Assignment 3

AI6123 Time Series Analysis

Junyu Yin
G2101985L

1 Data Examination

This assignment aims to design and develop the best model for forecasting the prices of Apple stock. The given data comprises of the open, close, high and the adjusted close prices of the stock from February 1, 2002 to January 31, 2017 (See Fig. 1).

Figure 1: Chart Series of AAPL.

1.1 Original Time Series Analysis

The close price of a stock is relatively important among all types of prices, so I choose to build a model to fit it. Fig. 2 shows the time series plot of this data. From this figure, we can observe that there is a generally non-linear increasing trend with a hint of higher variability with higher level of the stock value.

Moreover, it appears to have large changes followed by large changes and small changes tend to follow small changes. This phenomenon, called "volatility clustering", is commonly observed in data from financial markets. Volatility in a time series also refers to the phenomenon where the conditional variance of the time series varies over time.

It can be seen clearly that the data is non-stationary. We can further check it by plotting ACF/PACF plots and by doing unit root tests. The ACF plot in Fig. 3(a) shows that the ACF dies down very slowly and does not decay to zero. And

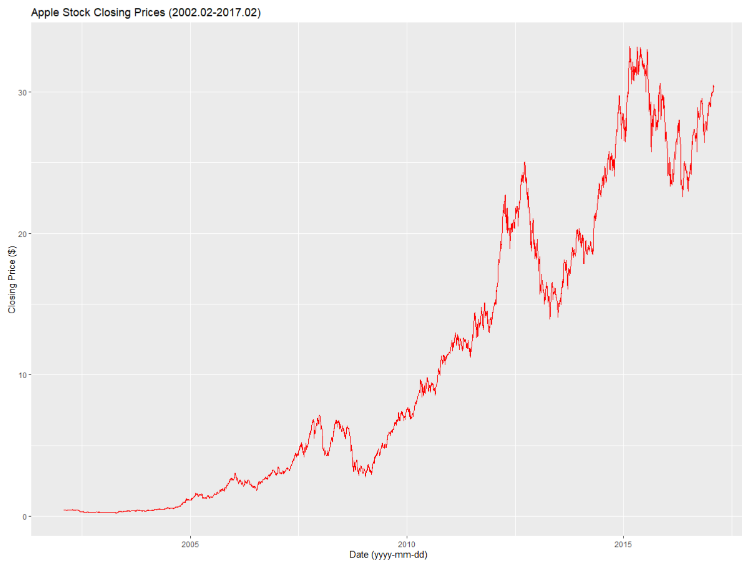
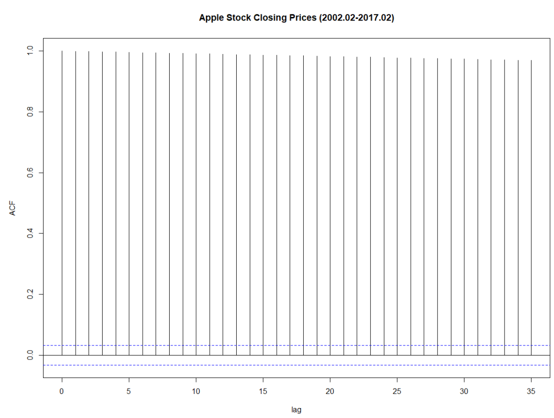
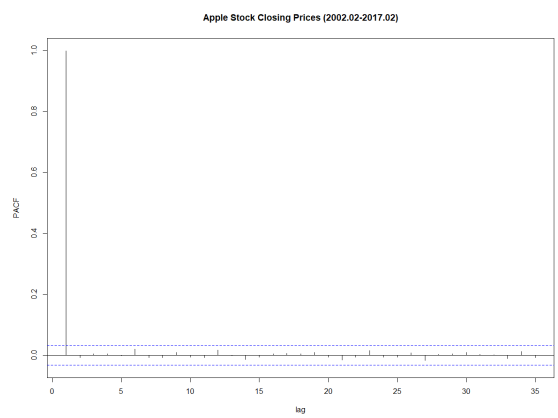


Figure 2: The close prices of Apple stock.



