Name: Sean Goh Ann Ray
Matric No: G2202190G

Data Analysis:

The purpose is to study is the Apple stock prices (open, close, high, low, and adjusted close) from 1st February 2002 to 31st January 2017, both dates inclusive. The data is downloaded from Yahoo Finance website and consists of 3776 daily records of Apple's stock prices. As per most financial studies, we are going to observe only the daily close price. Figure 1 shows Apple's closing stock prices over the period mentioned above.

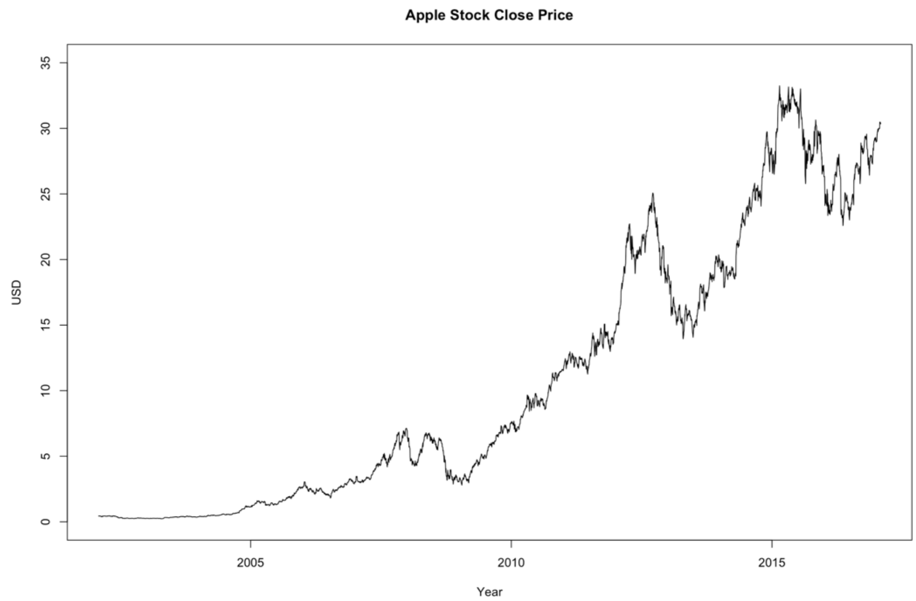

Figure 1. Apple Stock Close Price (01/02/2002 – 31/01/2017)

A clear upwards trend but no clear pattern in seasonality can be observed from Figure 1 and seems to have increasing variance over time. Therefore, the data can be considered a non-stationary time-series data, which is further supported by the slow convergence of the ACF as shown in Figure 2.

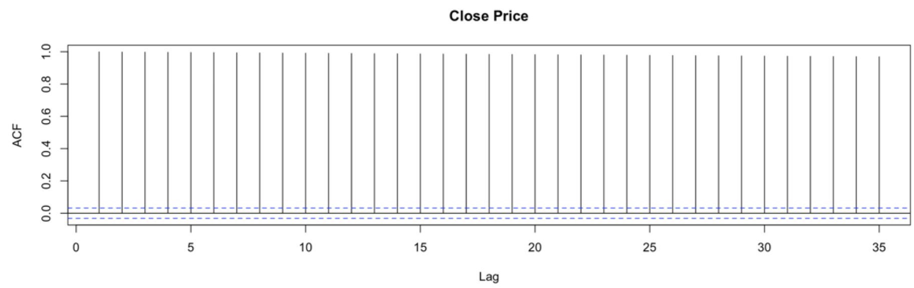

Figure 2. ACF of Close Price

In this study, 2 methodologies are used, the first being the ARIMA method and the other being the GARCH method, to find a suitable model and perform a 1 year forecast (up to 31st January 2018) of the stock price, which will then be compared to the actual data. A change-point analysis was also performed on the stock price up to the forecasted date as an additional study.

ARIMA Analysis:

To stabilize the variance of the data, BoxCox transformation is applied. Using the BoxCox.lambda function, the produced lambda value of 0.11464 suggests a transformation between a log transformation and a square root transformation. The data is transformed based on that lambda value and the transformed close price is shown in Figure 3.

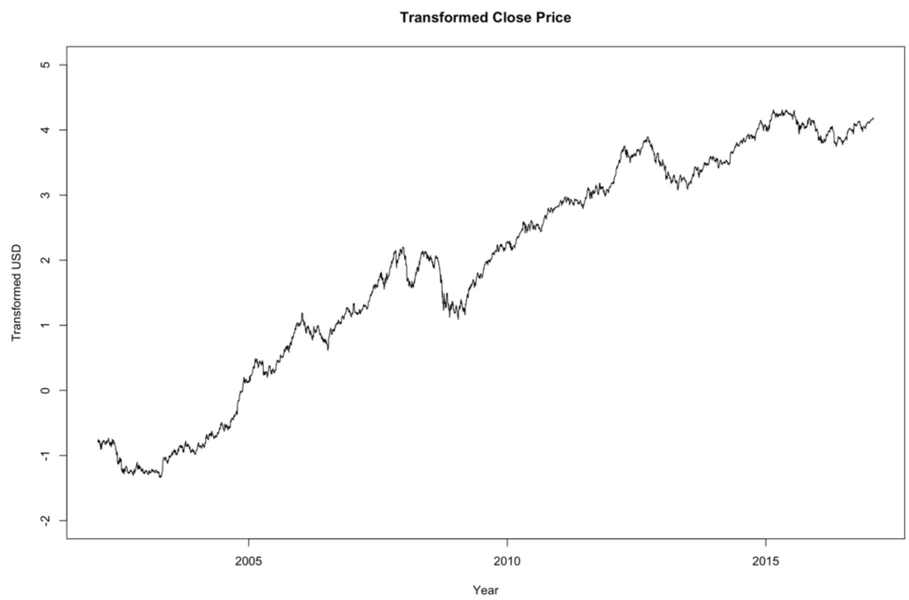

Figure 3. Transformed Close Price (01/02/2002 - 31/01/2017)

The transformed data does seem to have a constant variance over time, but the trend component is clearly there. Figure 4 below shows the ACF plot of the transformed data, which supports the non-stationary data hypothesis.

Figure 4. ACF of Transformed Close Price

Without any clear patterns in the trend, a differencing of 1 lag is done to the transformed data and is shown in Figure 5. With the differencing, the data seems to appear stationary. To support this, 2 unit root tests, ADF test and KPSS test, are performed on the differenced data. The ADF test produced a p-value of 0.01, meaning that the null hypothesis (data is non-stationary) must be rejected, and thus suggests that the data is stationary. The KPSS test produced a p-value of 0.1, which means that we cannot reject the null hypothesis, which also suggests that the data is stationary.

