# STOCK PREDICTION

Time Series A1

s3890479 Faith Ha

# INTRODUCTION

The primary objective of this report is to utilize a model-building approach to predict the stock prices for the next five trading days. Through analysis and testing, this report will endeavor to ascertain an optimal time series model that effectively captures the intricate dynamics of the market. The report will evaluate a diverse array of models, encompassing linear, quadratic, seasonal, and cylindrical variations. By using a three-step model-building strategy, which includes model specification, model fitting, and diagnostic checking,  the report aims to provide valuable insights into the most robust and reliable model for navigating the complexities of the financial markets.

# DESCRIPTIVE ANALYSIS

Figure 1 depicts the movement of stock prices across a certain period. Upon examination, it becomes evident that the trend undergoes a significant change, initially displaying a downward trajectory in the first half before transitioning into an upward trend. Additionally, there appears to be a pattern of seasonality present within the data, although further investigation is necessary to conclusively establish its presence. Notably, the graph exhibits an auto regressive trend without any discernible change points and a changing variance. This result seems to confirm the data is a deterministic trend as it follows a fixed pattern depending only on time.

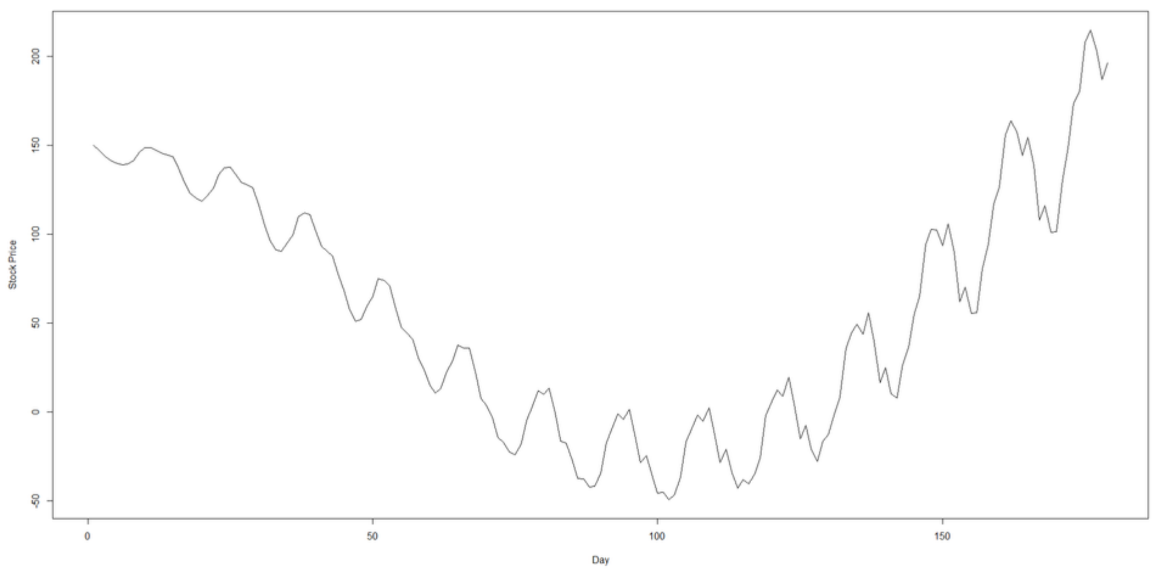



**Figure 1: Graph of 179 stock prices over time**

# SUMMARY OF STOCK PRICES

The summary statistics of the stock prices shown in Figure 2, depict a wide range of values. Prices vary from a minimum of -49.167 to a maximum of 214.611, with a median of 51.105 and a mean of 57.043. This suggests a diverse price distribution, highlighting positive and negative fluctuations over the observed period.

| Min | 1st Quartile | Median | Mean | 3rd Quartile | Max |
|---|---|---|---|---|---|
| -49.167 | -2.685 | 51.105 | 57.043 | 117.781 | 214.611 |

**Figure 2: Table of Stock Price Summary**

# AUGMENTED DICKEY-FULLER TEST

The Augmented Dickey-Fuller Test conducted on the stock time series indicates a test statistic of 0.24384 with a lag order of 5. The resulting p-value of 0.99 suggests a lack of evidence to reject the null hypothesis, indicating that the series is non-stationary. This finding suggests that the stock prices exhibit a trend over time. Accordingly, the result further confirms we use a deterministic model to make the series Trend-Stationary.

```
> adf_result <- adf.test(stock_ts)
> print(adf_result)

	Augmented Dickey-Fuller Test

data:  stock_ts
Dickey-Fuller = 0.24384, Lag order = 5, p-value = 0.99
alternative hypothesis: stationary
```

**Figure 3: Code for ADF Test**

# AUTOCORRELATION

The strength of autocorrelation between variables is evaluated using the zlag function, employing lag values of one and two. Figures 4 and 5 illustrate a significant autocorrelation between preceding variables, with a correlation of .985 observed between the first and .963 between the second. Both plots in figure 5 have strong linear relationships further confirming strong autocorrelation between stock prices.

```
> stock_price1 = stock_ts
> lagged_stock_price1 = zlag(stock_ts)
> index = 2:length(lagged_stock_price1) # Get rid of first NA value
> cor(stock_price1[index],lagged_stock_price1[index])
[1] 0.9868369

> stock_price2 = stock_ts
> lagged_stock_price2 = zlag(zlag(stock_ts))
> index = 3:length(lagged_stock_price2) # Get rid of first NA value
> cor(stock_price2[index],lagged_stock_price2[index])
[1] 0.963663
```

**Figure 4: Correlation Code**

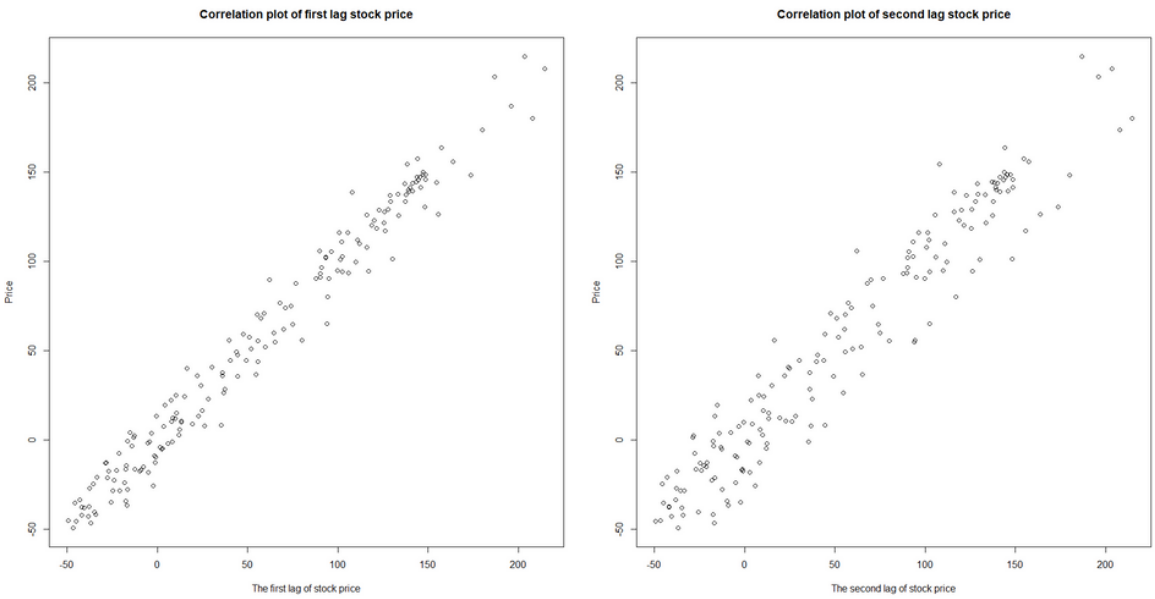



**Figure 5:  Correlation Plots**

# SEASONALITY

To investigate the presence of seasonality within the time series data, an Autocorrelation Function (ACF) plot was generated for the stock price column of the data frame, using a maximum lag of 60 (Figure 6). Analysis of the plot in Figure 7 suggests a recurring pattern with a frequency of approximately 13, inferred from the observed waves, despite some truncation at the beginning of the plot.

```
acf(stock_df$x, lag.max = 60, main = "ACF plot for Stock Data Frame")
```