(a) ACF



(b) PACF

Figure 3: ACF/PACF plots of the original data.

the ADF test, KPSS test and PP test all indicate that the data is non-stationary. As a result, we need to perform some transformations to make the data stationary.

```
        Augmented Dickey-Fuller Test

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

        KPSS Test for Trend Stationarity

data:  aapl.c
KPSS Trend = 5.4744, Truncation lag parameter = 9, p-value = 0.01

        KPSS Test for Level Stationarity

data:  aapl.c
```

```
KPSS Level = 34.617, Truncation lag parameter = 9, p-value = 0.01

        Phillips-Perron Unit Root Test

data:  aapl.c
Dickey-Fuller Z(alpha) = -9.7182, Truncation lag parameter = 9, p-value = 0.5678
alternative hypothesis: stationary
```

1.2 Data Transformation

According to the common practice in financial time series, I apply the logarithmic transformation first and take a first-order difference after. This gives the log-returns of the original data. What's more, the log-returns are then multiplied by 100 so that they can be interpreted as percentage changes in the price. The multiplication may also reduce numerical errors as the raw returns could be very small numbers and render large rounding errors in some calculations.

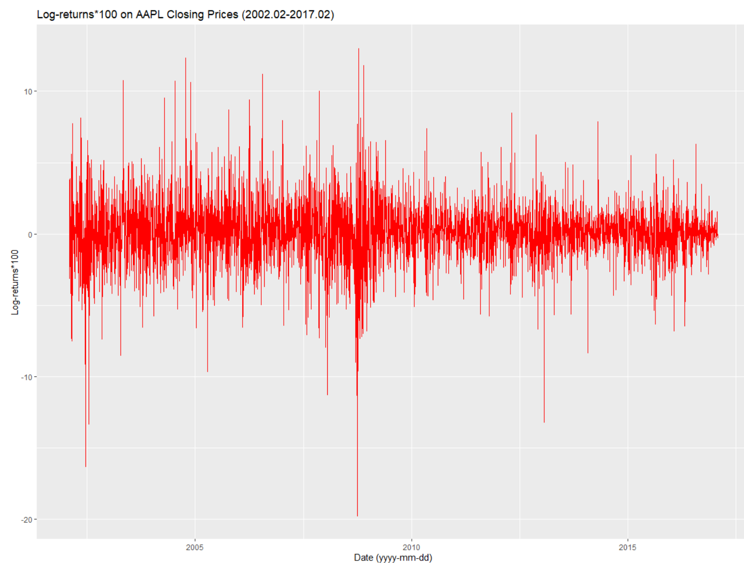
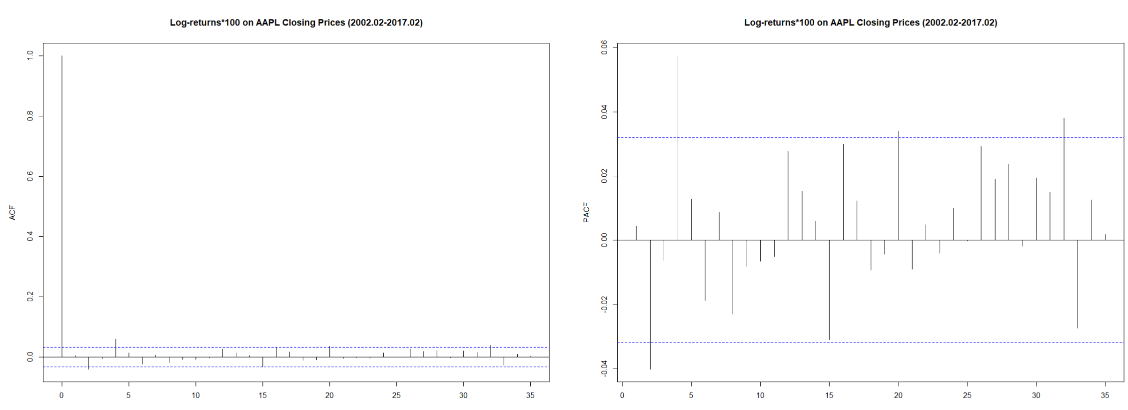


Figure 4: The log-returns (multiplied by 100) of Apple stock.



