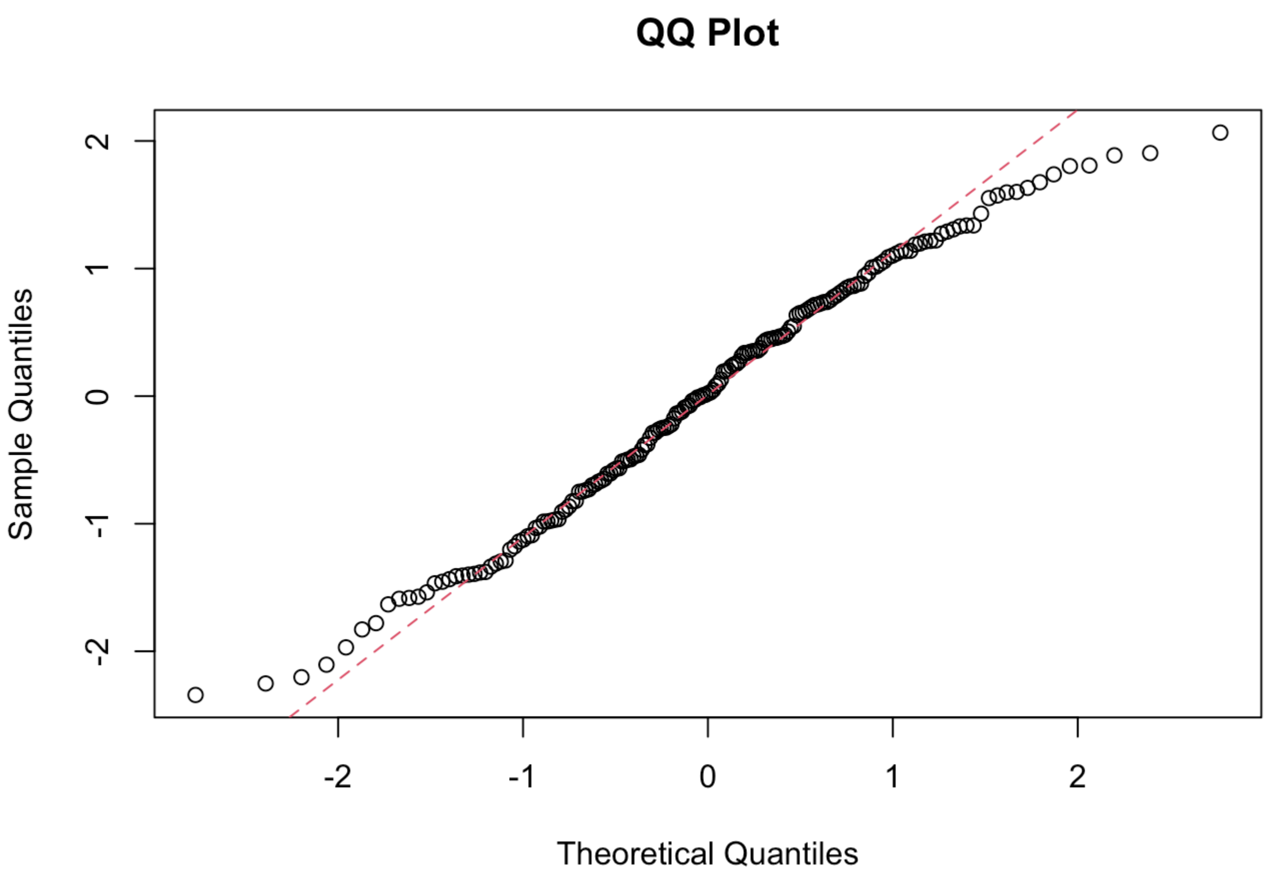

```
# Residual Analysis.
plot(y= rstudent(model2), x= as.vector(time(ReturnsTS)), xlab= 'Trading Days',
     ylab= 'Standardized Residuals', type= 'o', main = "Time Series Plot")
```



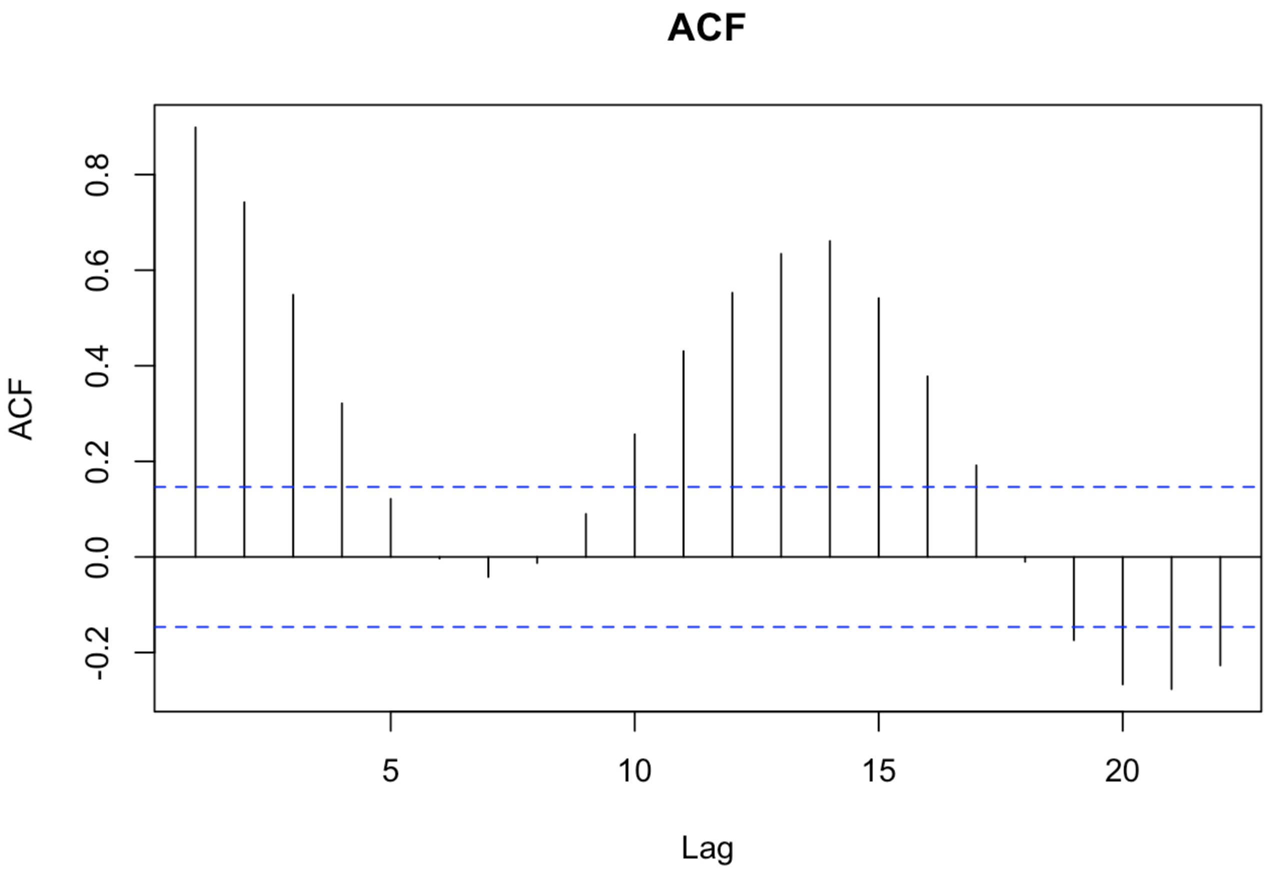

```
# Histogram Distribution.
hist(rstudent(model2), xlab= 'Standardized Residuals', main = "Histogram")
```



```
# Q-Q plots.
y = rstudent(model2)
qqnorm(y, main = "QQ Plot")
qqline(y, col= 2, lwd= 1, lty= 2)
```




```
# ACF of quadratic model
acf(rstudent(model2), main = "ACF")
```




```
# Shapiro-Wilk test
y = rstudent(model2)
shapiro.test(y)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  y
## W = 0.98397, p-value = 0.03799
```

