10 Days Forecast for AAPL Price based on SARIMA model*

Result based on estimated stock price with SARIMA model

Zhendong Zhang

2021/4/5

Abstract

Apple's (NASDAQ: AAPL) stock price is a significant factor in the US stock market, which also has influences on the whole US economy. In the report, I will use SARIMA Model for forcasting the next 10-days spot price, exploring its seasonality and trend. As the conclusion, the spot price for AAPL follows a 18-days periodicity with 256, 85.3 and 128 days first three dominant periods, and will achieve the highest at the 7th day among the next 10 days.

Contents

1 Introduction **1**

2 Statistical Methods **2**

3 Results **3**
- 3.1 ARIMA Models 4
- 3.2 SARIMA Models 7
- 3.3 Forecast 9
- 3.4 Spectral Analysis 9

4 Conclusion **10**

1 Introduction

Apple (NASDAQ: AAPL) is a great company in the US. By 2021, its market capital is 2.23 trillion US dollars, which is the largest company around the world. Apple is also a component of the S&P 500 index, which is the most weighted stock in this index. Because of its large

*Source Data is extracted from: https://finance.yahoo.com/quote/AAPL/

market value, a shock on AAPL price will always indicate a more severe shock on the overall market, and even the whole US economy.

The analysis are done on software **RStudio** [RStudio Team, 2019], based on language **R** [R Core Team, 2020]. In this paper, it will sufficiently use package **astsa** [David Stoffer, 2020], package **splus2R** [William Constantine and Tim Hesterberg, 2021], with package **bookdown** [Yihui Xie, 2020] in order to produce analysis. And package **forecast** [Hyndman R., 2020] in forecasting. The data is gained from Yahoo Finance through the package **BatchGetSymbols** [Marcelo Perlin, 2020], which contains 252 samples. This data-set contains daily prices from the past year's April 1st to this year's April 1s. The prices are divided into open and close price. The open price represents the beginning price for each trading days, while the close price means the last price it ends with in each days.

The seasonality of AAPL price is important for every investor since it will affect the whole market, and it could also be an indicator for the overall stock market. Thus, it is reasonable to investigate the trend and seasonality for the AAPL price, and trying to forecast its future prices. In this paper, we will use ARIMA or SARIMA models to fit the prices and use them to forecast the future. In the end, we will conduct a spectral analysis, decomposing its first 3 dominant periods.

2 Statistical Methods

Before the analysis, we will use the average of open and close price for each trading days in pre-processing the series. The formula is shown as following:

$$P_t = \frac{P_{Open} + P_{Close}}{2} \qquad\qquad P_t : Stock\ price\ for\ time\ t$$

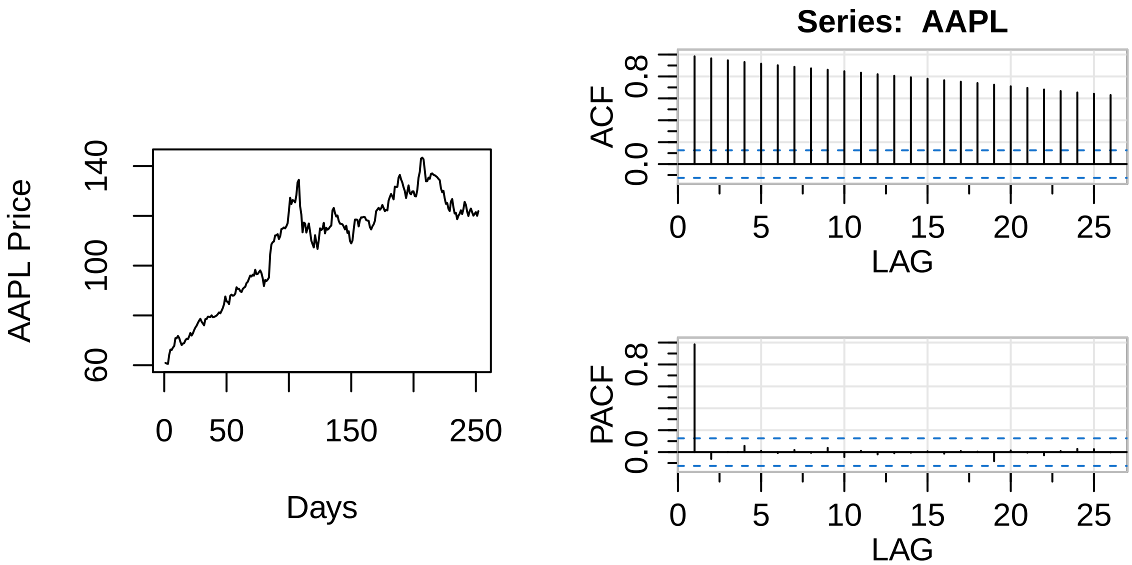
The strategy in this analysis is as the following:
1. Examine the data-set and its sample ACF visually, making it stationary through transform or differencing.
2. Determine the order of dependence (Might have several possible models) from the stationary ACF and PACF.
3. Conclude one model with the lowest AICc and the best significance.

