AI6123 TIME SERIES ANALYSIS

INDIVIDUAL PROJECT 2

School of Computer Science and Engineering

Programme: MSAI

Date: 2 May 2024

Authored By: Tan Jie Heng Alfred (G2304193L)

Contents

Part 1

Question 1

Let $k \in \mathbb{Z}$, then $\mathbb{E}(X_{2k}) = \mathbb{E}(Y_t) \neq \mathbb{E}(Y_t) + 1 = \mathbb{E}(X_{2k+1})$, X_t is not stationary.

Question 2

Since $X_t = (1 + 2t)S_t + Z_t$, we take

$$\begin{aligned}\nabla_{12}X_t &= X_t - X_{t-12} \\ &= (1 + 2t)S_t + Z_t - \big(1 + 2(t - 12)\big)S_{t-12} + Z_{t-12}) \\ &= 1 + 2tS_t + Z_t - 1 - 2tS_{t-12} + 24S_{t-12} - Z_{t-12} \\ &= 24S_t + Z_t - Z_{t-12}\ (\because S_t = S_{t-12})\end{aligned}$$

Assuming S_t is stationary, then we are done. Otherwise, we perform another seasonal differencing and have,

$$\begin{aligned}\nabla_{12}^2X_t &= \nabla_{12}(24S_t + Z_t - Z_{t-12}) \\ &= 24S_t + Z_t - Z_{t-12} - (24S_{t-12} + Z_{t-12} - Z_{t-24}) \\ &= Z_t - 2Z_{t-12} + Z_{t-24}\ (\because S_t = S_{t-12})\end{aligned}$$

which is stationary.

Question 3

(a) No, as there seems to be volatility clustering. For instance, for the time period between 1950 and 1952, the expected number of passengers is roughly 200. However, between 1960 and 1962, the expectation lies much higher, closer to 500. This means that the expected value of the time series is not constant across time, hence not stationary.

(b) It contains a clear (upward) trending component, seasonality component, where there is an increase in number of passengers at the start of each year, before a decline in the last few months of the year. Additionally, there is also an irregular fluctuation component, which is the inherent randomness in the time series.

(c) We could use the Box-cox transformation:

$$z_t = \frac{x_t^\lambda - 1}{\lambda}, \text{for } \lambda > 0,$$

where x_t are the observed data instances and z_t are the corresponding transformed data instances.

Part 2

1. Introduction

In this project, we will analyze the price of Apple stock (Ticker symbol: AAPL) over the period from 1 February 2002 to 31 January 2017. We will first extract the closing prices of the stock over the period from Yahoo! Finance[1], before analyzing the trend of the stock price, which is a time series, including fitting different types of models and using them for forecasting.

1.1 Exploratory Data Analysis

We first extract the stock prices from Yahoo! Finance using the `getSymbols` function from the `quantmod` library. Then, we extracted the closing price, which is the 5th column in the stock data. With this, we plotted the time series for the abovementioned duration in Figure 1.

Figure 1: Closing price of Apple stock (AAPL) from 1 February 2002 to 31 January 2017.

From the plot, we can see that there is an obvious trending component, and hence is not a stationary time series. Also, its seasonality component is not very prominent. Additionally, we can see that the phenomenon of volatility clustering is very prominent here, where a large (small) change in value typically follows from a large (small) change. More precisely, it means that the conditional variance of the time series varies over time. This also suggests that the use of GARCH models would be appropriate in modelling such time series.

As a more rigorous way of determining whether it is stationary, we plotted the sample ACF (SACF) and PACF (SPACF) in Figure 2 below. Clearly, the SACF is significantly large and not decaying quickly, while its SPACF cuts off after lag 1. This suggests that the time series is non-stationary. Indeed, performing an Augmented Dickey-Fuller (ADF) test on our time series yields a Dickey-Fuller statistics of -2.3941 and a p-value of $0.4114 > 0.05$. Hence, under 5% level of significance, we have insufficient evidence to reject the null hypothesis that a unit root exists for the characteristic polynomial of the underlying model of the time

[1] https://finance.yahoo.com/quote/AAPL/history?p=AAPL&period1=1012521600&period2=1485820800

series. As such, given that the p-value is so high, the ADF test corroborates the claim that the time series is indeed non-stationary.

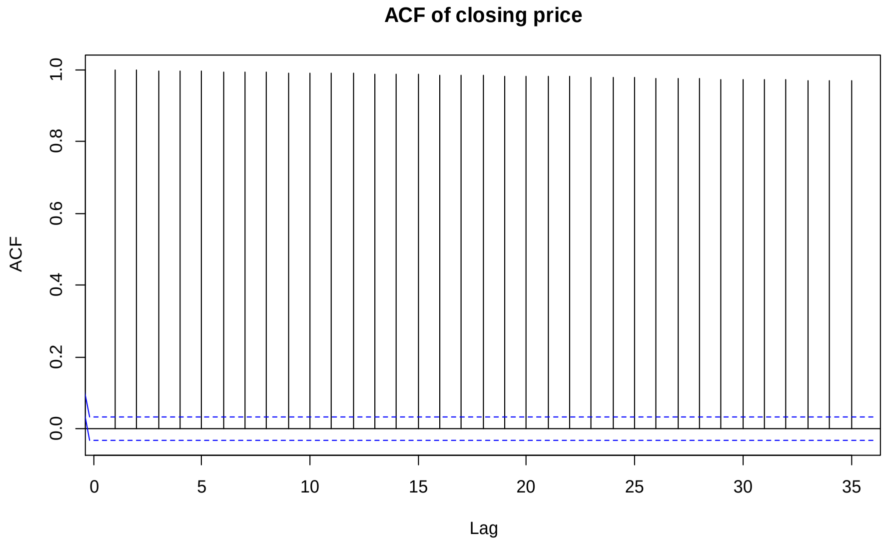

Figure 2: Sample ACF of closing price of Apple Stock, for time lags up to 35.

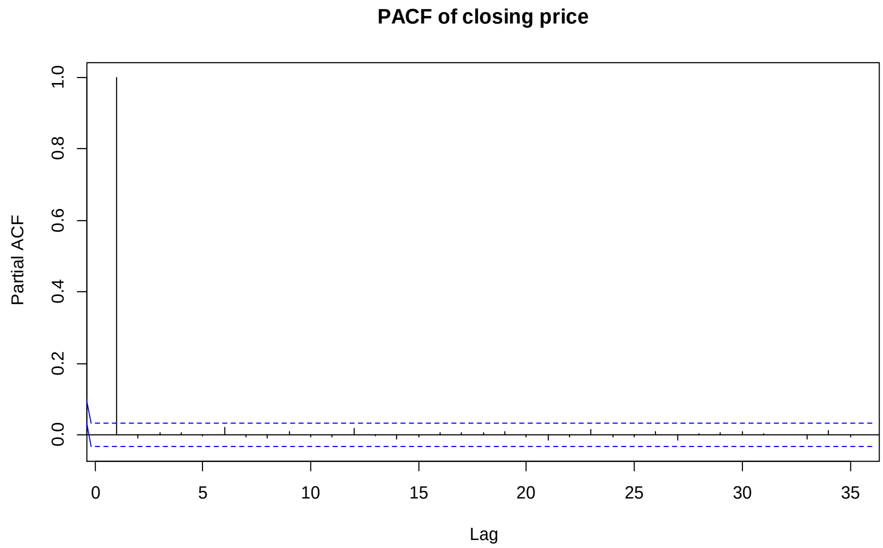

Figure 3: Sample PACF of closing price of Apple Stock, for time lags up to 35.

For completeness, we also performed KPSS test, which assumes as null hypothesis that the trending component of our time series is deterministic (i.e., trend-stationary). For our time series, the KPSS statistics returns to be 34.617 and a p-value of $0.01 \leq 0.05$. This means that we have sufficient evidence to reject the null hypothesis that the time series is trend-stationary and claim that it is indeed non-stationary.

To further dissect our time series, we performed a Seasonality decomposition of Time series by Loess (STL), which decomposes our time series into its trending, seasonality, and irregular components. This is done using Loess, which is a local regression method that locally fits a regression surface to the time series data. To implement this, we convert our time series into monthly observations, before setting a frequency of 12 (for the seasonality component). We include the output of the function in Figure 4 below.