(a) ACF (b) PACF

Figure 5: ACF/PACF plots of the transformed data.

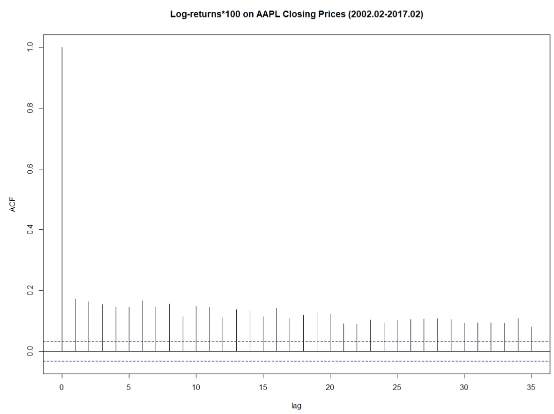
The results of the unit root tests (see Tab. 1) now suggest the log-returns (multiplied by 100) are likely stationary.

Test Type	p-value
ADF Test	0.01
KPSS Trend Test	0.1
KPSS Level Test	0.1
PP Test	0.01

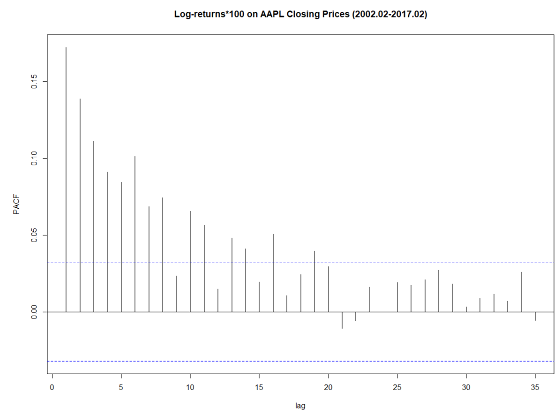
Table 1: The results of unit root tests.

The sample ACF and PACF of the log-returns (multiplied by 100) shown in Fig. 5, suggest that the log-returns have little serial correlation at all. But the volatility clustering observed in the data (See Fig. 4) gives us a hint that they may not be independently and identically distributed, otherwise the variance would be constant over time.

It can be further examined by looking at the ACF and PACF of the absolute log-returns and those of the squared log-returns. From Fig. 6 and Fig. 7, some significant autocorrelations can be observed and hence provide some evidence that the log-returns are not independently and identically distributed.

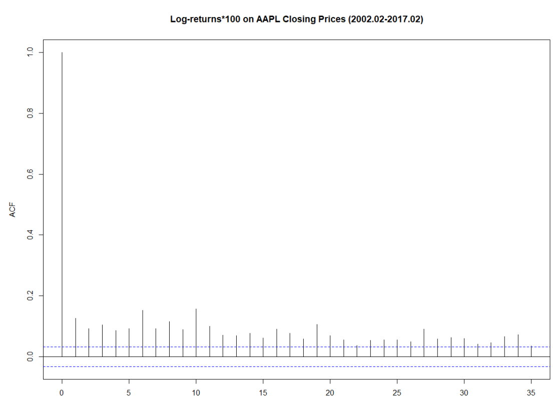


(a) ACF

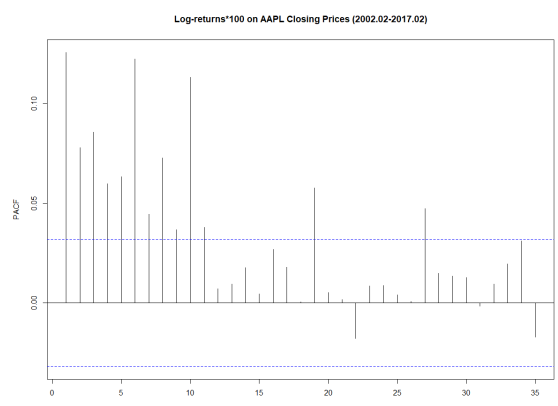


(b) PACF

Figure 6: ACF/PACF plots of the absolute data.