We would like to conclude our model from the order of dependence that is gained from the stationary ACF and PACF only. Since sometimes a higher order model would fit better but less precise in forecast, it is reasonable to avoid these over-fitted models to obtain a more precise forecast.

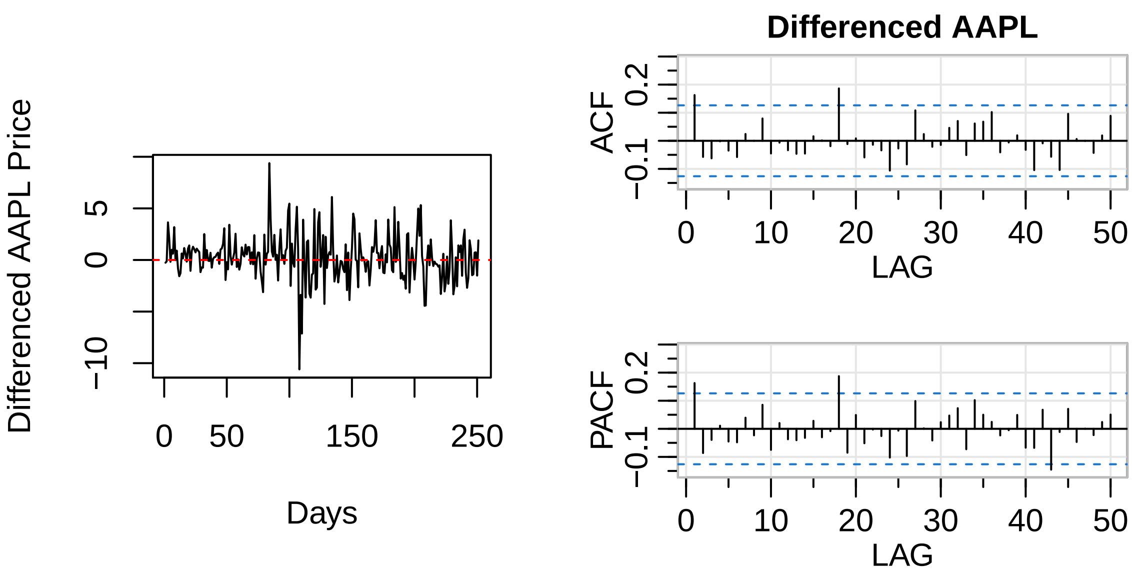
When determining the best model, we pick the model based on the lowest AICc, since the data-set only contains 252 samples, which is considerably small. The AICc is the AIC with a correction for small sample sizes, also the value is based on the variance of the estimators and the number of parameters in our model. Minimizing the AICc is the same as minimizing the variance of the estimators. Also, we performed a spectral analysis of the stock price. In the end, we could get the first 3 dominant periods, which means the price series could be divided into 3 series with the corresponding dominant period.

After which the best model is concluded, we will perform a forecast based on this model, obtaining the next 10 days' estimated stock prices. The estimations of the coefficients in our selection of models are accomplished by the *sarima* function, and the forecast is based on the *sarima.for* function.

3 Results



The variability is stable visually, thus there is no need to transform this data. (Except there are some outliers peaks, which do not affect the overall variability)
The sample ACF behaves a slow decay, so it may be needed to perform a first-order differencing.



We can see that after differencing, the data moves around zero. And the lags of the sample ACF are mostly within the boundary, which means it is zero. So, the differenced data appears to be stationary.

Non-Seasonal Component: From the sample ACF and PACF, we can see that it suggests the sample ACF cuts off at lag 1, and the sample PACF tails off. Another possibility is the sample ACF tails off, and the sample PACF cuts off at lag 1.

Seasonal Component: Also, it might also suggest the sample ACF tails off every 18 lags, and the sample PACF cuts off at lag 18. This suggests that it is a SAR (1)[18] model. Because of the behaviors within the seasons, which is suggested from the Non-Seasonal Component, we could try to combine this part into the non-seasonal part.