Analysis of Quadratic Model

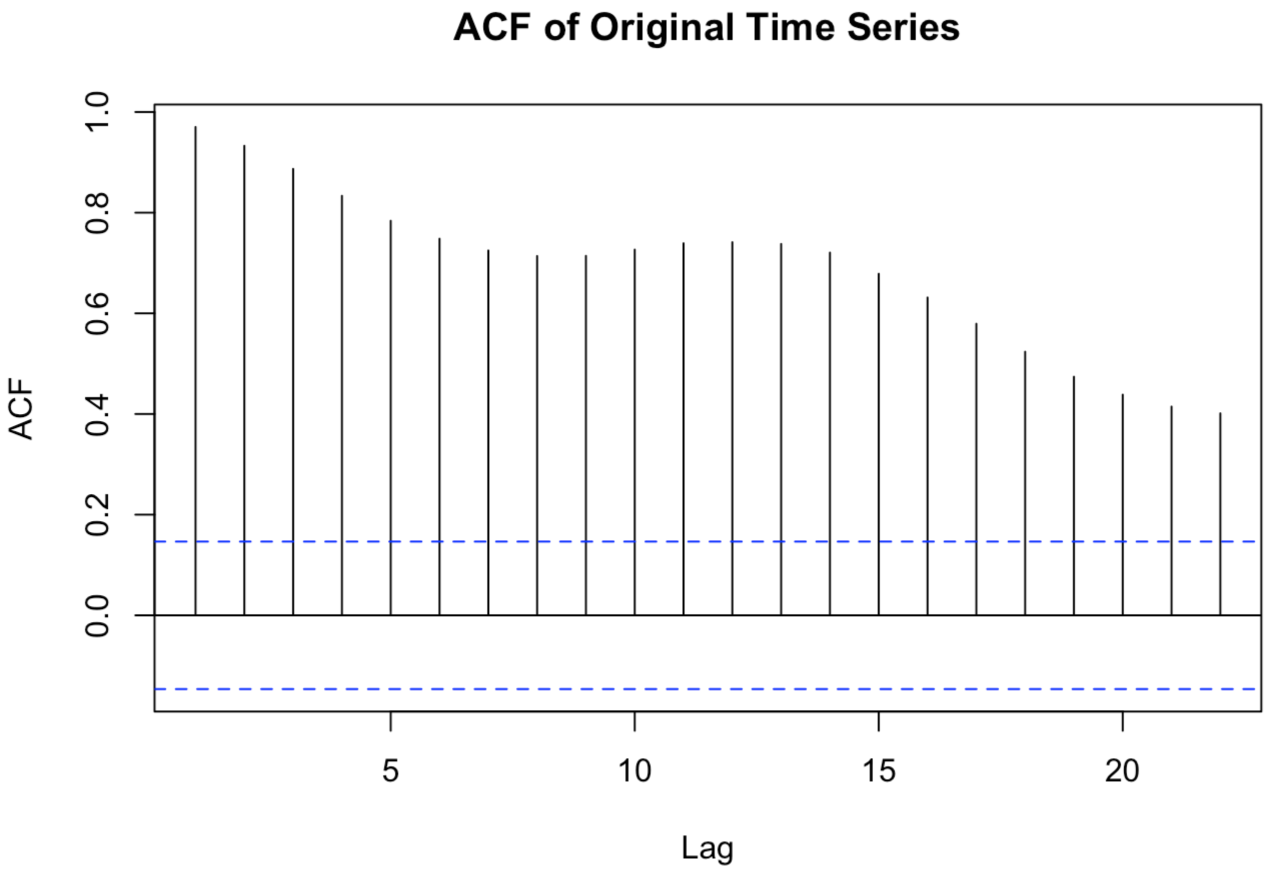
The p value of slope is less than 0.05 , we can say that the slope is significant. The R sqaured value is 0.8539 which is 85.39 % which is a good value observed .Time-series plot of the residuals represents randomness. The histogram plots a symmetrical distribution of data which lies between -3 and +3. The Q-Q plot appears almost like a straight line indicates that data points are normally distributed.

Next **Shapiro- Wilk test** is used. In this case, the p-value (0.03799) is less than the typical significance level of 0.05.If the p-value is less than the chosen significance level (e.g., 0.05), we reject the null hypothesis of normality.

The Acf plot shows that there are significant bars from lag 1 to lag 4 and from lag 10 to lag 16 indicating correlation in the residuals.
Hence Quadratic Model can be considered an ideal model for the given data.

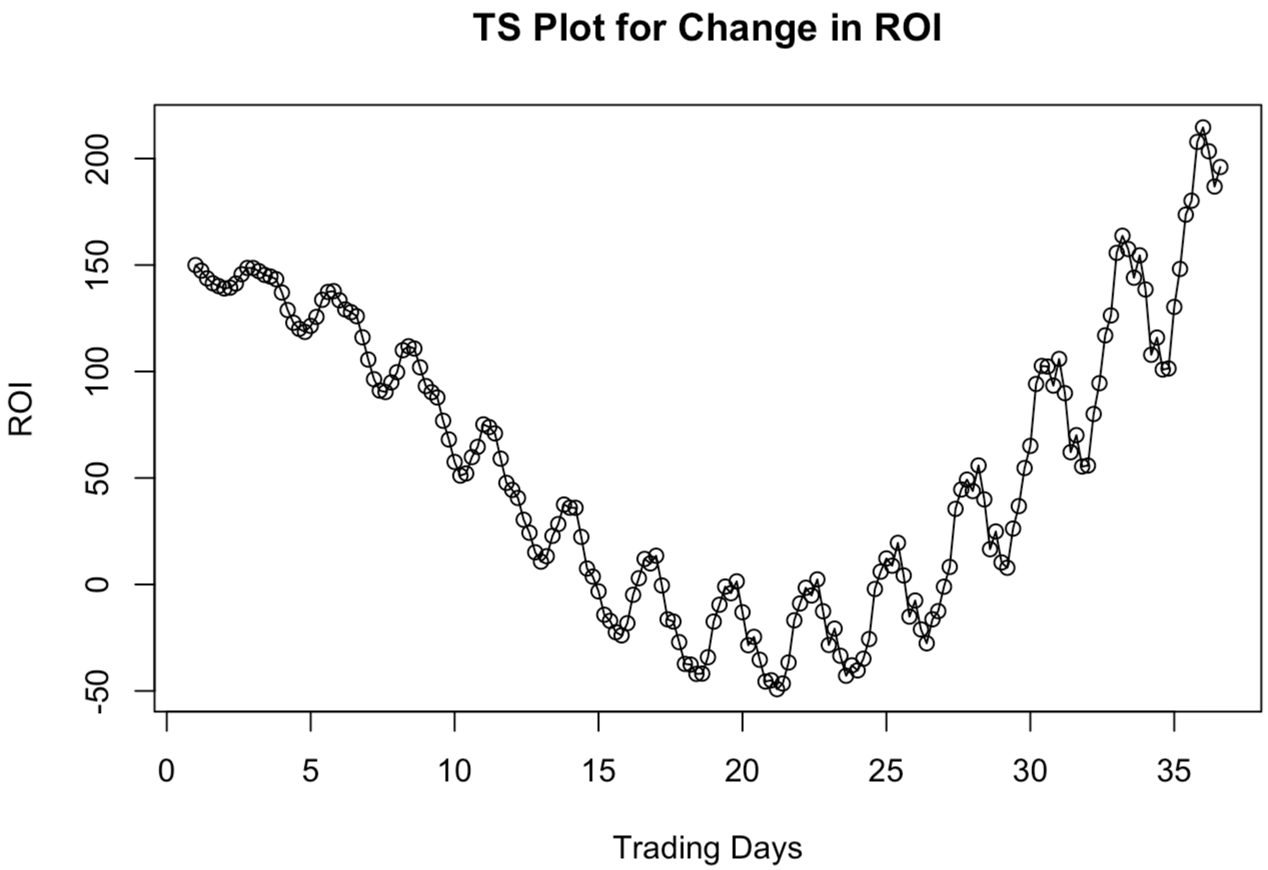
Seasonal Model

```
# ACF plot for the original time series to find its frequency for the Seasonal Model.
acf(returns_values, main = "ACF of Original Time Series")
```



Setting the frequency as 5 as the wave pattern repeats in every 5 lags.