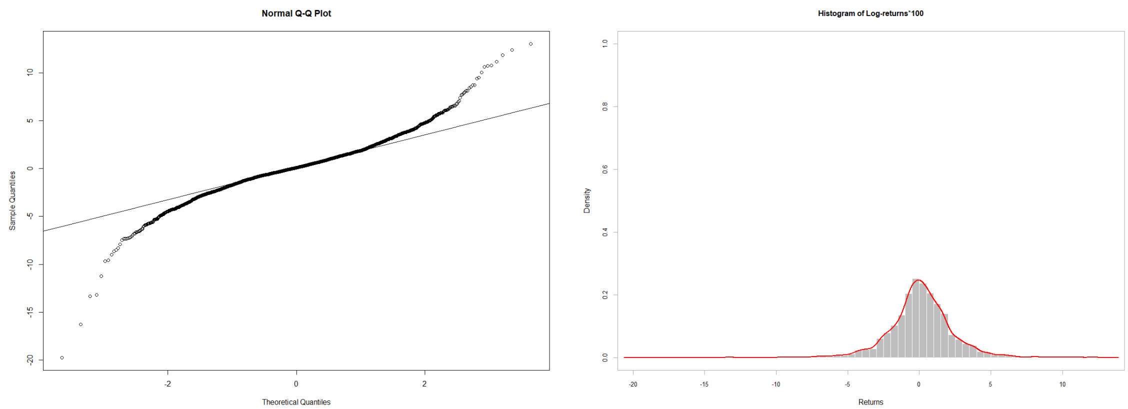
(a) ACF



(b) PACF

Figure 7: ACF/PACF plots of the squared data.

The distributional shape of the log-returns can be explored by constructing a QQ normal scores plot (see Fig. 8). The QQ plot suggests that the distribution of the log-returns may have a tail thicker than that of a normal distribution and may be somewhat skewed to the left. The sample kurtosis returns a positive value of 5.440092 and the sample skewness returns a negative value of -0.1901295. The former justifies a heavy-tailed distribution and the latter indicates the distribution is left-skewed. To give a better sense of this, I also show the histogram of the log-returns.



(a) QQ plot of the transformed data. (b) Histogram of the transformed data.

Figure 8: Data distribution exploration.

In summary, the log-returns of Apple stock are found to be serially uncorrelated but admit a higher-order dependence structure, namely volatility clustering, and a heavy-tailed left-skewed distribution. It is commonly observed that such characteristics are rather prevalent among financial time series data. The GARCH models provide a framework for modeling and analyzing time series that display some of these characteristics.

2 Model Selection

2.1 Train-Test Split

The log-returns contain 3775 observations and are split into training and test sets. Owing to the high variance of the data set, we only hold 30 data points for test set. The training set will be used to fit the model and the test set is for checking the forecasting ability.

2.2 EACF

The sample EACF of the log-returns is used to select the orders of a ARMA(p, q) model.

```
> eacf(aapl.logr.train)
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 result above indicates a ARMA(0, 4) model is possible for the log-returns.

What's more, it can be applied to the squared log-returns series and to the absolute log-returns series to access the orders of a GARCH(p, q) model.