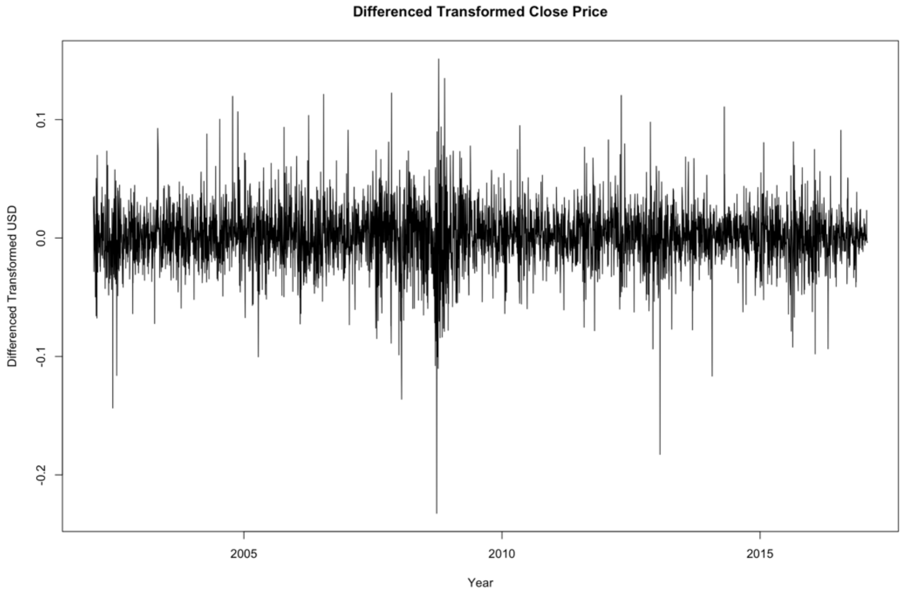

Figure 5. Differenced Transformed Close Price

Figure 6 shows the ACF and PACF plots of the differenced data, and p and q values are observed to be cut-off at lag 4 as it had the largest correlation value.

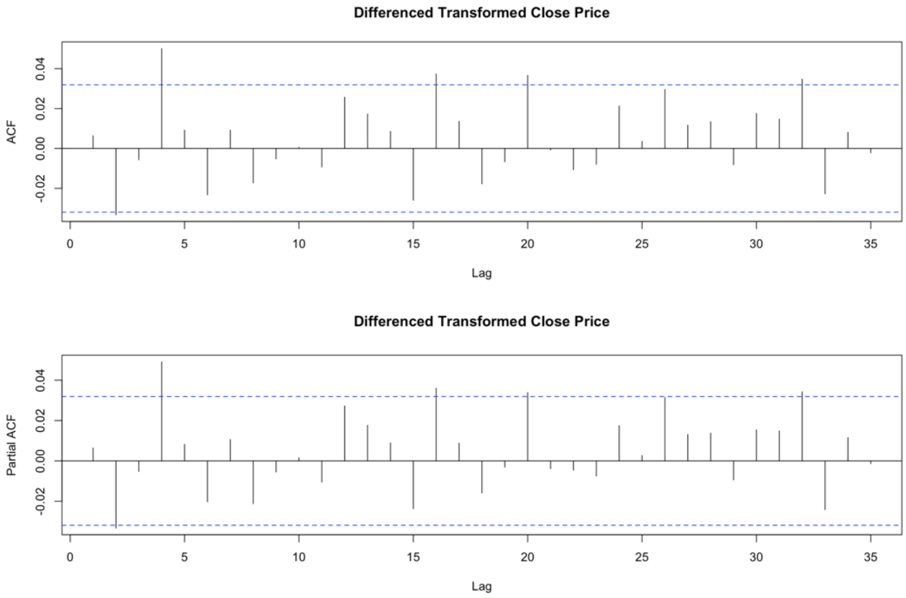

Figure 6. ACF and PACF Plots of Differenced Transformed Close Price

Therefore, ARIMA models of order (4,1,0) and (0,1,4) were checked and their diagnostics are shown in Figures 7 and 8 respectively. Both models are adequate as they both have random standardized residuals, their ACF cuts off at lag 0, and their p-values for the Ljung-Box statistics are above 0.05. Therefore, no further improvements to the model were made. Based on their AIC values of -16839.62 and -16839.49 respectively, the ARIMA(4,1,0) is marginally better and will be used in the forecasting of the stock price.

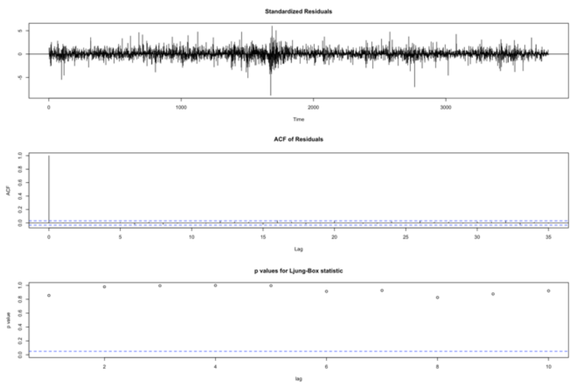

Figure 7. Diagnostics of ARIMA(4,1,0) Model

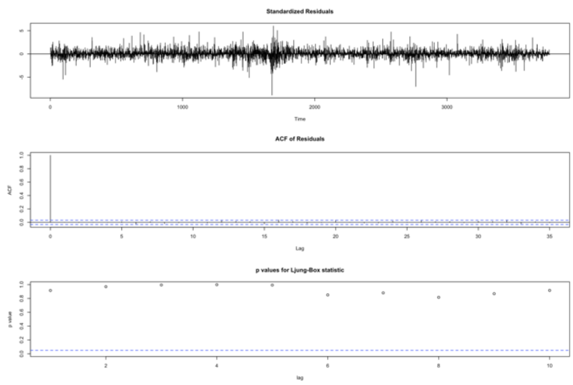

Figure 8. Diagnostics of ARIMA(0,1,4) Model

The models are then used to predict the transformed forecast, which are then inversely transformed to match the original data which is the close prices. The actual close price, forecasted close price and the upper and lower bounds of the forecast of each model are shown in Figures 9 and 10, up to the date of 31/01/2018, with a dashed vertical line showing the start of the forecast. Both models achieved almost the same forecasting, and although it can be observed that the forecasted price is very far off from the original close price, it is still within the bounds of the forecast.

Figure 9. ARIMA(4,1,0) Model Forecast of Close Price (01/02/2002 – 31/01/2018)

Figure 10. ARIMA(0,1,4) Model Forecast of Close Price(01/02/2002 – 31/01/2018)

GARCH Analysis:

In GARCH analysis, it is common to analyze the log returns of the stock prices, thus the stock price data is transformed by the following equation:

$$r_t = (\log p_t - \log p_{t-1}) * 100$$

where r_t is the returns, and p_t is the stock price at time t. The 100 multiplier is to see the returns as a percentage change in the price. This data is shown in Figure 11 with an additional red line to show the zero-change line. The returns data is comparable to the differenced transformed data in the ARIMA analysis, with a slightly different transformation and a 100 multiplier. The returns data seems to appear stationary, and the ADF test and KPSS test are conducted again. The same p-values of 0.01 and 0.1 respectively are produced, thus the same conclusion can be made where the returns data is stationary. However, volatility clustering can be observed as periods of high variance does seem to follow periods of high variance, and periods of low variance does seem to follow periods of low variance.

