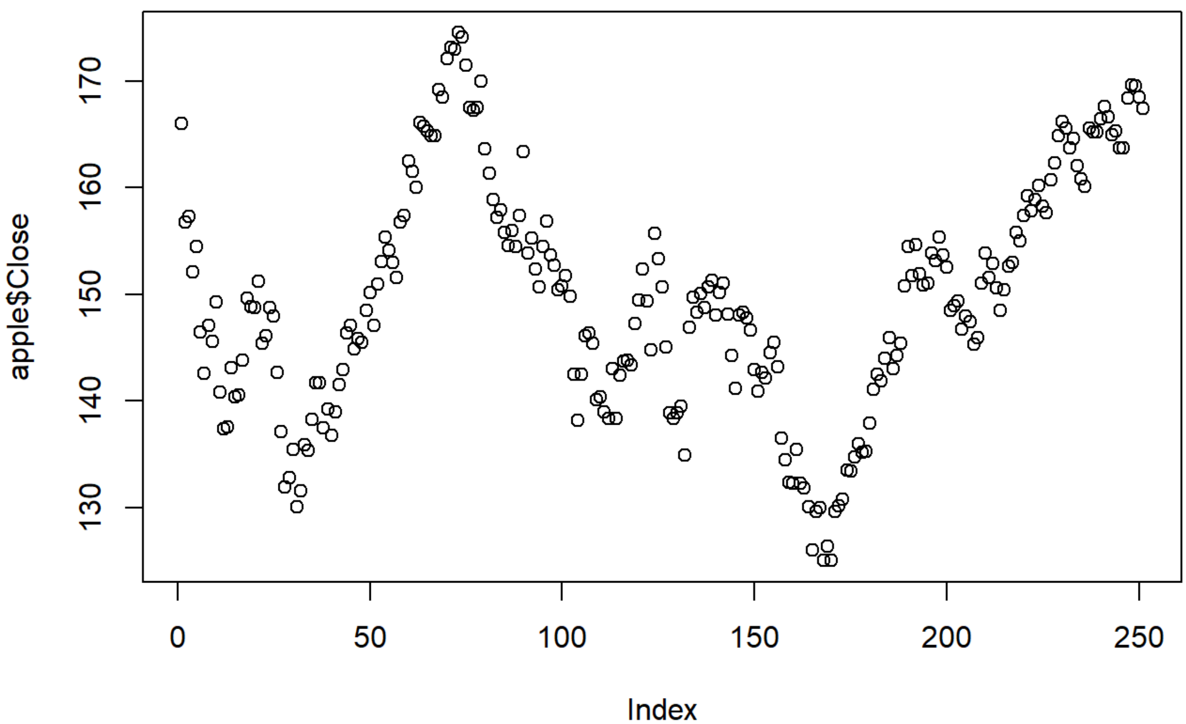
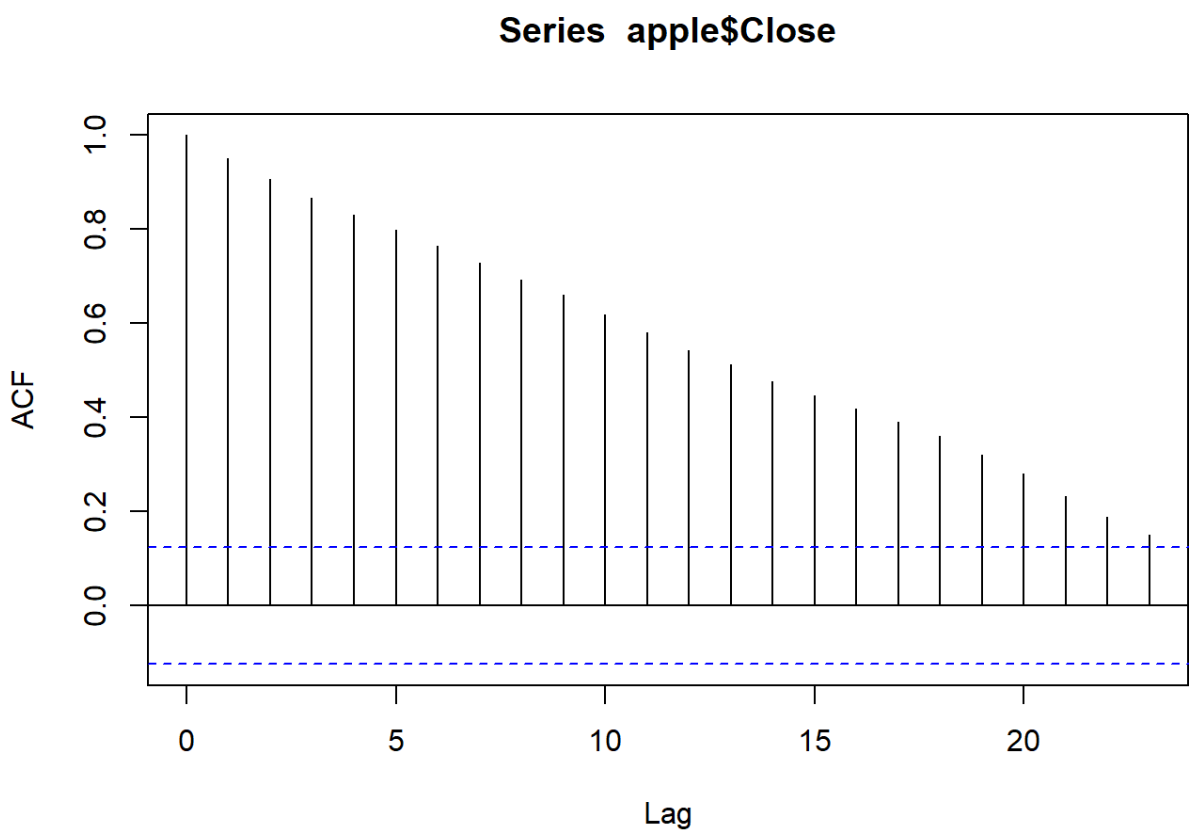
STA-750 Project