```
> eacf(aapl.logr.train^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 result above indicates ARMA(p, p) model (p > 0) are all possible for the squared log-returns.

```
> eacf(abs(aapl.logr.train))
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 result above indicates ARMA(1, 1), ARMA(2, 2) and ARMA(3, 3) are possible for the absolute log-returns.

Since an ARMA(max(p, q),p) model for the squared log-returns implies the GARCH(p,q) model for the original log-return series, we can choose from a ARMA(0, 4) model and a bunch of GARCH models.

2.3 Model Fitting

Given the results of the EACF analysis, I try the models in the following table.

Model Type	AIC
ARMA(0,4)	16691.97
GARCH(1,1)	**16133.41**
GARCH(1,0)	16548.67
GARCH(2,2)	16479.84
GARCH(2,1)	16495.32
GARCH(2,0)	16452.37
GARCH(3,3)	16412.57
GARCH(3,2)	16430.2
GARCH(3,1)	16446.18
GARCH(3,0)	16392.55

Table 2: AIC values of fitted models.

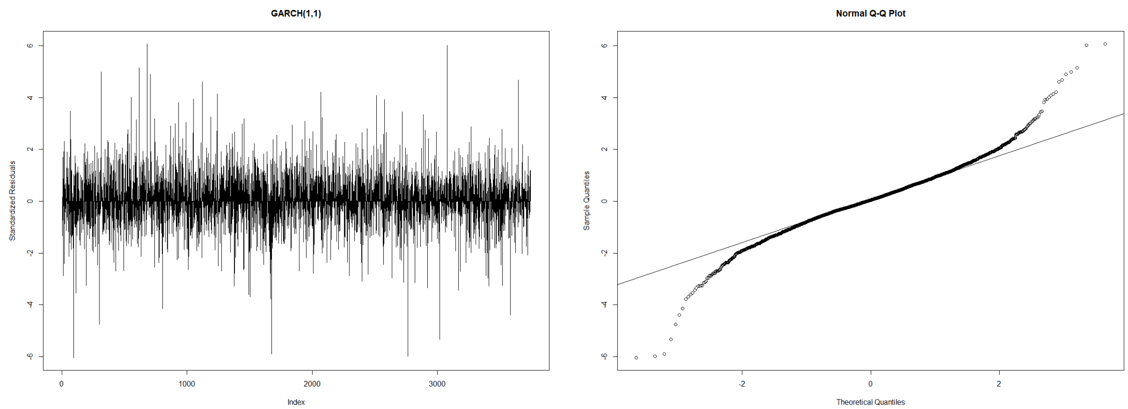
The AIC results indicate the GARCH(1,1) provides the best fit among all the models.

2.4 Model Diagnostic

Fig. 9 shows the standardized residuals for GARCH(1,1) and the QQ plot. The sample ACF of the squared residuals of GARCH(1,1) (see Fig. 10) shows they are serially uncorrelated. And the p-value of the Generalized Portmanteau Test are all higher than 0.05. This suggests that we can accept GARCH(1,1) as our fitted model.

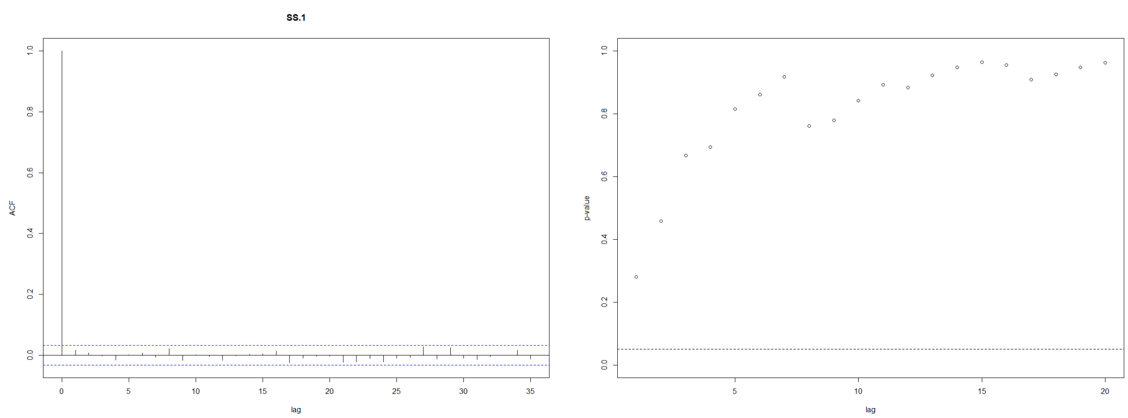
2.5 Model forecasting

In this section, I use a GARCH forecasting package in R, namely fGarch, to do the forecasting.



(a) Residuals (b) QQ plot

Figure 9: Residuals and QQ plots for GARCH(1,1).



(a) ACF of the squared residuals (b) Generalized Portmanteau Test

Figure 10: Diagnostic Checking for GARCH(1,1).