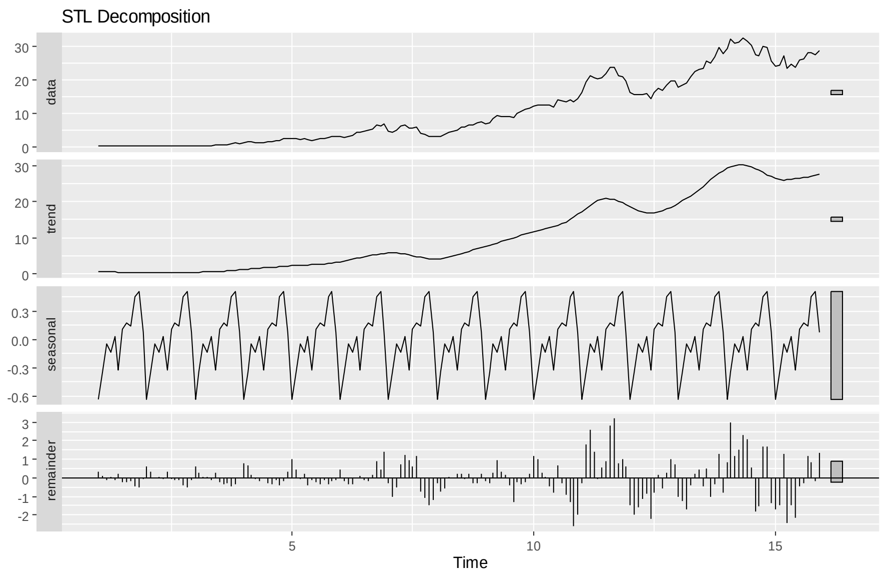

Figure 4: STL decomposition of time series into three components – trend, seasonal and remainder (i.e., irregular).

From the STL decomposition, we can see that the trend is indeed non-linear and, assuming a periodic seasonality component, we can see that the magnitude of the residuals (i.e., remainder) seems to increase with time.

We also plotted the periodogram for the time series in Figure 5 for a clearer picture of our time series. As we can see, the peak frequencies are close to 0 Hz. In fact, the peak frequency corresponds to the frequency 2×10^{-4} Hz, suggesting that there is a slow variation in the stock price, which happens over a long-time span.

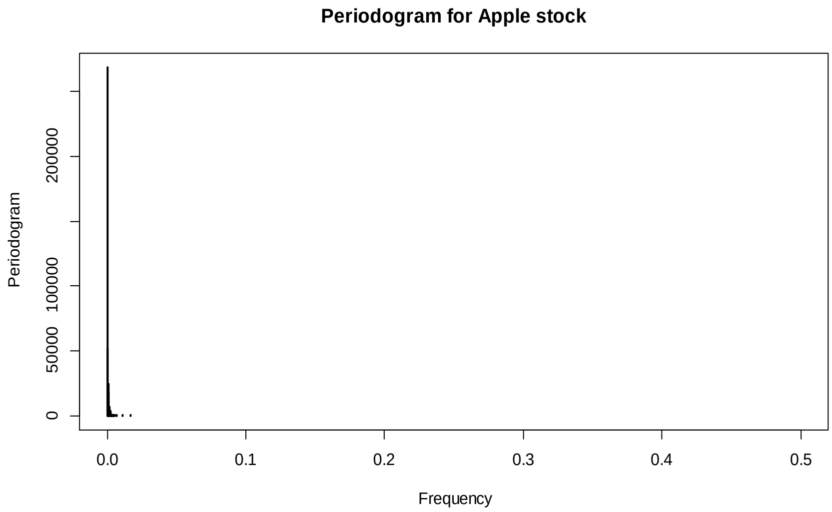

Figure 5: Periodogram for the closing price of Apple stock

1.2 Data Preparation: Train-test Split

Now, for us to model the time series, we will first split the time series into training and test sets. This allows us to fit the model to the training set and validate its forecasting ability on the test set. Because of the observed volatility clustering, we do not want to have too large a test set, as the conditional variance can swing wildly and hence a poorer predictive ability. As such, we set our test set to be the last 30 observations of our time series, while the remaining 3,746 observations form our training set.

2. ARIMA Models

Since there does not seem to be a clear seasonality component across fixed interval, we shall assume that an ARIMA model is sufficient. In what follows, we will attempt to fit the model to the training dataset, before using the fitted model to perform forecasting and validate it against the holdout test set.

2.1 Model Fitting

To smooth out the seasonality component, we shall apply a Box-Cox transformation, obtaining the time series $\{z_t\}_{t\in\mathbb{N}}$, for $z_t = \frac{x_t^\lambda - 1}{\lambda}, \lambda > 0$, for the initial time series $\{x_t\}_{t\in\mathbb{N}}$. In our implementation, λ is obtained using the default Guerrero's method, which minimizes the coefficient of variation for the subseries of the time series $\{x_t\}_{t\in\mathbb{N}}$. Through this, we found that $\lambda = 0.1126439$. We then plot the transformed time series in Figure 6.

Figure 6: Plot of Box-Cox transformed time series, $\lambda = 0.1126439$.

We can still see a clear trending component. Indeed, the SACF and SPACF of the transformed time series (Figures 7 and 8) further strengthen the point that this transformed time series is still non-stationary, with the SACF decaying very slowly. Hence, we perform a differencing on our Box-Cox transformed time series. We then plot the one-time differenced, Box-Cox transformed time series in Figure 9.

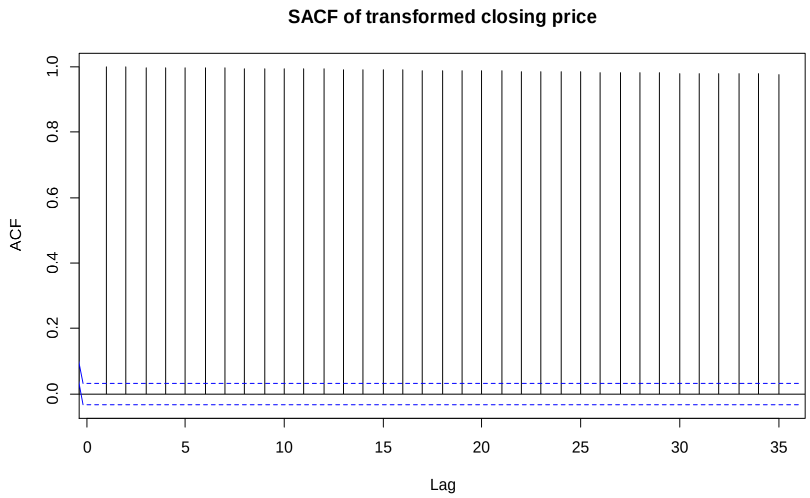

Figure 7: SACF of Box-Cox transformed time series.

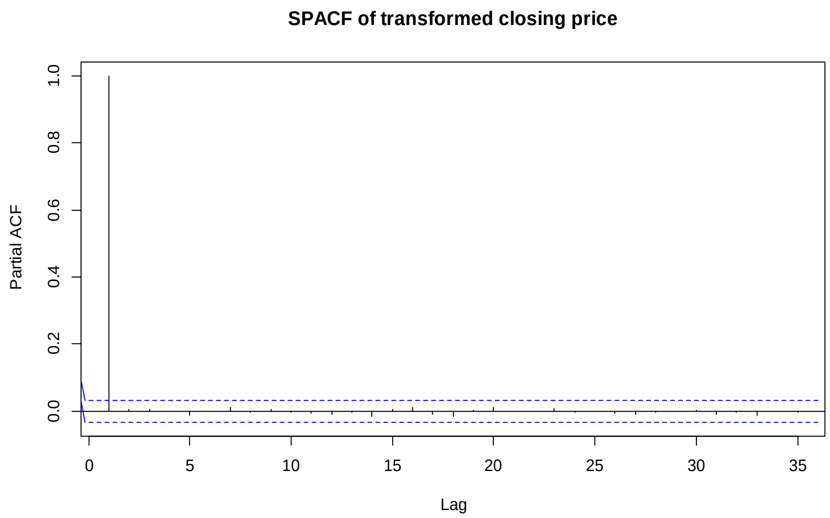

Figure 8: SPACF of Box-Cox transformed time series.

Figure 9: Plot of Box-Cox transformed, one-time differenced time series.

From the plot in Figure 9, the resulting time series seems to resemble a white-noise plot, with constant mean. In order to clarify this, we plotted the SACF and SPACF in Figures 10 and 11.

