AI6123 Time Series Analysis

Chen Yongquan (G2002341D)
Nanyang Technological University

Assignment 2

Table of Contents

Exploratory Data Analysis

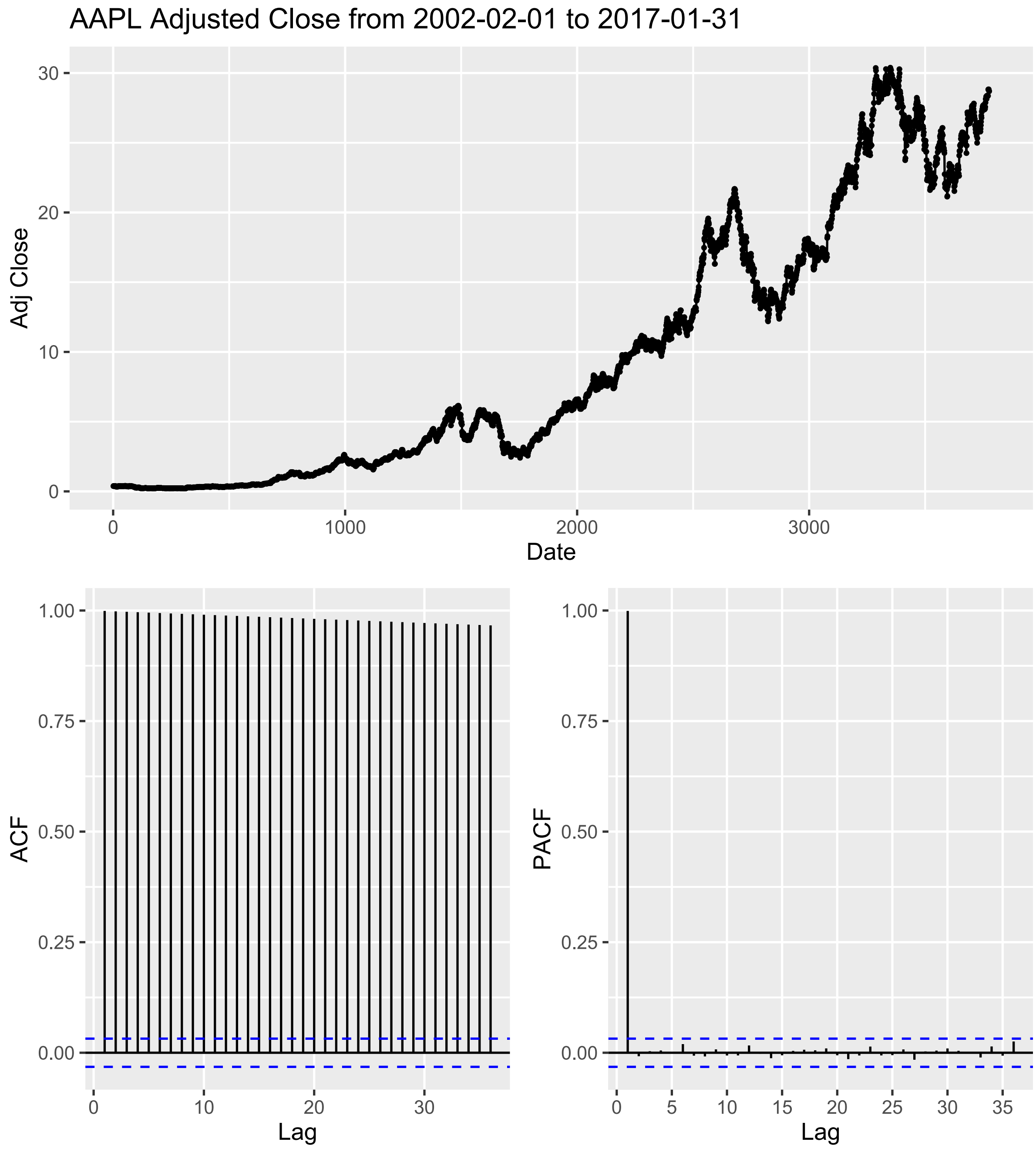

Figure 1

For this assignment we will be fitting a good model to forecast volatility for Apple stock prices. The data is extracted from the Yahoo Finance website and is accurate as of April 2021. The data contains the daily historical Apple stock prices (open, high, low, close and adjusted prices) from February 1, 2002 to January 31, 2017. We will be using the adjusted close prices for model fitting and forecasting.