As seen in Tab. 3, GARCH(1,1) has the slightly better RMSE and MPE, while GARCH(2,2) has the slightly better MAE and MAPE. However, their forecasting performances are about the same, we shoulld still choose the GARCH(1,1) model since it has the best fitting performance.

Model Types	RMSE	MAE	MPE	MAPE
GARCH(1,1)	**0.504715**	0.3774639	**96.05282**	224.1906
GARCH(2,2)	0.504744	**0.3773173**	96.07412	**223.3045**
GARCH(3,3)	0.5047166	0.3774556	96.05403	224.1403

Table 3: Forecast errors of various GARCH models using the fGarch package.

Fig. 11 shows the 30-days prediction plots (with confidence intervals) for the GARCH(1,1).

3 Conclusion

In this assignment, I analyze the close prices of the Apple stock from February 1, 2002 to January 31, 2017. The time plot of the original data shows the volatility clustering phenomenon and the ACF/PACF plots further show that this data is non-stationary. Then I take the log transformation and the first-order difference to make it stationary. The resulting data is called log-returns in financial analysis. By looking at the ACF/PACF plots of the squared and absolute log-returns, we can see that the log-returns are not independently and identically distributed.

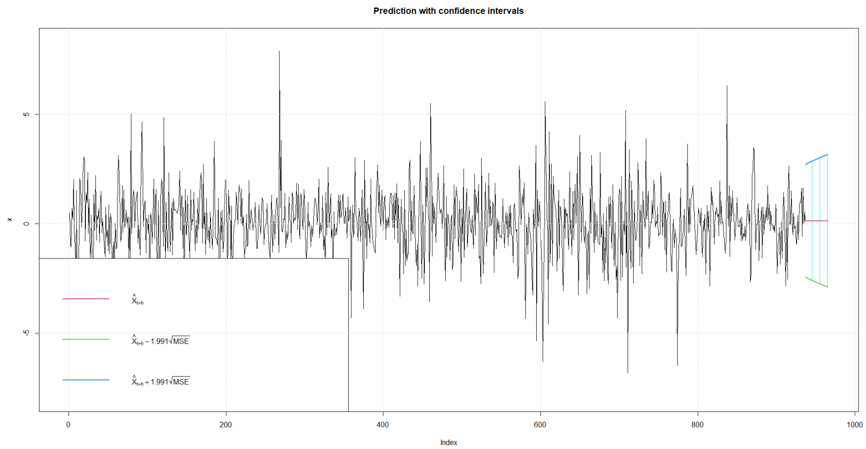


Figure 11: Forecasting plot for GARCH(1,1) using the fGarch package.

After the close examination of the dataset, I decide to use a GARCH model to fit this stock prices owing to its volatility clustering. By checking the EACF results, a bunch of possible GARCH models are shown to be the candidates. After examining them one by one, the GARCH(1,1) model is found to have the lowest AIC values. The diagnostic checking further indicates it is a suitable model for the data here.

At last, I use a R package, namely fGarch, to do the prediction. The forecasting performances of various GARCH models are nearly same. As a result, the GARCH(1,1) is the most appropriate model here since it has the best fitting performance.