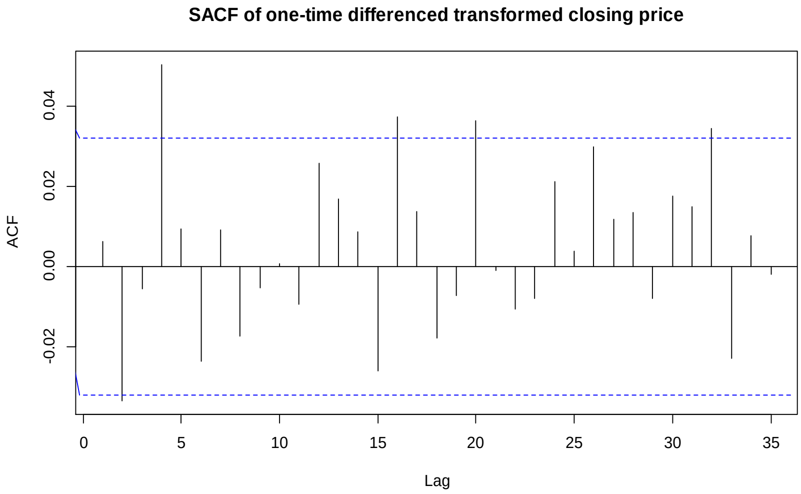

Figure 8: SACF of Box-Cox transformed, one-time differenced time series.

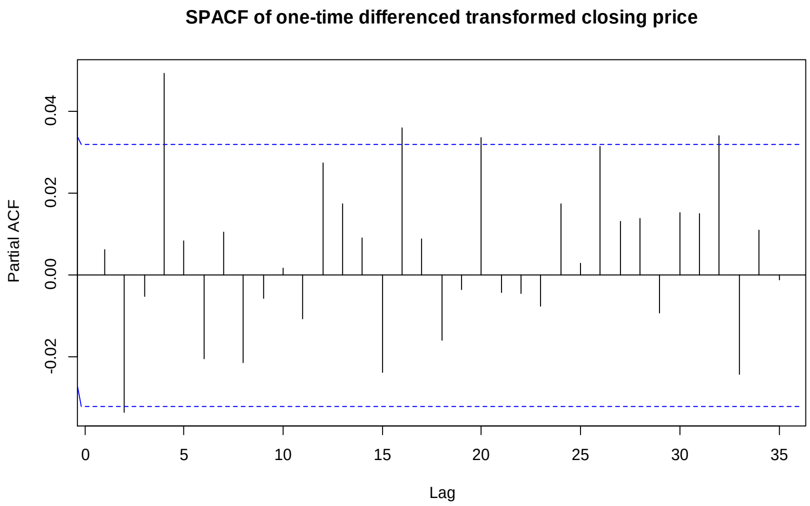

Figure 9: SPACF of Box-Cox transformed, one-time differenced time series.

Taking into consideration that we use a tight bound of $\frac{1.96}{\sqrt{n}}$, where n is the number of observations in our training set, as well as a confidence level of 95%, we will assume that the ACF cuts off after time-lag $k = 4$. To be more rigorous, we also performed an ADF test, obtaining a test statistic of -14.577 and p-value of $0.01 \leq 0.05$. This suggests that we can reject the null hypothesis of the existence of a unit root for the characteristic polynomial of this time series, under 5% level of significance. Hence, we can conclude, with 95% level of confidence, that the one-time differenced, transformed time series is stationary. Furthermore, from the SACF and SPACF plots, we can propose two models – ARIMA(0,1,4) and ARIMA(4,1,0).

2.2 Diagnostic test and Model Selection

Now, we shall perform diagnostic tests on both proposed models. In particular, we first fit both ARIMA models to our training time series data and look at the residuals. Additionally, we performed a Ljung-Box statistic test on each of the fitted model. If the models are adequate, then the residues should look like white noise. Figures 10 and 11 show the results of the diagnostic test performed on ARIMA(0,1,4) and ARIMA(4,1,0) respectively.

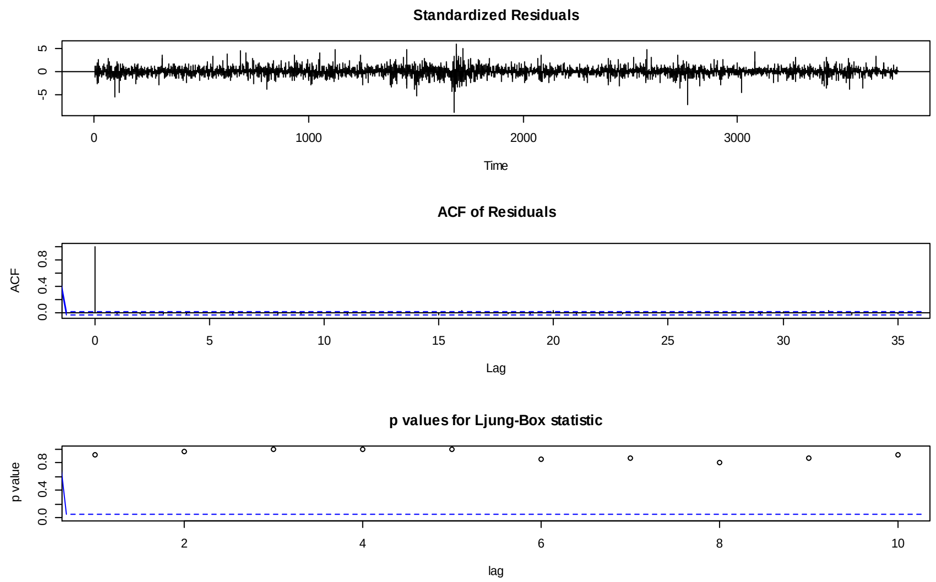

Figure 10: Diagnostic test for ARIMA(0,1,4).

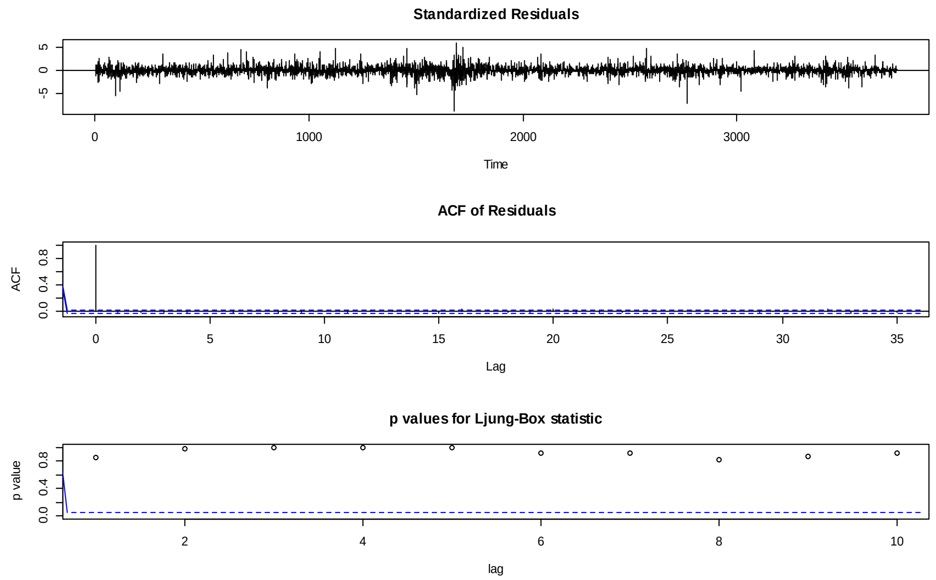

Figure 11: Diagnostic test for ARIMA(4,1,0).

In our test, we plotted the standardized residuals to visually verify that it indeed looks like a white noise distribution. We also plotted the SACF of the residuals – if the SACF at lag $k > 0$ is significantly larger than 0, then it is very likely that the residuals are not following a white noise distribution, as the latter has zero covariance between observations across time. Finally, we perform a Ljung-Box statistic test, which assumes the null hypothesis that our model is adequate and the errors from our model are random and independent up to a certain time lag. Then, we test the likelihood of observing the Ljung-Box statistic, which follows a Chi-squared distribution under our null hypothesis, under a 5% level of significance.