From the plot of the adjusted close price in Figure 1 above, we can see a non-linear increasing trend with alternating quiet and volatile periods. This pattern is referred to as volatility clustering. Volatility in a time series refers to the phenomenon where the conditional variance of the time series varies over time.

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

The ACF plot in Figure 1 dies down slowly indicating that the data is non-stationary. The Augmented Dickey-Fuller Test yields a p-value of 0.4848 which is more than 0.05 and is not statistically significant, indicating strong evidence for the null hypothesis that the series contains a unit-root. Thus, we reject the alternative hypothesis that the series is stationary.

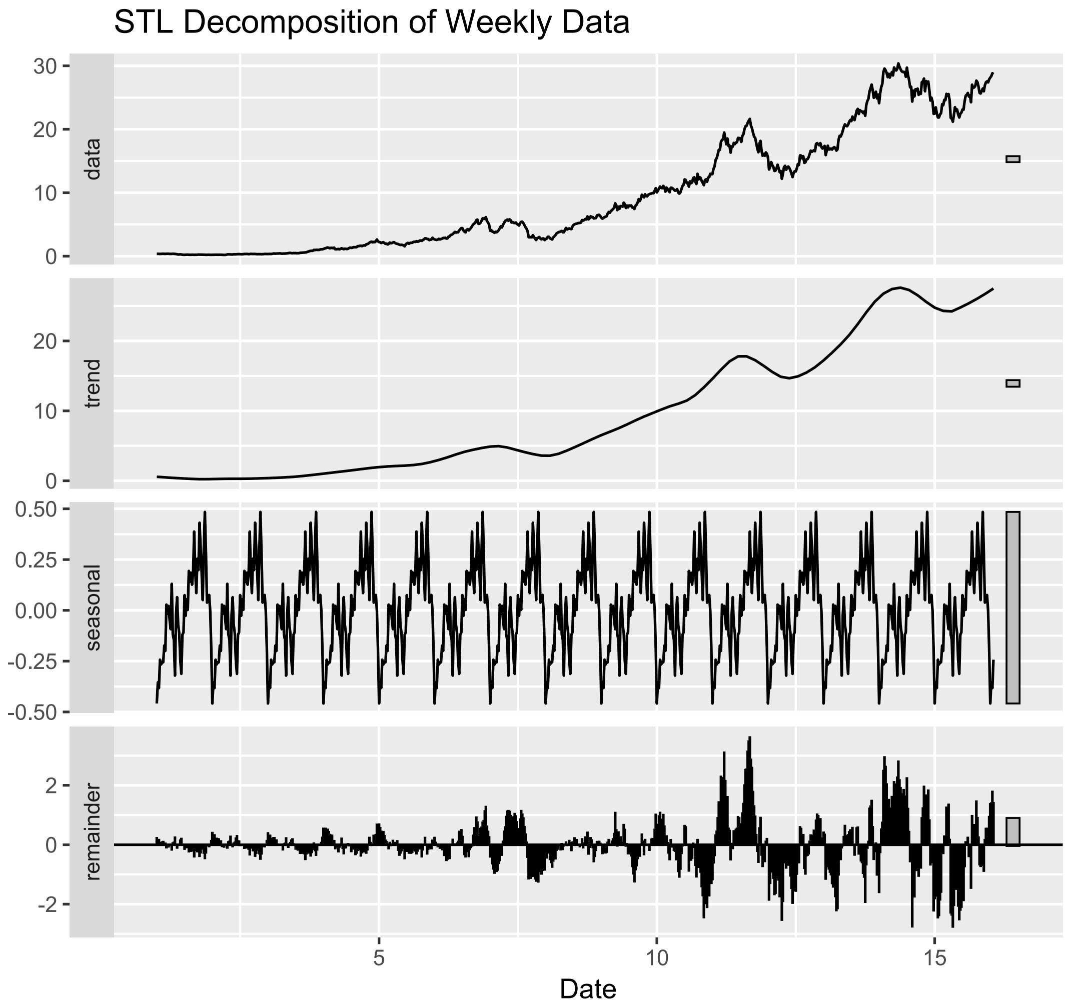

Figure 2

The seasonal and trend decomposition plot of the data in Figure 2 more clearly shows a non-linear trend component, and a seasonal component with increasing variance.