Fall-2023

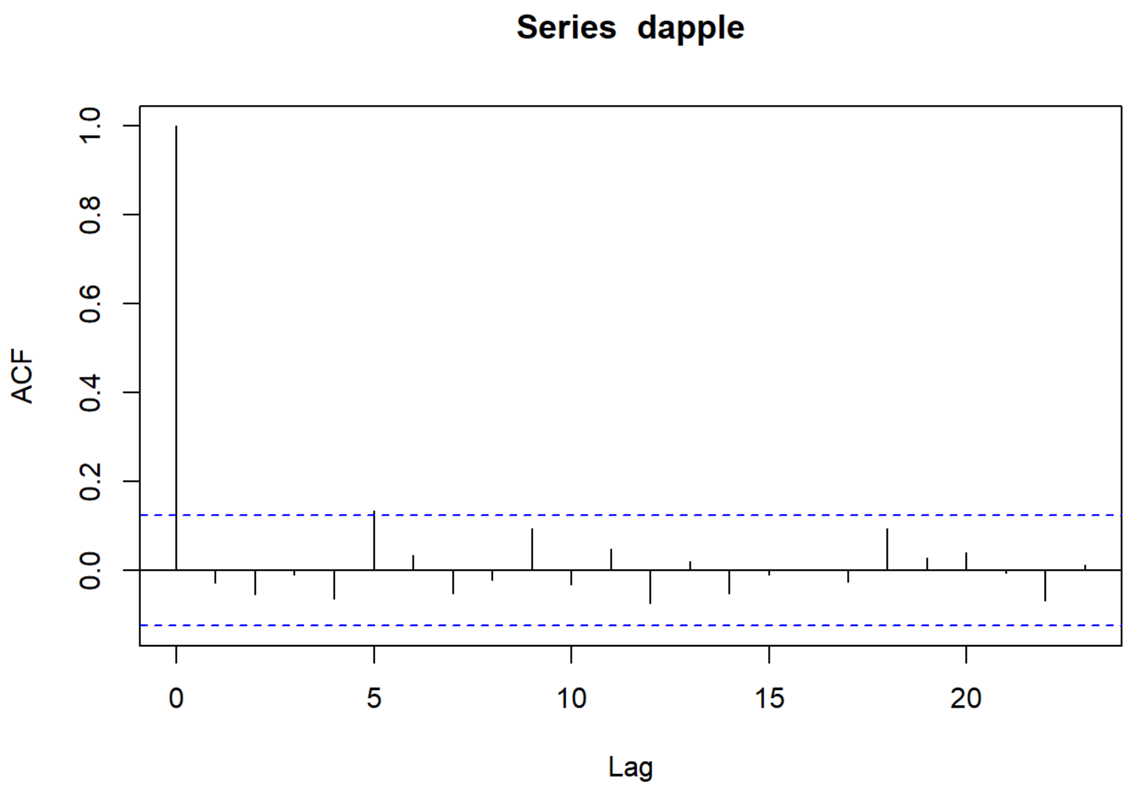
The data is collected from Yahoo Finance. We have selected Apple stock closing price, and we use one year of Monday to Friday data starting from 3rd May 2022 to 3rd May 2023. Apple's daily closing price data fails to satisfy the properties of stationarity. After plotting the data points initially, we found out that the mean of the series is not constant.