**Figure 6: Code for ACF**

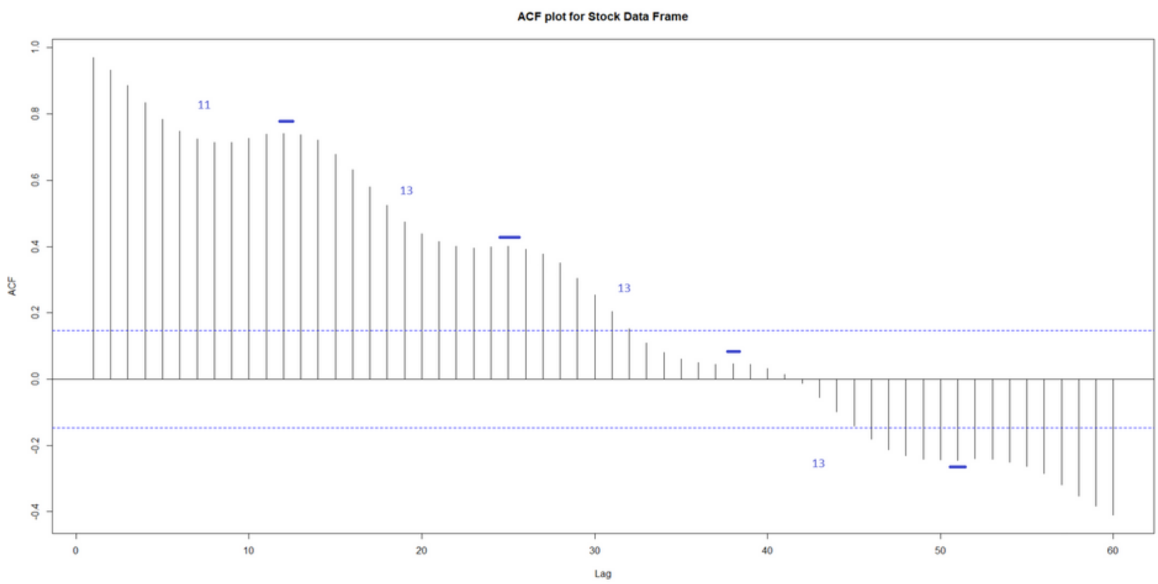



**Figure 7: ACF Plot for Stock Price**

# TIME SERIES

To further analyze and model the data, the data frame was converted into a time series plot with a frequency of 13 shown in Figures 8 and 9.

```
stock_ts <- ts(stock_df$x,frequency =13)
```

**Figure 8: Code for Time Series**

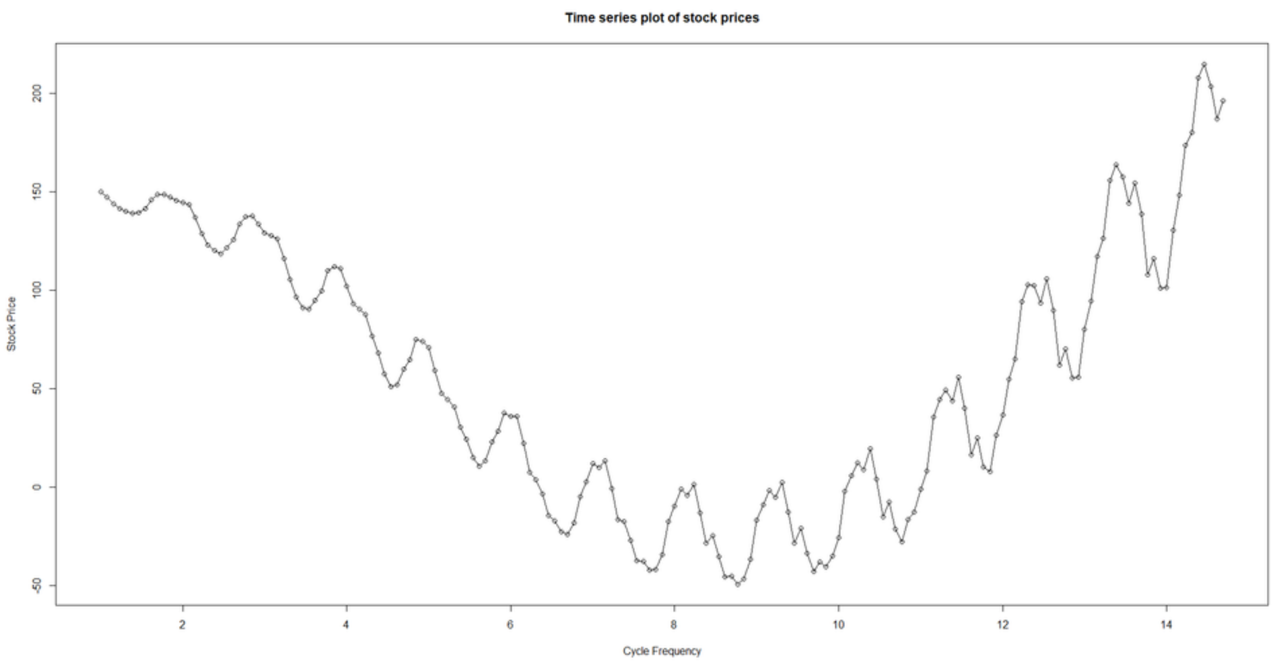
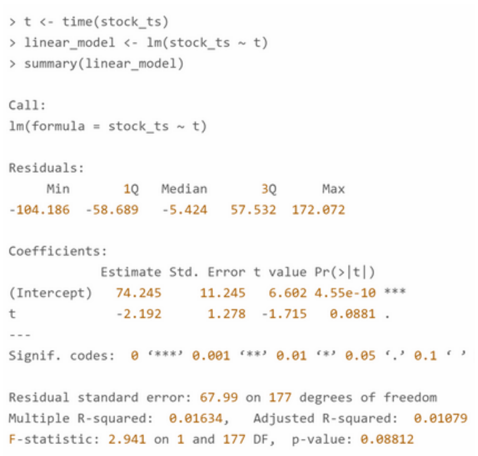



**Figure 9: Time Series Plot**

# MODEL ANALYSIS

## MODEL 1: LINEAR MODEL



```
> t <- time(stock_ts)
> linear_model <- lm(stock_ts ~ t)
> summary(linear_model)

Call:
lm(formula = stock_ts ~ 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)   74.245     11.245   6.602 4.55e-10 ***
t             -2.192      1.278  -1.715   0.0881 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’

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
```