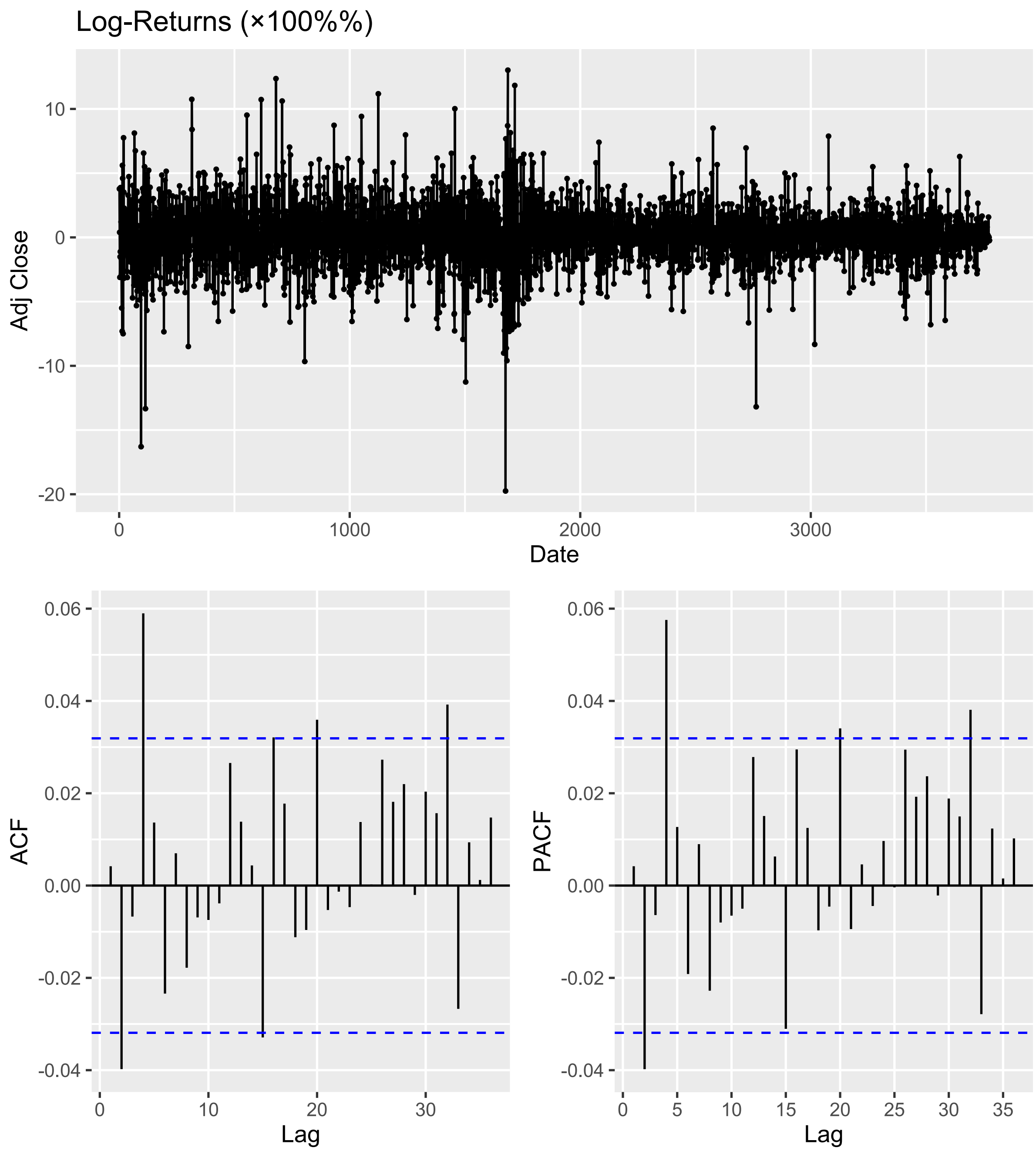

Figure 3

Figure 3 shows the log-return series, multiplied by 100 so that they can be interpreted as percentage changes in the price. The multiplication also reduce numerical errors as the raw returns could be very small numbers and cause rounding errors in some calculations. We can see a large spike in volatility around 2008 and 2009, which might be triggered by the subprime mortgage crisis leading up to the global financial crisis back then.

The sample ACF and PACF suggests that the log-returns have little serial correlation at all. The average log-return equals 0.114864 with a standard error of 0.036475. Thus, the mean of the return process is not statistically significantly different from zero.

