AI6123 Time Series Analysis Assignment 3

Liu Yaohong(G2203319C)

Introduction

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.

Original Data Analysis

First, we load AAPL data and plot the adjusted close price, ACF and PACF of the AAPL data, which is shown in Figure 1.

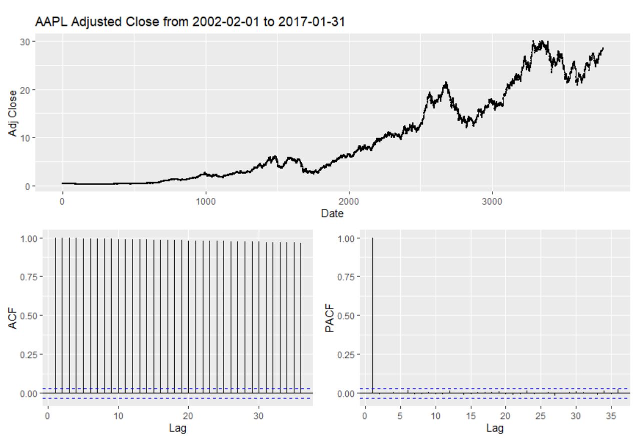

Figure 1: Original data plot

The minimum value of the data is 0.1994051 and the maximum value is 30.02323. From the plotting, on the one hand, a non-linear trending is going up, and the mean of the data is change with volatile period. On the other hand, we can see the variance of the data is changing over different time period. From the ACF plot, AAPL data dose not cut off or die down, even though PACF is cutting off after lag 1.Hence, ADF test and KPSS is applied to the data.

```
        Augmented Dickey-Fuller Test

data:  dataset$AAPL.Adjusted
Dickey-Fuller = -2.2209, Lag order = 15, p-value = 0.4848
alternative hypothesis: stationary
```

```
        KPSS Test for Trend Stationarity

data:  dataset$AAPL.Adjusted
KPSS Trend = 6.0844, Truncation lag parameter = 9, p-value = 0.01
```

From the result of Augmented Dickey–Fuller(ADF) test, the p-value is 0.4848 which is bigger than 0.05, it means we cannot reject the null hypothesis that a unit root is presented in data and the data is non-stationary. The p-value of KPSS test is 0.01 which is lesser than 0.05. Therefore, we reject the alternative hypothesis, the time series data is not trend stationary.

We then use the *stl()* function to decomposite the seasonality of the original data. We can see a more obvious nonliner trend and a clear seasonal component with increasing variance in Figure 2.

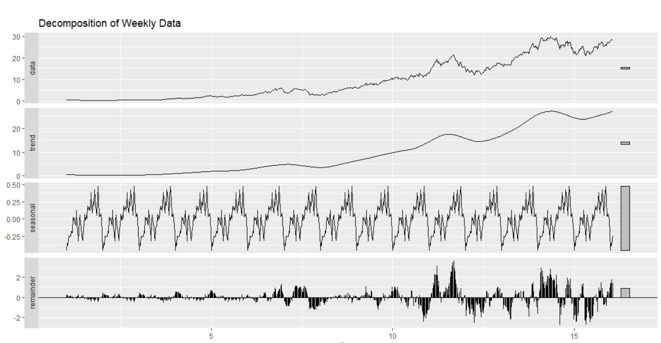

Figure 2: STL decomposition

Transformation

After original data analysis, we know the AAPL data contains mean and variance changing over the time. Log transformation and 1st order difference is implemented. For clear visualization, the data is multiplied by 100 and it can be interpreted as the price changing percentage (shown in Figure 3). Hence, we fit the log-return data in to ADF and KPSS test. the p-value of ADF test is 0.01 and the p-value of KPSS test equals 0.1. Based on the results of both test that the data is stationary.