So, the suggested models are (1,1,0), (0,1,1), (1,1,0)x(1,0,0)[18] and (0,1,1)x(1,0,0)[18].

3.1 ARIMA Models

3.1.1 Order (1,1,0)

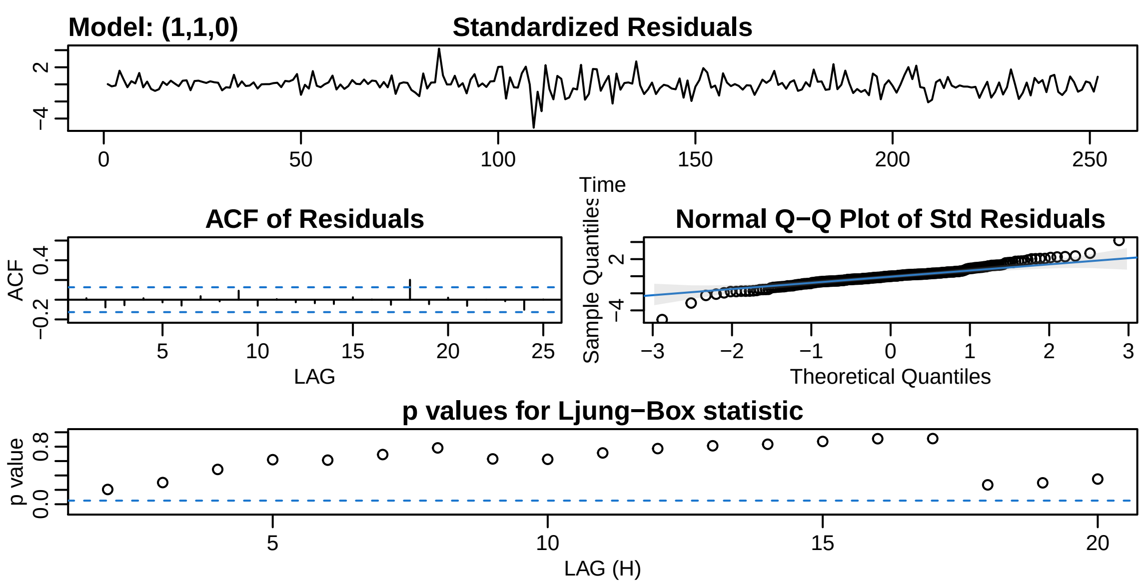


Figure 1: Model with Order (1,1,0)

For this model, the estimate of coefficient is 0.1627, with its P-value under t-test(H0: Coefficient = 0) equals to 0.0095. And the estimated constant term is 0.2438, with a P-value for t-test(H0: Coefficient = 0) of 0.1356. Under 5% significance level, the coefficient estimate is significant, but the constant term is not. The AICc is 4.403716. This model suggests that the a increase of 1 for the yesterday's price will increase today's price by 1.1627. An increase of 1 for the price 2 days ago will decrease today's price by 0.1627. The constant term means today's price will at least be 0.2438 even if the yesterday and 2 days ago price are zero.

By observing the standardized residuals plot, we can see the residual moves around zero, and there is no clear trend. The ACF of fitted residual is mostly within the boundary, which means it is mostly zero. So, the residual is stationary. From the Normal Q-Q plot of standard residual, we can see most points are close to the diagonal line, this means the residual follows normal distribution. Except there are some possible outliers at two tails, but we can still conclude that it is approximately normal. By observing the P-values from Ljung-Box statistics, most of the values are above 0, which means that we could accept the null hypothesis.(H0 for Ljung-Box: The data is i.i.d/the residual is white noise). Except the significance of the constant term, this is a good fit.