Figure 10: Linear Model Results

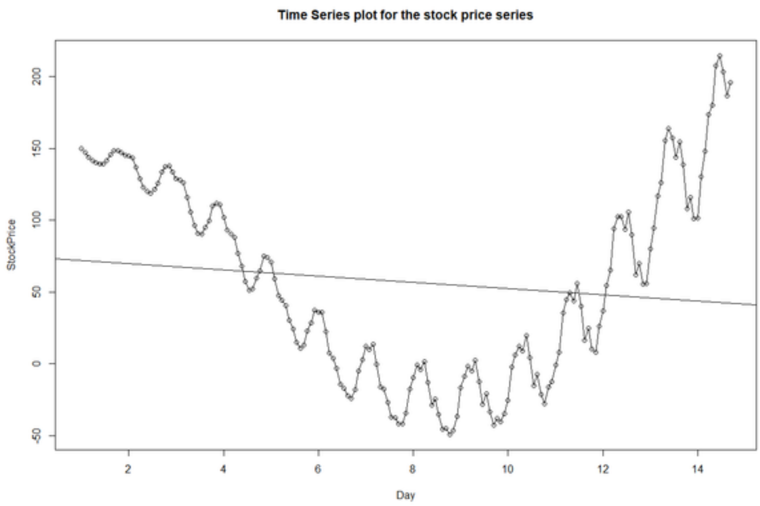



Figure 11: Linear Model Graph

### MODEL SPECIFICATION & FITTING

A linear model was constructed to evaluate its ability to model the dataset. However, analysis of figures 10 and 11 reveals that the model failed to capture the data adequately, evident from its relatively high p-value of 0.0881. This value exceeds the typical threshold of 0.05 for statistical significance, indicating that the model failed to reject the null hypothesis and lacks significance. The model estimated that the intercept is 74.245, which represents the expected stock price at the beginning of the time series. The coefficient for the time variable (t) is -2.192, suggesting a slight negative trend over time, which is incorrect as seen in Figure 11 where there is a changing trend. Additionally, the adjusted R-squared value is 0.01079. This indicates that after adjusting for the time variable approximately 1.1% of the variability in the stock prices can be explained by the linear trend in time. The model's multiple R-squared value of 0.01634 indicates that only approximately 1.6% of the variance in the stock prices can be explained by the linear trend in time. The residual standard error which represents the typical deviation of observed values from the fitted values is quite high, indicating that the model may not be capturing all the underlying patterns. Overall, while the linear model suggests a slight decreasing trend in stock prices over time, the model is not statistically significant and only explains a small amount of the variability in the stock prices.

## DIAGNOSTIC CHECK FOR LINEAR MODEL

```
> residual_analysis(linear_model, "linear model")
Shapiro-Wilk test p-value: 1.022466e-05
```

**Figure 12: Code for residual analysis**

Figures 12 and 13, show the residual analysis of the linear model. The first plot shows that the linear model failed to capture the underlying trend in the data. The histogram is not symmetrical, indicating that the residuals were not standardised. Furthermore, the QQ plot shows that the model fails to capture normality and the Shapiro-Wilk test rejected the null hypothesis that there is normality, as the p-value was extremely small (Figure 12). Lastly, the ACF shows that large amounts of autocorrelation are still around. Accordingly, the linear model is soundly rejected as a possible model based on it lack of significance and failed residual tests.

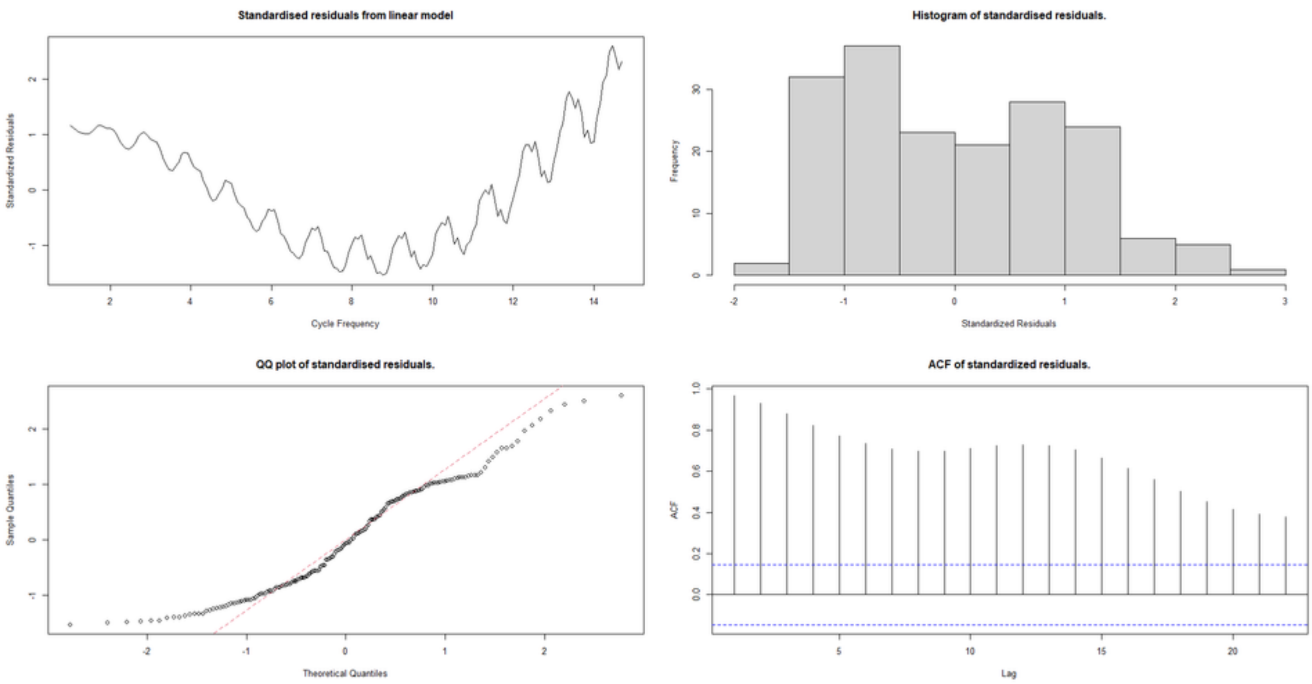



**Figure 13: Residual Analysis Plots**

# MODEL 2: QUADRATIC MODEL

## MODEL SPECIFICATION & FITTING

Next, a quadratic model is used to determine the model's ability to model the data. It performed significantly better than the previous model. As seen in Figures 14 and 15, the p-value is less than the null hypothesis of 0.05 indicating the model is significant. While the linear variable (t) captures the overall downward trend, the quadratic variable (t2) accounts for the curvature in this trend, showing that the rate of decrease in stock prices slows over time. The high value of the adjusted R-squared (0.8539) indicates that approximately 85.39% of the variability in stock prices can be explained by the quadratic model, which is a substantial improvement over the linear model. The multiple R-squared value of 0.8523 reveals that approximately 85.23% of the variance in the stock prices can be explained by the quadratic trend in time. The residual error of 26.27 is also significantly smaller indicating that the model is accounting for the trends signficantly better. Additionally, the F-statistics large value and associated small p-value, reinforce the model's significance, indicating that the overall regression model is statistically significant. Overall, the quadratic model provides a much better fit to the data than the linear model and is a more appropriate choice for modeling the relationship between time and stock prices.

