Assignment 3

Ong Jia Hui JONG119@e.ntu.edu.sg
G1903467L

1. Problem

This project is to analyze financial data. The data are from the daily historical Apple stock prices(open, high, low, close and adjusted prices) from February 1, 2002 to January 31, 2017 extracted from the Yahoo Finance website. The data has logged the prices of the Apple stock everyday and comprises of the open, close, low, high and the adjusted close prices of the stock for the span of 15 years. The goal of the project is to discover an interesting trend in the apple stock prices over the past 15 years (3775 attributes) and to design and develop the best model for forecasting.

1.1 Dataset

The AAPL stock data are fetched from Yahoo Finance using the following R codes.

```
# Fetch apple stock prices from 2002-02-01 to 2017-01-31
getSymbols('AAPL', from='2002-02-01', to='2017-02-01') #"to" is not inclusive
aapl.c <- AAPL[,4] # extract Close prices
```

Listing 1: Fetch AAPL stock data

2. Examine Dataset

This section describes the inital steps taken to analyze the time series data described in Section 1.1 in order to derive the appropriate model for forecasting.

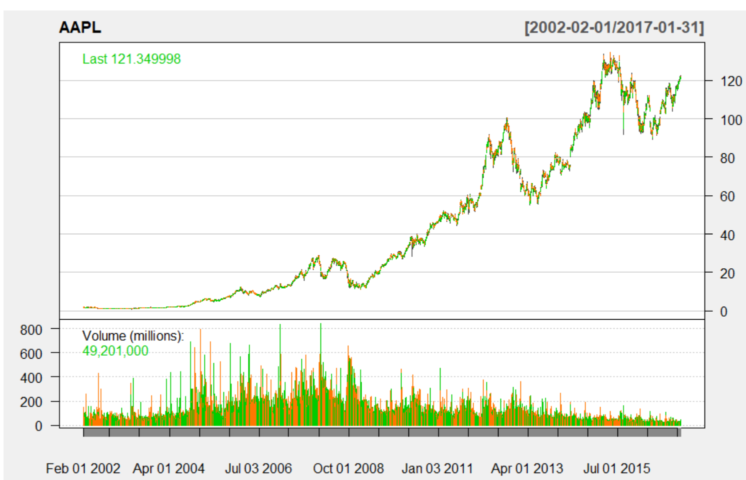


Figure 1: Chart Series of AAPL

2.1 Original Time Plot

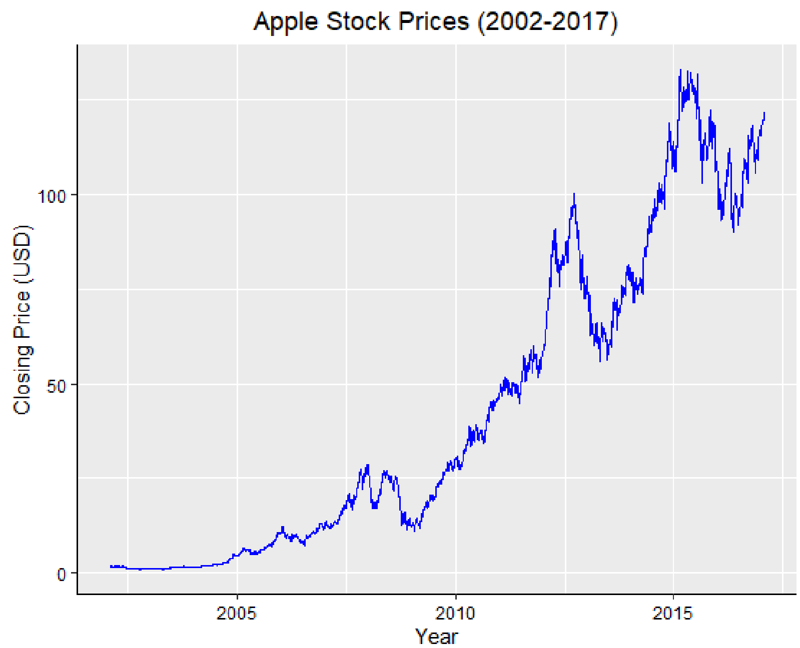


Figure 2: Time Series Plot of Original Data

Data from financial markets are commonly known to show ”volatility clustering”. This can also be observed in Figure 2, which appears to have large changes followed by large changes and small changes tend to follow small changes. This volatility clustering also statistically implies time-varying conditional variance, which is a property that can be captured by ARCH models. A non-linear upward trend can also be observed with a hint of higher variability at higher stock values.