For both models, we can see that the standardized residues resemble a white-noise distribution. Their SACFs also cut off after lag 0, further corroborating the point that the residues are likely following a white-noise distribution. Furthermore, we can see that the p-values of the Ljung-Box statistic test for each model, and all time lag $k \in \{1, 2, \dots, 10\}$, are significantly higher than the 5% level of significance. This means that we have insufficient evidence to reject the hypothesis that the residues are indeed random and independent, up to time lag 10. Altogether, we can safely assume that both proposed models are adequate in modelling our one-time differenced and Box-Cox transformed time series.

Since both models are adequate, we will look at their Akaike's Information Criterion (AIC) score to determine which of the ARIMA models are more accurate, while also considering its complexity (i.e., degree of our ARIMA model). For ARIMA(0,1,4), we obtain an AIC score of -16701.04, while ARIMA(4,1,0) yields an AIC score of -16701.17, when fitted to our training data. Since a lower AIC score suggests that a model is better able to model the training data while balancing its complexity, we will select ARIMA(4,1,0) as the ARIMA model for further forecasting. However, note that both ARIMA models are of similar complexity, hence their slight difference in AIC score suggests that ARIMA(4,1,0) is only very slightly better fitted to our training data. Finally, we plot the fitted ARIMA(4,1,0) together with the ground-truth training observations in Figure 12. We can see that the model fitted very well to the training set, which explains the low AIC score.

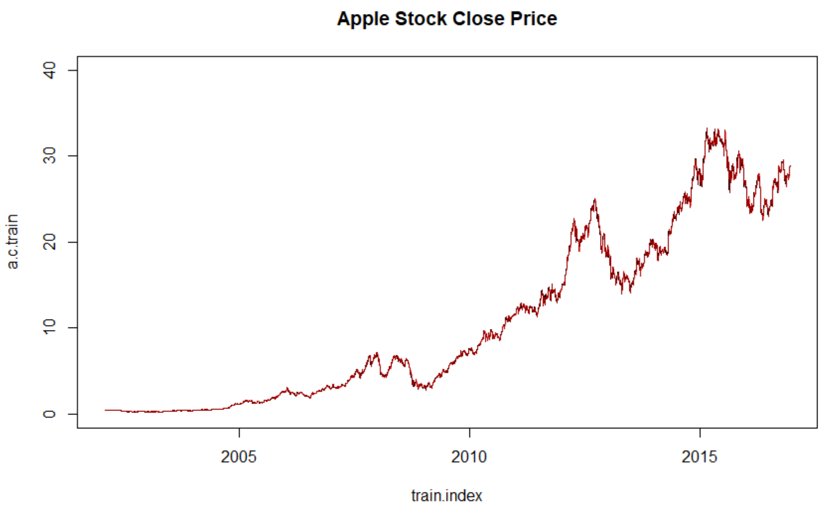

Figure 12: Fitted ARIMA(4,1,0) to training set (in red) and actual ground-truth training set.

2.3 Forecasting and Comparison to `auto.arima`

With the fitted ARIMA(4,1,0), we forecasted the closing price of the Apple stock for the next 30 days, before comparing it to the ground-truth observations (i.e., test set). However, because of the Box-Cox transformation, after obtaining the 30-day prediction, we have to apply an inverse transformation using the same value of $\lambda = 0.1126439$. After accounting for the inverse-transform, we plotted the prediction in red, with the lower and upper bound in blue (i.e., 95% confidence interval), and the ground-truth observation in black (Figure 13).

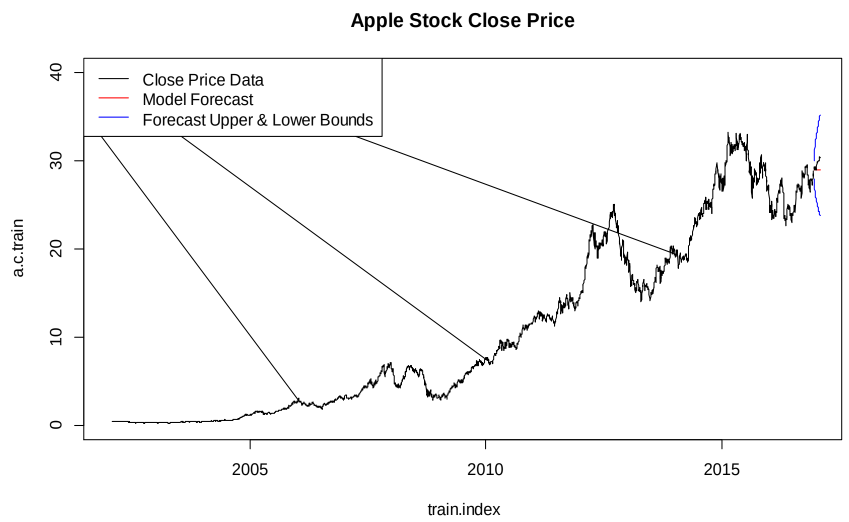

Figure 13: Forecasting using fitted ARIMA(4,1,0) model.

We can see that, despite being so well fitted to the training set, our ARIMA(4,1,0) could not provide a similarly accurate forecast. In fact, the forecast of our ARIMA(4,1,0) has an mean squared error (MSE) of 0.7047028.

As a point of comparison, we fitted an ARIMA(2,1,2) with drift to our training data. It turns out to be adequate, following an identical diagnostic test, and yielding an AIC score of -16708.19, outperforming the ARIMA(4,1,0). More impressively, it could produce forecasts that have smaller errors (see Figure 14). In fact, the MSE of this model is a mere 0.1826923, significantly lower than that from ARIMA(4,1,0). This outperformance likely has a huge attribution to the additional drift component, which accounts for the changing mean of the time series over time. This is a much better attempt at accounting for the trending component of our time series, hence a better fitting and forecasting ability as well.

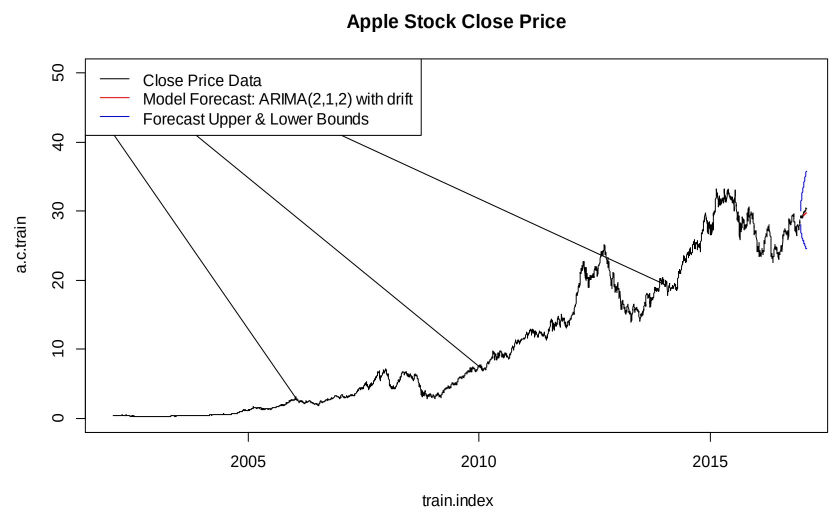

Figure 14: Forecasting using fitted ARIMA(2,1,2) with drift model.

3. GARCH Models

We see that the ARIMA models, without accounting for drift, does not perform well on our time series. This prompted us to look for alternatives, one of which is Generalized Auto-Regressive Conditional Heteroscedasticity (GARCH). As mentioned above, we observed volatility clustering in the time series, which ARIMA models do not deal with. In fact, the time series exhibits varying conditional mean and conditional variance over time, and ARIMA models are able to only model conditional mean well. To model the conditional variance, we appeal to GARCH models.

3.1 Exploratory Data Analysis on Relative Returns

We first apply a log-transformation, equivalently a Box-Cox transformation of $\lambda = 0$, to our initial time series. Following this, we similarly performed a one-time differencing to remove the trend component. Using a log-transformation, instead of Guerrero's method, allows for a more interpretable result of this differencing as well, since,

$$\begin{aligned}\ln x_t - \ln x_{t-1} &= \ln\left(1 + \frac{x_t - x_{t-1}}{x_{t-1}}\right) \\ &= \frac{x_t - x_{t-1}}{x_{t-1}} - \left(\frac{x_t - x_{t-1}}{x_{t-1}}\right)^2 + \left(\frac{x_t - x_{t-1}}{x_{t-1}}\right)^3 - \cdots \\ &\approx \frac{x_t - x_{t-1}}{x_{t-1}},\end{aligned}$$

where the second equality follows from the Maclaurin series of $\ln(1+z)$, and it is true for $-1 < \left|\frac{x_t - x_{t-1}}{x_{t-1}}\right| \leq 1$. The third approximation is reasonable for small $\frac{x_t - x_{t-1}}{x_{t-1}}$, such that powers more than 1 are negligible. Furthermore, we can interpret $\frac{x_t - x_{t-1}}{x_{t-1}}$ as the relative return of the stock at time t. After performing these transformations, we further multipled by 100, before plotting the resulting time series in Figure 15, which can be interpreted as percentage return. Notice that the values are rather small, with majority of the observations being between 5% and -5%. Hence, we can assume that this plot is for the relative return of the stock. Furthermore, the transformed time series looks stationary.