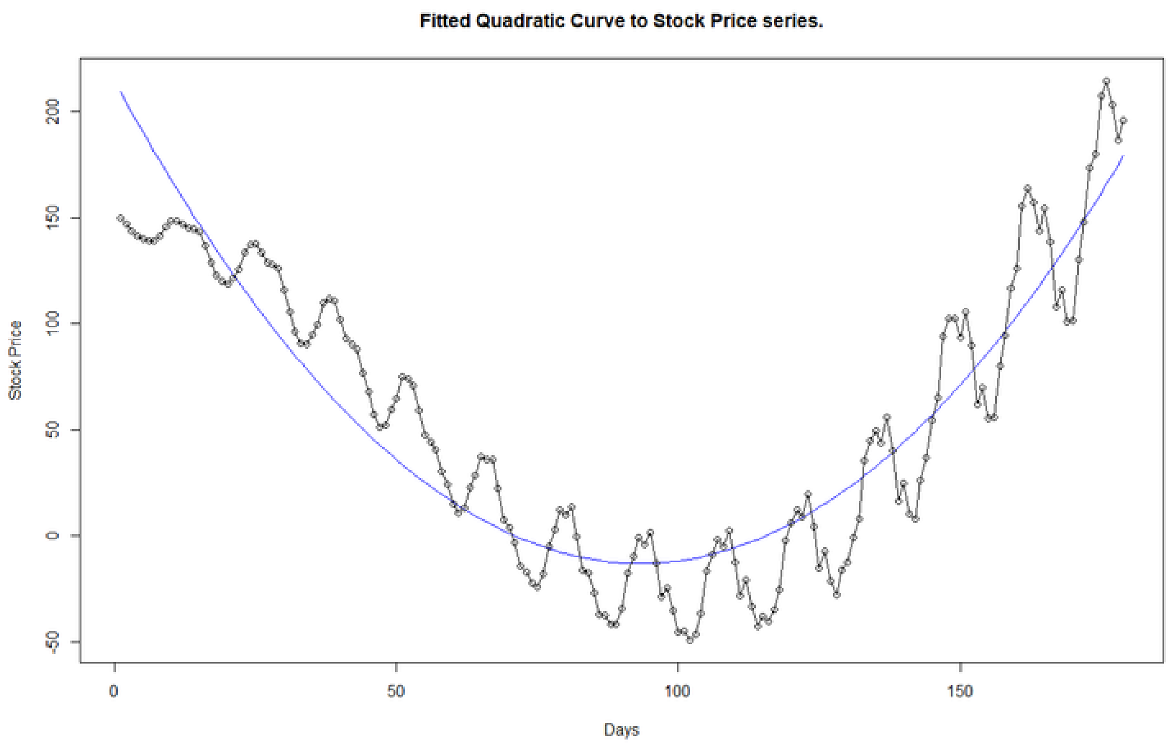



**Figure 14: Quadratic Model Graph**

```
> quadratic_model <- lm(stock_ts ~ t + t2)
> summary(quadratic_model)

Call:
lm(formula = stock_ts ~ 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)  276.295      7.703   35.87   <2e-16 ***
t            -71.476      2.236  -31.96   <2e-16 ***
t2             4.415      0.139   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
```

**Figure 15: Code for Quadratic Model Analysis**

```
> residual_analysis(quadratic_model, "quadratic model")
Shapiro-Wilk test p-value: 0.03798675
```

**Figure 16: Code for Residual Analysis**

# DIAGNOSTIC CHECK FOR QUADRATIC MODEL

The residual analysis also performed much better. From plot 1 of figure 17, we see that the seasonality seems to be captured much better with most of the graph looking random. However, it does dip down to -2 at the start, indicating there is still trend in the data. The histogram is also symmetrical which means the residuals are standardised. However, the QQ plot and Shapiro-Wilk test show the data is not normalized enough with a p-value of 0.038. There is also significant autocorrelation observed in the ACF plot. Accordingly, while the quadratic model performs significantly better than the linear model, it is still not able to account fully for normality, seasonality, and autocorrelation.

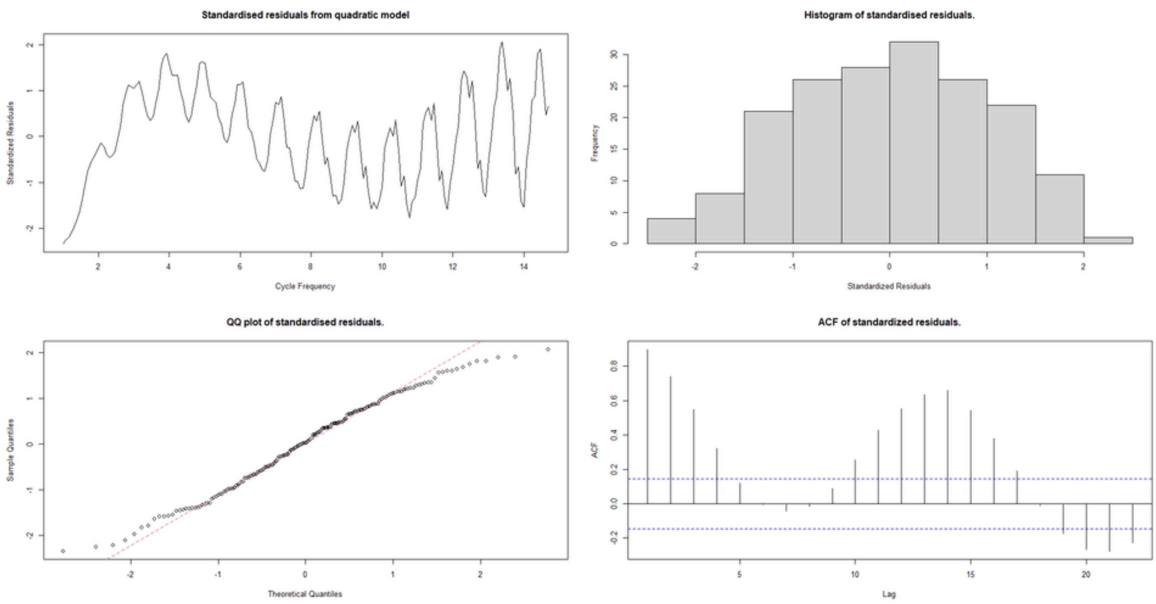



**Figure 17: Residual Analysis Plots for Quadratic Model**

# MODEL 3: SEASONAL MODEL

## MODEL SPECIFICATION & FITTING

The seasonal model fitted to the stock price time series aims to capture the seasonal variation in stock prices by including categorical variables representing each cycle. As seen in figure 18 and 19, The summary output of the model reveals that the coefficients for each cycle, except the 11th cycle coefficent which is .0506, are statistically significant, with p-values less than 0.05. The overall model fit is moderate, as indicated by the multiple R-squared value of 0.4249, which implies that approximately 42.49% of the variance in stock prices can be explained by the seasonal factors. However, the adjusted R-squared value of 0.3798, which adjusts for the number of predictors in the model, suggests that around 37.98% of the variance can be explained, indicating a slightly lower model fit. The residual standard error of 70 indicates high deviation of observed stock prices from the fitted values. The F-statistic of 9.433 and its associated p-value of 1.747e-14 suggest that the model as a whole is statistically significant. Overall, the seasonal model provides valuable insights into the seasonal variation in stock prices, with most seasons showing significant effects on stock prices compared to the reference season. However, while it seems to caputure the trend of each cycle, it fails to capture the overall day to day trend.