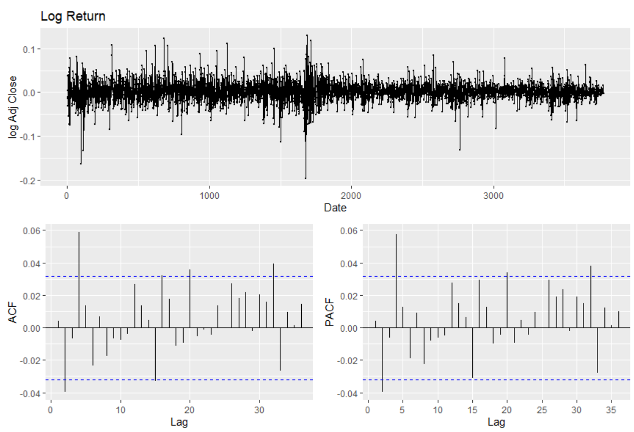

Figure 3: Log-return Data

Furthermore, we have to investigate the absolute value and the square of the log-return data, which are shown in Figur 4 and Figure 5 respectively. From the sample ACR and PACF, we know the data of log-returns is not indecently and identically distributed(*i.i.d*). Therefore, we plot the log-return data by using QQ-plot. From Figure 6, the data is heavy-tailed distribution whihc has thick tail with skewing to left. After applying *basicStats()* function, the skewness of log-return data is -0.192333. skewness value is smaller than 0 which means the distribution is left-skewed. The kurtosis equals to 5.441402. the posive kurtosis value tell us the data is heavy tailed. Thus, to modeling this pattern data GARCH model is a suitable choice.

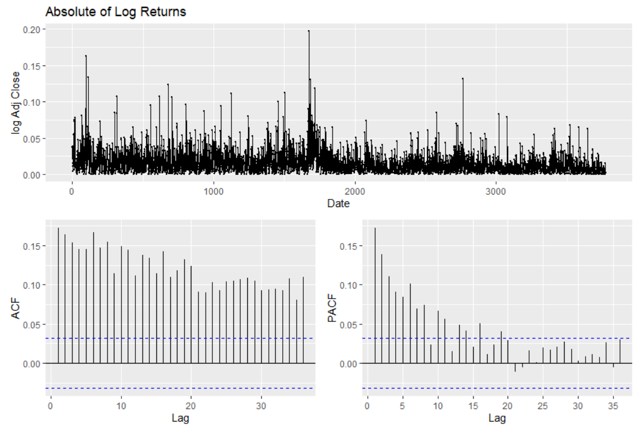

Figure 4: Absolute of Log-return Data

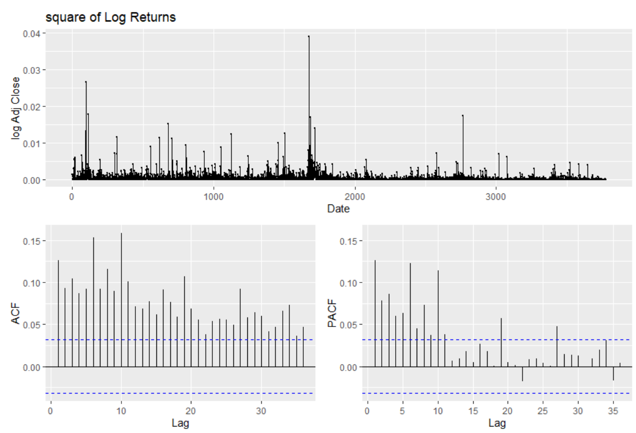

Figure 5: Square of Log-return Data

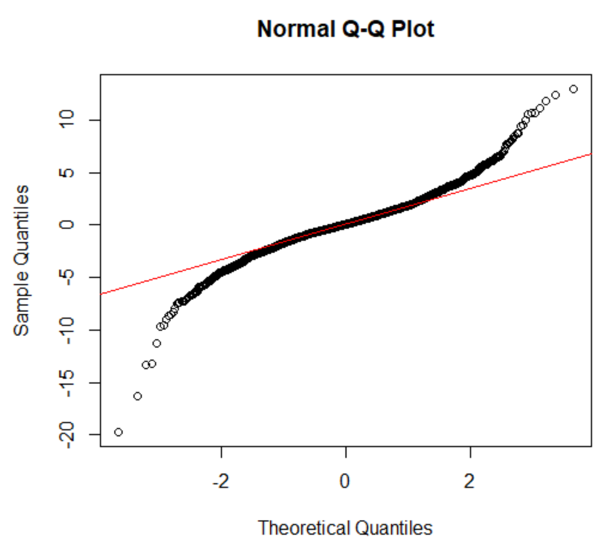

Figure 6: QQplot of Log-return Data

GARCH Model Fitting

First of all, we use *EACF()* function to get the optimal *max(p,q)* value of ARMA model. the EACF of log return data, absolute log return data and log return square data are shown in Figure7,Figure8 and Figure9.

```
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 7: EACF of Log-return Data