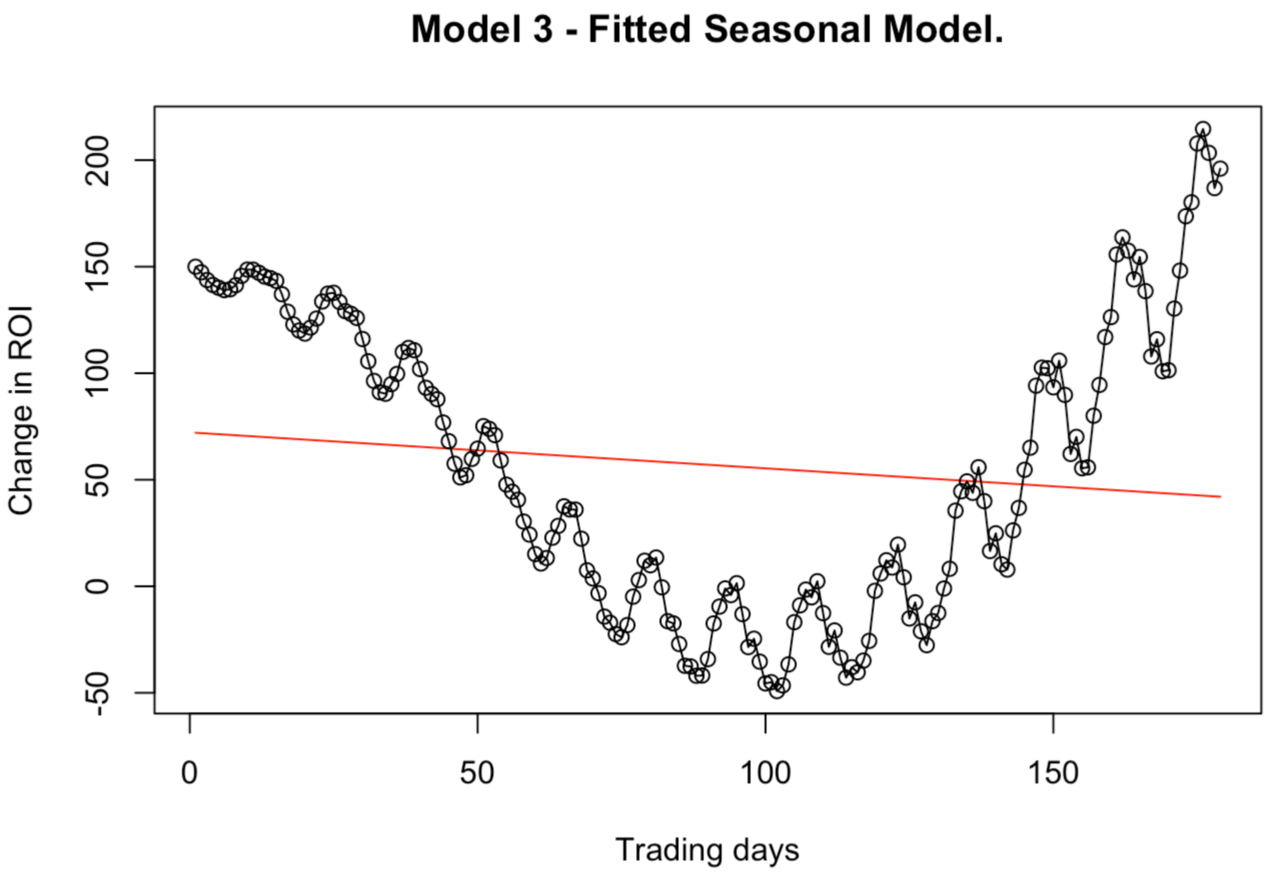
```
newTS<- ts(dataset$Returns,frequency = 5)
plot(newTS, ylab='ROI', xlab='Trading Days', type='o', main ="TS Plot for Change in R
OI")
```



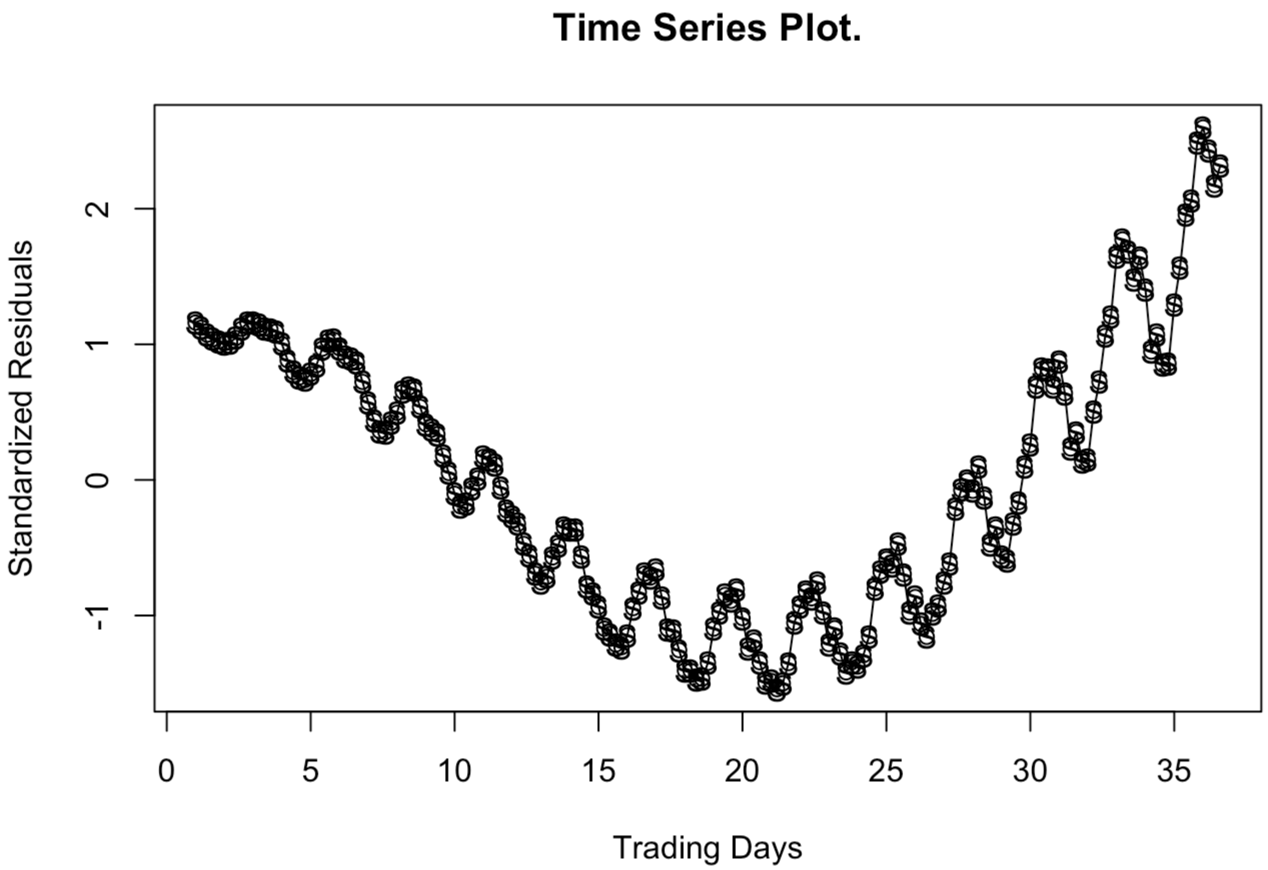

```
# Seasonal Trend
trend = season(newTS)
t1 <- time(newTS)

# Model 3 :Seasonal Model
model3= lm(newTS ~ t1)
summary(model3)
```

```
##
## Call:
## lm(formula = newTS ~ t1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -104.186  -58.689   -5.424   57.532  172.072
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  72.8955    10.5492   6.910 8.42e-11 ***
## t1           -0.8432     0.4917  -1.715   0.0881 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 67.99 on 177 degrees of freedom
## Multiple R-squared:  0.01634,    Adjusted R-squared:  0.01079
## F-statistic: 2.941 on 1 and 177 DF,  p-value: 0.08812
```