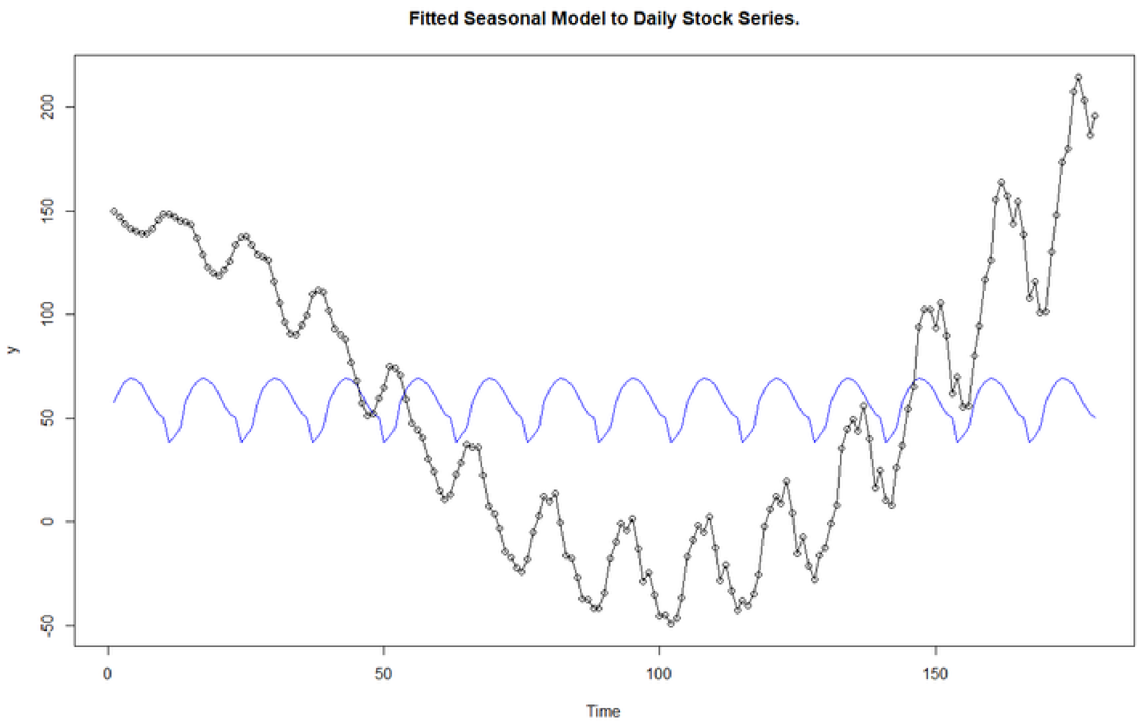



**Figure 18: Seasonal Model Graph**

```
> cycle. = season(stock_ts)
> seasonal_model= lm(stock_ts ~ cycle. -1)
> summary(seasonal_model)

Call:
lm(formula = stock_ts ~ cycle. - 1)

Residuals:
    Min      1Q  Median      3Q     Max
-97.730 -64.795  -7.509  58.863 153.035

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
cycle.Season-1     57.85      18.71   3.092 0.002332
cycle.Season-2     63.74      18.71   3.407 0.000824
cycle.Season-3     67.61      18.71   3.614 0.000399
cycle.Season-4     69.48      18.71   3.714 0.000278
cycle.Season-5     68.52      18.71   3.663 0.000335
cycle.Season-6     66.39      18.71   3.549 0.000504
cycle.Season-7     61.58      18.71   3.291 0.001218
cycle.Season-8     56.22      18.71   3.005 0.003068
cycle.Season-9     52.16      18.71   2.788 0.005924
cycle.Season-10    50.12      18.71   2.679 0.008123
cycle.Season-11    38.22      19.41   1.969 0.050660
cycle.Season-12    41.32      19.41   2.128 0.034778
cycle.Season-13    45.03      19.41   2.320 0.021577

cycle.Season-1  **
cycle.Season-2  ***
cycle.Season-3  ***
cycle.Season-4  ***
cycle.Season-5  ***
cycle.Season-6  ***
cycle.Season-7  **
cycle.Season-8  **
cycle.Season-9  **
cycle.Season-10 **
cycle.Season-11 .
cycle.Season-12 *
cycle.Season-13 *
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 70 on 166 degrees of freedom
Multiple R-squared:  0.4249,    Adjusted R-squared:  0.3798
F-statistic: 9.433 on 13 and 166 DF,  p-value: 1.747e-14
```

**Figure 19: Seasonal Model Results**

```
> residual_analysis(seasonal_model, "Seasonal Model", stock_ts)
Shapiro-Wilk test p-value: 3.716058e-07
```

**Figure 20: Residual Analysis Code**

## RESIDUAL ANALYSIS OF SEASONAL MODEL

The analysis of residuals, depicted in Figures 20 and 21, reveals limitations in the seasonal model's ability to adequately represent the underlying data trends. While the model effectively captures the average behavior for each cycle, it falls short in capturing the overall trend of the data. The skewed distribution observed in the histogram indicates that the residuals are not fully standardized. Additionally, the QQ plot and Shapiro-Wilk test results further confirm the residuals do not have normality. The ACF plot also shows large amounts of autocorrelation which the model is unable to explain. In summary, while the seasonal model adequately captures cycle behavior, there is still residual seasonality, unstandardised data, normality, and autocorrelation left over.

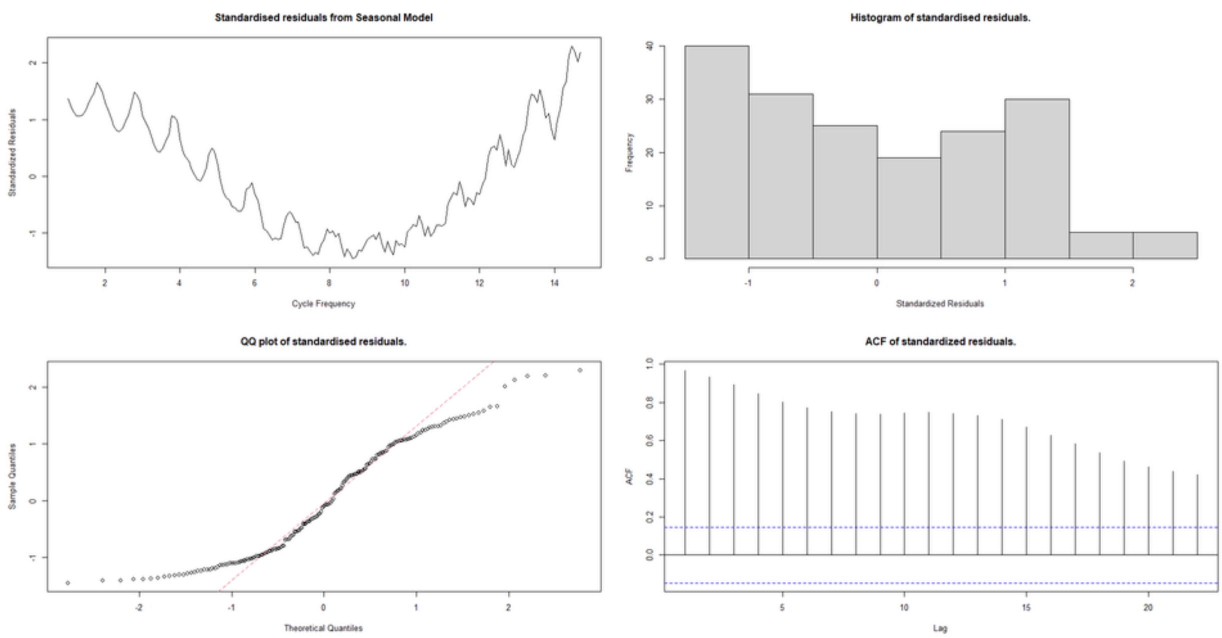



**Figure 21: Residual Analysis Plots for Seasonal Model**

# MODEL 4: CYCLIC MODEL

## MODEL SPECIFICATION & FITTING

As seen in Figures 22 and 23, a cyclic model was constructed to explore the relationship between stock prices and both the cosine and sine components of a periodic function, allowing for a more flexible representation of cyclic patterns in the data. However, upon examination of the model's coefficients, it appears that neither the cosine term nor the sine term significantly contributes to explaining the variability in stock prices. This conclusion is supported by the relatively high p-values associated with both coefficients, indicating that they are not statistically significant. Additionally, the low Multiple R-squared value of 0.02049 suggests that only a small proportion of the variance in stock prices is explained by the cosine model. The adjusted R-squared value is even lower, further indicating the limited explanatory power of the model after accounting for the number of predictors. Overall, while the cosine model was constructed to capture cyclic patterns in the data, the results suggest that it does not provide a meaningful explanation of the relationship between stock prices and time, as evidenced by the non-significant coefficients and low R-squared values.