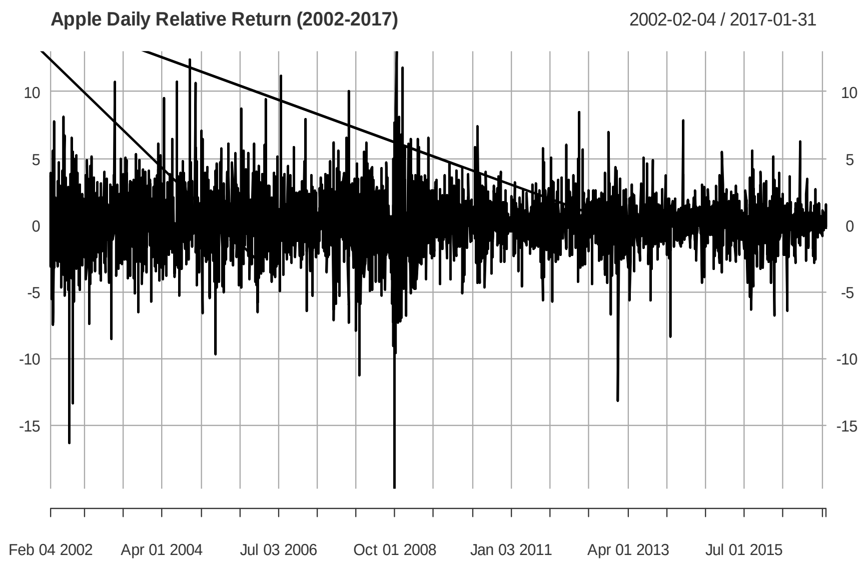

Figure 15: Plot of log-transformed and one-time differences Apple stock closing price.

We included the SACF (Figure 16) and SPACF (Figure 17) plots of the relative return to check for stationarity.

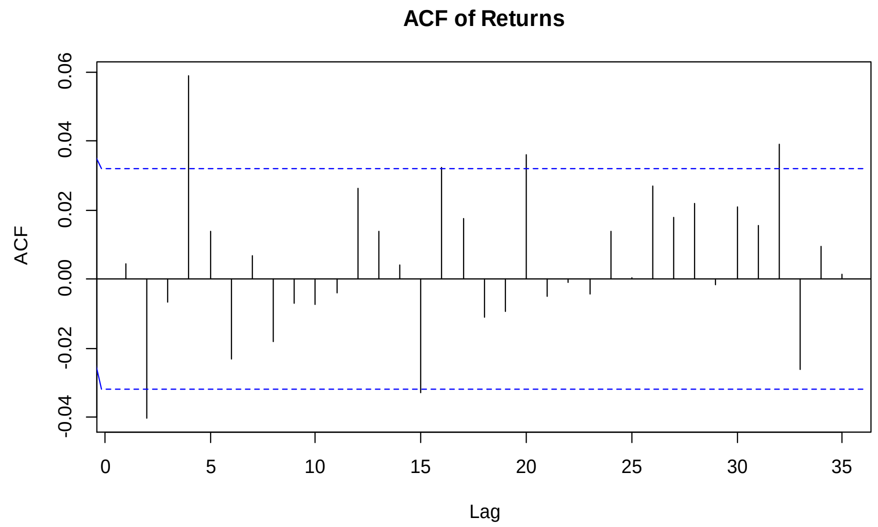

Figure 16: SACF of relative return of Apple stock.

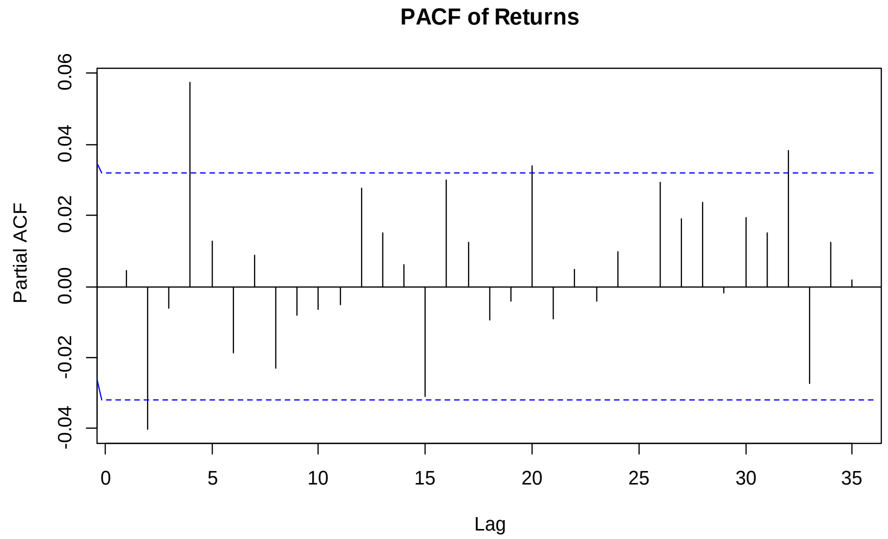

Figure 17: SPACF of relative return of Apple stock.

From the plots, we could assume that ACF and PACF both cut off after lag 4 (similar to what was seen in Section 2.1). With a further ADF test yielding a statistic of -14.931 and p-value of $0.01 \leq 0.05$, we can reject the null hypothesis, under 5% level significance, of ADF and assume that the time series of the Apple stock return is indeed stationary.

Now, we can still see, from Figure 15, the presence of volatility clustering in periods like around June 2002, October 2008, and April 2013, to name a few. To observe this, we plotted the absolute and squared returns, with their respective SACFs, in Figures 18 and 19.

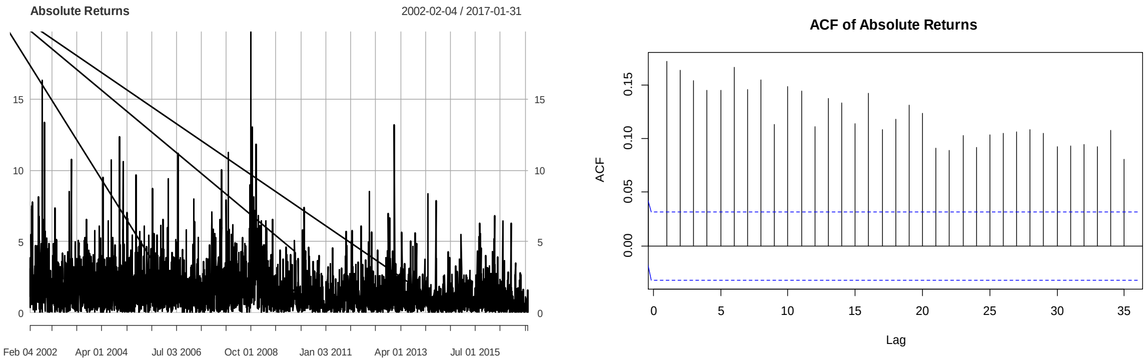

Figure 18: Plots of absolute returns (left) and its SACF (right).

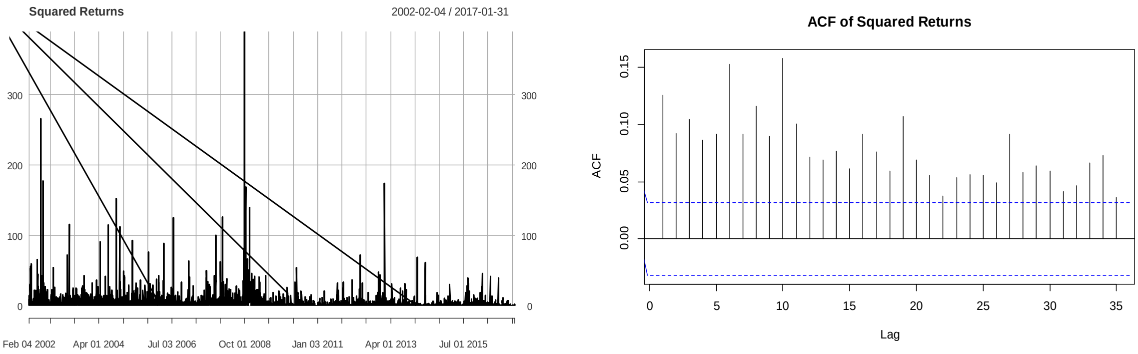

Figure 19: Plot of squared returns (left) and its SACF (right).