3.1.2 Order (0,1,1)

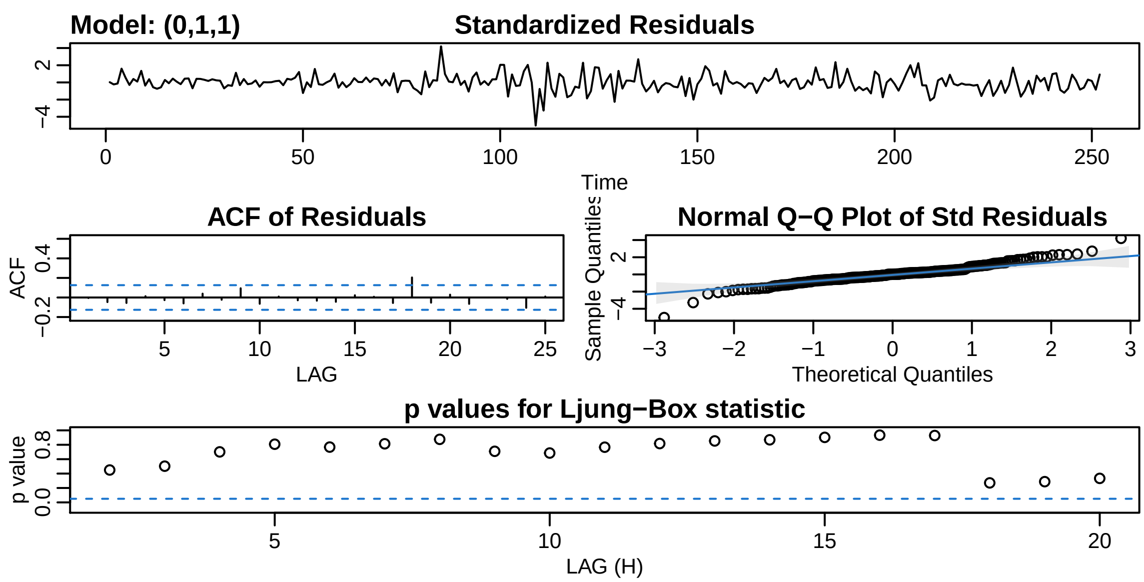


Figure 2: Model with Order (0,1,1)

For this model, the estimate of coefficient is 0.1832, with its P-value under t-test(H0: Coefficient = 0) equals to 0.0043. And the estimated constant term is 0.2441, with a P-value for t-test(H0: Coefficient = 0) of 0.1308. Under 5% significance level, the coefficient estimate is significant, but the constant term is not. The AICc is 4.399894. This model suggests that today's price will be based on the yesterday's price plus the 0.1832 times of the yesterday's white noise with the constant 0.2441. The corresponding white noise series is actually the fitted residuals.

This model's diagnostic graphs are similar with the model with order (1,1,0). The P-values from Ljung-Box statistics(H0 for Ljung-Box: The data is i.i.d/the residual is white noise) are larger than the model with order (1,1,0), which is better. There are still some outliers in the normal Q-Q plot, but it would not affect the conclusion (Residual follows normal distribution approximately). The residual ACF is stationary as well. And this model's AICc is smaller than the order (1,1,0), which means it is better comparing with order (1,1,0). Except the significance of the constant term, this is also a good fit.

3.1.3 Order (1,1,0) without Constant Term

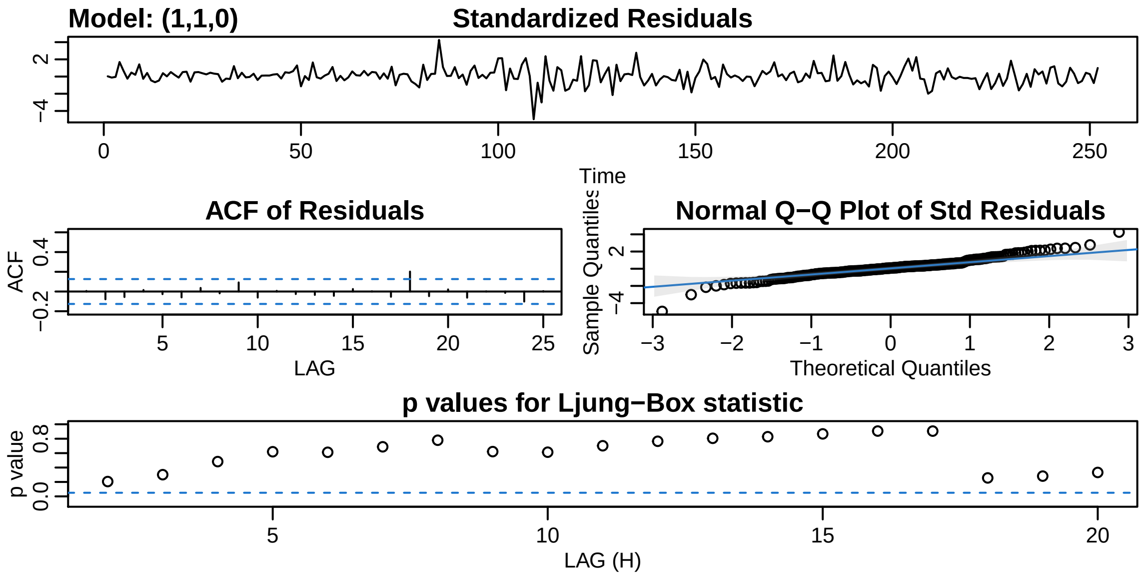


Figure 3: Model with Order (1,1,0) without Constant Term

Since the constant term is not significant, it is better to fit an order without the constant term.