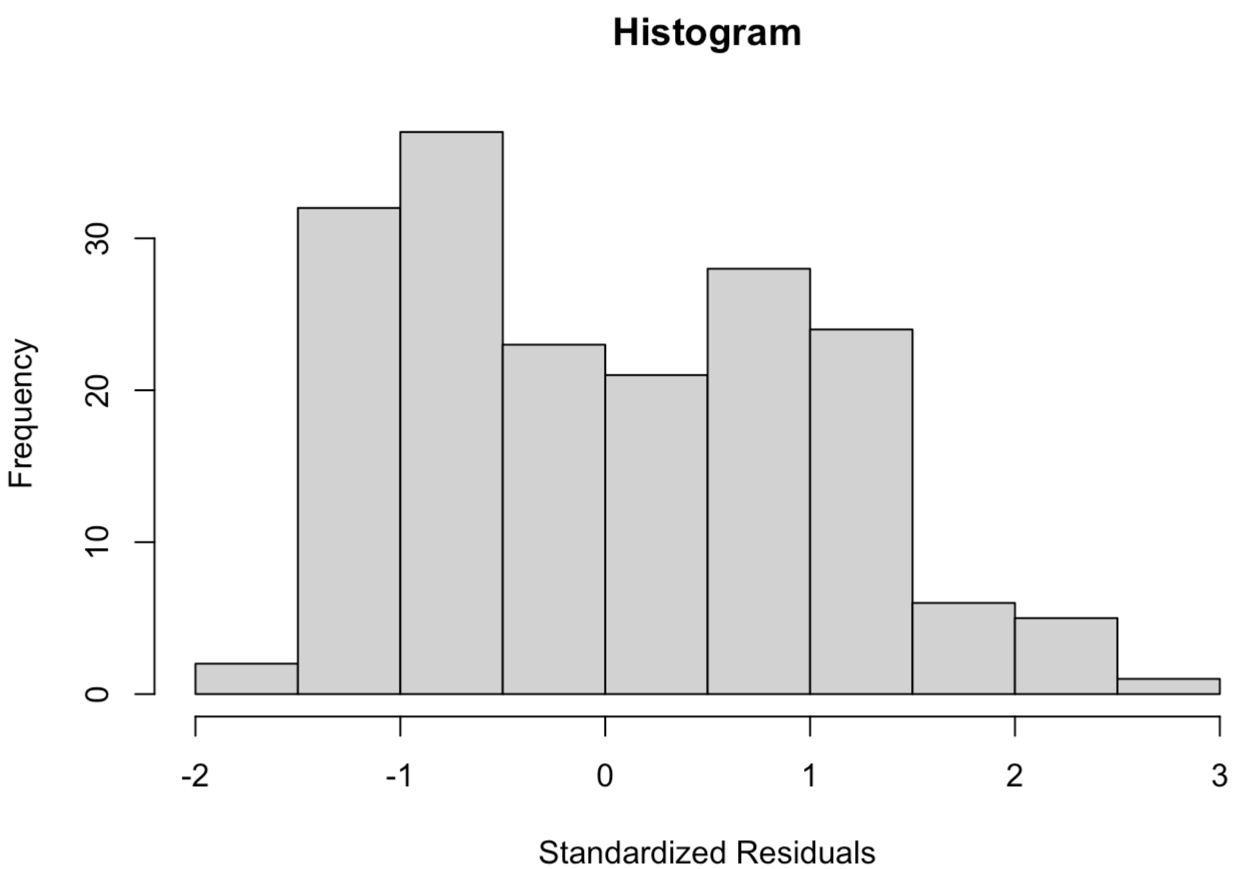
```
# Convert fitted values to numeric
fitted_values3 <- as.numeric(fitted(model3))

# Convert newTS to numeric
newTS_values <- as.numeric(as.vector(newTS))

# Calculate the limits for the y-axis
y_min <- min(c(fitted_values3, newTS_values), na.rm = TRUE)
y_max <- max(c(fitted_values3, newTS_values), na.rm = TRUE)

# Plot the fitted values
plot(ts(fitted_values3), ylab = 'Change in ROI',xlab='Trading days', main = "Model 3
- Fitted Seasonal Model.",
     ylim = c(y_min, y_max), col = "red")

# Add lines for the actual values
lines(newTS_values, type = "o")
```



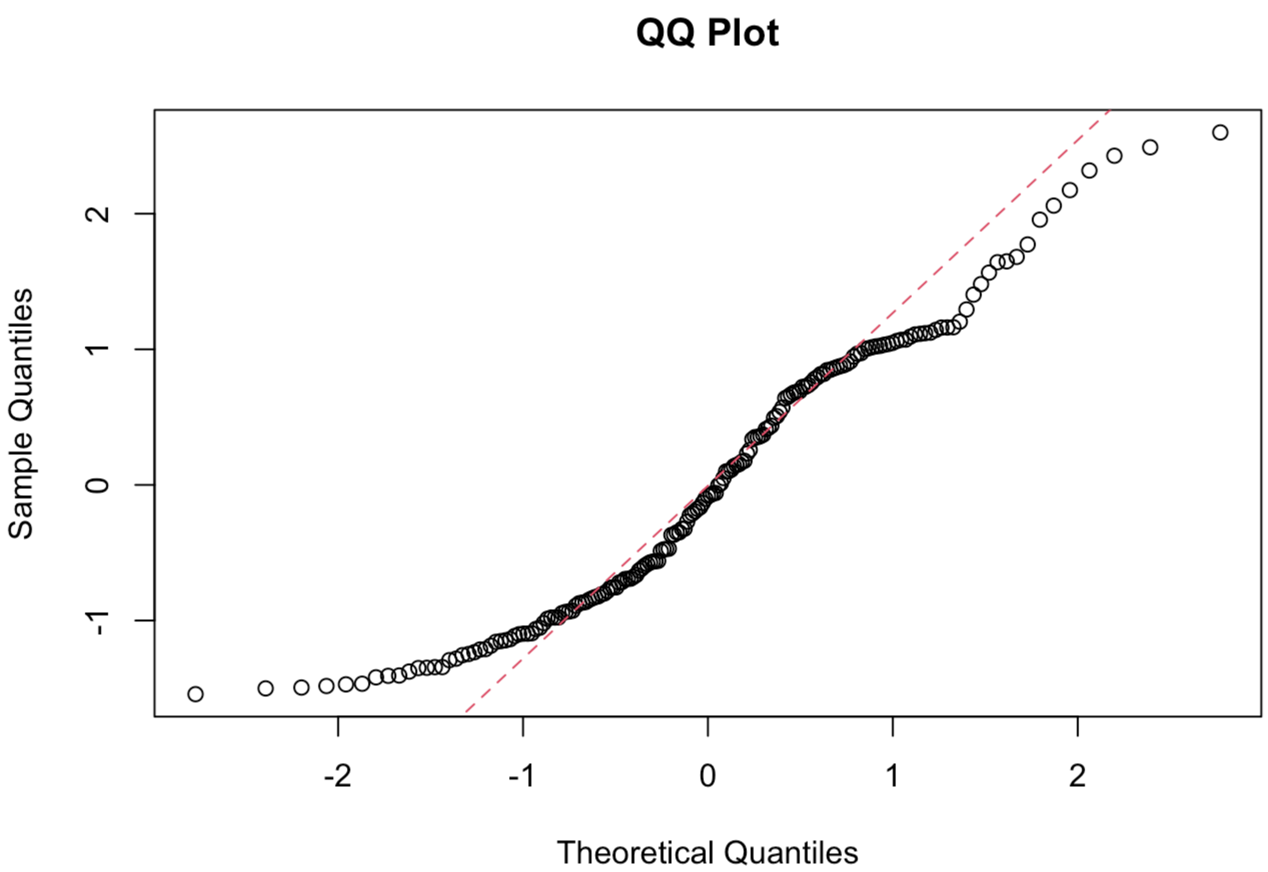

```
# Residual Analysis.
plot(y= rstudent(model3), x= as.vector(time(newTS)), xlab= 'Trading Days',
     ylab= 'Standardized Residuals', type= 'o', main ="Time Series Plot.")
points(y= rstudent(model3), x= as.vector(time(newTS)), pch= as.vector(season(newTS)))
```