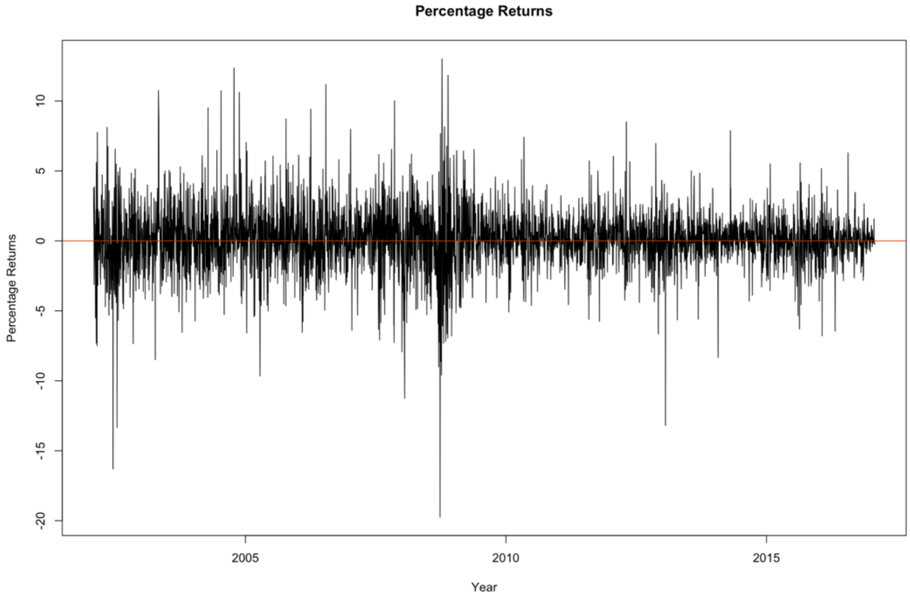

Figure 11. Percentage Returns (01/02/2002 - 31/01/2017)

The ACF and PACF plots of the returns are shown in Figure 12, which suggests some correlations at lags 2 and 4. However, observing the ACF and PACF plots of the absolute returns and squared returns as shown in Figures 13 and 14 respectively shows significant autocorrelations, further supporting that the returns data is not independent and identically distributed.