For order (1,1,0) without constant term, the estimate of coefficient is 0.1727, with its P-value under t-test(H0: Coefficient = 0) equals to 0.0059. Under 5% significance level, the coefficient estimate is significant. The AICc is 4.40441, which becomes larger than the identical with-constant model. This model suggests that the a increase of 1 for the yesterday's price will increase today's price by 1.1727. An increase of 1 for the price 2 days ago will decrease today's price by 0.1727. The diagnostic graphs have slightly improvement, but is almost the same as the identical with-constant model, which means this is still a good fit.

3.1.4 Order (0,1,1) without Constant Term

For order (0,1,1) without constant term, the estimate of coefficient is 0.1905, with its P-value under t-test(H0: Coefficient = 0) equals to 0.0027. Under 5% significance level, the coefficient estimate is significant. The AICc is 4.40086, which becomes larger than the identical with-constant model. This model suggests that today's price will be based on the yesterday's price plus the 0.1905 times of the yesterday's white noise. The corresponding white noise series is actually the fitted residuals.

The diagnostic graphs have slightly improvement, but is almost the same as the identical with-constant model, which is similar with the behavior of the order (1,1,0) without constant term. Comparing with the with-constant order (0,1,1) model, the AICc does increase slightly. When comparing the AICc, it is not recommended to choose the without-constant models. But, since the constant term is not significant for the first two models, these models should be excluded before we comparing the models. So, from the compare between the two without-constant models, the one with the lowest AICc is the order (0,1,1).

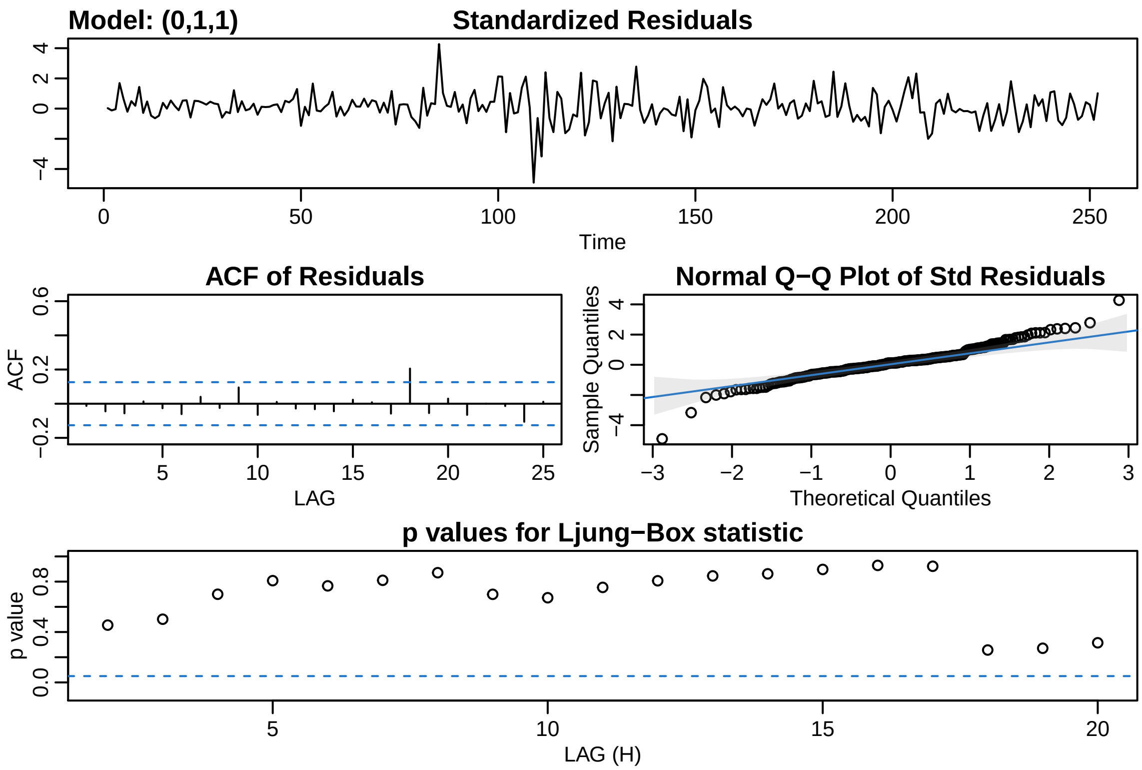
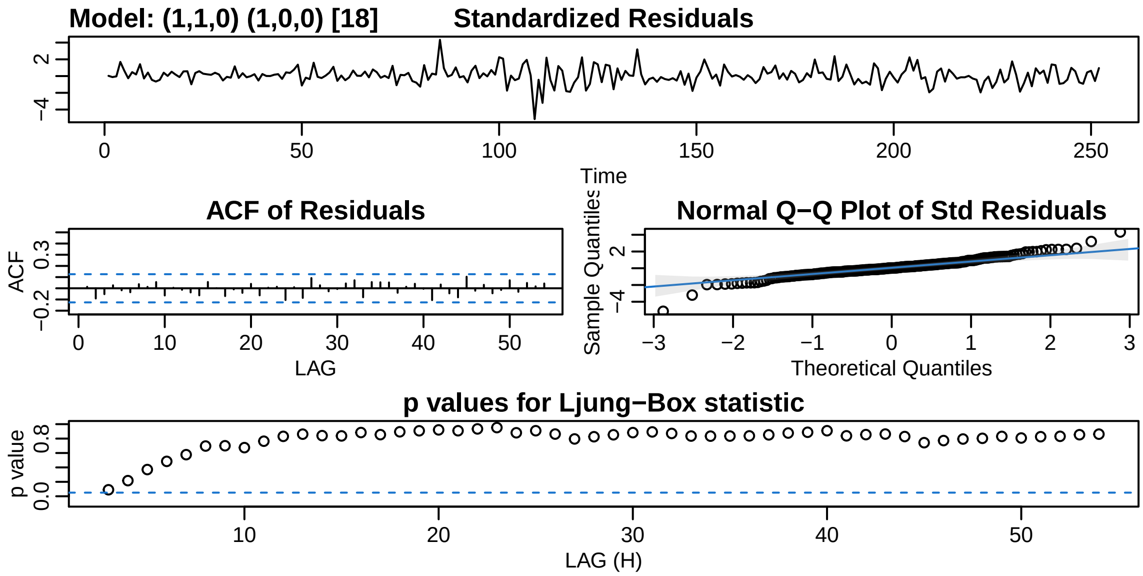