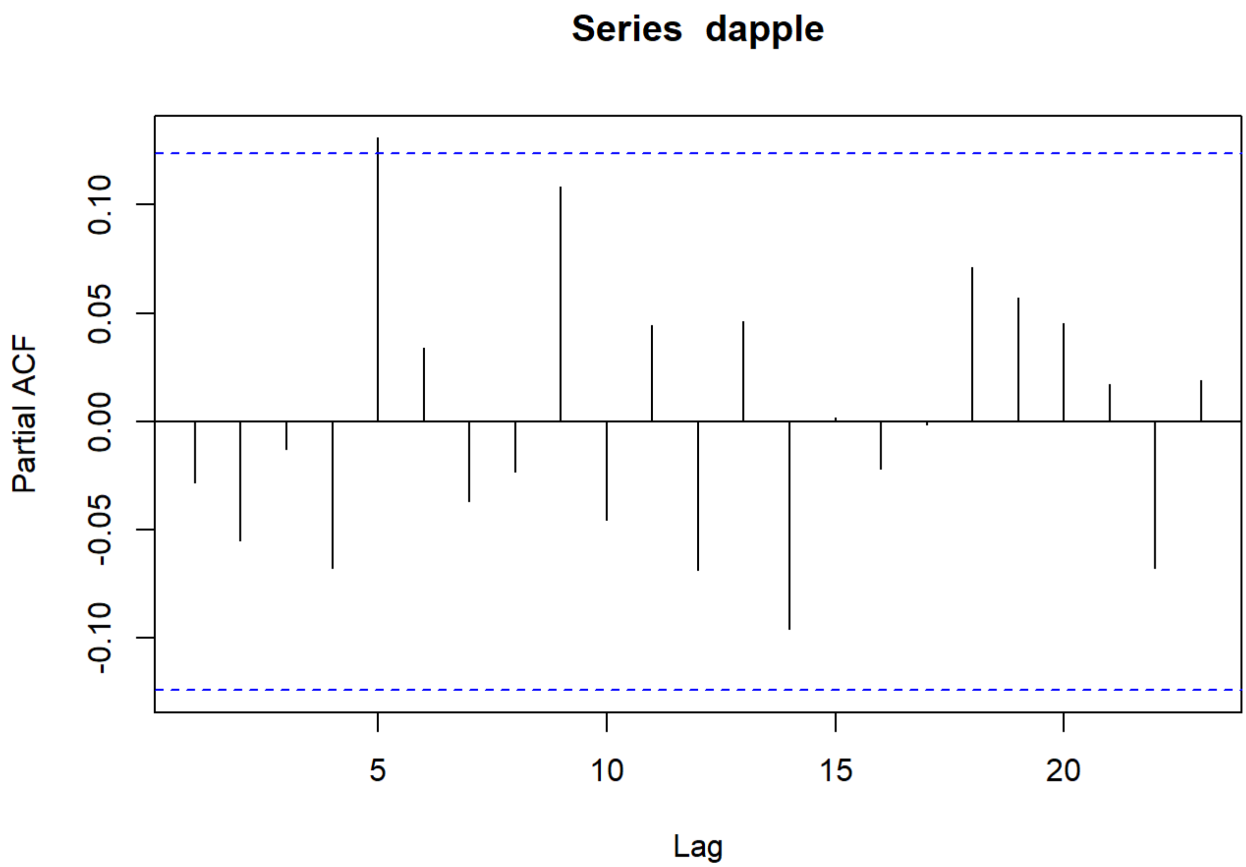

Then plot the ACF for the apple data and observe that ACF decays slowly.