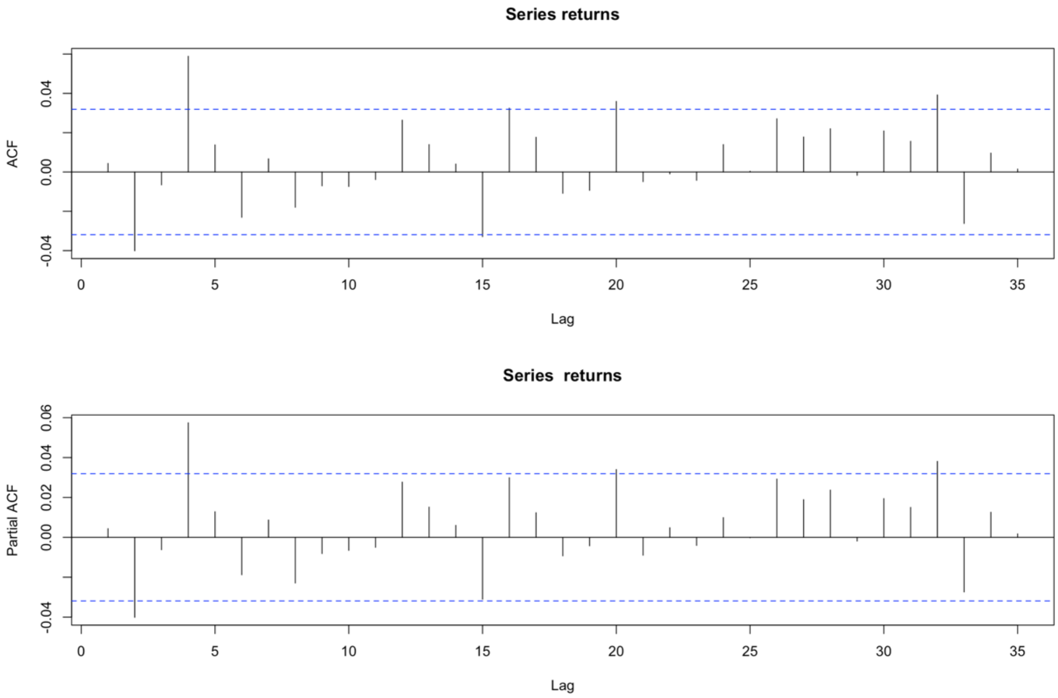

Figure 12. ACF and PACF of Returns

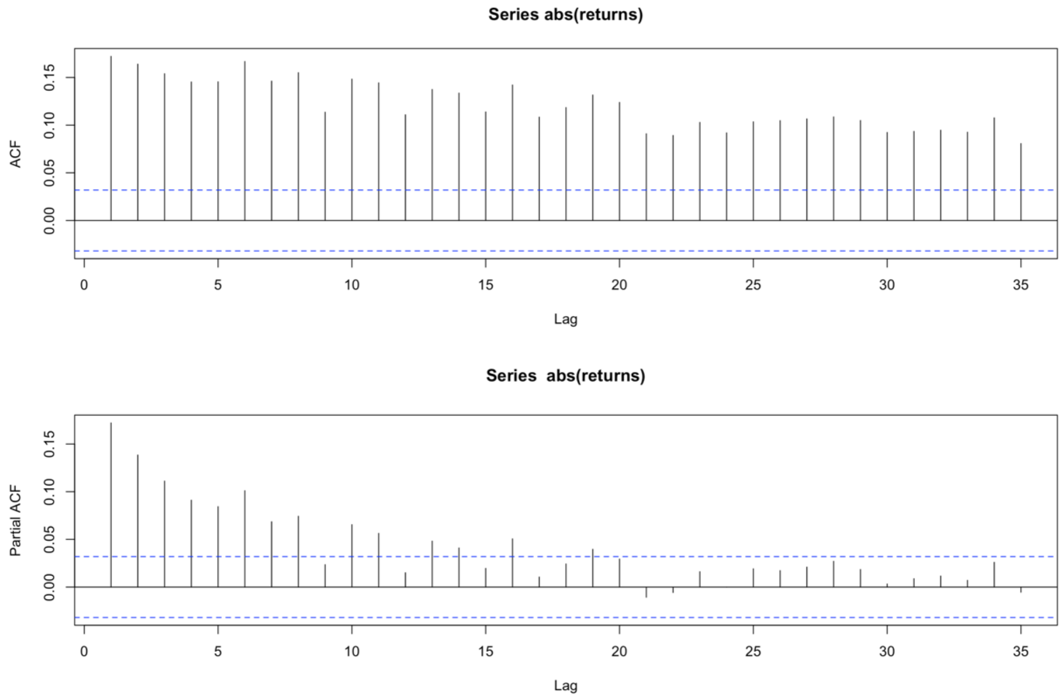

Figure 13. ACF and PACF of Absolute Returns

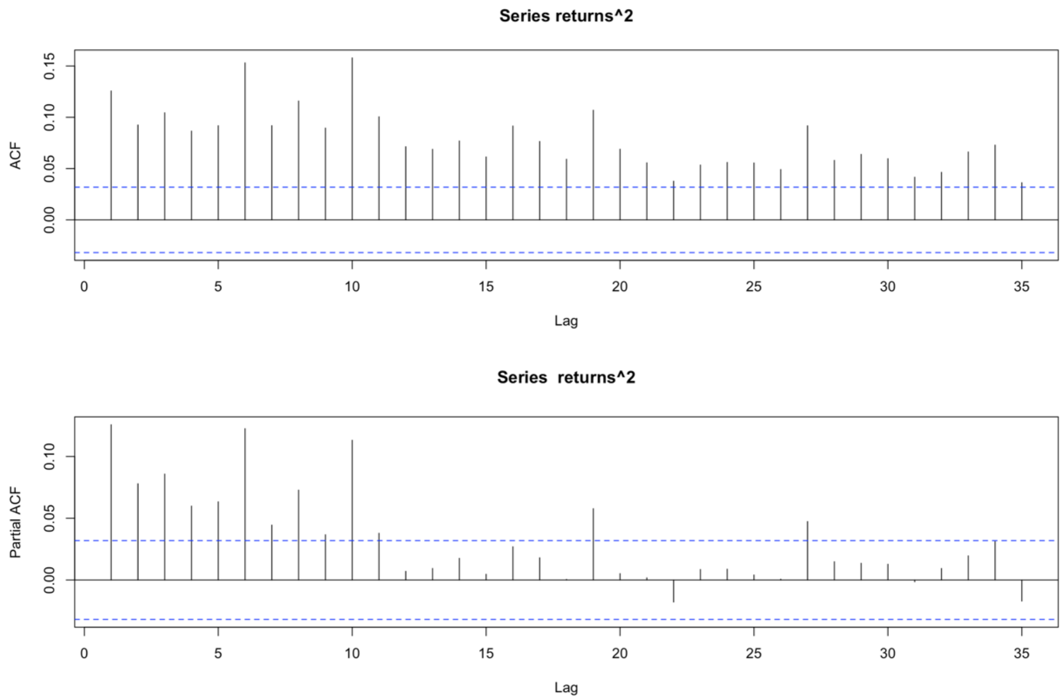

Figure 14. ACF and PACF of Squared Returns

The QQ plot of the returns in Figure 15 suggests that the data has a thicker tail than a normal distribution as the data is far away from the theoretical quantiles at the tails, thus suggesting a heavy-tailed distribution, a common characteristic of stock price data. The kurtosis value of 5.440092 represents the thickness of the tail of the data distribution relative to that of a normal distribution.

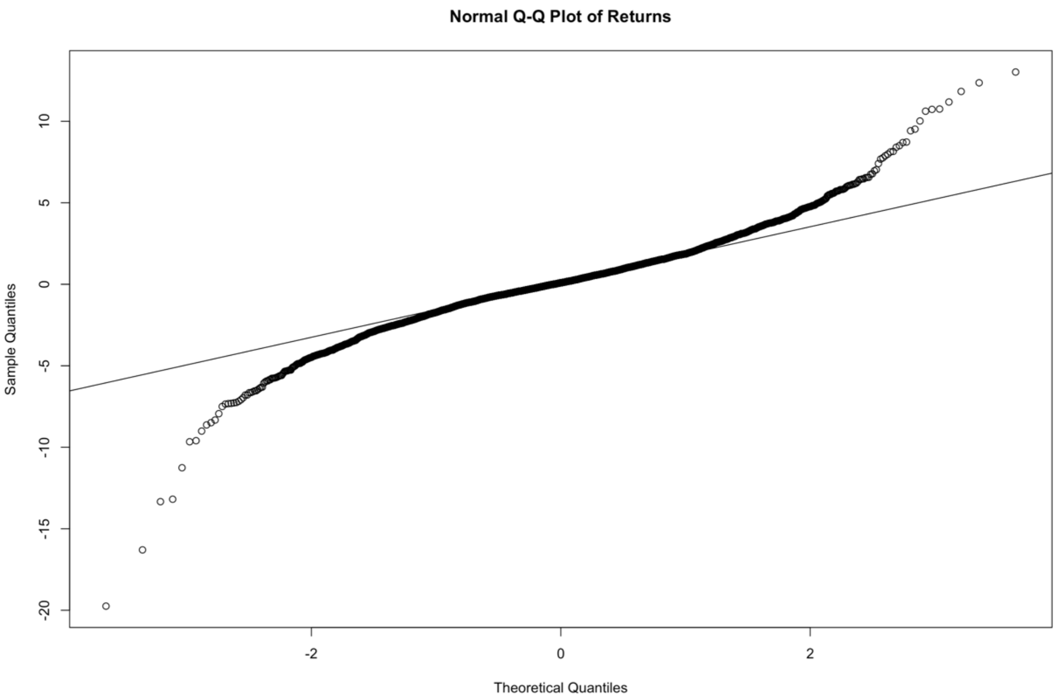

Figure 15. Normal QQ Plot of Returns

Therefore, although the return data seems to be serially uncorrelated (as per Figure 12), there exist higher order dependence structure, such as volatility clustering and a heavy-tail distribution. The EACF approach is used to identify a suitable GARCH model for the return data, with Figure 16, 17 and 18 showing the EACF of the returns, absolute returns, and squared returns respectively.

AR/MA	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	o	x	o	x	o	o	o	o	o	o	o	o	o	o
1	x	x	o	x	o	o	o	o	o	o	o	o	o	o
2	x	x	o	x	o	x	o	o	o	o	o	o	o	o
3	x	x	o	o	o	x	o	o	o	o	o	o	o	o
4	x	x	x	x	o	x	o	o	o	o	o	o	o	o
5	x	x	x	x	x	x	o	o	o	o	o	o	o	o
6	x	x	o	x	x	x	o	o	o	o	o	o	o	o
7	x	x	x	x	x	x	x	o	o	o	o	o	o	o

Figure 16. EACF of Returns

AR/MA	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	x	x	x	x	x	x	x	x	x	x	x	x	x	x
1	x	o	o	o	o	o	o	x	x	o	o	x	o	o
2	x	x	o	o	o	o	o	o	x	o	o	x	o	o
3	x	o	x	o	o	o	o	o	o	o	o	o	o	o
4	x	x	o	o	o	o	o	o	o	o	o	o	o	o
5	x	x	x	x	x	o	o	o	o	o	o	o	o	o
6	x	x	x	x	x	x	o	o	o	o	o	o	o	o
7	x	x	x	x	x	x	x	o	o	o	o	o	o	o

Figure 17. EACF of Absolute Returns

AR/MA	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	x	x	x	x	x	x	x	x	x	x	x	x	x	x
1	x	o	o	o	o	x	x	o	o	x	o	o	o	o
2	x	x	o	o	o	x	o	o	o	x	o	o	o	o
3	x	x	o	o	o	x	o	o	o	x	o	o	o	o
4	x	x	x	x	o	x	o	o	o	x	o	o	x	o
5	x	x	x	x	x	o	o	o	o	x	o	x	o	o
6	x	x	x	x	x	x	o	o	o	x	o	o	x	o
7	x	x	x	x	x	x	x	o	o	x	o	o	x	o

Figure 18. EACF of Squared Returns

Figure 16 suggests a GARCH(0,2) model, with some lags showing correlation, and a GARCH(0,4) model if strictly no correlation allowed. Figure 17 clearly shows a GARCH(1,1) model with some lags showing correlation, but these are within 5% allowance of random correlations occurring. Figure 18 also suggests a GARCH(1,1) model, but there is a lot of lags showing correlations. A GARCH(1,8) model is also suggested if within the 5% allowance, and a GARCH(1,10) model if strictly no correlation allowed. These 5 GARCH models are explored and fitted to the returns data using the garch command with their respective orders, and the result is as shown in Table I.