2.2 ACF and PACF of Original Data

The ACF plot in Figure 3 also shows that the ACF dies down very slowly and does not decay to zero, which indicate that the data is non-stationary. With a p-value of 0.4114 from Augmented Dickey-Fuller (ADF) Test also shows that the null hypothesis that the data is non-stationary cannot be rejected.

```
Augmented Dickey-Fuller Test
data:  aapl.c
Dickey-Fuller = -2.3941, Lag order = 15, p-value = 0.4114
alternative hypothesis: stationary
```

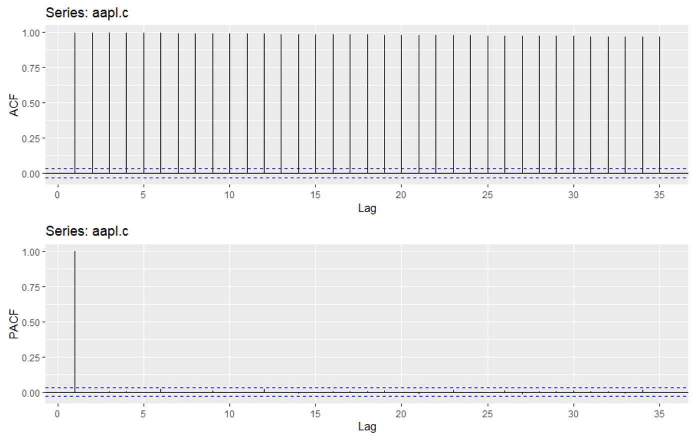


Figure 3: ACF and PACF Plots of Original Data

2.3 Seasonal Decomposition

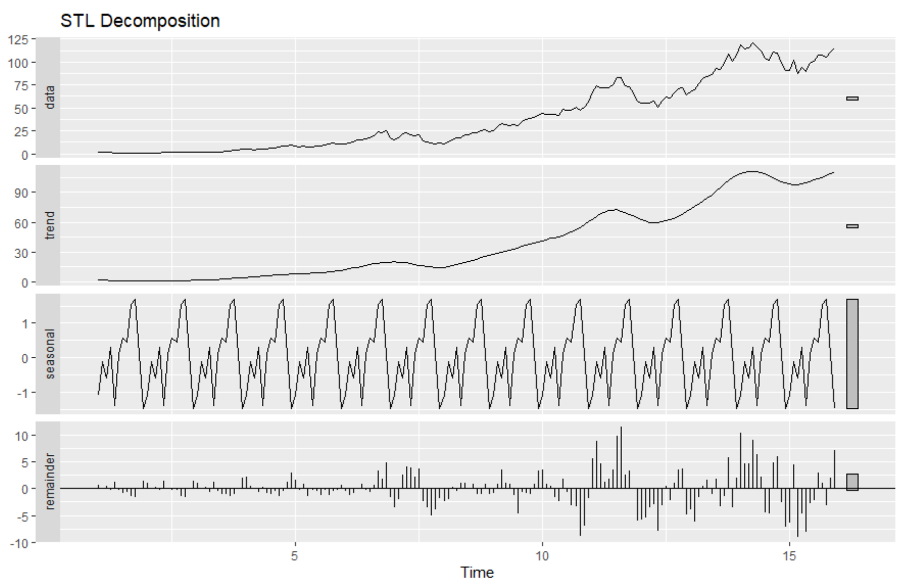


Figure 4: STL Decomposition of converted Monthly Data

To analyze the data using STL (Figure 4), the daily stock data is first converted to a monthly data. After which, a frequency of 12 is set so that STL can be used to decompose the seasonality in the data. It can be observed that there is indeed a need to perform log or BoxCox transformation on the data. Once again, a clear non-linear upward trend can be observed.

3. Data Transform

3.1 Train-Test Split

The data of 3,776 observations are split into training and test set. Due to the high variance of the data set, the test set is preset to 30 days. The training set will be used for fitting the model and the testing set to be used for calculating the accuracy of the model's forecasts. As a result of the split, the training set contains 3,746 observations, while the test set contains the remaining 30 observations.

3.2 Log Transform

Box-Cox transformation is to stabilize the variance of the data. Using a zero lambda will mean a pure logarithmic function applied to the dataset.

3.3 Trend Differencing

As the data appears to have a trending component, one time differencing is applied to remove the trending component. This is done after performing logarithmic transformation on the data as described in Section 3.2.