To overcome non-stationarity, we took the first difference of the apple data and then checked the ACF again. This time the ACF decays fast.



Series dapple
ACF
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
Lag

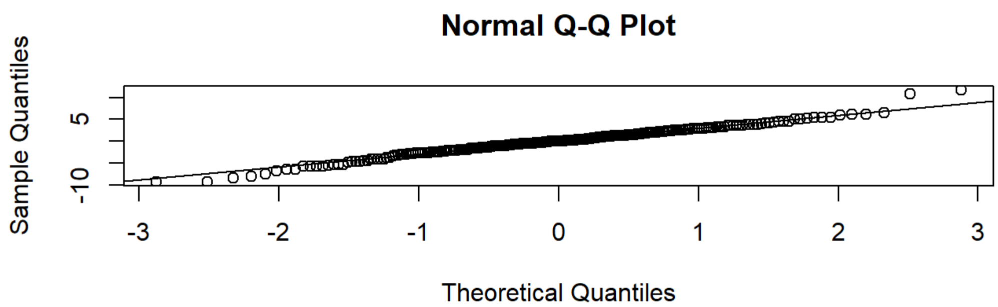


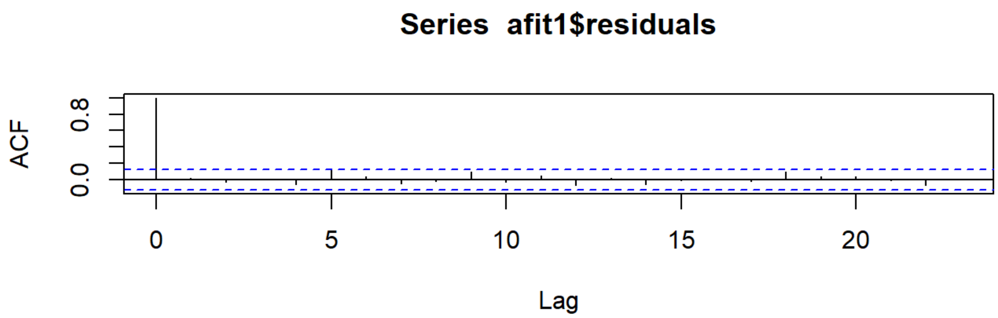

After observing ACF and PACF, we decided to check a few models. ARIMA(1,1,1), ARIMA(0,1,1), ARIMA(1,1,0). For all these three models the coefficients and standard error and selection criterion is given below.

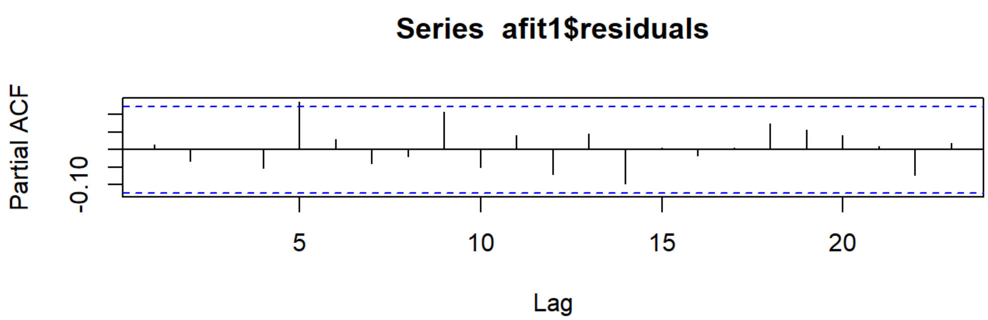
Where the associated standard errors (s.e.) for the autoregressive (ar1) and moving average (ma1) coefficients are listed directly below their associated coefficient, and Sigma^2 is the square of variance. There is no trend detected in standardized residuals. ACF of Residuals - All the lags are within significance.

```
Coefficients:ARIMA(1,1,1)
##          ar1      ma1
##       0.4814  -0.5277
## s.e.  0.4744   0.4578
##
## sigma^2 estimated as 9.658:  log likelihood = -638.21,  aic = 1282.42

Coefficients:ARIMA(0,1,1)
##           ar1
##      -0.0326
## s.e.  0.0683

##
## sigma^2 estimated as 9.675:  log likelihood = -638.43,  aic = 1280.86

Coefficients:ARIMA(1,1,0)
##           ar1      ma1
##      -0.0289
## s.e.  0.0643
## sigma^2 estimated as 9.676:  log likelihood = -638.44,  aic = 1280.89
```