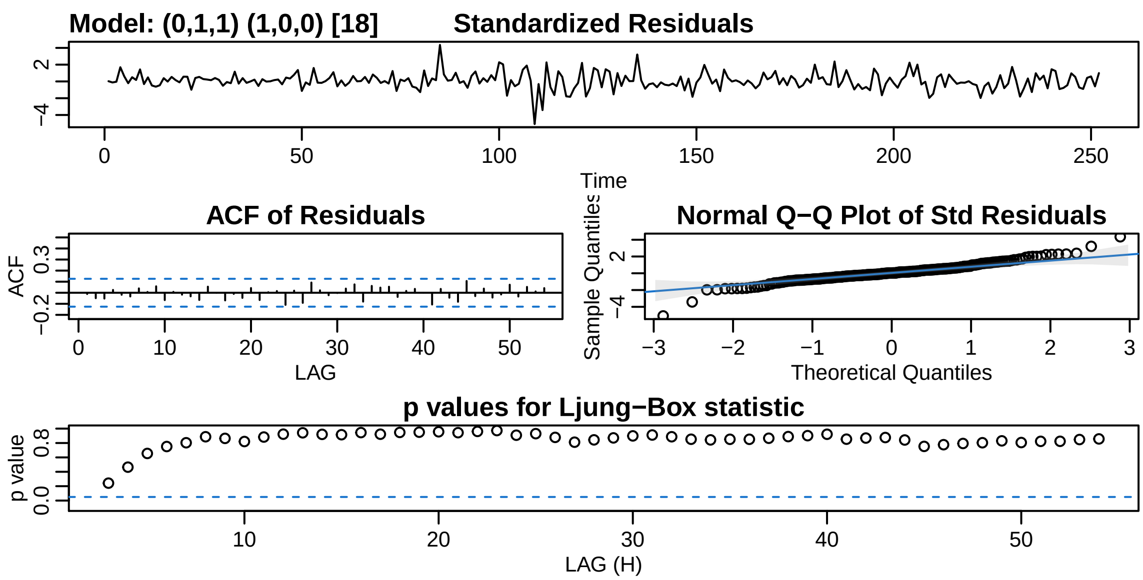
Figure 4: Model with Order (0,1,1) without Constant Term

3.2 SARIMA Models

For all the SARIMA models, we do not consider constant term for them, since it is not significant in ARIMA models.