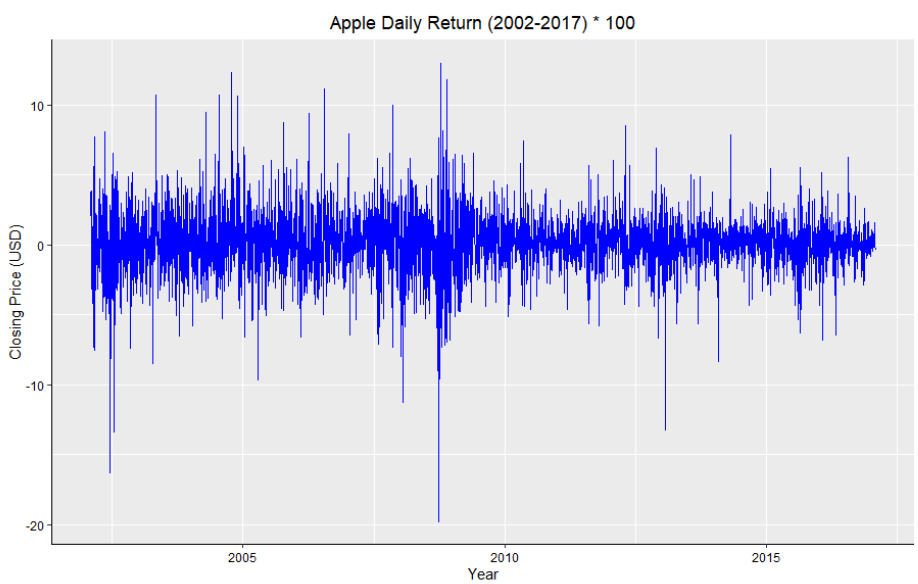


Figure 5: Time Plot after Transformation

The time plot after the data transformation multiplied by 100 can be seen in Figure 5. It can be interpreted as the percentage changes in the closing price. It also indicates that the conditional variance varies over time. The ADF test of p-value 0.01 (< 0.05) also shows the data is now likely stationary.

3.4 ACF and PACF of Transformed Data

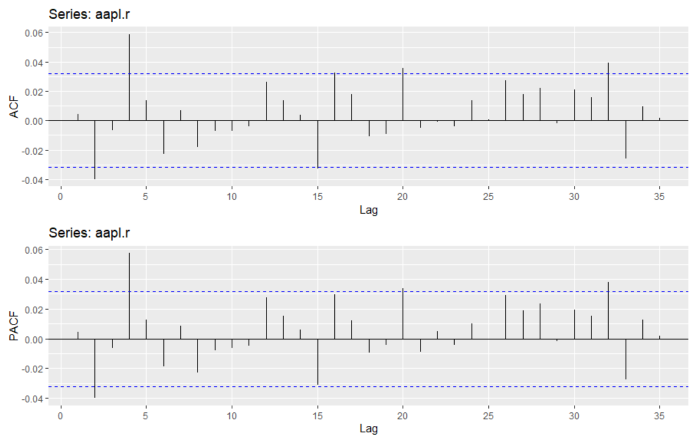
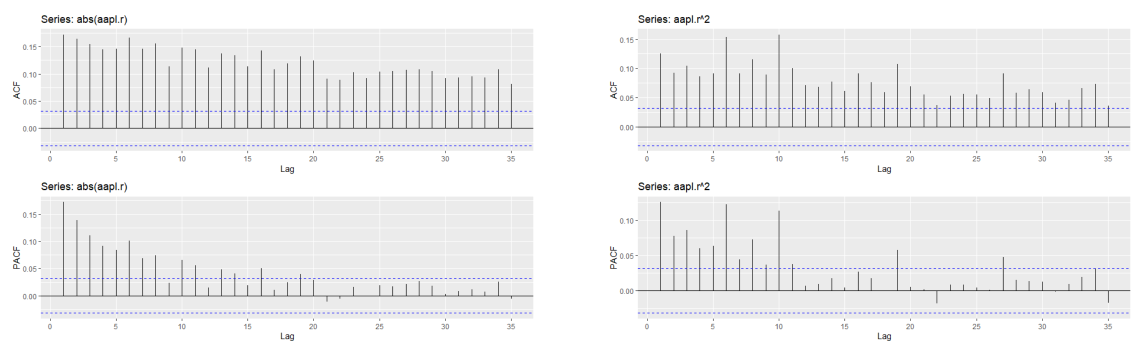


Figure 6: ACF and PACF Plots of Transformed Data



(i) Absolute Data

(ii) Squared Data

Figure 7: ACF and PACF Plots of Absolute or Squared Data

The ACF and PACF plots in Figure 6 shows that the closing prices have little serial correlation. However, with the volatility clustering observed in the time plots, it indicates signs of non-constant variances. Therefore, the ACF and PACF plots of the absolute and squared returns were also plotted. Both sample ACF and PACF plots showed high level of correlation, and thus provided evidence that the returns are not independently and identically distributed.