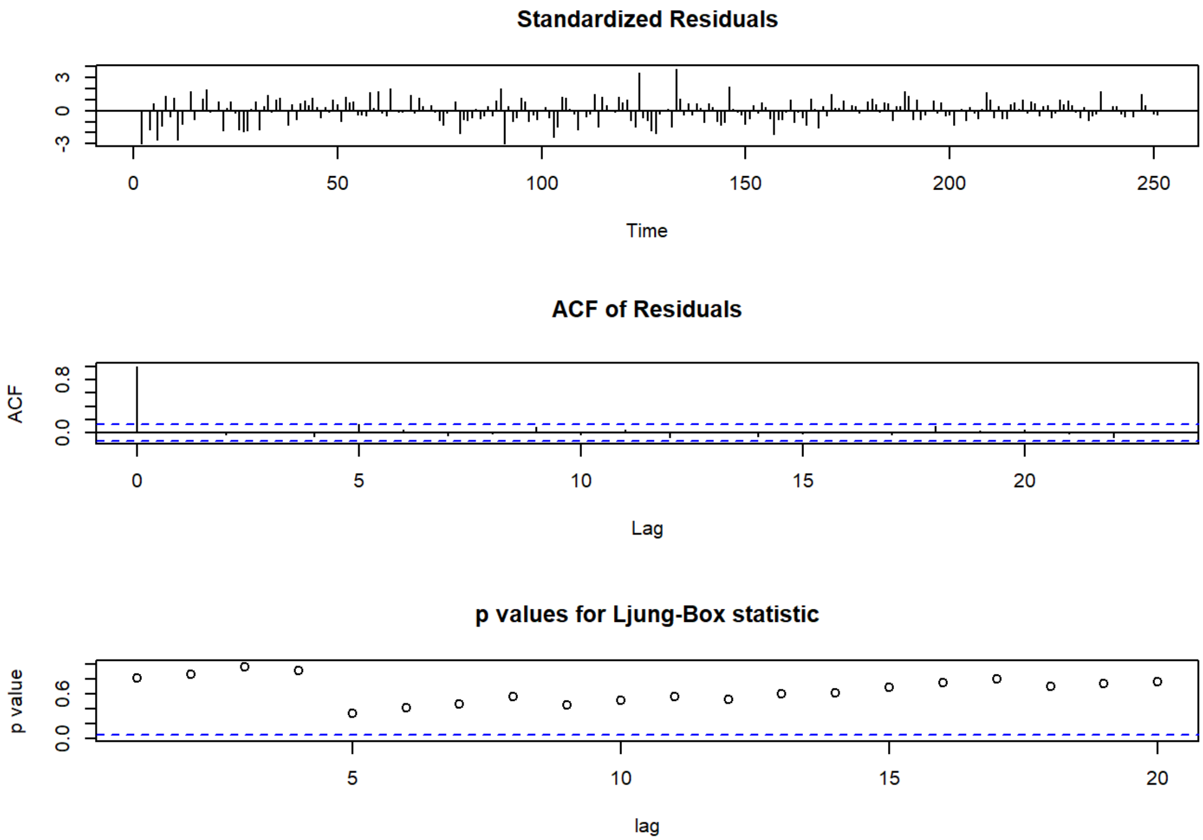
Using the AIC, since all of these models have AIC very close to each other, I choose ARIMA (1,1,1).