Table I. Diagnostics of GARCH Models

GARCH Model	Box-Ljung p-value of fit	AIC
0, 2	0.4253	16547.11
0, 4	0.3304	16447.81
1, 1	0.2878	**16205.51**
1, 8	0.8546	16313.33
1, 10	0.838	16272.13

The GARCH(1,1) model appears to be the best model based on its lowest AIC score, thus it will be used for the forecasting. The standardized residuals of the fitted GARCH(1,1) model to the returns data is shown in Figure 19, while the QQ plot of the standardized residuals w.r.t. normal distribution is shown in Figure 20.

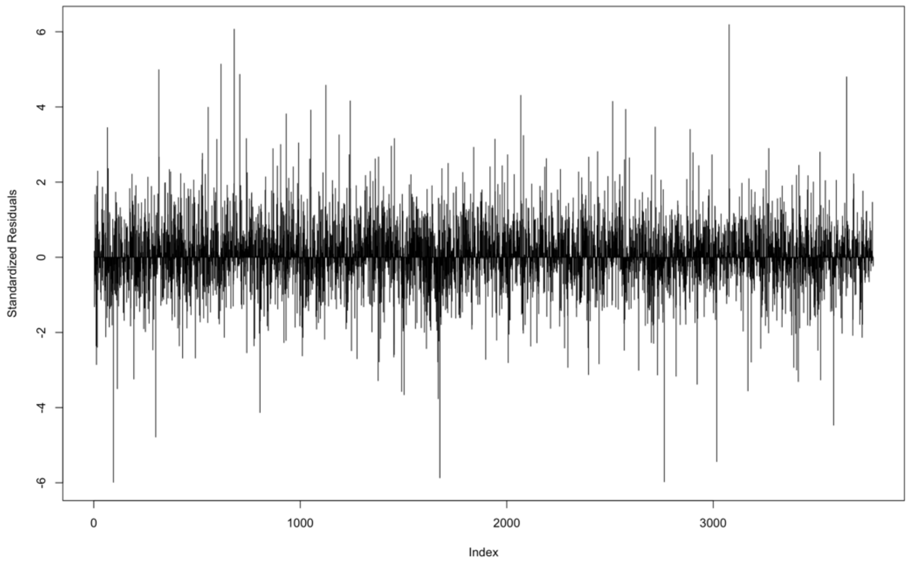

Figure 19. Standardized Residuals from Fitted GARCH(1,1) Model

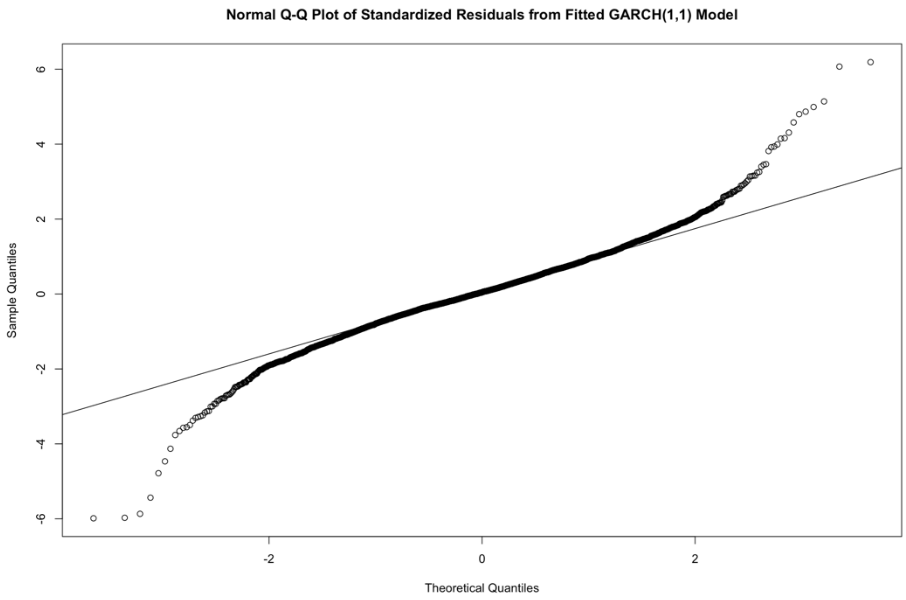

Figure 20. Normal QQ Plot of Standardized Residuals from Fitted GARCH(1,1) Model

As no clear pattern can be observed in Figure 19, it suggests that the standardized residuals from the fitted model is white noise. However, the QQ plot in Figure 20 still shows a heavy tailed distribution, with a kurtosis value of 3.540995, which is smaller than that of the original returns data. The ACF of the squared residuals of the fitted model also suggests that there is no correlation between the standardized residuals from the fitted model, as shown in Figure 21. Figure 22 shows the p-values of the generalized portmanteau tests with the squared standardized from the fitted GARCH(1,1) model, which are all above 0.05, thus further suggesting that there is no correlation between the standardized residuals.