3.5 QQPlot of Transformed Data

To explore the distributional shape of the AAPL returns, the QQ normal plot is plotted as shown in Figure 8. The QQ plot suggested a heavy-tailed distribution, whereby the distribution has a tail thicker than that of a normal distribution, and is somewhat skewed to the left. The thickness of the tail is measured by kurtosis, which is defined as:

$$E(x - EX)^4/\sigma^4 - 3$$

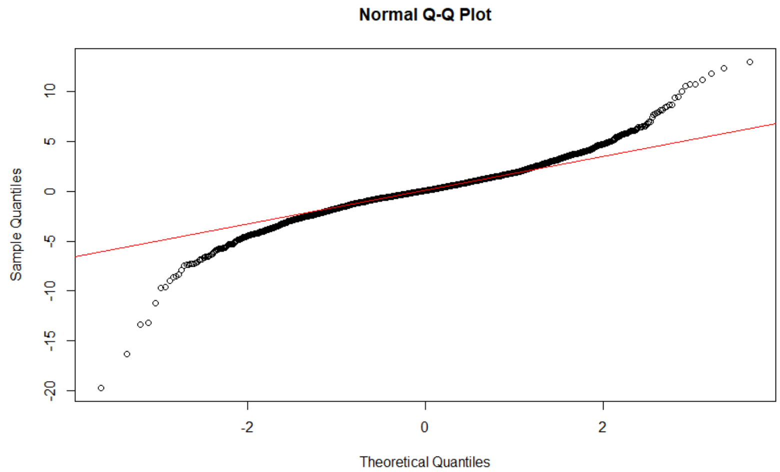


Figure 8: QQ Plot of Transformed Data

The sample kurtosis of the AAPL data returns a positive value of 5.440086. This positive kurtosis estimate further justifies a heavy-tailed distribution. The skewness of the AAPL data returns a negative value of -0.1901286. This negative skewness estimate indicates that the mean of the data values is less than the median, thus the data distribution is left-skewed.

In summary, the AAPL daily returns from 2012 to 2017 are found to be serially uncorrelated, but has a high level of volatility clustering, and a heavy-tailed and left-skewed data distribution. When a data exhibits these characteristics, we could use ARCH/GARCH model to fit the data.

4. GARCH Models

4.1 EACF

The sample EACF function is used to access the orders of the ARMA(p,q) models.

```
> eacf(aapl.r)
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
```

The sample EACF of the daily returns of the AAPL indicates a possible p=4 and q=0.

```
> eacf(abs(aapl.r))
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
```

The sample EACF of the absolute returns indicates a possible p=1 and q=1. Alternatively, p=2 and q=2 or p=3 and q=3 pairs seem to be viable as well.

```
> eacf(aapl.r^2)
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
```

The sample EACF of the squared returns indicates a possible p=1 and q=1.

4.2 Model Fitting

Given the results of the EACF analysis, several GARCH models were experimented. Table 1 shows the AIC values of the different GARCH models fitted.

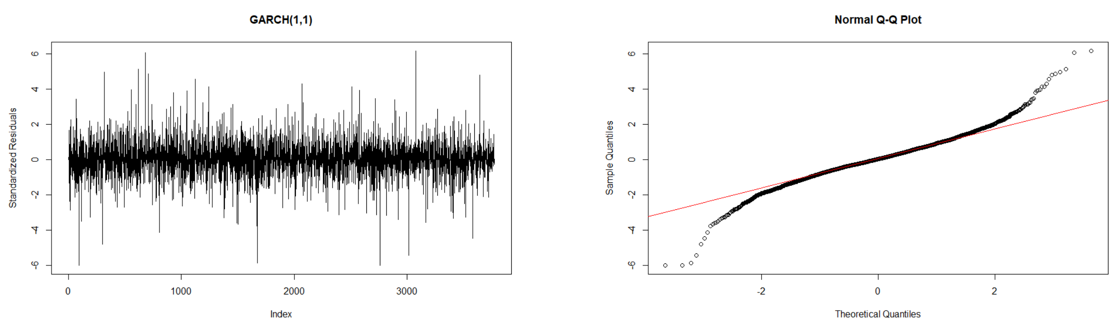
Models/ Data	aapl.r	appl.r *100
GARCH(4,0)	-17926.01	16806.18
GARCH(1,1)	**-18554.31**	16205.51
GARCH(2,2)	-18446.87	16577.53
GARCH(3,3)	-18472.86	16507.91
GARCH(2,1)	-18549.33	16201.29
GARCH(3,1)	-18512.5	**16195**

Table 1: AIC values of GARCH models using daily returns and daily returns * 100

The AIC results indicate that both GARCH(1,1) and GARCH(3,1) models provide a better fit over the other models.