Normal Q-Q Plot
Sample Quantiles
-10
5
-3
-2
-1
0
1
2
3
Theoretical Quantiles


Series afit1$residuals
ACF
0.0
0.8
0
5
10
15
20
Lag


Series afit1$residuals
Partial ACF
-0.10
5
10
15
20
Lag

The QQ-plot shows that the data is normally distributed. The ACF and PACF of the residuals show that there is no dependency in the randomness and the residuals are white noise.



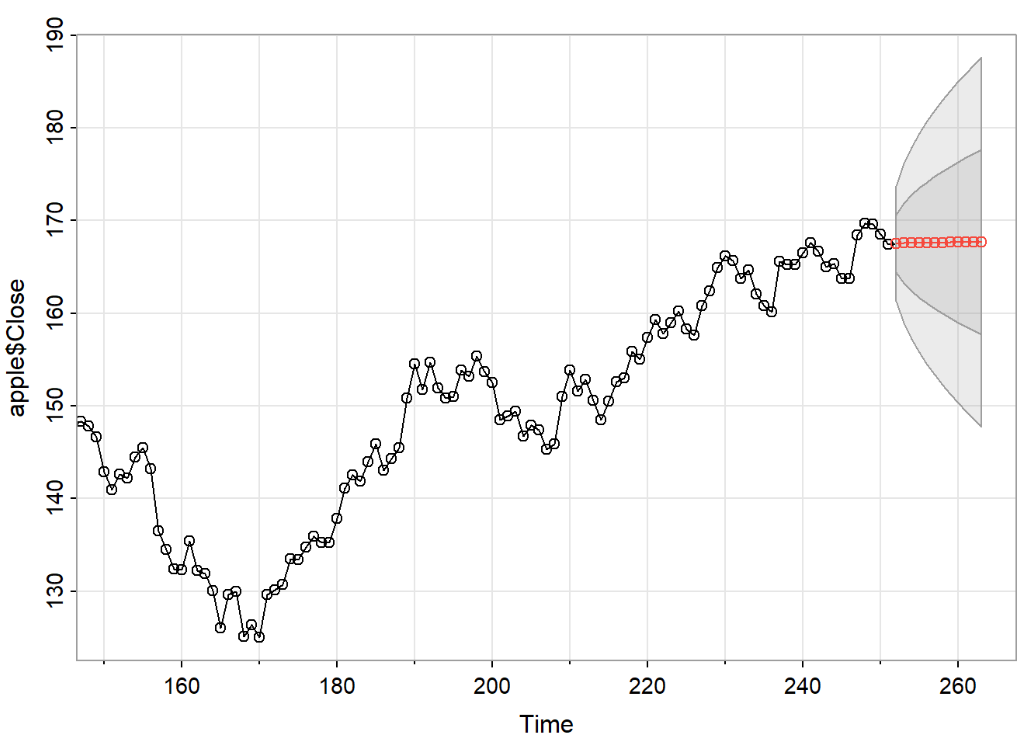

The Standardized residual plot seems to be constant. The p-values high indicate that the residuals are random. And there exists no correlation between the residuals.

Now we can use this model to forecast the future data points. Here in this analysis, we have predicted twelve data points.

