STAT 426 Final Project: L'Oréal Close Prices

Elizabeth Knatcher

Spring 2023

1 Introduction

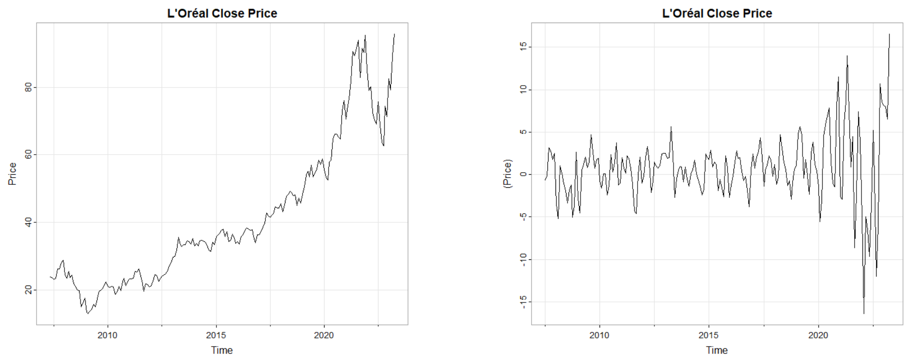
The stock market is an important part of the economy that impacts our daily lives. The L'Oréal stocks are considered in this paper, and an ARIMA model is fitted to the data to predict the L'Oréal close prices.

2 Data

The data used is L'Oréal Stock data. L'Oréal is a major cosmetics company with products found in drugstores as well as more high end beauty stores. They have multiple subsidiaries that allow for L'Oréal to have products for skin care, hair care, cosmetics, and perfume. The data set consists of the Close price of stocks from May 2007 to April 2023, for each month. In other words, the data has a frequency of 12. This data set effectively consists of the past 16 years of close price data for L'Oréal. This data was found on Yahoo Finance.

3 Time Series Model

First, the data was simply plotted. After doing this, it can be observed from the time series plot that there is a positive quadratic trend. To correct for this trend, the data is has a difference of degree two applied. After the differencing, the time series plot no longer displays a trend.



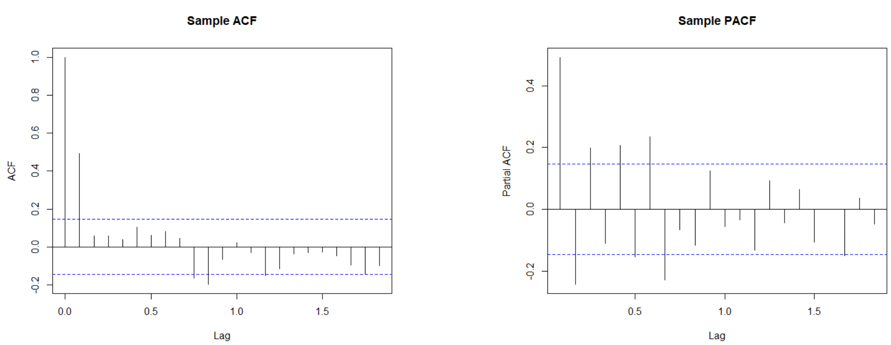
(a) Time Series Plot Before Differencing of Price

(b) Time Series Plot After Differencing of Price

Figure 1: Time Series Plots

The data was then divided into training and test subsets, where the training subset consists of the first 181 entries and the test subset contains the last 9 entries. This effectively divides the data where the training subset consists of 95% of the original data, as $n = 190$.

The Auto-correlation Function and Partial Auto-correlation Function plots are then rendered. It can be observed in the Auto-correlation Function (ACF) plot that the first two lags are significant, as ACF is significant at lag 0 and lag 1. The Partial Auto-correlation Function (PACF) plot demonstrates that the first three lags are significant. The analysis so far suggests an ARIMA(3, 2, 2) model should be used, based off the ACF and PACF of the data after having the difference of degree 2 applied.