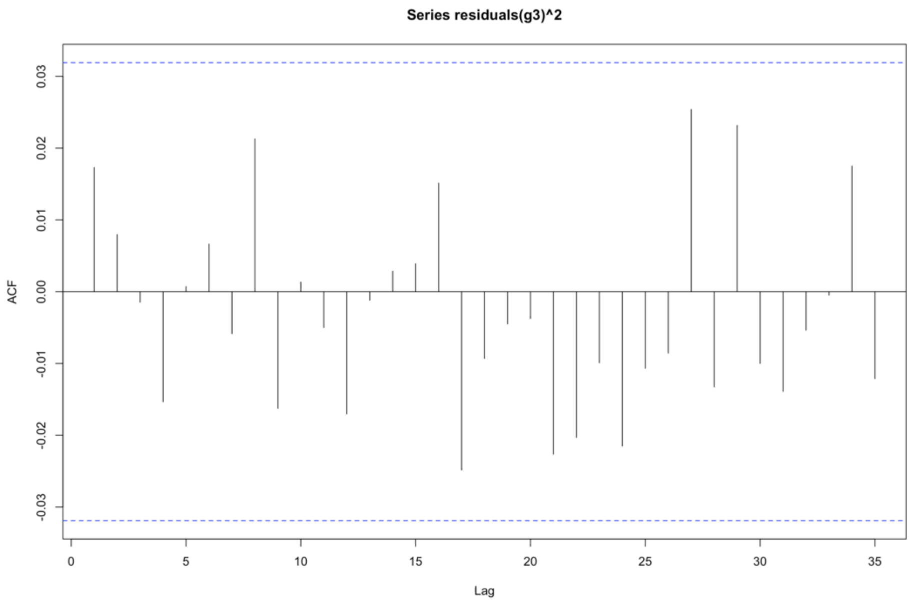

Figure 21. ACF of Squared Residuals from Fitted GARCH(1,1) Model

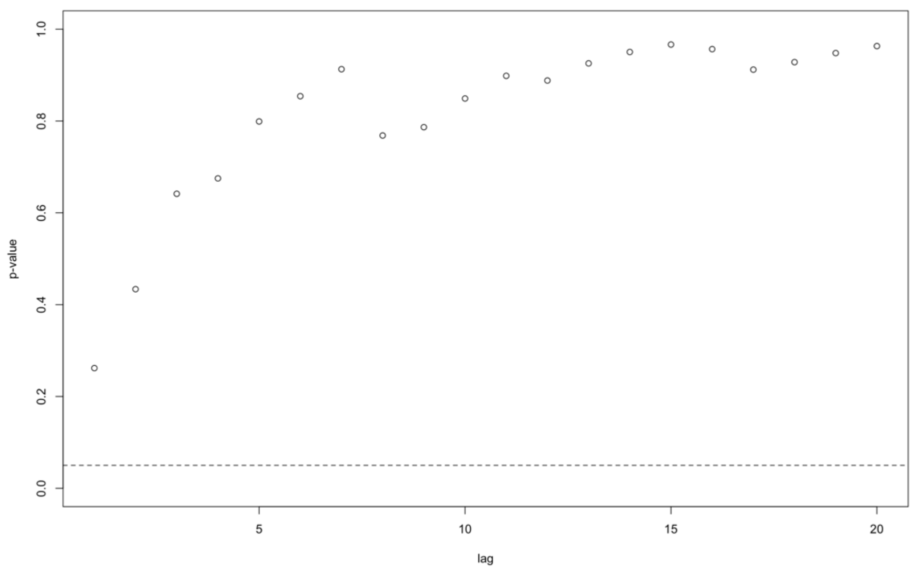

Figure 22. Generalized Portmanteau Test Plot of Squared Residuals from Fitted GARCH(1,1) Model

The next step involved the fitting of the returns data using the rugarch library in R. It allows us to specify the distribution type as well as the variance model type to fit to the returns data. The distribution types are explored first, and includes 'norm' for normal, 'snorm' for skew-normal, 'std' for student-t, and 'sstd' for skew-student distributions. The results are shown in Table II, which shows that the 'std' distribution best reflects the distribution of the returns data as it achieved the lowest AIC score. The variance model types are then explored based on the 'std' distribution, and includes 'sGARCH' for standard GARCH, 'gjrGARCH' for GJR GARCH, and 'eGARCH' for exponential GARCH. The results are shown in Table III, which shows that the 'eGARCH' model of order (1,1) best reflects the variance of the returns data.

Table II. Distribution Exploration Results

Variance Model	Distribution	LogLikelihood	AIC
sGARCH	norm	-8088.507	4.2874
sGARCH	snorm	-8088.274	4.2878
sGARCH	std	-7912.203	**4.1945**
sGARCH	sstd	-7911.74	4.1948

Table III. Variance Model Exploration Results

Variance Model	Distribution	LogLikelihood	AIC
sGARCH	std	-7912.203	4.1945
gjrGARCH	std	-7902.228	4.1898
eGARCH	std	-7888.421	**4.1825**

The specification of 'std' distribution and 'eGARCH'(1,1) variance model is then used to fit to the returns data. The QQ plot w.r.t. 'std' distribution, ACF of standardized residuals and ACF of squared standardized residuals are shown in Figures 23, 24 and 25 respectively. The model fits more closely to the 'std' distribution as compared to the 'norm' distribution previously observed in Figure 20. The ACF of the standardized residuals and squared standardized residuals also shows little correlation except at lag 25 for the squared standardized residuals, which may suggest that there is some monthly seasonality pattern in the data.