3.2.1 Order (1,1,0)x(1,0,0)[18] and (0,1,1)x(1,0,0)[18]





For order (1,1,0)x(1,0,0)[18], the non-seasonal AR coefficient is 0.1889, with P-value for t-test of 0.0026. And the seasonal AR coefficient is 0.2105, with P-value for t-test of 0.0008. Both of these coefficients are significant under 5% significance level. From its diagnostic graphs, every lag of the residuals ACF is within the boundary, which means it is stationary. The normal Q-Q plot concludes the residuals are approximately normal (Except some outliers). The P-values from Ljung-Box statistics are above zero, with the first one close to the boundary as well. The AICc of this model is 4.367635. This model suggests that today's price is based on 1.1889 times of yesterday's price, plus 0.1889 times of 2 days ago price, plus 0.2105 times of 18 days ago price, minus 0.25026 times of 19 days ago price, and plus 0.03976 times of 20 days ago price.

For order (0,1,1)x(1,0,0)[18], the non-seasonal MA coefficient is 0.2137, with P-value for t-test of 8e-04. And the seasonal AR coefficient is 0.2152, with P-value for t-test of 6e-04. Both of these coefficients are significant under 5% significance level. From its diagnostic graphs, every lag of the residuals ACF is within the boundary, which means it is stationary. The normal Q-Q plot concludes the residuals are approximately normal (Except some outliers). The P-values from Ljung-Box statistics are above zero, and this non-seasonal MA model also outperform the non-seasonal AR model. The AICc of this model is 4.362266. This model suggests that today's price is based on yesterday's price, plus 0.2152 times of 18 days ago price, minus 0.2152 times of 19 days ago price, plus 0.2137 times of yesterday's fitted residual.

By choosing the lowest AICc model, the conclusion is the order (0,1,1)x(1,0,0)[18].

3.3 Forecast

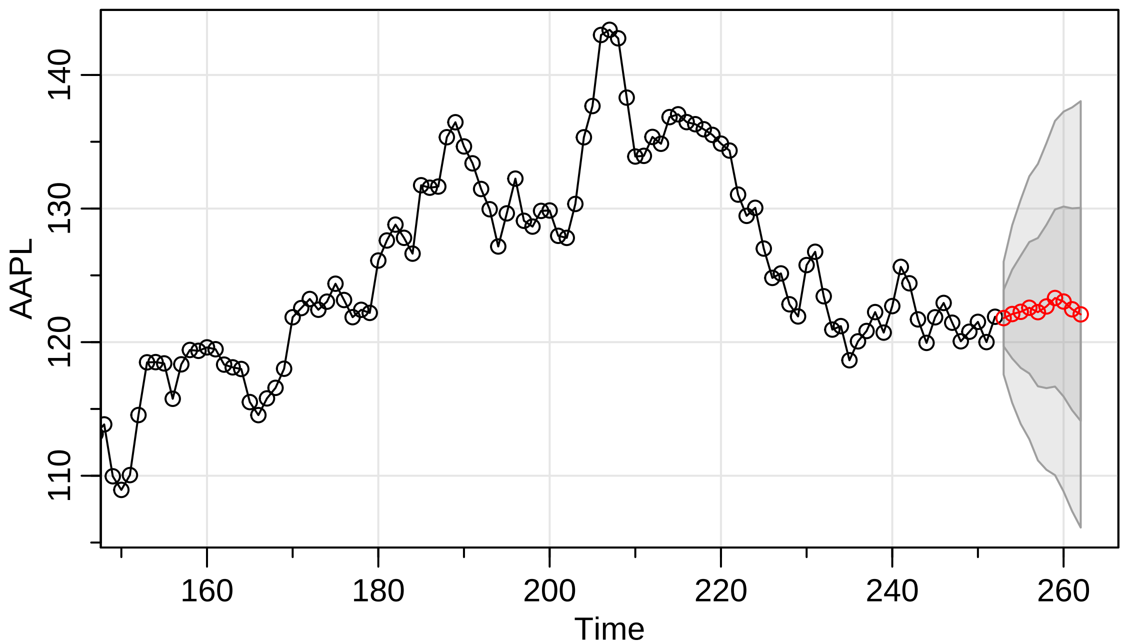


Figure 5: 10 Days AAPL Price Forecast

The estimated forecast values are 121.8003791, 122.1048249, 122.2715722, 122.5760181, 122.245753, 122.6728388, 123.3043226, 123.0396809, 122.4587574, 122.0800824.