(a) ACF Plot

(b) PACF Plot

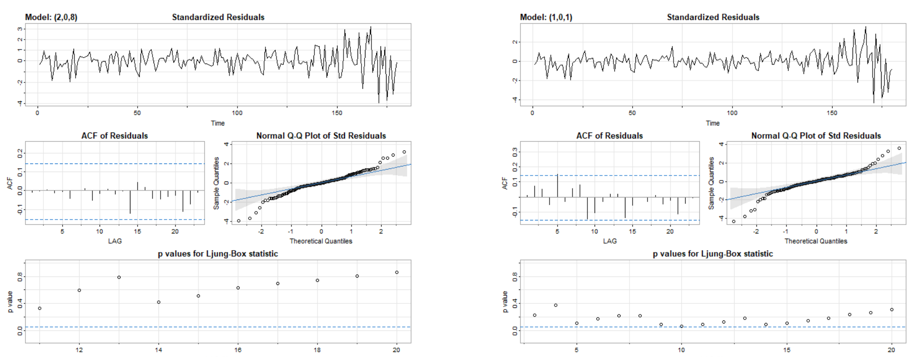
Figure 2: ACF and PACF Plots

Further analysis of the data is conducted. An optimal model is fitted by computing the AIC, AICc, and BIC of possible ARIMA(p, d, q) models. By computing these values, it is found that the AIC chooses an ARIMA(6, 2, 7) model for the data. This model results in an AIC value of 4.7224, an AICc value of 4.7365, and a BIC value of 4.9884. the AICc chooses an ARIMA(2, 2, 3) model for the data. This model results in an AIC value of 4.7325, an AICc value of 4.7352, and a BIC value of 4.8567. The BIC chooses an ARIMA(0, 2, 1) model. This is logical as the BIC penalizes additional terms to a higher degree than the AIC or AICc. The ARIMA(0, 2, 1) model results in an AIC value of 4.7575, an AICc value of 4.7579, and a BIC value of 4.8108.

The Ljung-Box test shows that for the ARIMA(6, 2, 7), ARIMA(2, 2, 3) and ARIMA(0, 2, 1) models the auto-correlation of the residuals is statistically significant. This means that there is evidence of residual patterns that have not been accounted for by the model. Therefore none of these models can be used.

Further analysis is therefore conducted and the ARIMA(2, 2, 8) and ARIMA (1, 2, 1) models are chosen based off the AIC, AICc, and BIC.

Both the AIC and AICc choose the ARIMA(2, 2, 8) model. This model results in an AIC value of 4.7342, an AICc value of 4.7429, and a BIC value of 4.9470. The ARIMA(1, 2, 1) model results in an AIC value of 4.7606, an AICc value of 4.7614, and a BIC value of 4.8316. It is important to note that these models have the smallest respective AIC, AICc, and BIC values while still passing the Ljung-Box test.



(a) ARIMA(2, 2, 8) Diagnostic Plots

(b) ARIMA(1, 2, 1) Diagnostic Plots

Figure 3: Diagnostic Plots

It should be observed that the figures label the models as (2, 0, 8) and (1, 0, 1), respectively. This is due to the training data coming from data that was differenced, and therefore does not require further differencing.

It can be observed from the Standardized Residuals plot of the ARIMA(2, 2, 8) model that the values are mostly random, and no discernible pattern can be detected. Some of the values towards the far right have an absolute value larger than 2, which indicates there may be outliers. Overall, this plot indicates random variance. The ACF plot demonstrates that all values are between the two blue bands, meaning there are no significant auto-correlations for the standardized residuals. The Q-Q Plot indicates that the residuals demonstrate normality. This supports the model assumptions of normality. The Ljung-Box test shows that the auto-correlation of the residuals is not statistically significant. This means that there is not evidence of residual patterns that have not been accounted for by the model.

It can be observed from the Standardized Residuals plot of the ARIMA(1, 2, 1) model that the values are mostly random, and no discernible pattern can be detected. This indicates random variance. The ACF plot demonstrates that almost all values are between the two blue bands, meaning there are no significant auto-correlations for the standardized residuals. There is one value slightly above the blue band, however the Ljung-Box test is later considered for model validation. The Q-Q Plot indicates that the residuals demonstrate normality. This supports the model assumptions of normality. The Ljung-Box test shows that the auto-correlation of the residuals is not statistically significant. This means that there is not evidence of residual patterns that have not been accounted for by the model. Both models are used for forecasting.