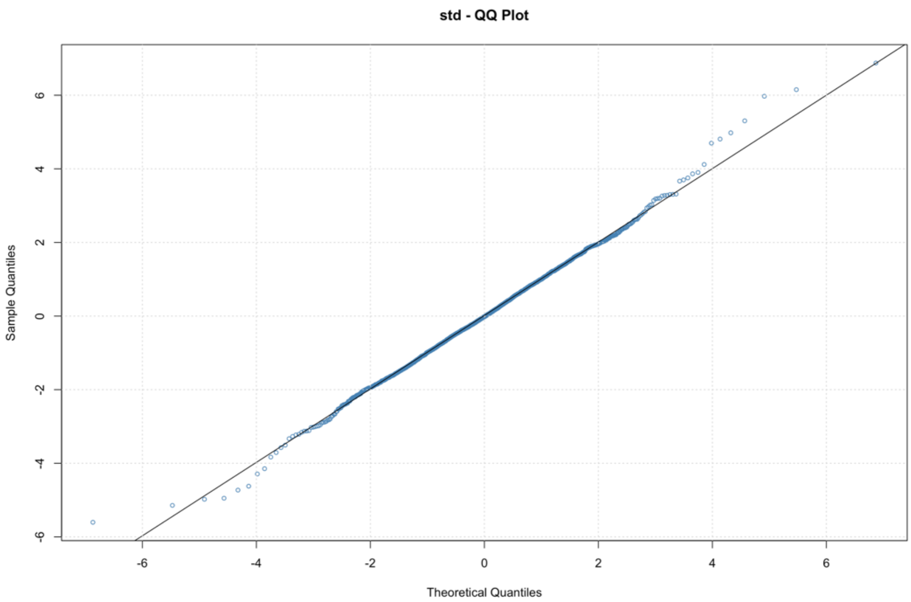

Figure 23. Std QQ Plot of Standardized Residuals from Fitted eGARCH(1,1) Model

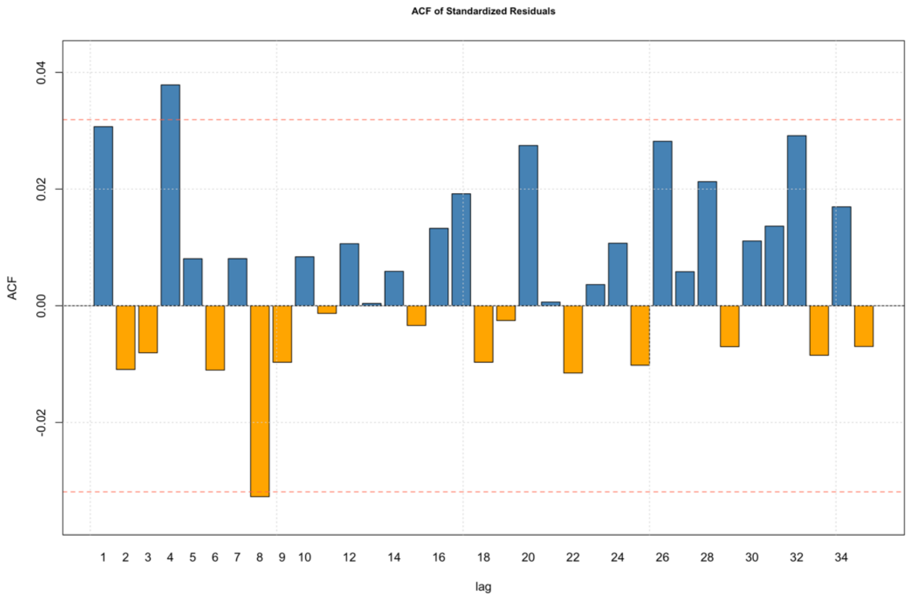

Figure 24. ACF of Standardized Residuals from Fitted eGARCH(1,1) Model

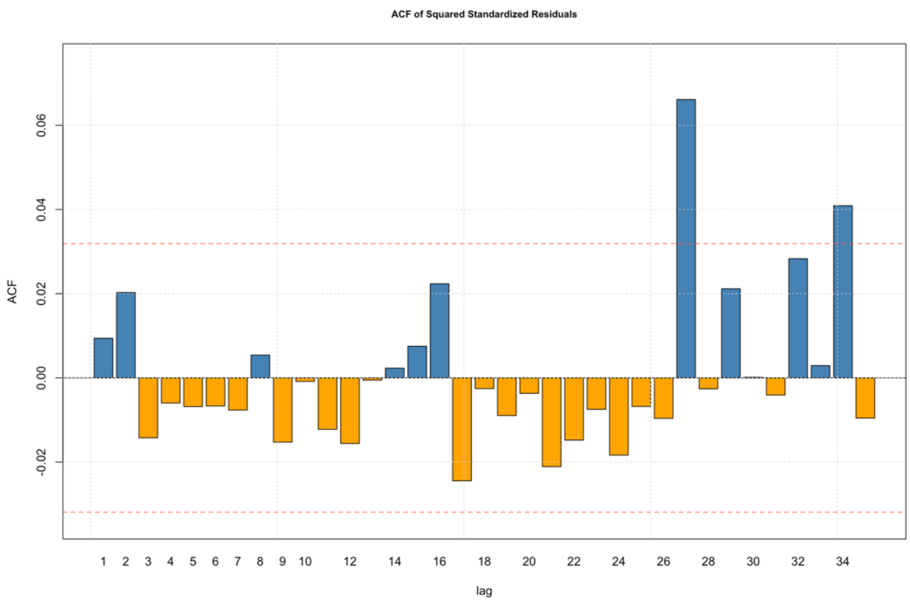

Figure 25. ACF of Squared Standardized Residuals from Fitted eGARCH(1,1) Model

The fitted model is then used to make a forecast of 1 year ahead, up to 31st January 2018, and the forecasted returns is shown in Figure 26. The forecasted returns shows a flat line, similar to that of the ARIMA models. The forecasted variance is shown by the yellow area, which covers most of the actual returns of the 1 year ahead. However, there are some discrepancies where the actual returns is much more than the upper and lower bounds of the forecasted returns. This is likely due to a sudden jump in the stock price after 31/01/2017, where Figure 27 shows the zoomed in close price, with the vertical red dashed line showing 01/02/2017. There is an obvious jump in the close price from this day to the next, thus resulting in larger variance in the returns data which the fitted eGARCH(1,1) model was not expecting.

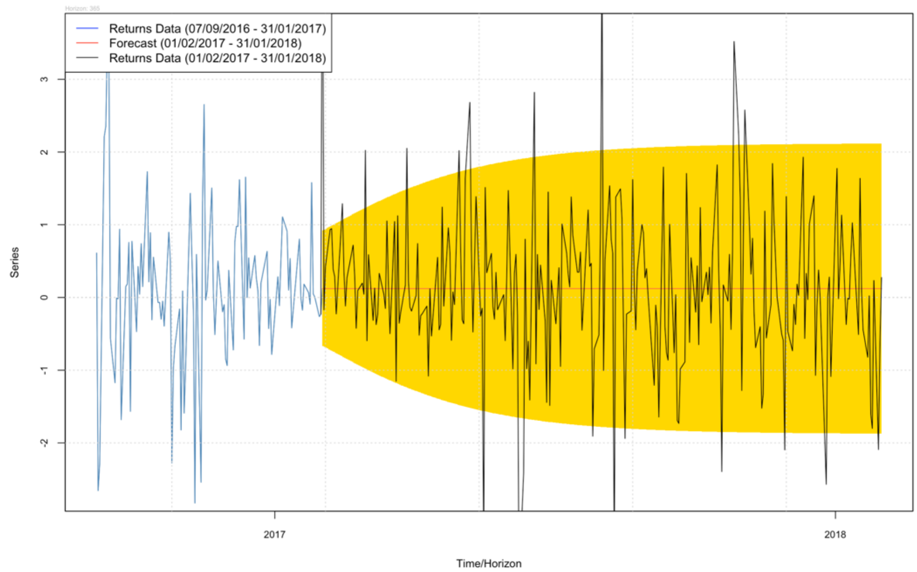

Figure 26. Forecasted Returns from Fitted eGARCH(1,1) Model

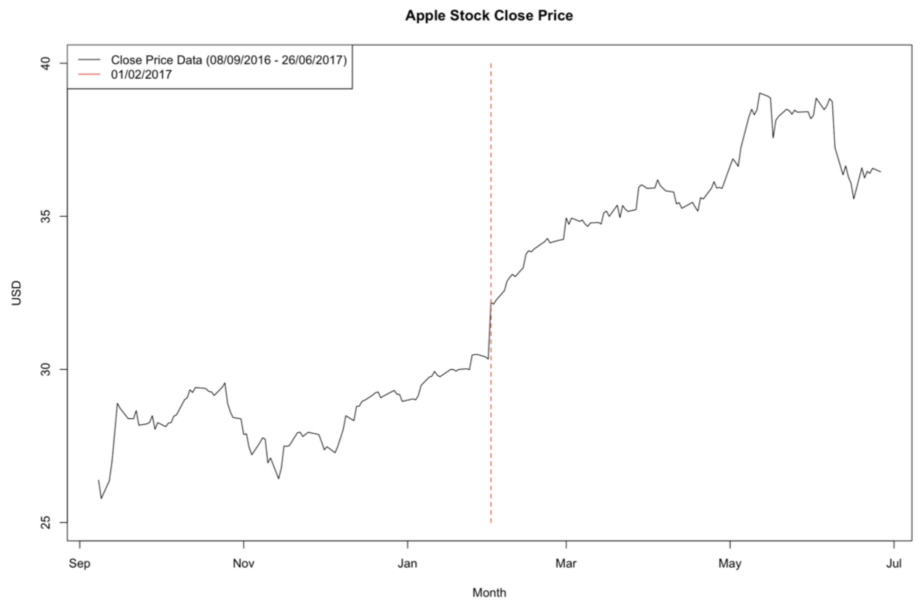

Figure 27. Apple Stock Close Price (08/09/2016 - 26/06/2017)