From both plots, we can tell that period of high (low) variance typically follows a period of high (low) variance as well. Their ACFs also suggest that the absolute and squared returns have non-negligible correlations across time, hence the returns are not independent and identically distributed. To get a clearer picture of its distribution, we plotted the Normal Q-Q plot (Figure 20) of the returns.

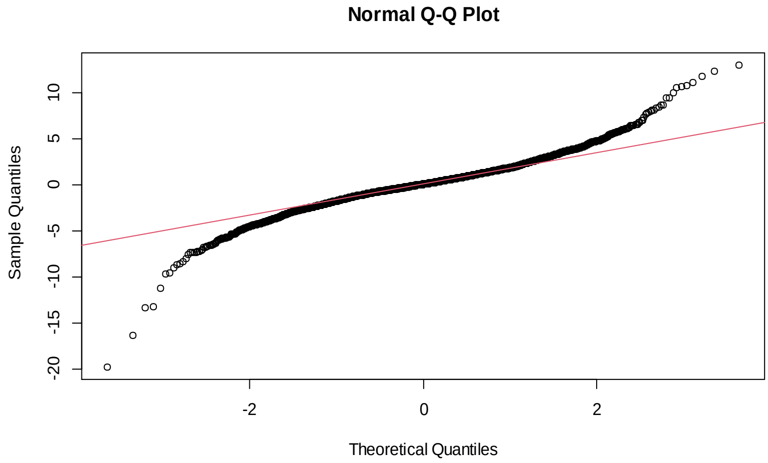

Figure 20: Normal Q-Q plot of returns of Apple stock

From the plot, we can observe that the returns distribution is not Gaussian but a heavy-tailed one, as the Q-Q plot curves outwards, for theoretical quantiles far from 0. Additionally, we can calculate the (sample) excess kurtosis as $\frac{\mu_4}{\sigma^4}$, where $\mu_k = \mathbb{E}((X-\mu)^k)$ is the k^{th} central moment, and the skewness as $\frac{\mu_3}{\mu_2^{3/2}}$. We found that the kurtosis has a positive value of

$5.4409 > 0$, suggesting that the returns distribution is indeed a heavy-tailed one, and a skewness of -0.1901292, suggesting that it is a left-tailed distribution (i.e., the mean value is smaller than the median). This corroborates with the plots we see in Figure 18 (Figure 19), where majority of the absolute (squared returns) are below 5% (below 25%), suggesting that most of the returns are congregated near smaller values.

3.2 Model Fitting

Now that we have seen that the returns are indeed non-Gaussian and not identically and independently distributed, we will attempt to fit a GARCH model that takes these into account. The GARCH(p, q) model is defined by,

$$x_k = \sigma_k \epsilon_k, \text{ where}$$
$$\sigma_k^2 = \omega + \sum_{i=1}^{p} \alpha_i x_{k-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{k-j}^2$$

Here, $\{x_t\}_{t \in \mathbb{N}}$ is the time series, with ϵ_k being i.i.d and having a zero mean and unit variance, and $\omega, \alpha_i, \beta_j \geq 0$, for all $i \in \{1, \ldots, p\}$, and $j \in \{1, 2, \ldots, q\}$. We can see that the GARCH model allows the current conditional variance, σ_k^2, to regress on its squared historical observations, $\{x_t^2\}_{t<k}$ and on historical variances, $\{\sigma_s^2\}_{s<k}$. This structure allows the GARCH model to capture the varying conditional variance.

Before fitting in our GARCH model, we will again split our returns time series into training and test set, with the latter having, again, 30 observations and the former now having 3,745 observations. The split was performed similar to that done for the closing prices. Again, for the fitting of GARCH models, we will fit it to the training observations.

Now, to find a reasonable GARCH model to fit, we have to make use of extended ACF (EACF). First, note that assuming a GARCH(p, q) model for the returns implies that the squared returns follows an ARMA$(\max(p, q), p)$ model. As such, the candidate GARCH model can be realized by finding suitable p and q for the ARMA$(\max(p, q), p)$ model of the squared returns. To that end, we shall use EACF to effectively and efficiently pick the values of $p'$and q' for an ARMA(p', q') model. We found the sample EACF (SEACF) of the squared returns and displayed it in Figure 21.

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 21: SEACF of squared returns.

From the SEACF, we can see that the squared returns most likely follows an ARMA(1,1) or ARMA(2,2) model. However, notice that although we can recover the value of p, the value of q is hidden in the sense that, an ARMA(1,1) model could mean that $q = 0$ or 1. In this case, we will first fit a GARCH(1,1) and the value of q can be changed according to the significance of the coefficients in the GARCH model.

It turns out that we could also study the absolute returns to determine the GARCH order in similar manner. As such, we also computed the SEACF of the absolute returns and displayed it in Figure 22.

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 21: SEACF of absolute returns.

From the computed SEACF of the absolute returns, we can see that a GARCH(1,1) and GARCH(2,2) are the most likely, similar to what we saw in the SEACF of the squared returns. As such, we have two candidates – GARCH(1,1) and GARCH(2,2).

3.3 Diagnostic Test and Model Selection

Similar to checking whether an ARIMA model is adequate, we have to do the same for our GARCH candidates. Again, if a model is adequate (or correctly specified), the standardized residuals are approximately i.i.d. We plot the standardized residuals of GARCH(1,1) in Figure 22 (left). It looks like a white-noise distribution, but not a Gaussian one, as suggested from the QQ-plot (Figure 22, right) and a sample excess kurtosis of $3.485912 > 0$.

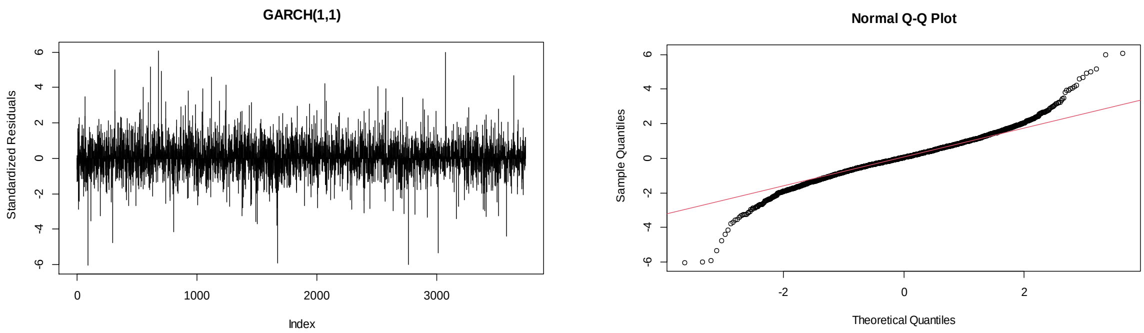

Figure 22: Standardized residuals of GARCH(1,1) (left) and its Q-Q plot (right).

Now, we will look at the squared residuals to observe whether volatility clustering is still present, as well as if the residuals are indeed i.i.d. We plotted the SACF of the absolute residuals and squared residuals, in Figures 23 and 24, to observe if there are any correlations within them. As can be seen, both SACFs of the absolute and squared residuals point to a lack of autocorrelation between the standardized residuals, with both SACFs cutting off after lag 0. Finally, performing a generalized portmanteau test, which assumes the null hypothesis of the standardized residuals being random, yields p-values above the significance level of 5% (see Figure 25), providing us insufficient evidence to reject the hypothesis that the standardized residuals are indeed random. Hence, we can assume that GARCH(1,1) is indeed adequate.