The corresponding 95% Confidence Intervals are (117.657886, 125.9428722), (115.5902161, 128.6194337), (114.0423555, 130.5007889), (112.9323733, 132.219663), (111.3701029, 133.1214031), (110.691201, 134.6544766), (110.3104962, 136.2981489), (109.107007, 136.9723549), (107.6466245, 137.2708904), (106.4378587, 137.7223061).

From the forecast, it suggests that the price will reach the highest in the 7th day among the next 10 days, which is 123.3043226.

3.4 Spectral Analysis

```
##    Series Dominant.Periods      Spec      Lower      Upper
```

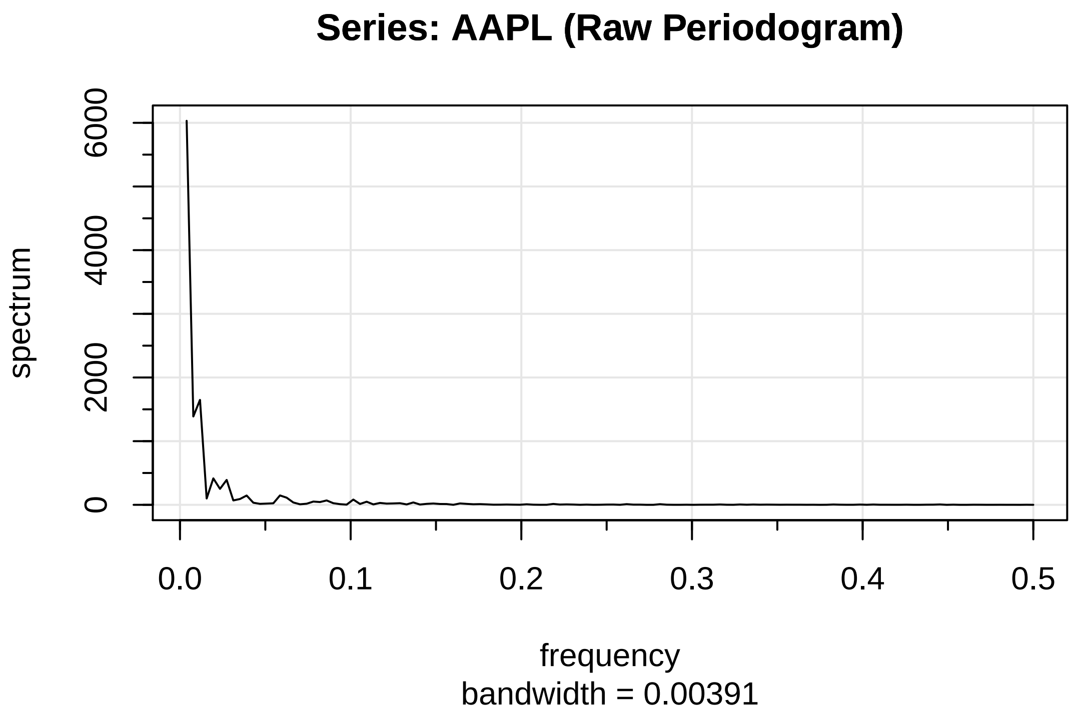


Figure 6: Raw Periodogram for AAPL Price

```
## 1   AAPL         256.0000 6031.708 1635.1058 238239.74
## 2   AAPL          85.3333 1648.294  446.8279  65104.14
## 3   AAPL         128.0000 1387.878  376.2329  54818.24
```

Here is a table containing the first 3 dominant periods, with its corresponding spectrum and 95% Confidence Interval. All of the 3 dominant periods' confidence intervals (95% Level) are too wide to distinguish the significance.

4 Conclusion

The analysis suggests that the price has periodicity for a 18 days period. Based on this periodicity, it indicates the highest price appears at the 7th day in the next 10 days forecast. The AAPL price series can be decomposed into 3 series with periods 256, 85.3333 and 128 days, respectively. Since we used the seasonal model, which has more variables than a normal ARIMA model, we may be still suffered from the over-fit issue. Also based on the Q-Q normal plot, the fit residuals do not follow a strictly normal distribution because of the outliers at two tails. This will lead our forecast less precise, and our sample size is small as well. Our model is based on the prices gained from the average of a day's open and close prices, which might not be the best way to represent a day's stock price. Also, every model for the stock price is wrong, but some are useful. For further researches, it is recommended to keep a track of the seasonality of the AAPL price and be attention to any shock on the stock price.