The forecasted variance shown in Figure 28 predicts that the variance will increase over the 1 year ahead forecast. Take note that the plot shows the variance in the percentage returns, hence should be divided by 100 for actual standardized variance. Since no exact close price can be predicted due to the variance, 3 simulated forecasts were performed using the rugarch library and is shown in Figure 29, with the vertical dashed line showing the start of the forecasts. This is done by simulating 3 forecasted returns, followed by a division by 100, cumulative sum, and then exponential function to reverse the differenced log transformation of the returns (was in percentage) data to the close price data. The 3 simulated forecast of the close price (shown in red, blue and orange) all have smaller close prices as compared to the actual data shown in black, which may be a result of the random seed used in the simulation and the smaller variance of the forecasted returns data as compared to the actual returns data.

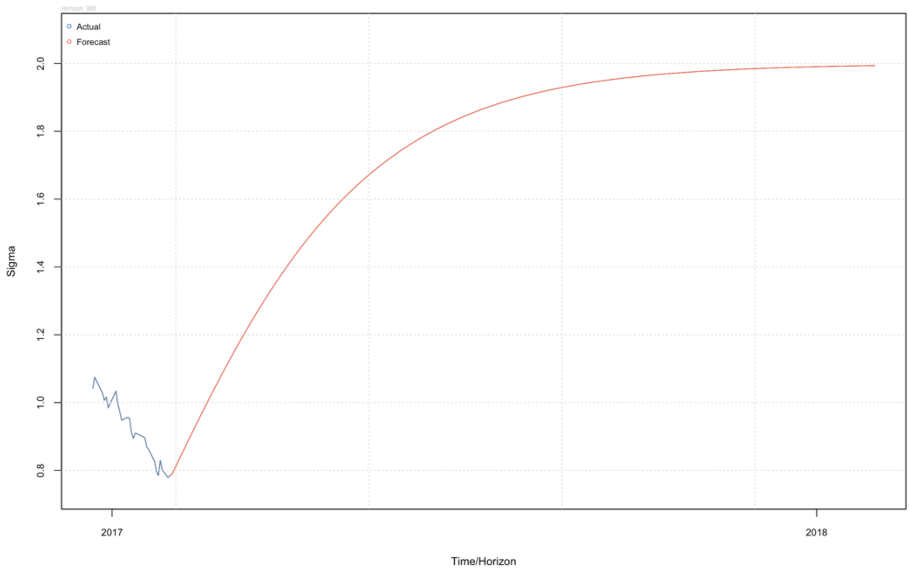

Figure 28. Forecasted Variance from Fitted eGARCH(1,1) Model

Figure 29. Simulated Forecasted Close Price from Fitted eGARCH(1,1) Model (01/02/2002 – 31/01/2018)

Change-point Analysis:

As the actual close price has a large increase after the 01/02/2017 date, a change-point analysis was conducted on by detecting the changes in the mean. With the cpt.mean function from the changepoint library, the PELT method was utilised to detect the changepoints, and the plot is shown in Figure 30, with the dashed vertical line on the date of 01/02/2017. The detected changepoints are shown in Table IV, with the datapoint range and the respective mean of the close price. The bolded datapoints starts from the date of 01/02/2017, which further supports that there is a much higher mean close price in the period between 01/02/2017 and 31/01/2018.

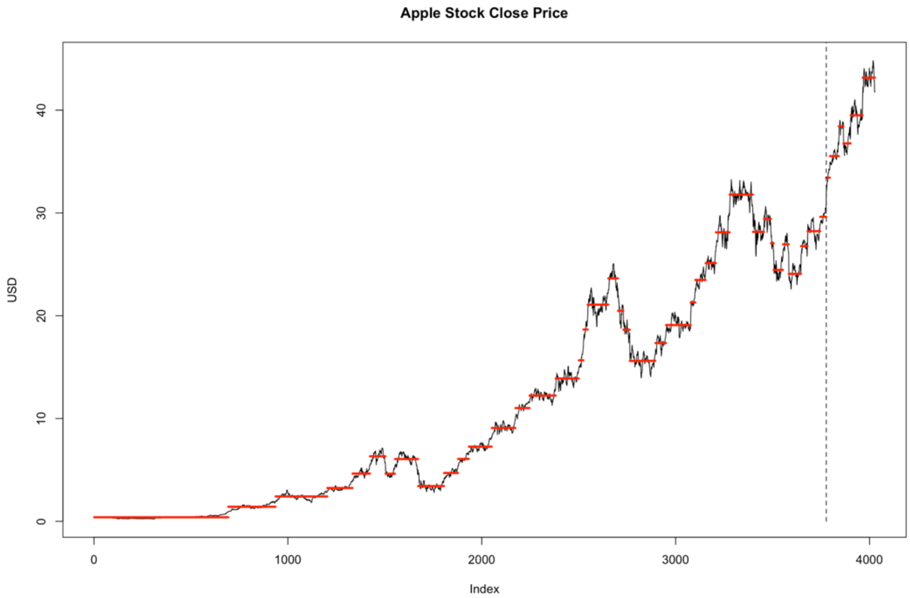

Figure 30. Change-points Plot of Stock Price

Table IV. Changepoints and Respective Mean Close Price

Datapoint	Mean (USD)	Datapoint	Mean (USD)	Datapoint	Mean (USD)
0 - 692	0.396	2382 - 2500	13.896	3399 - 3455	28.159
693 - 936	1.420	2501 - 2524	15.666	3456 - 3491	29.419
937 - 1203	2.421	2525 - 2546	18.659	3492 - 3505	27.036
1204 - 1333	3.240	2547 - 2652	21.075	3506 - 3553	24.456
1334 - 1423	4.646	2653 - 2702	23.633	3554 - 3583	26.944
1424 - 1503	6.327	2703 - 2730	20.475	3584 - 3646	24.072
1504 - 1552	4.629	2731 - 2763	18.639	3647 - 3680	26.767
1553 - 1671	6.061	2764 - 2895	15.609	3681 - 3745	28.222
1672 - 1805	3.430	2896 - 2950	17.341	3746 - 3776	29.623
1806 - 1877	4.702	2951 - 3077	19.094	**3777 - 3795**	**33.429**
1878 - 1933	6.068	3078 - 3101	21.302	3796 - 3841	35.526
1934 - 2051	7.265	3102 - 3154	23.479	3842 - 3865	38.424
2052 - 2173	9.084	3155 - 3207	25.121	3866 - 3902	36.763
2174 - 2246	11.018	3208 - 3278	28.108	3903 - 3964	39.499
2247 - 2381	12.236	3279 - 3398	31.785	3965 - 4028	43.157