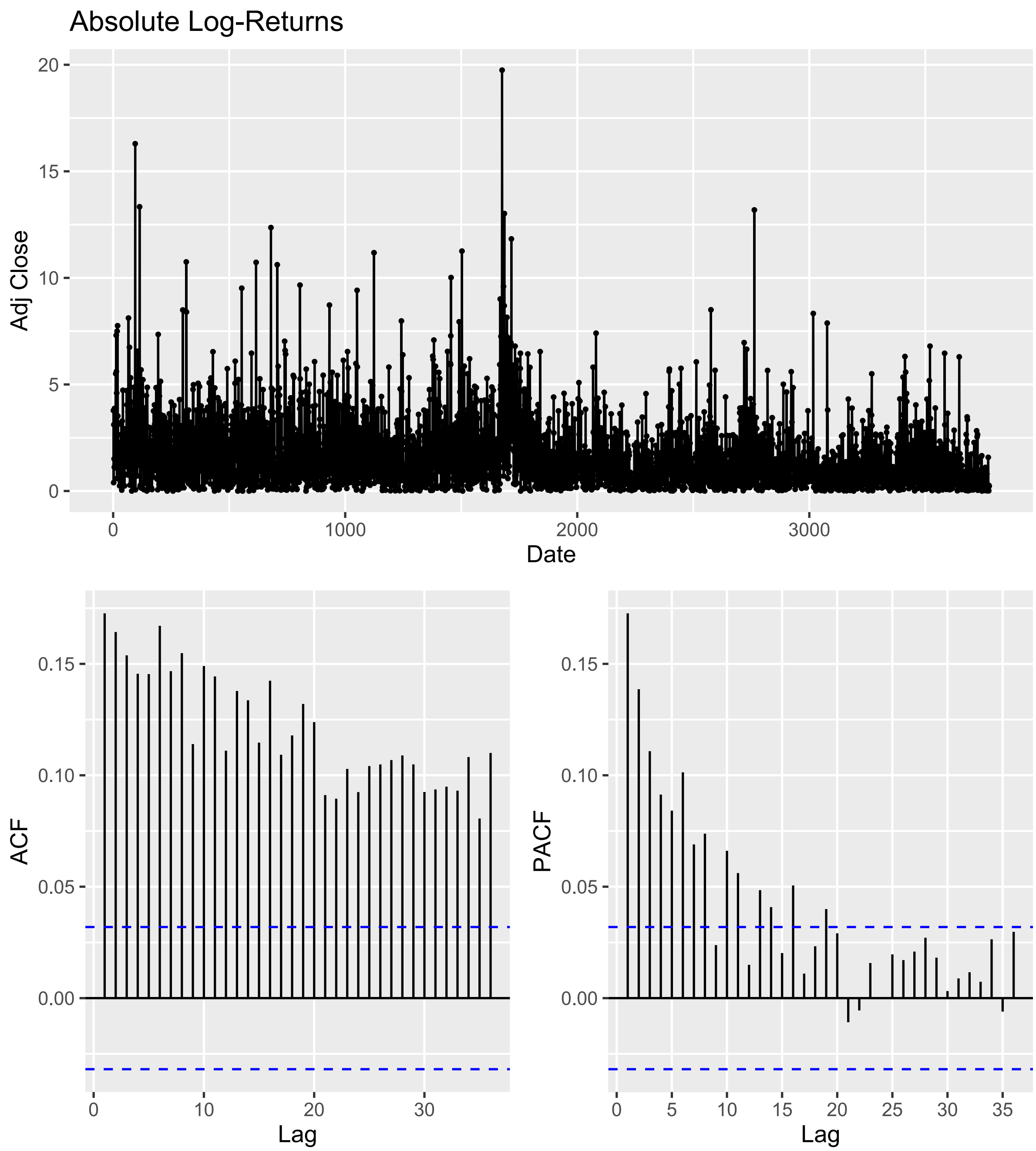

Figure 4

However, the volatility clustering observed in the log-return data gives us a hint that they may not be independently and identically distributed, otherwise the variance would be constant over time. We can look at the plots for the absolute and squared log-returns for stronger evidence that the series is autocorrelated and not $i.i.d.$

Figure 4 shows the absolute log-returns and its corresponding ACF and PACF plots. Indeed, the sample ACF and PACF of the absolute log-returns display some significant autocorrelations and hence provide some evidence that the log-returns of the Apple adjusted close prices are not $i.i.d.$