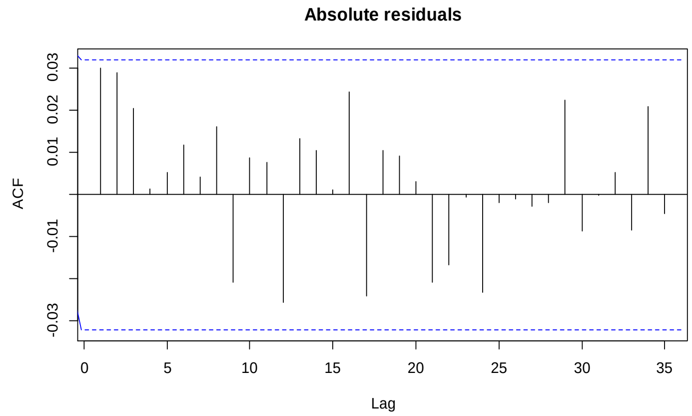

Figure 23: SACF of absolute residuals for GARCH(1,1).

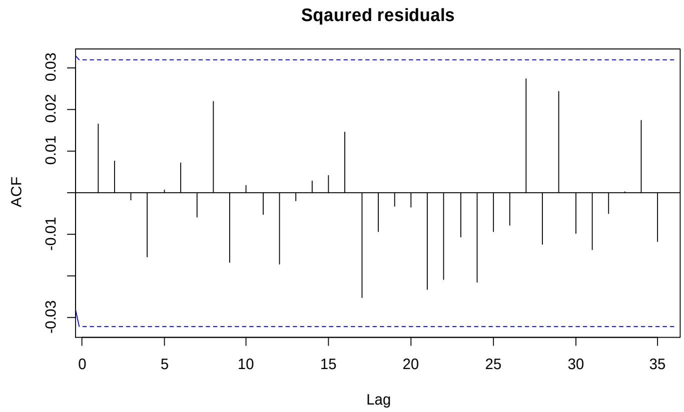

Figure 24: SACF of squared residuals for GARCH(1,1).

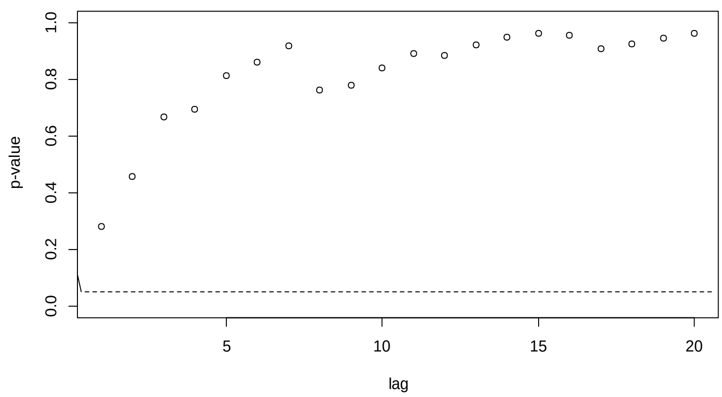

Figure 25: p-values of generalized portmanteau test on GARCH(1,1).

We did the same diagnostic on GARCH(2,2), yielding a white-noise-looking standardized residuals (Figure 26, left) with excess kurtosis of 4.916129, and a Q-Q plot (Figure 26, right) that suggests the standardized residuals are again not following a Gaussian distribution. However, the SACFs for both the absolute (Figure 27) and squared residuals (Figure 28), seem to suggest that the standardized residuals are indeed correlated, with both SACFs decaying really slowly. A further generalized portmanteau test (Figure 29), performed up to lag 20, shows that the p-values after lag 2 are all below the significance level of 0.05. As such, we have sufficient evidence, under 5% level significance, to conclude that the standardized residuals are indeed non-random and the GARCH(2,2) is not an adequate model.

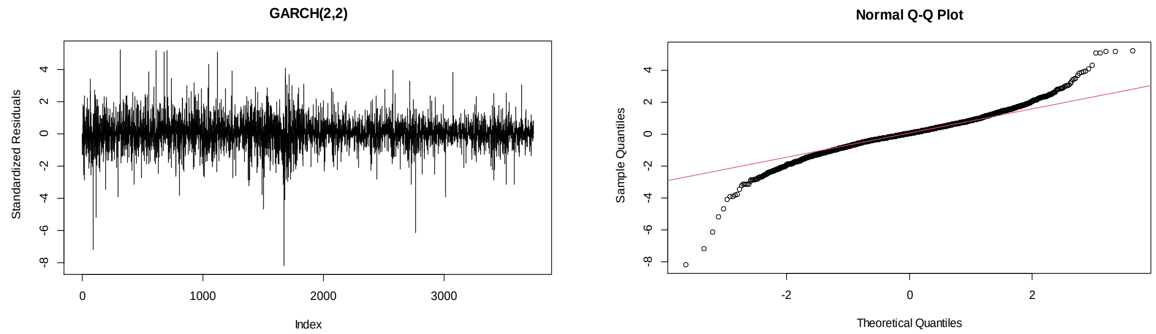

Figure 26: Standardized residuals of GARCH(2,2) (left) and its Q-Q plot (right).

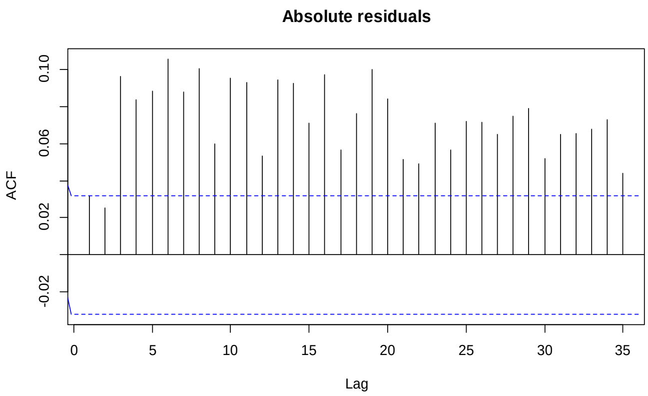

Figure 27: SACF of absolute residuals of GARCH(2,2).

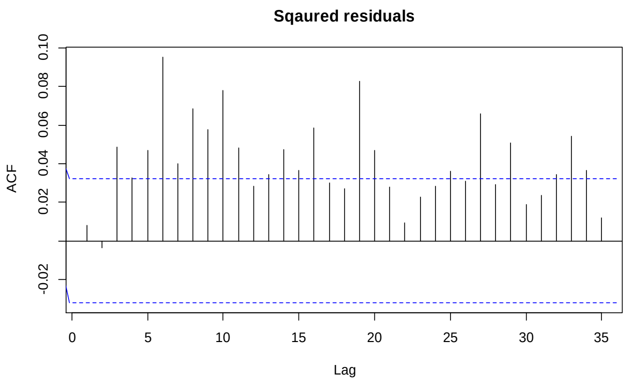

Figure 28: SACF of squared residuals of GARCH(2,2).

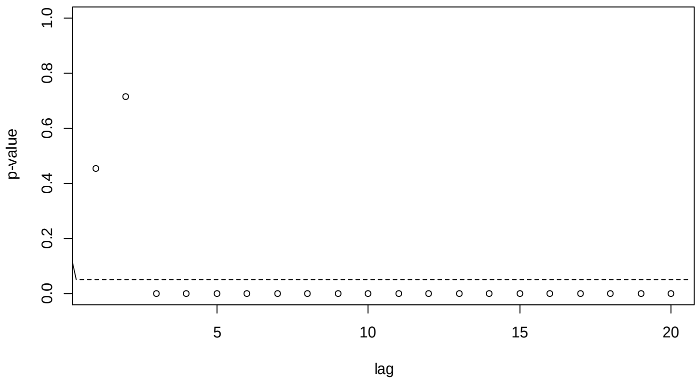

Figure 29: p-values of generalized portmanteau test on GARCH(2,2).

Now, as we mentioned previously, since we obtained the GARCH(1,1) model from EACF, the value of q can be either 0 or 1. To figure this out, we printed the coefficients of our GARCH(1,1) model in Figure 30.

```
Coefficient(s):
    Estimate  Std. Error  t value Pr(>|t|)
a0  0.053283    0.007770    6.858 6.99e-12
a1  0.047270    0.003538   13.361  < 2e-16
b1  0.942597    0.004463  211.218  < 2e-16
```

Figure 30: Coefficients of fitted GARCH(1,1) model.

Here, `a0` refers to ω, `a1` refers to β_1 and `b1` refers to α_1, referring to the definition of GARCH(p, q) model above. Note that the t-values of all the coefficients are all very high; in particular, the extremely small p-value corresponding to `a1` suggests that we can reject the null hypothesis, under 5% level of significance, that the coefficient `a1` is actually 0. As such, we can conclude that q is indeed 1, since the coefficient `a1` is statistically significant.