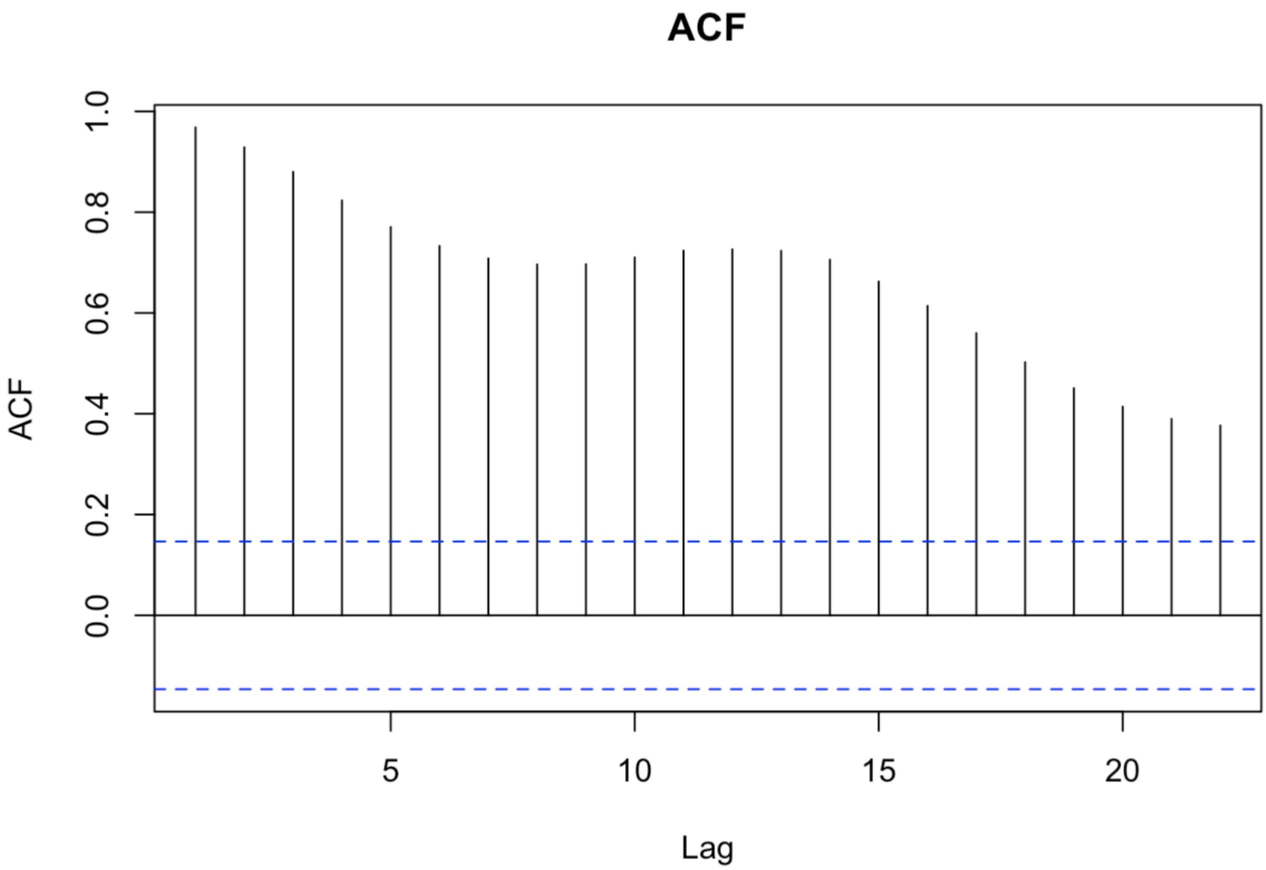

```
# Histogram Distribution
hist(rstudent(model3),xlab='Standardized Residuals', main ="Histogram")
```



```
# Q-Q Plot
y = rstudent(model3)
qqnorm(y, main ="QQ Plot")
qqline(y, col =2, lwd =1, lty =2)
```




```
# ACF
acf(rstudent(model3), main ="ACF")
```




```
# Shapiro-Wilk test of seasonal model ####
y = rstudent(model3)
shapiro.test(y)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  y
## W = 0.9525, p-value = 1.022e-05
```

Seasonal Model Analysis

The p value of the slope 0.0812 is more than the commonly used significant level of 0.05, we can say that the slope is not significant.. Time-series plot of the residuals represents randomness. The multiple R-squared value is 0.01634, indicating that only approximately 1.634% which is very less to be considered an ideal model for this data. The Histogram does not show a symmetrical distribution of data.The Q-Q plot appears to be nearly like a straight line with slight deviations at the ends which indicates that the data points are normally distributed.

Shapiro Wilk Test with p value (1.022e-05 or 0.00001022) is significantly smaller than the typical significance level of 0.05.Therefore, we reject the null hypothesis of normality.The data does not appear to be consistent with a normal distribution at the 0.05 significance level.

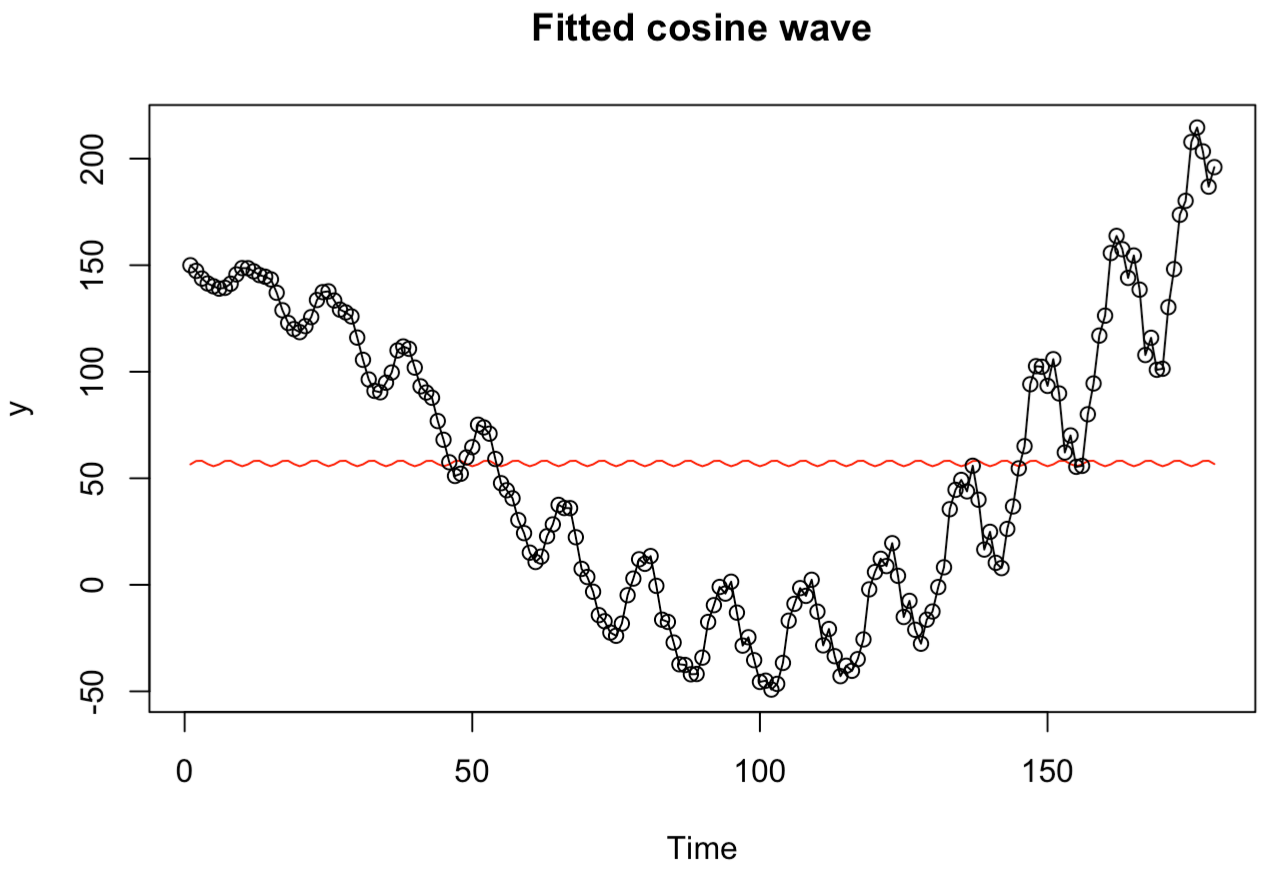
Overall this model cannot be considered an ideal match for this data