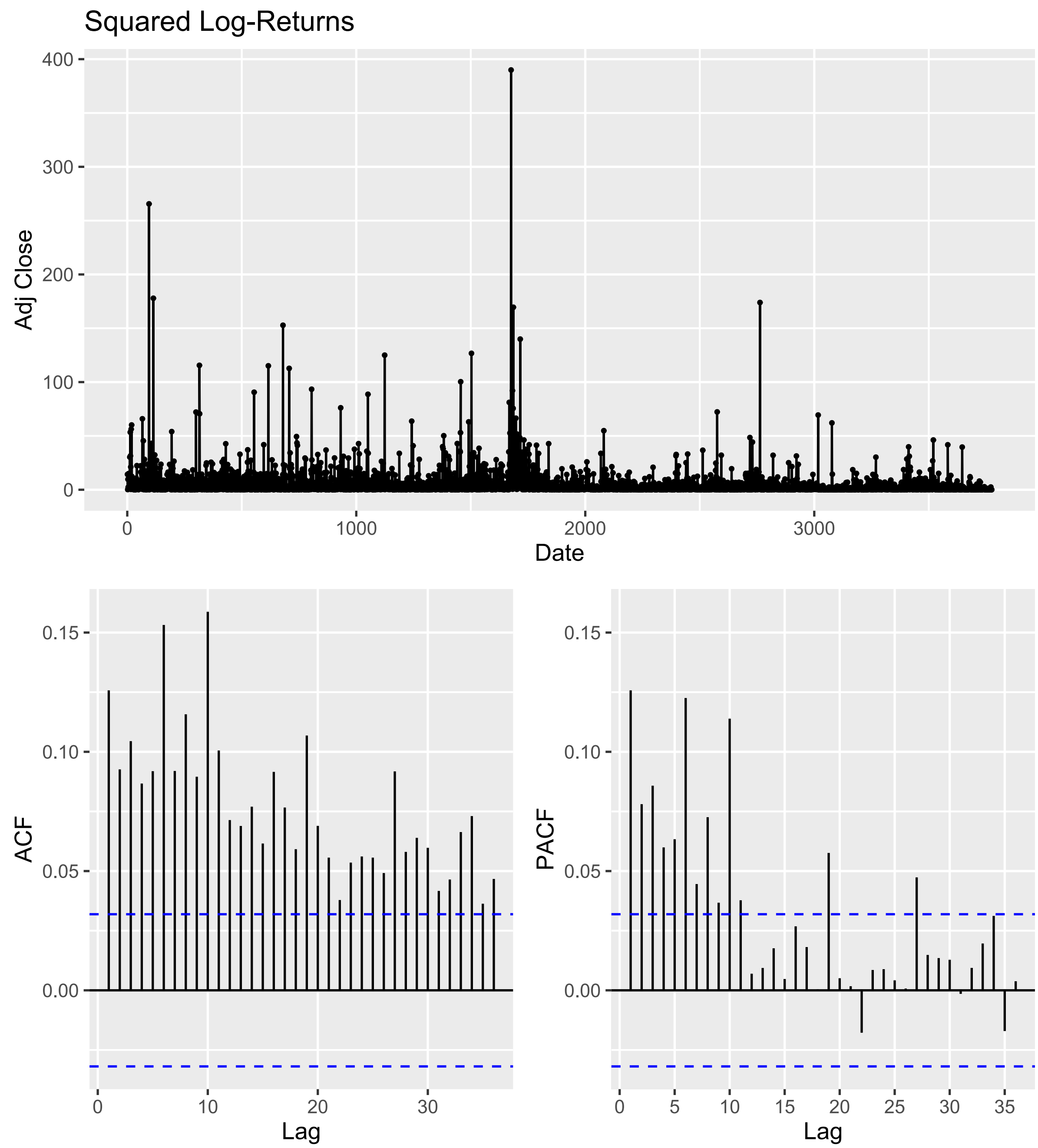

Figure 5

Figure 5 shows the squared log-returns and its corresponding ACF and PACF plots. The sample ACF and PACF of the squared log-returns too display some significant autocorrelations and hence we have strong evidence that the log-returns of the Apple adjusted close prices are not $i.i.d.$