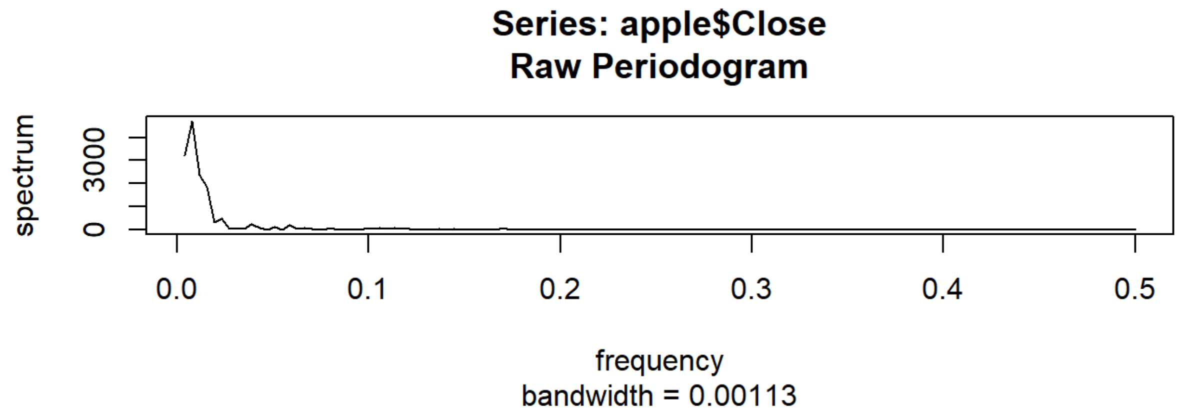
```
## $pred
## Time Series:
## Start = 252
## End = 263
## Frequency = 1
##  [1] 167.5038 167.5297 167.5422 167.5482 167.5511 167.5525 167.5532 167.55
35
##  [9] 167.5536 167.5537 167.5537 167.5538
##
## $se
## Time Series:
## Start = 252
## End = 263
## Frequency = 1
##  [1] 3.107757 4.294438 5.178828 5.916613 6.565113 7.152007 7.692891 8.1975
49
##  [9] 8.672609 9.122837 9.551807 9.962296
```

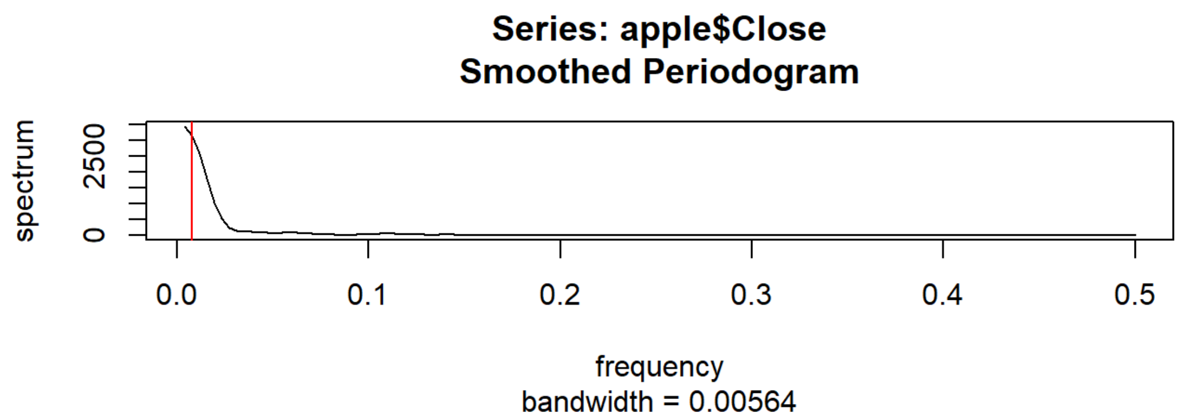
Then We found out the Confidence Interval for these twelve predicted values. The 95% Confidence Interval contains the predicted value.

```
##    Lower Limit  Upper Limit
[1,] 161.4126     173.5950
[2,] 159.1126     175.9468
[3,] 157.3917     177.6927
[4,] 155.9516     179.1448
[5,] 154.6835     180.4187
[6,] 153.5346     181.5704
[7,] 152.4751     182.6312
[8,] 151.4863     183.6207
[9,] 150.5553     184.5520
[10,] 149.6730    185.4345
[11,] 148.8322    186.2753
[12,] 148.0277    187.0799
```



Then we did the spectral analysis to find out any peak in the data set.




As we can see that there is a peak at .016 in the apple data sets. Which is the true dominance frequency. Later, using the periodogram smoother we smoothed the periodogram. the spectral peaks to see how many, if any, of the peaks match up. The raw data periodograms are shown in the figure above. Since the data has rapid transitions at the beginning to evaluate the periodograms are smoothed over points within 5 days (1 business week), then we Identify the predominant peaks. Since each plot contains 223 data points R uses 225 as a highly composite number.