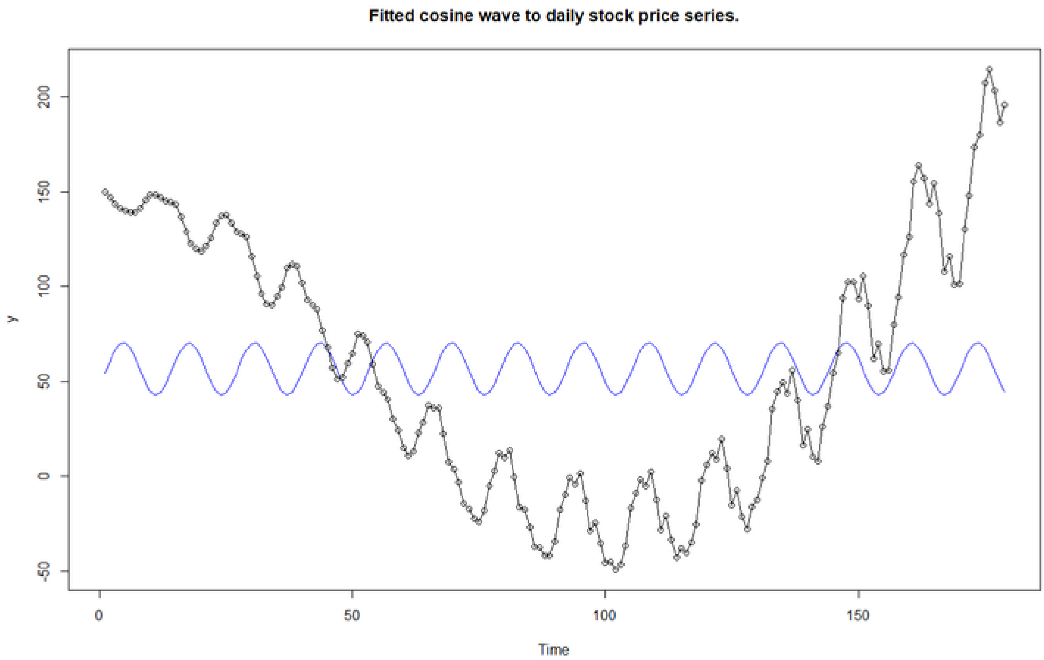



**Figure 22: Cyclic Model Graph**

```
> har. <- harmonic(stock_ts, 1) # calculate cos(2*pi*t) and sin(2*pi*t)
> stock <- data.frame(stock_ts,har.)
> cyclic <- lm(stock_ts ~ cos.2.pi.t. + sin.2.pi.t. , data = stock)
> summary(cyclic)

Call:
lm(formula = stock_ts ~ cos.2.pi.t. + sin.2.pi.t., data = stock)

Residuals:
    Min      1Q  Median      3Q     Max 
-96.240 -63.559  -6.592  59.015 152.142 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   56.847      5.087  11.176   <2e-16 ***
cos.2.pi.t.   -2.427      7.178  -0.338   0.7356    
sin.2.pi.t.   13.641      7.209   1.892   0.0601 .  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 68.04 on 176 degrees of freedom
Multiple R-squared:  0.02049,	Adjusted R-squared:  0.009354 
F-statistic:  1.84 on 2 and 176 DF,  p-value: 0.1618
```

**Figure 23: Cosine Model Results**

```
> residual_analysis(cyclic, "Cyclic Model", stock_ts)
Shapiro-Wilk test p-value: 3.07705e-07
```

**Figure 24: Cyclic Residual Analysis Code**

## RESIDUAL ANALYSIS OF CYCLIC MODEL

Upon examining the residual plots for the cyclic model, depicted in Figure 25, it becomes evident that the model falls short in capturing both the seasonality and trend of the data. Additionally, the histogram reveals that the data is not standardized, while the QQ-plot and Shapiro-Wilk test indicate that it is not normally distributed (Figure 24). Furthermore, the presence of significant autocorrelation, as depicted in the ACF plot, suggests that the model fails to adequately account for temporal dependencies in the data. Consequently, based on these findings, the cyclic model is deemed unsuitable for use as it lacks statistical significance and fails all four residual tests.

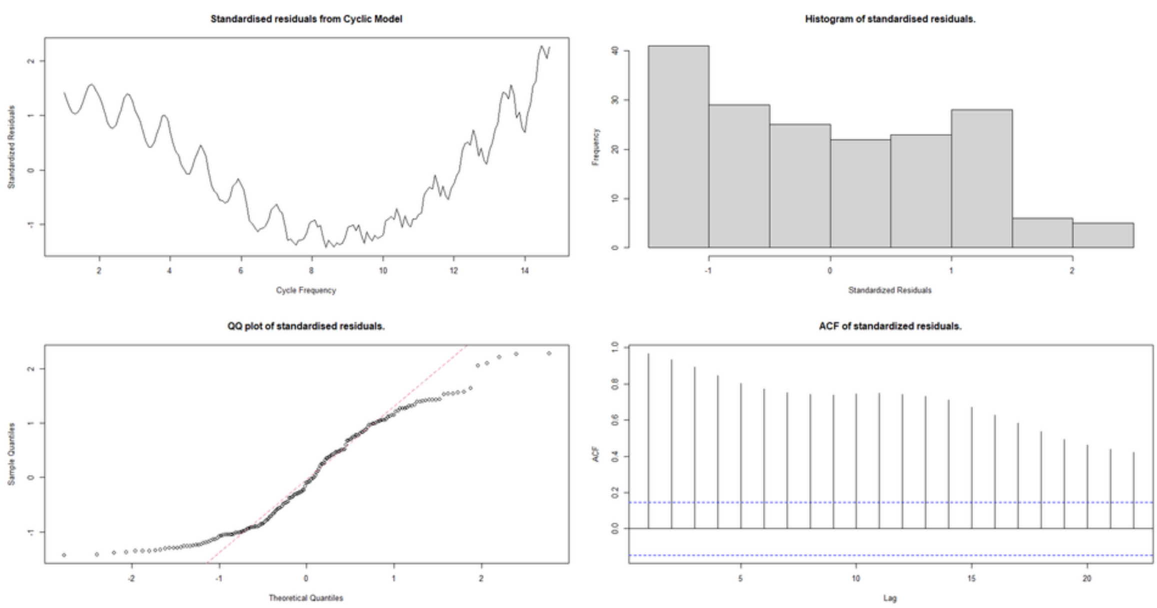



**Figure 25: Residual Analysis Plots for Cyclic Model**

# MODEL 5: QUADRATIC + SEASONAL MODEL

## MODEL SPECIFICATION & FITTING

Finally, the quadratic and seasonal model are combined to test its combined predictive powers on the stock series. As seen in Figures 26 and 27, the combined model demonstrates strong statistical significance, as indicated by the very low p-values associated with all coefficients. The coefficients for the cycle season effects show consistent and substantial estimated changes in stock prices across different cycle days, with all t-values highly significant. Similarly, the coefficients for the linear and quadratic trends (t and t^2) also display significant effects on stock prices, with extremely low p-values and high t-values. The Multiple R-squared value of 0.9205 suggests that approximately 92.05% of the variability in stock prices is explained by the model, while the Adjusted R-squared value of 0.9132 accounts for the number of predictors in the model. The low residual standard error of 26.18 indicates that the model is accounting for the pattern. However, figure 24 shows that although the model seems a good fit the model still has trouble fitting the data. Overall, these results suggest that the quadratic trend model provides a highly significant and explanatory framework for understanding the relationship between stock prices and both linear and quadratic trends.