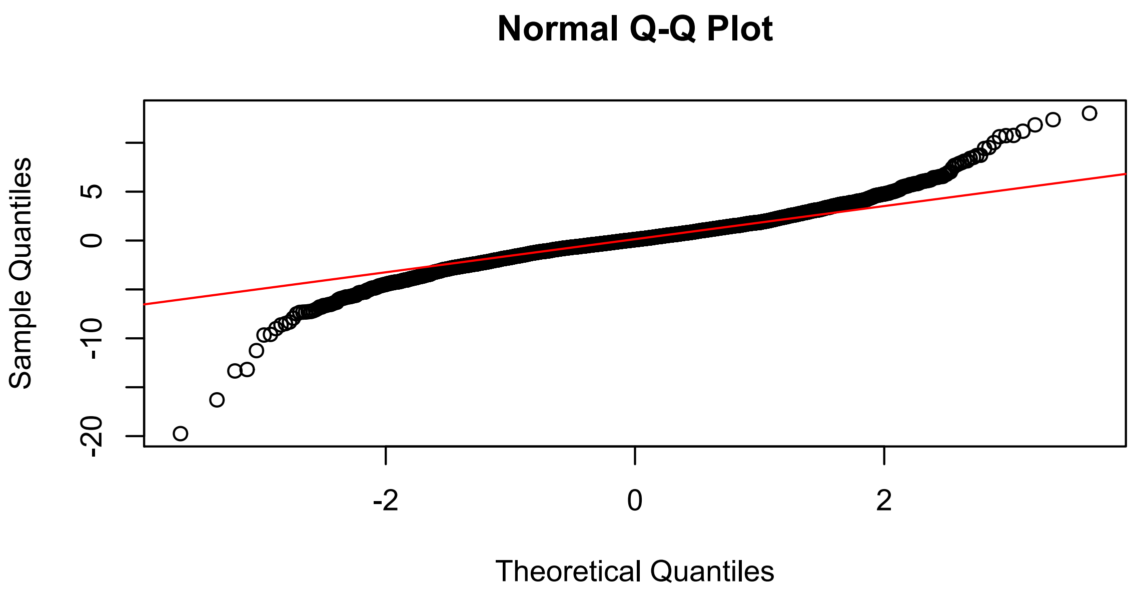

Figure 6

Figure 6 shows the QQ normal scores plot for the log-returns. The QQ plot suggests that the distribution of returns may have a tail thicker than that of a normal distribution and may be somewhat skewed to the left. This is referred to as a heavy-tailed distribution. The thickness of the tail of a distribution relative to that of a normal distribution is often measured by the kurtosis. The sample kurtosis of the log-returns equals 5.441412. The positive kurtosis tallies with our observation that this is a heavy-tailed distribution. The skewness of the log-returns is -0.192333. The negative skewness estimate indicates the mean of the log-returns is less than the median, and both are less than the mode. Thus, the data distribution is left-skewed.

```
AAPL.Adjusted
nobs           3775.000000
NAs               0.000000
Minimum         -19.746992
Maximum          13.019399
1. Quartile      -1.000549
3. Quartile       1.283712
Mean              0.114864
Median            0.094746
Sum             433.612252
SE Mean           0.036475
LCL Mean          0.043352
UCL Mean          0.186376
Variance          5.022248
Stdev             2.241037
Skewness         -0.192333
Kurtosis          5.441412
```

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

Model Fitting

We can apply model identification techniques for ARMA models to the squared return series to identify p and $max(p, q)$ for an $\mathrm{ARMA}(max(p, q), p)$.

We use the extended autocorrelation function (EACF) to assess the orders of the ARMA model for the squared return series. Then we can fit a $\mathrm{GARCH(p, p)}$ model and estimate q by examining the significance of the resulting ARCH coefficient estimates to get the final $\mathrm{GARCH}(p, q)$ model for the log-returns series.

```