```
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 x 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 8: EACF of Absolute Log-return Data

```
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 9: EACF of Square Log-return Data

From Figure7, by choosing parameter *(p,q) = (0,2)*, the triangle contains x is less than 5 percent, which is acceptable. From Figure8, the best parameter is *(p,q) = (1,1)*. The EACF of square Log return data contains too many 'x', *(p,q) = (5,9)* is suitable parameter. Since from Figure7 and Figure8, we have better choices, we omit the *(p,q) = (5,9)*. Then we fit the data into GARCH(1,1) and GARCH(0,2) model. By implementing the AIC, AIC of GARCH(1,1) = -18556.44 and AIC of GARCH(0,2) = -18205.25. Therefore, GARCH(1,1) has the better fit to our data.Then we plotting the residuals of GARCH(1,1) model and the QQ-plot,which are shown in 11 and 11 respectively.

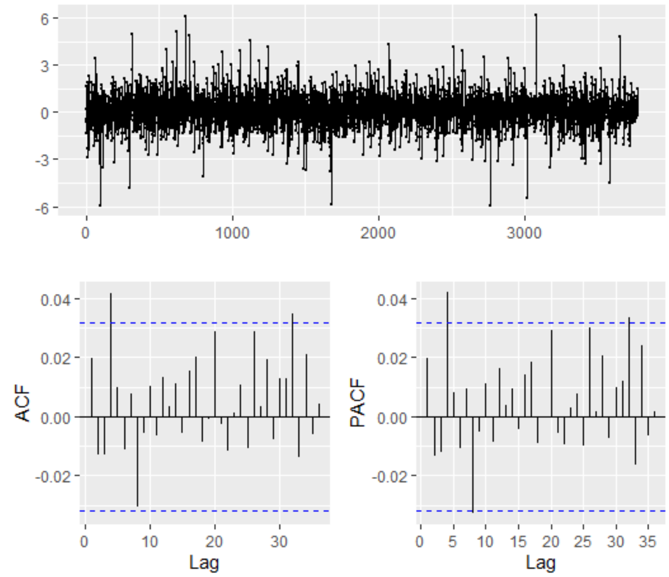

Figure 10: Residual Plot of GARCH(1,1)

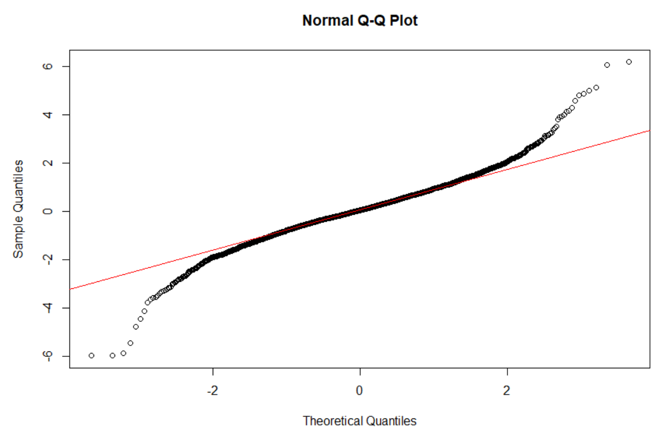

Figure 11: QQ-Plot of GARCH(1,1)