4.3 GARCH(1,1) Diagnostic Check

The standardized residuals for GARCH(1,1) model and the QQ plot are plotted in Figure 9 respectively.

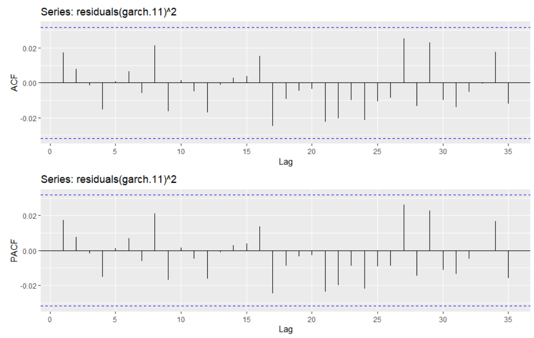


(i) Residuals from fitted GARCH(1,1) Model

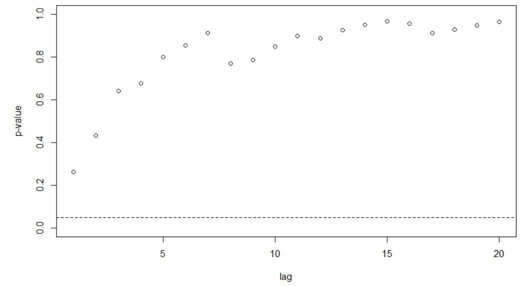
(ii) QQ Plot

Figure 9: Standardized Residuals Plots of GARCH(1,1) Model

As seen in Figure 12, the ACF plot suggests that the squared residuals of GARCH(1,1) model are serially uncorrelated. The p-values of the generalized Portmanteau test on the squared standardized residuals from the fitted GARCH(1,1) model are all higher than 0.05. This suggests that them are uncorrelated over time and hence the squared standardized residuals may be independent.



(i) Sample ACF and PACF

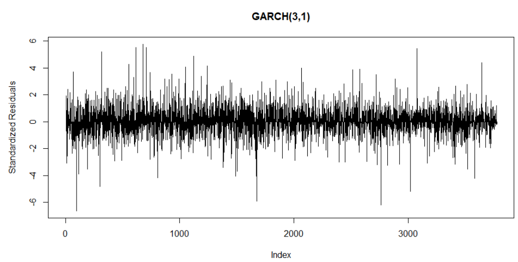


(ii) Generalized Portmanteau Test

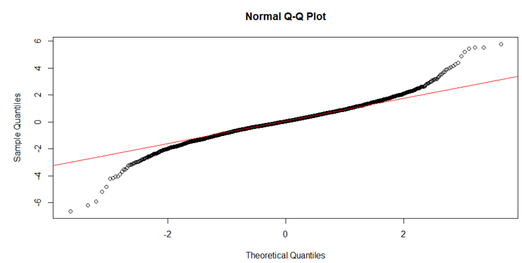
Figure 10: Squared Residuals of GARCH(1,1) Model

4.4 GARCH(3,1) Diagnostic Check

The standardized residuals for GARCH(3,1) model and the QQ plot are plotted in Figure 11 respectively.

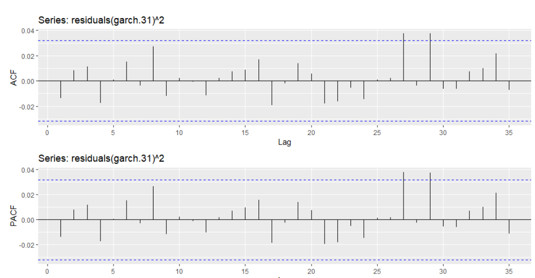


(i) Residuals from fitted GARCH(3,1) Model

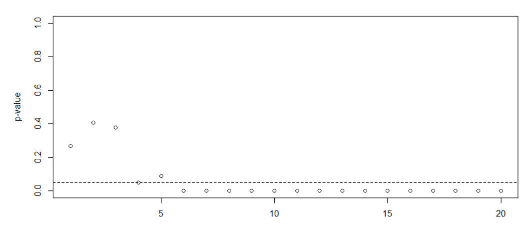


(ii) QQ Plot

Figure 11: Standardized Residuals Plots of GARCH(3,1) Model



(i) Sample ACF and PACF



(ii) Generalized Portmanteau Test

Figure 12: Squared Residuals of GARCH(3,1) Model

As seen in Figure 12, the p-values of the generalized Portmanteau test on the squared standardized residuals from the fitted GARCH(3,1) model are all lower than 0.05. This suggests that GARCH(3,1) may not be a good fit when the daily returns are not multiplied by 100.