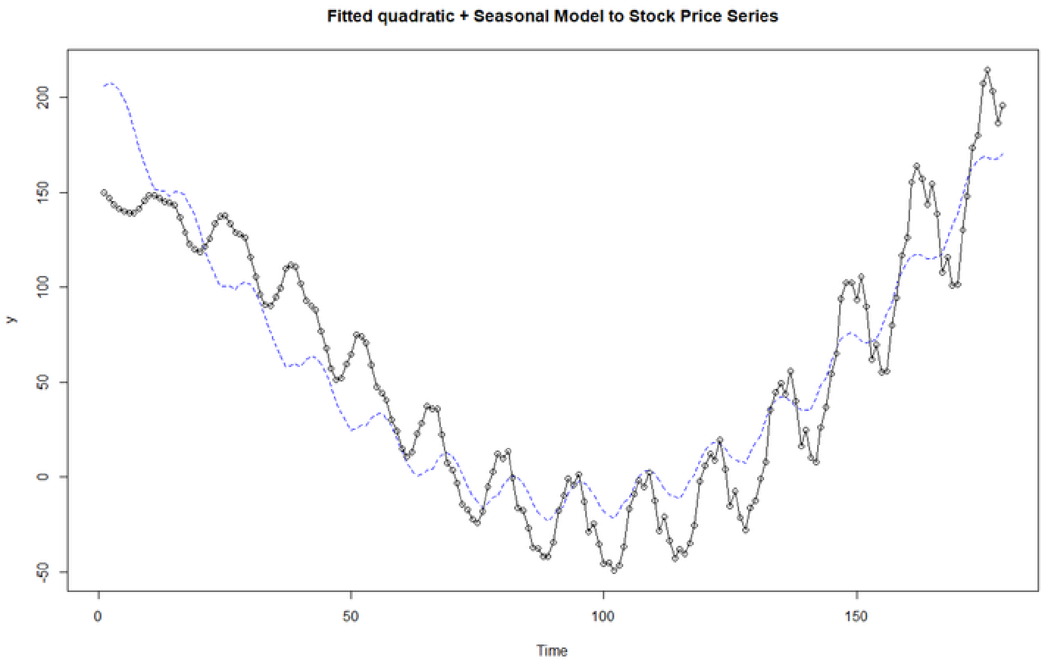



**Figure 26: Combined Model Graph**

```
> summary(combined_model)

Call:
lm(formula = stock_ts ~ cycle. + t + t2 - 1)

Residuals:
    Min      1Q  Median      3Q     Max 
-63.196 -18.423  -2.084  18.663  53.120 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
cycle.Season-1  272.7195    10.0982   27.01   <2e-16 ***
cycle.Season-2  278.9824    10.1163   27.58   <2e-16 ***
cycle.Season-3  283.1832    10.1334   27.95   <2e-16 ***
cycle.Season-4  285.3199    10.1496   28.11   <2e-16 ***
cycle.Season-5  284.5758    10.1649   28.00   <2e-16 ***
cycle.Season-6  282.6070    10.1793   27.76   <2e-16 ***
cycle.Season-7  277.9117    10.1928   27.27   <2e-16 ***
cycle.Season-8  272.6147    10.2055   26.71   <2e-16 ***
cycle.Season-9  268.5631    10.2172   26.29   <2e-16 ***
cycle.Season-10 266.4862    10.2281   26.05   <2e-16 ***
cycle.Season-11 264.0656    10.5339   25.07   <2e-16 ***
cycle.Season-12 267.3595    10.5483   25.35   <2e-16 ***
cycle.Season-13 271.2121    10.5617   25.68   <2e-16 ***
t               -71.1527     2.2332  -31.86   <2e-16 ***
t2                4.3969     0.1388   31.68   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 26.18 on 164 degrees of freedom
Multiple R-squared:  0.9205,	Adjusted R-squared:  0.9132 
F-statistic: 126.6 on 15 and 164 DF,  p-value: < 2.2e-16
```

**Figure 27: Combined Model Results**

```
> residual_analysis(combined_model, "Combined Model", stock_ts)
Shapiro-Wilk test p-value: 0.0665463
```

**Figure 28: CombinedResidual Analysis Code**

## RESIDUAL ANALYSIS OF COMBINED MODEL

The analysis of residuals, depicted in Figure 29, indicates that while the model is improving in capturing seasonality, it still encounters challenges, particularly at the start. The graph displays a more random pattern overall, suggesting better modeling of seasonality. The histogram exhibits a relatively symmetrical distribution, yet a longer tail suggests potential non-standardization of the data. Notably, the model passes the Shapiro-Wilk normality test, as evidenced by the QQ plot in Figure 28. However, substantial autocorrelation persists in the ACF plot, indicating that temporal dependencies are still present in the residuals. In summary, although the model demonstrates reasonable performance in capturing the data, there remains room for enhancement, particularly in addressing residual autocorrelation.

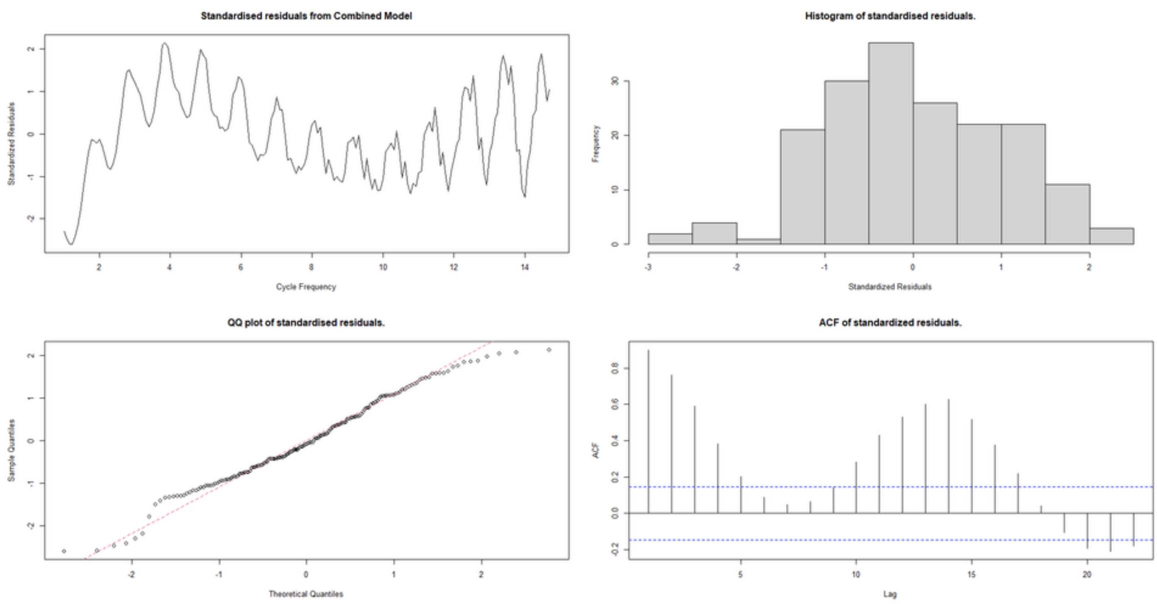



**Figure 29: Residual Analysis Plots for Combined Model**

# PREDICTION

After analyzing five models, the combined model was chosen for its strengths in achieving normality in residuals, highest R-squared score, and effective handling of seasonality. However, residual autocorrelation and incomplete seasonal variation adjustment were noted. Despite its advantages, predictions from the combined quadratic and seasonal model exhibit significant variability, indicating uncertainty in estimates. Forecasts for the next 5 days integrate both trend and seasonal variations, providing insights with 95% prediction intervals (Figures 30 and 31). These forecasts acknowledge inherent modeling uncertainties, offering a range for future observations and showcasing the model's confidence level.

| | Fit | Lower | Upper |
|---|---|---|---|
| **1** | 172.2842 | 117.4134 | 227.1551 |
| **2** | 180.1213 | 125.1997 | 235.0430 |
| **3** | 188.5693 | 133.5953 | 243.5433 |
| **4** | 194.7239 | 139.8827 | 249.5652 |
| **5** | 205.6862 | 150.7890 | 260.5834 |