Even though there are 2 lags is out of spec after lag 30 on ACF and PACF plot of 11, which are acceptable. The Ljung–Box test p-value of GARCH(1,1) model squared residuals is 0.2859, which means the squared residuals are *i.i.d.* And we use *gBox()* to diagnostic check GARCH(1,1) model, the result shown in Firgure12. All the p-values are greater than 0.05, which means all sqaured of residuals are not correlated. Therefore, the standard residual may be also *i.i.d.* The residuals plot of squared residuals is shown in 13, and all the lag of APC and PACF are within the boundary.

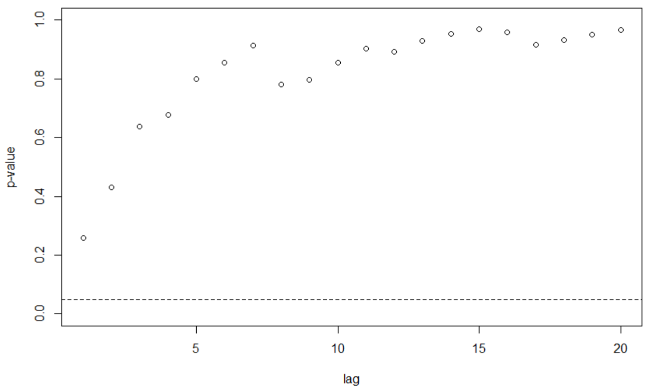

Figure 12: Generalized Portmanteau Tests of GARCH(1,1)

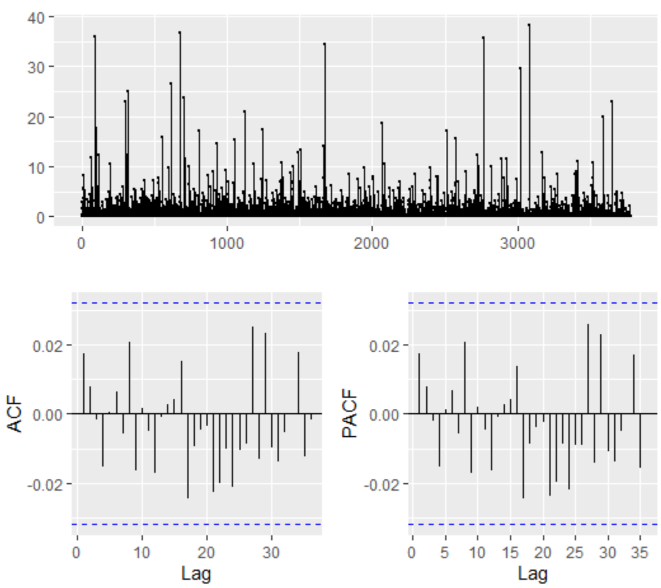

Figure 13: Square Residual Plots of GARCH(1,1)

We fit various distribution into standard GARCH(1,1) model to get the best fitted distribution of AAPL data set.The AIC of different distributions are shown in Table1 and T-Distribution has the best AIC score.

Distribution	AIC
Normal Distribution	-4.9238
Skew Normal Distribution	-4.9234
T-Distribution	**-5.0171**
Skew T-Distribution	-5.0167
Generalized Error Distribution	-5.0048
Skew Generalized Error Distribution	-5.0049