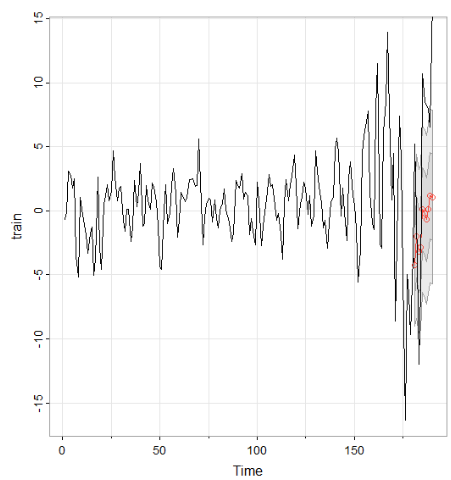
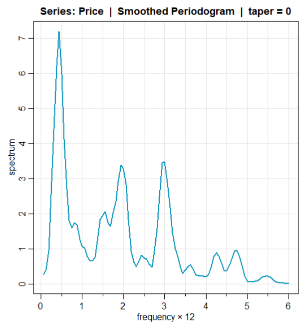


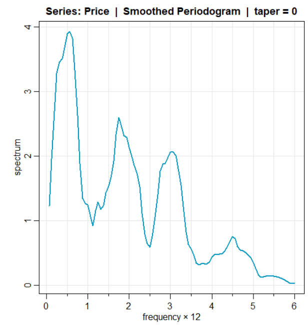
Figure 4: ARIMA(2, 2, 8) Forecast Plot with Original Data

The ARIMA(2, 2, 8) model is chosen. This is due to the ARIMA(2, 2, 8) model delivering the minimal forecast error, as it has a forecast error of 776.1737. This is compared to the ARIMA(1, 2, 1) model having a forecast error of 818.2837. The ARIMA(2, 2, 8) model forecast predictions can be compared with the original data in the above plot.

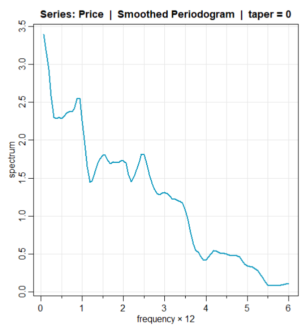
Lastly, the power spectrum estimate is computed. Four smoothed periodograms are computed to estimate the power spectrum. These are computed with four different bandwidths of 0.2857, 0.6579, 1.2821, and 2.5316. They are computed with 5, 10, 20, and 40 spans, respectively. It is observed that with bandwidths of 0.2857 and 0.6579 that the periodogram has several peaks. As the bandwidth increases, there are less peaks. When the bandwidth is 2.5316 a decreasing linear trend is observed, however the plot is smooth and there are no significant peaks to be observed.



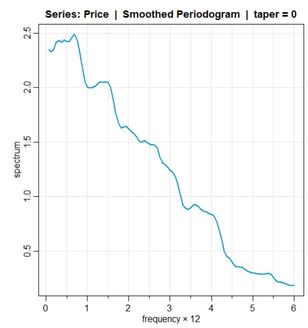
(a) Periodogram with 5 Spans



(b) Periodogram with 10 Spans



(c) Periodogram with 20 Spans



(d) Periodogram with 40 Spans

Figure 5: Periodogram Plots

4 Conclusion

The model that best forecasted the data was an ARIMA(2, 2, 8) model. This model is chosen based off the AIC, AICc, and BIC, while also ensuring that model assumptions are validated. The original data exhibits a sharp increase in close prices. The forecast data also demonstrates an increase, but not to the same extent. Therefore, the model is somewhat accurate and captures the overall trend of L'Oréal close prices.

References

Yahoo Finance (2023). Yahoo finance. `https://finance.yahoo.com/quote/LRLCY?p=LRLCY&.tsrc=fin-srch`, Last accessed on 2023-04-23.