**Figure 30: Table of Prediction Results**

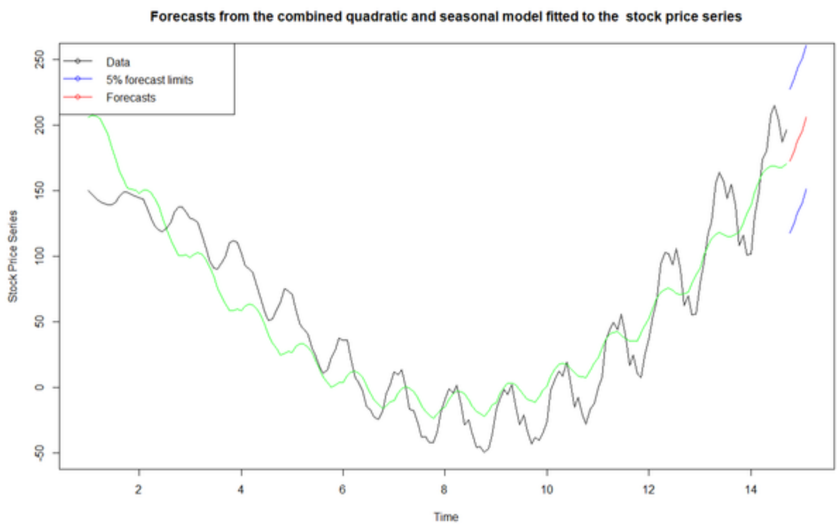



**Figure 31: Prediction Graph**

# CONCLUSION

After utilizing a systematic model-building strategy and analyzing all five models with residual analysis, the combined model outperformed the rest in terms of r-square score, seasonality, and normality. Despite its strengths, it's important to note that none of the models adequately addressed significant autocorrelation issues and all models struggled to address seasonality. Accordingly, while the combined model offers valuable insights and predictive capabilities, it's important to exercise caution, recognizing that no model is without its shortcomings and should be used with caution.

# CODE APPENDIX

```
rm(list=ls()) # Clean R's memory!

# Load libraries
library(TSA)
library(ggplot2)
library(tseries)
# Function for residual analysis
residual_analysis <- function(model, title, data = NULL) {
  res <- rstudent(model)
  par(mfrow=c(2,2))
  plot(y = res, x = as.vector(time(stock_ts)), xlab = 'Cycle Frequency', ylab='Standardized Residuals',
 type='l', main = paste("Standardised residuals from", title))
  hist(res, xlab='Standardized Residuals', main = "Histogram of standardised residuals.")
  qqnorm(y=res, main = "QQ plot of standardised residuals.")
  qqline(y=res, col = 2, lwd = 1, lty = 2)
  sw_test <- shapiro.test(res)
  cat("Shapiro-Wilk test p-value:", sw_test$p.value, "\n")
  acf(res, main = "ACF of standardized residuals.")
  #pacf(res, main = "PACF of standardized residuals.")
  par(mfrow=c(1,1))
}
# Read csv file
stock_df <- read.csv("assignment1Data2024.csv", header = TRUE)

# Plot stock price
plot(stock_df, xlab = "Day", ylab = "Stock Price", type = "l")

# Check seasonality with ACF
acf(stock_df$x, lag.max = 60, main = "ACF plot for Stock Data Frame")

# Convert dataframe to time series
stock_ts <- ts(stock_df$x, frequency = 13)

plot(stock_ts, xlab = "Cycle Frequency", ylab = "Stock Price", type = "o",
     main = "Time series plot of stock prices")

# Test for stationarity
adf_result <- adf.test(stock_ts)
print(adf_result)

# Linear model
t <- time(stock_ts)
linear_model <- lm(stock_ts ~ t)
summary(linear_model)
plot(stock_ts, ylab = 'StockPrice', xlab = 'Day', type = 'o', main = "Time Series plot for the stock
price series")
abline(linear_model)

# Residual analysis for linear model
residual_analysis(linear_model, "linear model")

# Quadratic model
t2 <- t^2
quadratic_model <- lm(stock_ts ~ t + t2)
summary(quadratic_model)

# Fitted Quadratic Curve to Stock Price series
plot(ts(fitted(quadratic_model)), ylab = 'Stock Price', xlab = 'Days', main = "Fitted Quadratic Curve
to Stock Price series.", col = "blue")
lines(as.vector(stock_ts), type = "o")
```

```r
# Residual analysis for quadratic model
residual_analysis(quadratic_model, "quadratic model")

# Seasonal Model
cycle. = season(stock_ts)
seasonal_model= lm(stock_ts ~ cycle. -1)
summary(seasonal_model)

plot(ts(fitted(seasonal_model)), ylab='y', main = "Fitted Seasonal Model to Daily Stock Series.",
     ylim = c(min(c(fitted(seasonal_model), as.vector(stock_ts))) ,
              max(c(fitted(seasonal_model), as.vector(stock_ts)))
     ), col = "blue" ) +
  lines(as.vector(stock_ts),type="o")

# residual analysis
residual_analysis(seasonal_model, "Seasonal Model", stock_ts)

# cyclic model
har. <- harmonic(stock_ts, 1) # calculate cos(2*pi*t) and sin(2*pi*t)
stock <- data.frame(stock_ts,har.)
cyclic <- lm(stock_ts ~ cos.2.pi.t. + sin.2.pi.t. , data = stock)
summary(cyclic)

plot(ts(fitted(cyclic)), ylab='y', main = "Fitted cosine wave to daily stock price series.",
     ylim = c(min(c(fitted(cyclic), as.vector(stock_ts))) ,
              max(c(fitted(cyclic), as.vector(stock_ts)))
     ), col = "blue" )
lines(as.vector(stock_ts),type="o")

# residual analysis
residual_analysis(cyclic, "Cyclic Model", stock_ts)

# Combined Model Quadratic + Seasonal
combined_model = lm(stock_ts~ cycle. + t + t2 -1)
summary(combined_model)

fitted.combined_model <- fitted(combined_model)

plot(ts(fitted(combined_model)), ylim = c(min(c(fitted(combined_model), as.vector(stock_ts))),
max(c(fitted(combined_model),as.vector(stock_ts)))),
     ylab='y' , main = "Fitted quadratic + Seasonal Model to Stock Price Series",
type="l",lty=2,col="blue")
lines(as.vector(stock_ts),type="o")
# residual analysis
residual_analysis(combined_model, "Combined Model", stock_ts)

# prediction
day_forecast <- 5 # 5 day forecast
t <- time(stock_ts)
dataFreq <- frequency(stock_ts)
last_time_point <- t[length(t)]
aheadTimes <- data.frame(cycle. = c("Season-11", "Season-12", "Season-13", "Season-1", "Season-2"),
                         t = seq(last_time_point+(1/dataFreq),
last_time_point+day_forecast*(1/dataFreq), 1/dataFreq),
                         t2 =  seq(last_time_point+(1/dataFreq),
last_time_point+day_forecast*(1/dataFreq), 1/dataFreq)^2)

frc_combined_model <- predict(combined_model, newdata = aheadTimes, interval = "prediction")
```

```
plot(stock_ts, xlim= c(t[1],aheadTimes$t[nrow(aheadTimes)]), ylim = c(-50,250), ylab = "Stock Price
Series", xlab = "Time",
     main = "Forecasts from the combined quadratic and seasonal model fitted to the  stock price
series")
lines(ts(fitted.combined_model,start = t[1],frequency = dataFreq), col = "green") #  Alternatively
you can use abline(model1)
lines(ts(as.vector(frc_combined_model[,3]), start = aheadTimes$t[1],frequency = dataFreq),
col="blue", type="l")
lines(ts(as.vector(frc_combined_model[,1]), start = aheadTimes$t[1],frequency = dataFreq), col="red",
type="l")
lines(ts(as.vector(frc_combined_model[,2]), start = aheadTimes$t[1],frequency = dataFreq),
col="blue", type="l")
legend("topleft", lty=1, pch=1, col=c("black","blue","red"),
       c("Data","5% forecast limits", "Forecasts"))
```