4.5 GARCH Forecast Models

This section describes the two GARCH volatility forecasting packages used, namely fGarch and rugarch to find the most suitable GARCH model configuration for each package.

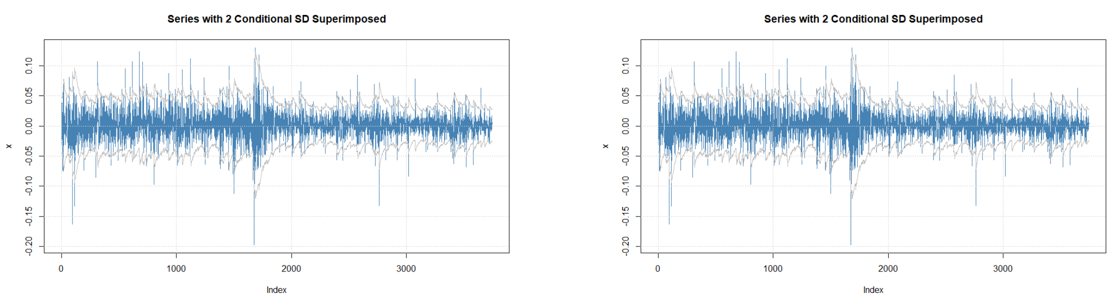
4.5.1 FGarch package Forecasting

Two GARCH models were fitted using garchFit function from the fGarch R package. The conditional distribution is set to "standardized Student-t distribution". Ljung-Box test was performed on both models' residuals, both have p-value $>$ 0.05.

```
fit.garch11 <- garchFit(formula = ~garch(1, 1),
                         data = aapl.r.train, trace = F, cond.dist = "std")
fit.garch21 <- garchFit(formula = ~garch(2, 1),
                           data = aapl.r.train, trace = F, cond.dist = "std")
```

Listing 2: R Codes to fit fGarch model

Figure 13 shows the series plots (with 2 conditional standard deviation superimposed) for the two Garch models fitted.



(i) fGarch(1,1)

(ii) fGarch(2,1)

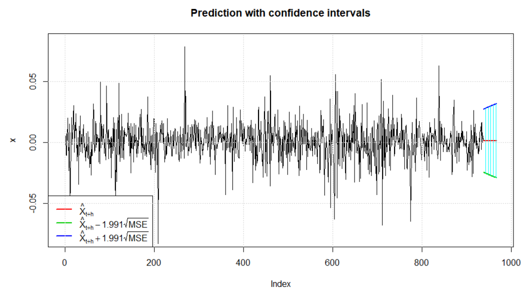
Figure 13: Series Plots with 2 Conditional SD Superimposed using fGarch package

Figure 14 shows the 30-days prediction plots (with confidence intervals) for the two Garch models fitted.

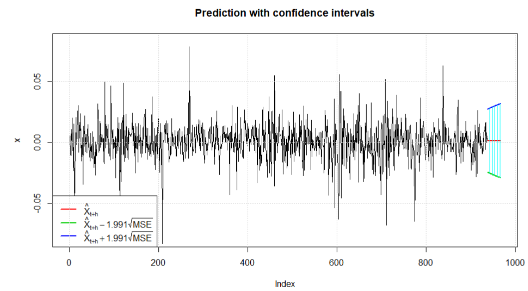
	RSME	**MAE**	**MPE**	**MAPE**
fGarch(1,1)	0.00504715	0.003774639	96.0805	224.1855
fGarch(2,1)	0.005047115	0.003774817	96.07794	224.293

Table 2: Forecast Accuracy of fGarch models on test set

As seen in Table 2, fGarch(2,1) has a slightly better forecast accuracy on the test set than fGarch(1,1) with a lower RSME and MPE value.



(i) fGarch(1,1)



(ii) fGarch(2,1)

Figure 14: Forecast Plots using fGarch package

4.5.2 ruGarch package Forecasting

Table 3 shows the various GARCH models fitted using the ruGarch package. sGARCH denotes standard Garch models, while gjrGARCH denotes the GJR GARCH model of Glosten et al. (1993). fGARCH stands for family GARCH model of Hentschel (1995) with Threshold Garch (TGARCH) model of Zakoian as submodel. eGARCH model of Nelson (1991) is a exponential Garch model. They are all fitted using ARMA(0,0) and GARCH(1,1) orders.