Cosine Model

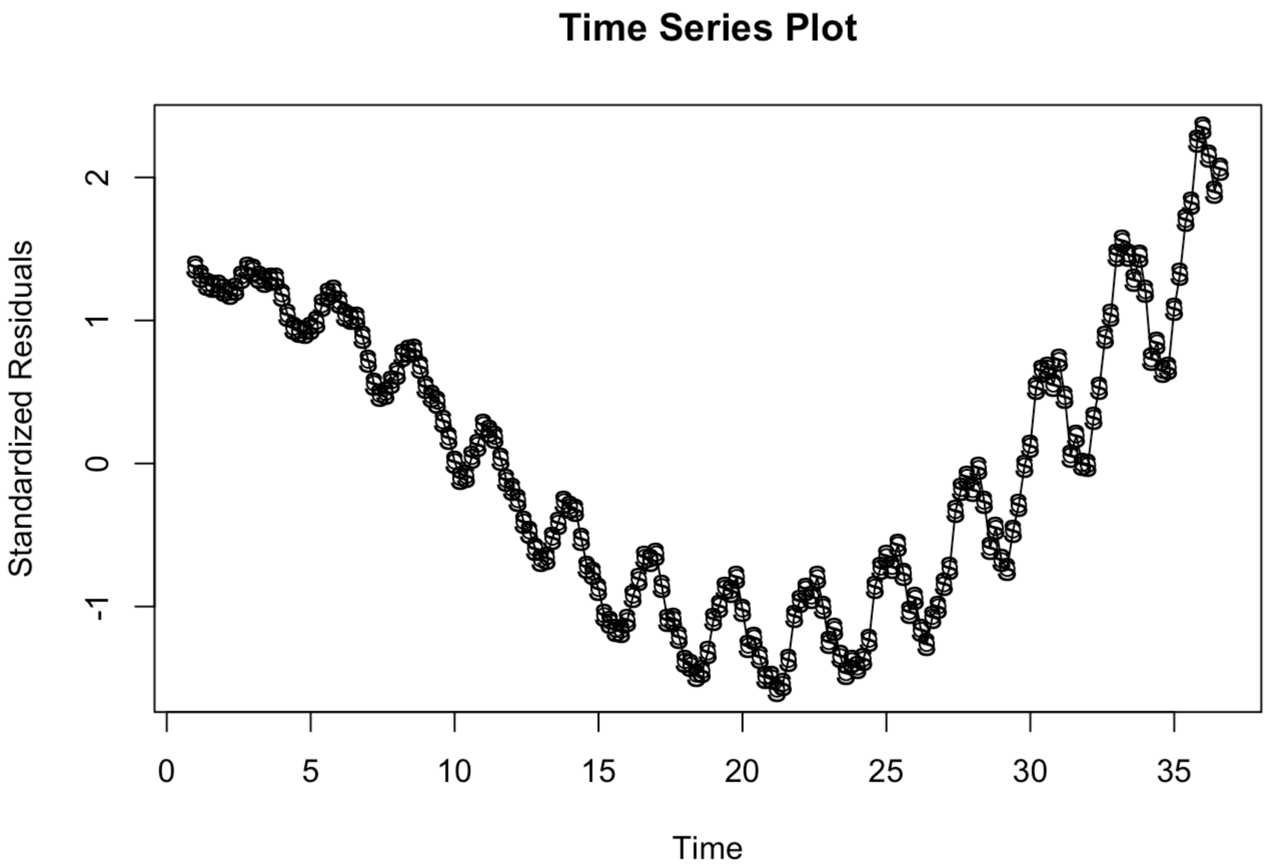
```
# Model 4 :Cosine Model
har <- harmonic(newTS, 1)
modelData <- data.frame(newTS, har)

model4 <- lm(newTS ~ cos.2.pi.t. + sin.2.pi.t. , data = modelData)
summary(model4)
```

```
##
## Call:
## lm(formula = newTS ~ cos.2.pi.t. + sin.2.pi.t., data = modelData)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -107.278  -59.716   -6.408   61.674  158.081
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  57.0348     5.1379  11.101   <2e-16 ***
## cos.2.pi.t.  -0.5054     7.2496  -0.070    0.945
## sin.2.pi.t.   1.2955     7.2826   0.178    0.859
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 68.74 on 176 degrees of freedom
## Multiple R-squared:  0.0002069,  Adjusted R-squared:  -0.01115
## F-statistic: 0.01821 on 2 and 176 DF,  p-value: 0.982
```

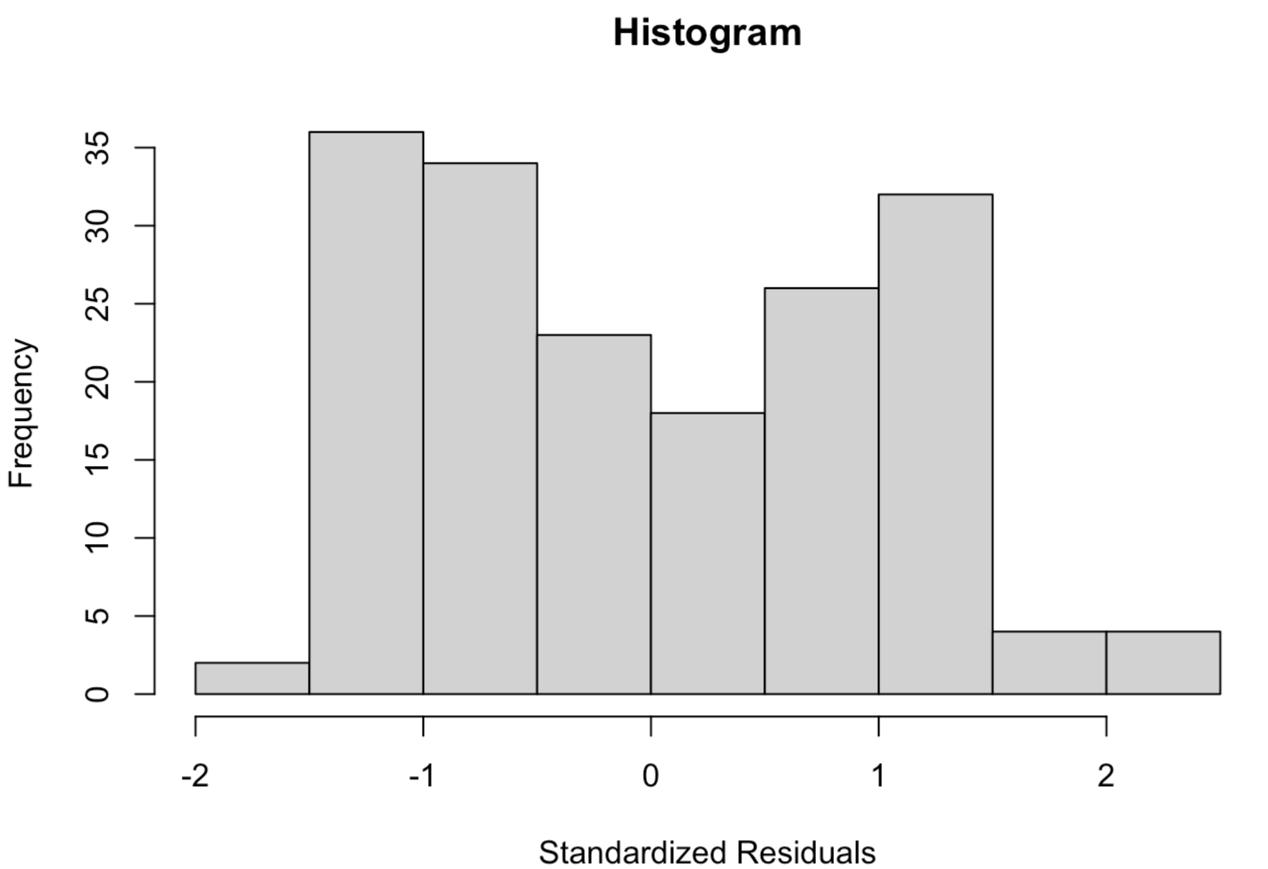
```
# Convert fitted values to numeric
fitted_values4 <- as.numeric(fitted(model4))

# Convert newTS to numeric
newTS_values4 <- as.numeric(as.vector(newTS))

# Calculate the limits for the y-axis
y_min <- min(c(fitted_values4, newTS_values4), na.rm = TRUE)
y_max <- max(c(fitted_values4, newTS_values4), na.rm = TRUE)

# Plot the fitted values
plot(ts(fitted_values4), ylab = 'y', main = "Fitted cosine wave",
     ylim = c(y_min, y_max), col = "red")

# Add lines for the actual values
lines(newTS_values, type = "o")
```