Normal Inverse Gaussian Distribution	-5.0144
Generalized Hyperbolic Distribution	-5.0161

Table 1: AIC of Different Distributions

Similar to previous step, we using T-Distribution fit into various GARCh model. From Tabel2, the best model is eGARCH(1,1) model.

Model, Submodel	AIC
fGARCH	-5.0171
fGARCH, TGARCH	-5.0282
fGARCH, AVGARCH	-5.0280
fGARCH, NAGARCH	-5.0246
fGARCH, APARCH	-5.0277
fGARCH, GJRGARCH	-5.0219
fGARCH, ALLGARCH	-5.0278
eGARCH	**-5.0291**
gjrGARCH	-5.0219
apARCH	-5.0277
iGARCH	-5.0174
csGARCH	-5.0233

Table 2: AIC of Different GARCH Model

Forecasting

To monitor model performing, we use last 30 days data as testing data. At same time, we are forecasting the 30 days ahead from 1/Feb/2017. The yellow area represents 95 percents confidence where the real data will fall on.

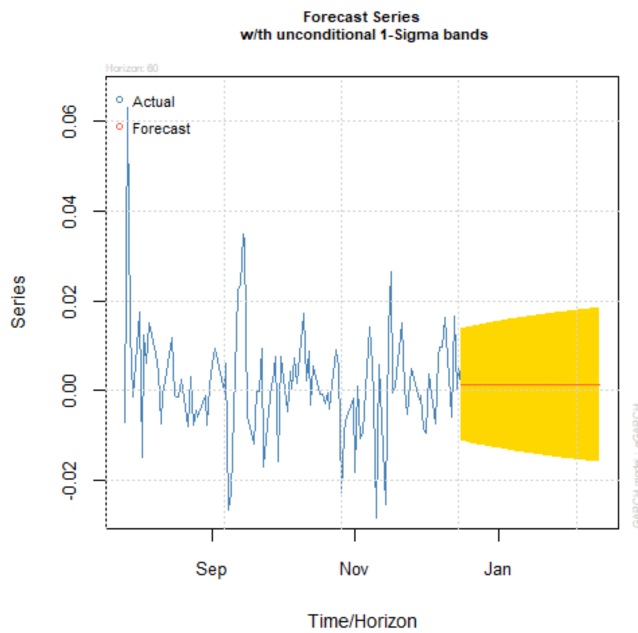

Figure 14: Unconditional 1-Sigma bands

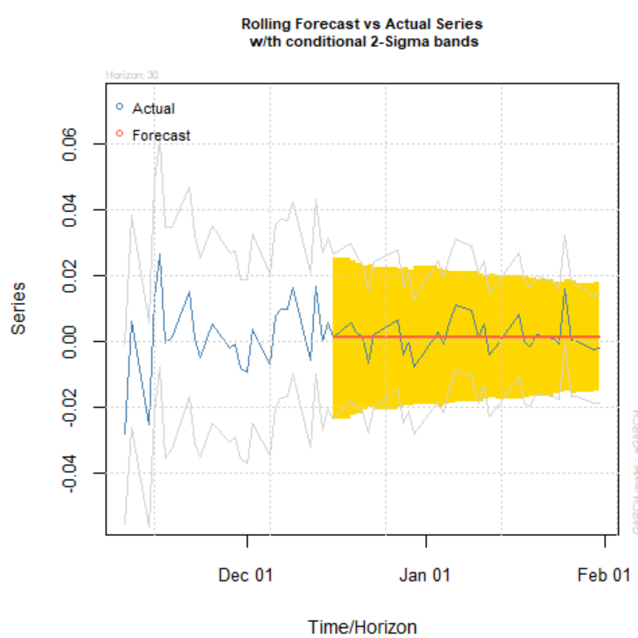

Figure 15: Conditional 2-Sigma bands

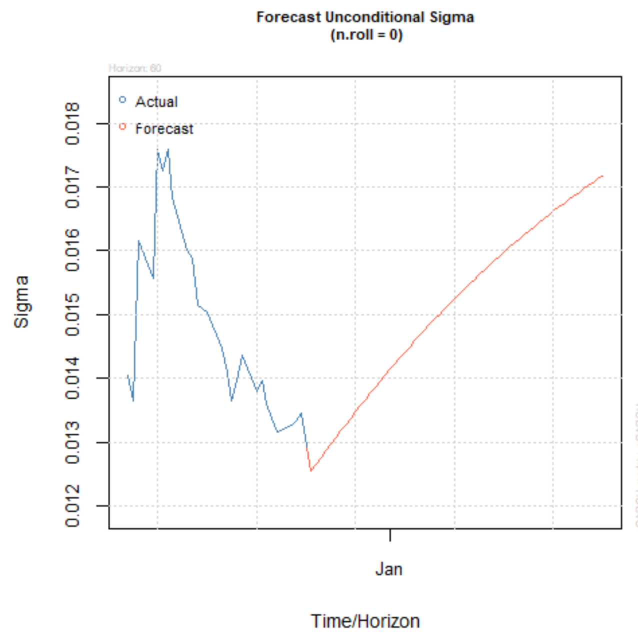

Figure 16: Unconditional Sigma (n.roll=0)

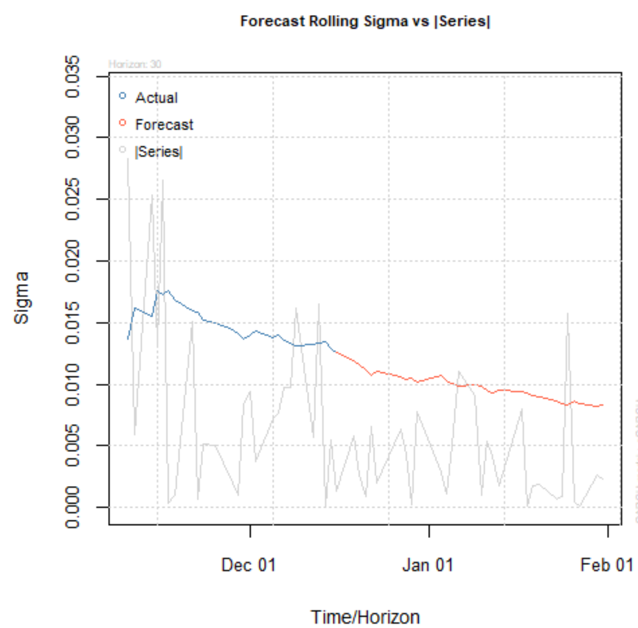

Figure 17: Rolling sigma