rugarchspec	**Distribution**	**Variance Targeting**	**Coefficients**	**Likelihood**	**Akaike**
sGARCH(1,1)	std	F	6	9472.777	-5.0158
sGARCH(1,1)	std	T	5	9471.963	-5.01614
gjr GARCH(1,1)	std	F	7	9482.752	-5.02027
fGARCH(1,1) -TGARCH	std	F	7	9494.399	-5.02697
eGARCH(1,1)	std	F	6	**9496.096**	**-5.02787**

Table 3: GARCH models fitted using ruGarch package

The likelihood of the sample is based on the product of all predicted densities by the GARCH models. It measures how likely it is to that the observed returns that come from the estimated GARCH model. The higher the likelihood, the better the model fits with the Apple Daily Return data. A GARCH model is parsimonious when it has a high likelihood and a relatively low number of parameters (coefficients). Furthermore, a model that has the lowest information criterion (Akaike) should be the preferred GARCH model to be used. As shown in Table 3, eGARCH(1,1) has a relatively low number of coefficients, the highest likelihood as well as the lowest Akaike value, hence can be said to be the best fit among the rest. Diagnostic checking on all the fitted models' residuals using Ljung-Box test have their p-values more than 0.05 as shown in Table 4.

rugarchspec	**Ljung-Box Test**
sGARCH(1,1)	0.173
sGARCH(1,1)	0.1375
gjrGARCH(1,1)	0.6314
fGARCH(1,1)-TGARCH	0.6208
eGARCH(1,1)	0.644

Table 4: GARCH models Ljung-Box Test p-values

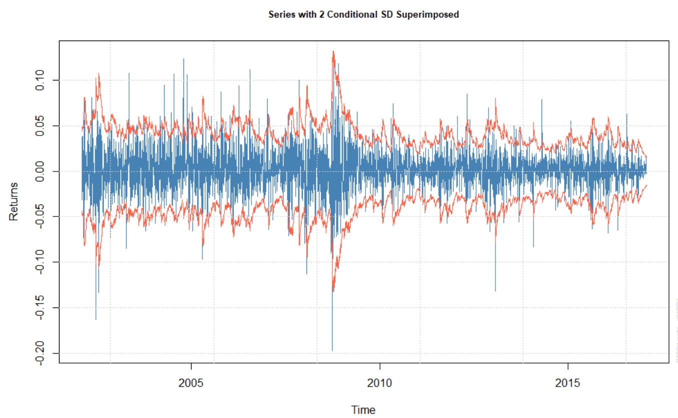
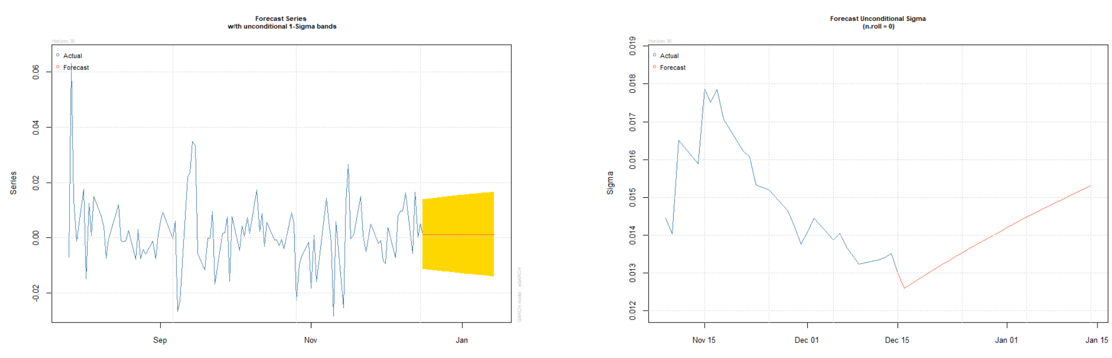


Figure 15: Series Plot of eGARCH(1,1) using ruGarch package



(i) Unconditional Series

(ii) Unconditional Sigma

Figure 16: Forecast Plots of eGARCH(1,1) using fGarch package

Figure 15 shows the series plot (with 2 conditional standard deviation superimposed) for the eGARCH(1,1) model fitted. Figure 16 shows the 30-days forecast series with Unconditional 1-Sigma and forecast unconditional sigma plots of a fitted eGARCH(1,1) model. Using an n-roll configuration, the e-Garch(1,1) fitted model will produce volatility forecasting graphs shown in 17.