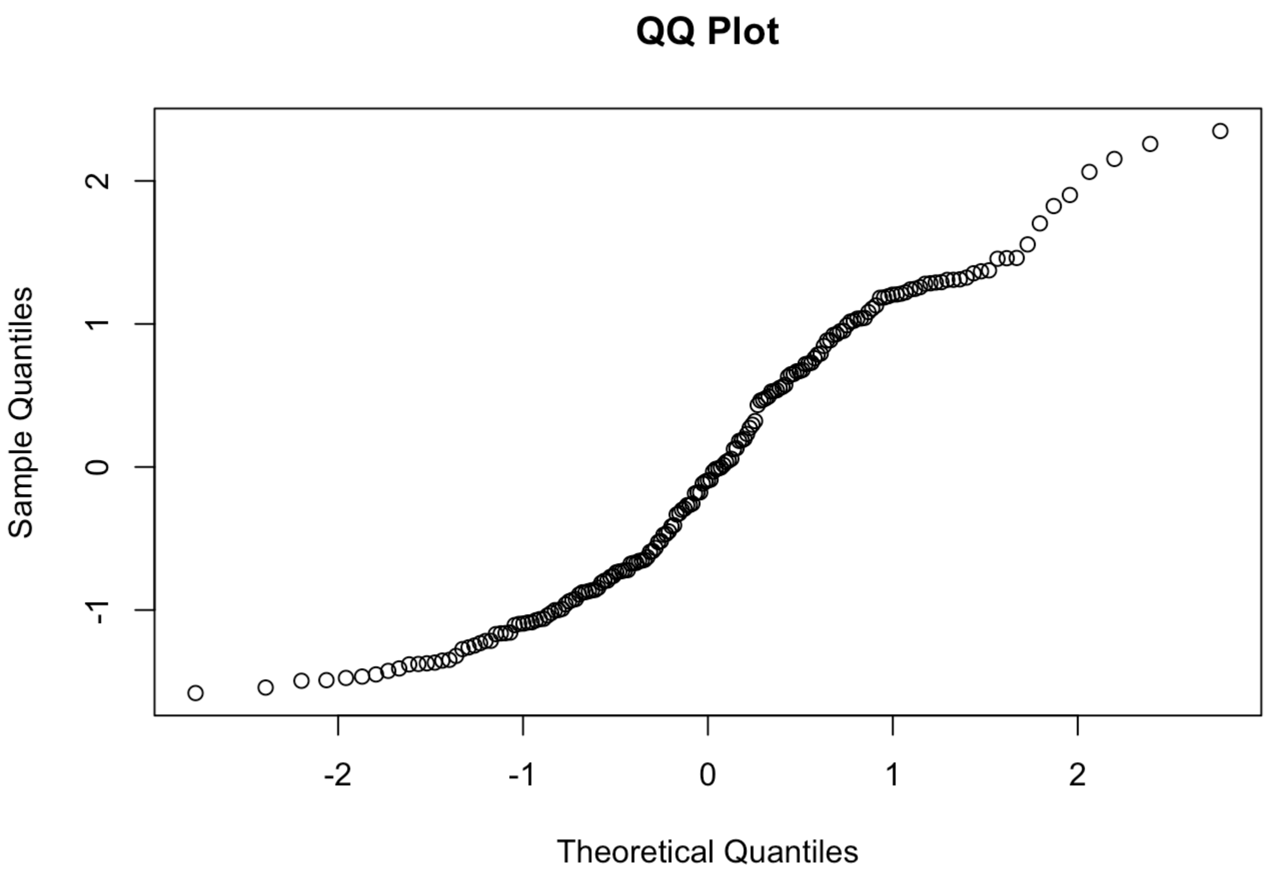

```
# Residual Analysis.
plot(y= rstudent(model4), x= as.vector(time(newTS)), xlab= 'Time',
     ylab='Standardized Residuals', type='o', main ="Time Series Plot")
points(y=rstudent(model4),x=as.vector(time(newTS)), pch=as.vector(season(newTS)))
```



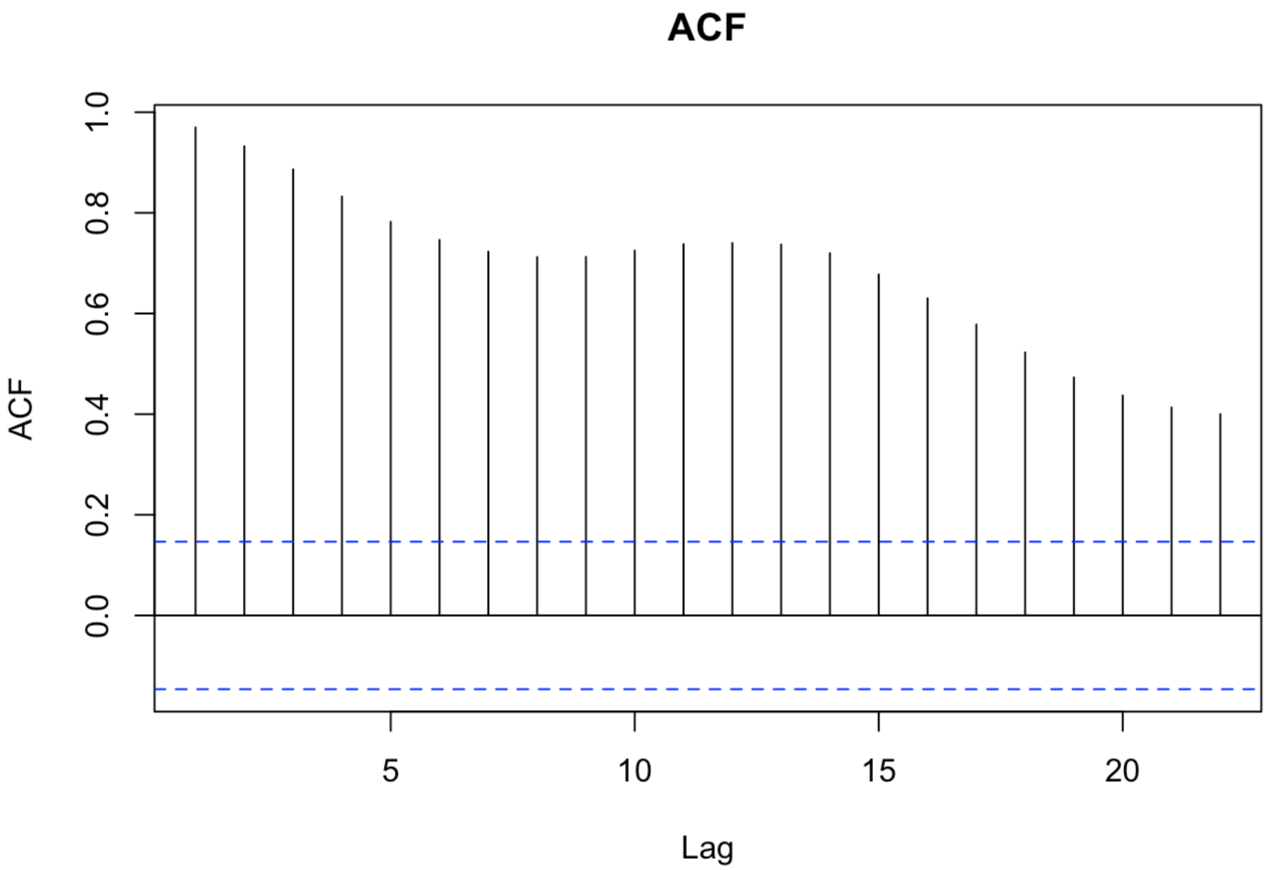

```
# Histogram distribution
hist(rstudent(model4),xlab='Standardized Residuals', main ="Histogram")
```



```
# Q-Q plots
y = rstudent(model4)
qqnorm(y, main ="QQ Plot")
```




```
# ACF
acf(rstudent(model4), main ="ACF")
```




```
# Shapiro-Wilk test
y = rstudent(model4)
shapiro.test(y)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  y
## W = 0.94756, p-value = 3.628e-06
```