However, note that the GARCH(1,1) model is not modelling the actual closing price, but models the conditional variance, hence we have to convert the returns to the closing prices.

3.4 Forecasting with GARCH(1,1)

For us to forecast using GARCH(1,1), we will use the `rugarch` library. We specify the ARMA order as (0,0), using exponential GARCH as our variance model, and setting the GARCH order to be (1,1), which we found. We made use of a base distribution model of `'std'`, corresponding to the student-t distribution, since the Q-Q plot, with `'std'` as the base distribution, seems to match (i.e., straight line, Figure 31). Then, we run a series of simulations, first forecasting the conditional variances for 30 observations, before recovering the closing price by performing the inverse-transform outlined in section 3.1. We can perform multiple simulations – as an example, we show the closing prices following three simulations in Figure 32. We can see that the results can vary wildly; in particular, the closing prices of the simulation in green (Figure 32) diverges from the other two simulations in black and red. This is inherent in GARCH models, which aim to capture varying conditional variance. Seeing that the simulations can vary so much, we might consider taking the mean. However, as we can see in Figure 32, the mean simulation (in blue) is much smoother than the other individual simulations. This is because, taking the expectation over multiple observations would smooth out the varying conditional variances, hence defeating the point of using GARCH models in the first place.

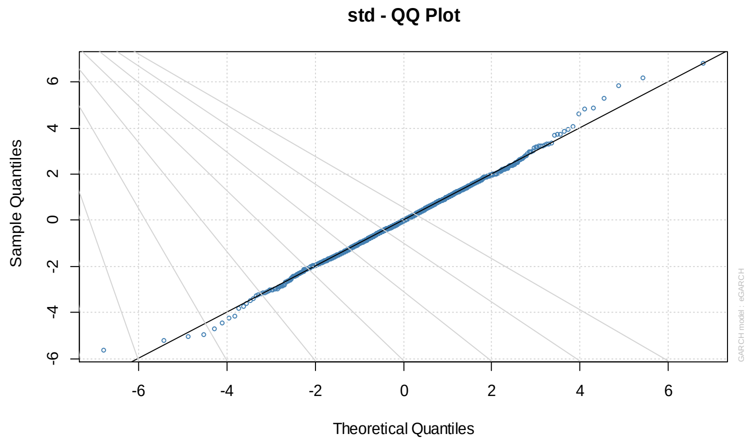

Figure 31: Q-Q plot for fitted model on training set, against student-t distribution.

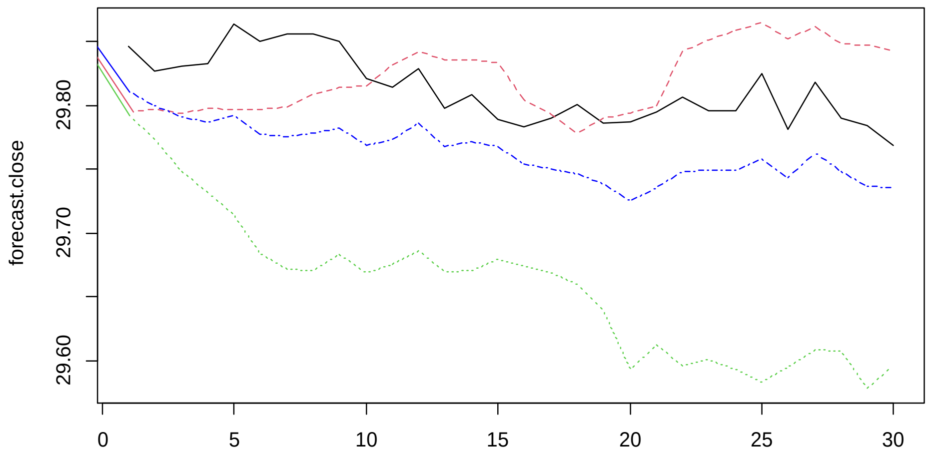

Figure 32: Three simulated forecasts using GARCH(1,1) model. Blue line is the mean of all three simulations.

With these in mind, we attempt to run multiple simulations, such that the maximum forecasted observations hit a certain threshold t, which we set to be the maximum value of the test set, perturbed by some 2ϵ, where $\epsilon \sim \mathcal{N}(0,1)$. For our run, we had $t = 30.5235$, and ran a total of 12897 simulations. We then plotted the forecast, together with the actual time series in Figure 33. Finally, we computed an MSE of the prediction error, yielding a value of 0.2996097, much lower than the ARIMA(4,1,0) model found in Section 2. However, note that our simulation is biased, in the sense that we restrict our forecast search to have at least an observation that is within a margin from the actual maximum of the test set. Notwithstanding that, the inherent modelling of varying conditional variance allows the GARCH(1,1) model to produce a better forecast.

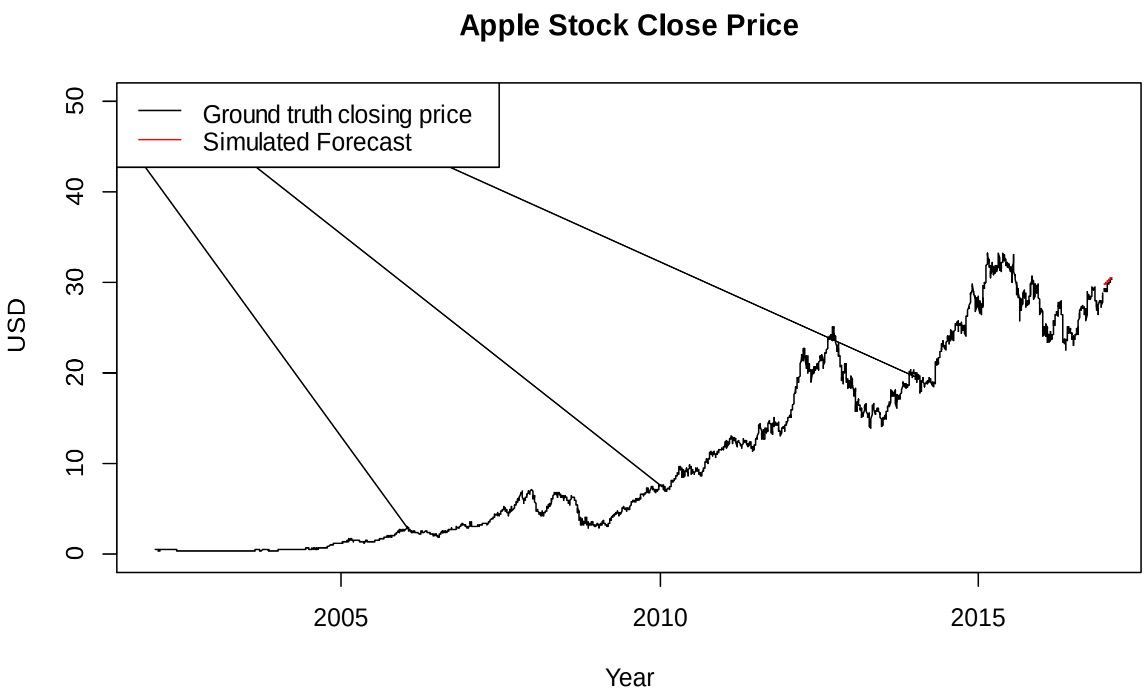

Figure 33: Forecasting with GRACH(1,1) (in red), against ground truth (black).

4. Conclusion

In this project, we looked analysed the closing price of Apple stock from 1 February 2002 to 31 January 2017. We first performed some preliminary data exploration and realized that there are generally only trending and irregular components, and a lack of clear seasonality component. With that, we attempted to fit two models – ARIMA and GARCH models. We found an ARIMA(4,1,0) that seemed to fit the training set well, but not the test set, and this was likely due to the changing conditional variance. For the GARCH model, we found GARCH(1,1) to be suitable in modelling the returns on the stock, including the changing conditional variance. With that, we also came to a forecast, but with some biases to reduce the search space of a decent forecast. If we remove this biased assumption, the forecast can vary wildly. Therefore, if we have some domain knowledge that could be applied to reduce the search space, we could ideally remove the biased assumption, and obtain a much better and generalizable forecast using the GARCH model, since the time series has both varying conditional variance and mean (i.e., non-linear, upward trend).