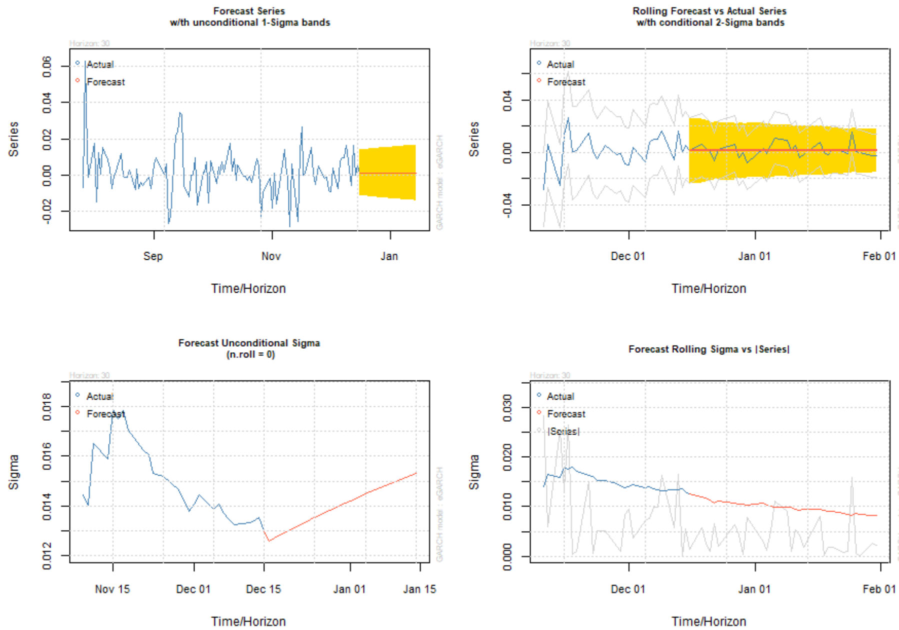


Figure 17: Forecast Plots of eGARCH(1,1) model with N-Roll

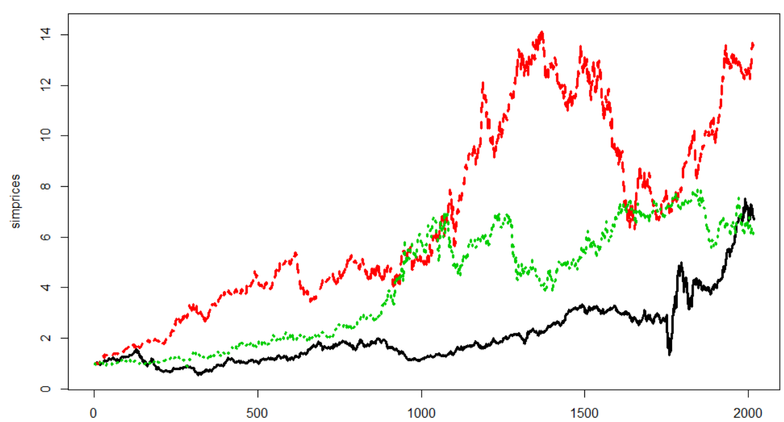


Figure 18: Three 8 year simulations of eGARCH(1,1) model using ruGarch package

The ruGarch package also allows simulation of GARCH models. Figure 18 shows three 8-years simulation plots of eGARCH(1,1) model using ruGarch package. All three simulation plots show a upward trend.

5. Conclusion

The goal for this project is to analyze Apple's financial stock data from February 1, 2002 to January 31, 2017. After retrieving the dataset from Yahoo Finance, the dataset was first examined by plotting a time plot. The timeplot showed volatility clustering, which statistically implied time-varying conditional variance. Furthermore, a non-linear upward trend can also be observed. By looking at the ACF plot, it also indicated that the stock data was non-stationary. Hence, both log transformation and differencing were applied on the close prices to make the data stationary. Revisiting the ACF and PACF plots of the absolute and squared data provided evidence that the returns were not independently and identically distributed. Furthermore, analyzing the kurtosis and skewness estimate also revealed that the data was heavy-tailed and left-skewed.

After close examination of the dataset, it is clear that a ARCH/GARCH model will be suitable to fit this financial dataset due to its volatility clustering. Using the EACF function, it provided a few possible candidate p and q values for the models. After fitting them using t-series GARCH, The GARCH(1,1) and GARCH(3,1) were found to have the lowest AIC values. However, after diagnostic checking, GARCH(3,1) was deemed as not suitable fit and GARCH(1,1) model was a more suitable fit for the dataset.

Two R GARCH packages were explored for this assignment, namely fGarch and ruGarch. Using the fGarch package, fGarch(1,1) was the model that had the best forecast accuracy among the other models fitted. On the other hand, the ruGarch package provides a variety of different GARCH models to try. The exponential GARCH model (eGARCH(1,1)) was found to be the best fit based on likelihood and information criterion.