Cosine Model Analysis

Coefficients:

The estimated intercept is 57.0348.The estimated coefficient for the predictor variable cos.2.pi.t. is -0.5054.The estimated coefficient for the predictor variable sin.2.pi.t. is 1.2955.

Significance of Coefficients:

The intercept coefficient is highly statistically significant with a very small p-value (essentially zero), indicating strong evidence against the null hypothesis that the intercept is zero.

The coefficients for both cos.2.pi.t. and sin.2.pi.t. are not statistically significant, as indicated by the high p-values (0.945 and 0.859, respectively). This suggests that neither of these predictor variables has a significant effect on newTS.The multiple R-squared value is very close to zero (0.0002069), indicating that only a very small proportion of the variance in newTS is explained by the predictor variables cos.2.pi.t. and sin.2.pi.t.The Adjusted R-squared value is negative (-0.01115), which is unusual and suggests that the model is not wellfitted to the data.The F-statistic tests the overall significance of the model. Here, the F-statistic is 0.01821 with a p-value of 0.982. Since the p-value is much greater than 0.05, we fail to reject the null hypothesis that all coefficients in the model are zero. This suggests that the overall model is not statistically significant.

The p value in the shapiro-wilk test which is 0.000003628 is less than 0.05 where we have to reject the null hypothesis of normality.The data does not appear to be consistent with a normal distribution at the 0.05 level

Overall, the analysis suggests that the linear regression model with cos.2.pi.t. and sin.2.pi.t. as predictors does not provide a good fit to the data. None of the predictor variables are statistically significant, and the model explains very little variance in the response variable newTS.

Result

In conclusion , we find that the quadratic model is the best fitting model to this data compared to the other models.

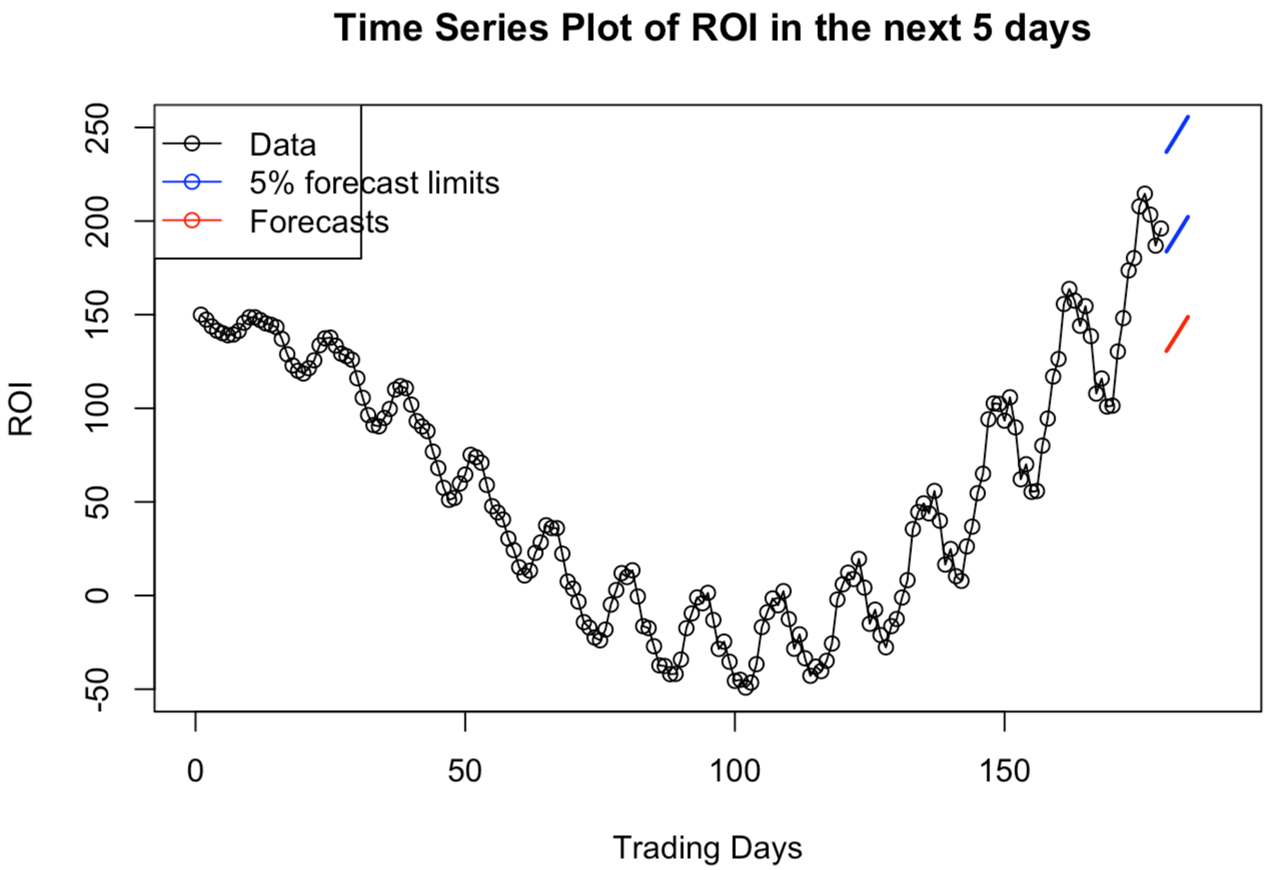
Task 2

Predictions for the next 5 trading days using the Quadratic Model.

```
t = time(ReturnsTS)
t2 = t^2

# Predict the data for the next five days
forecasts <- predict.lm(model2, data.frame(t=c(180,181,182,183,184), t2=c(180^2,181^
2,182^2,183^2,184^2)), interval = "prediction" )
print(forecasts)
```

```
##        fit      lwr      upr
## 1 183.7226 130.5525 236.8927
## 2 188.2826 135.0535 241.5117
## 3 192.8948 139.6044 246.1852
## 4 197.5593 144.2053 250.9132
## 5 202.2760 148.8561 255.6958
```

```
#Plot of forecasting for next 5 days
plot(ReturnsTS, ylab = "ROI",xlim=c(0,190),ylim=c(-50,250),type="o",xlab ="Trading Da
ys",
     main = " Time Series Plot of ROI in the next 5 days")

lines(ts(as.vector(forecasts[,1]), start = 180), col="blue", type="l", lwd=2)
lines(ts(as.vector(forecasts[,2]), start = 180), col="red", type="l", lwd=2)
lines(ts(as.vector(forecasts[,3]), start = 180), col="blue", type="l", lwd=2)
legend("topleft", lty=1, pch=1, col=c("black","blue","red"),
text.width = 18,
c("Data","5% forecast limits", "Forecasts"))
```



Result

From the above plot we can see that there is increment of Return of investment which follows the same pattern for the next five days.

Conclusion

Task 1: We can conclude that the best fitting model for the given dataset is the quadratic model as it satisfies the appropriate factors related to it

Task 2: The prediction of the next five trading days is an increment which follows the same pattern .

REFERENCES

https://otexts.com/fpp2/autocorrelation.html (https://otexts.com/fpp2/autocorrelation.html)