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 7 EACF for squared log-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 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 for absolute log-returns

The pattern in the EACF table for the squared log-returns in Figure 7 is not very clear, with more crosses than expected within the right vertex, although an $\mathrm{ARMA}(1,1)$ model seems to be suggested. We can look at the EACF for the absolute returns too as to a first-order approximation, a power transformation preserves the theoretical autocorrelation function and hence the order of a stationary ARMA process. The sample EACF table for the absolute returns shown in Figure 8 convincingly suggests an $\mathrm{ARMA}(1,1)$ and therefore a $\mathrm{GARCH}(1,1)$ model for the original data.

GARCH(1,1) Diagnostic Check

Since we only have one model suggested, we can accept the model as the best model if it passes the Ljung-Box diagnostic test.

```
garch11 = garch(AAPL.Adjusted_Log_Return, order = c(1,1))
summary(garch11)
AIC(garch11)
```

```
Model:
GARCH(1,1)

Residuals:
     Min       1Q   Median       3Q      Max
-5.98525 -0.48689  0.04786  0.63970  6.19130

Coefficient(s):
    Estimate  Std. Error  t value Pr(>|t|)
a0 4.361e-06   6.737e-07    6.473 9.59e-11 ***
a1 4.757e-02   3.455e-03   13.768  < 2e-16 ***
b1 9.444e-01   4.198e-03  224.964  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Diagnostic Tests:
     Jarque Bera Test

data:  Residuals
X-squared = 1980.8, df = 2, p-value < 2.2e-16


     Box-Ljung test

data:  Squared.Residuals
X-squared = 1.1399, df = 1, p-value = 0.2857
---
AIC = -18556.44
```

The p-value for the Ljung-Box test is higher than 0.05 indicating that the model is an adequate fit for our data. However, we can look at plot of the generalized portmanteau test in Figure 12 below to more clearly evaluate. As seen, the p-values in the plot are all higher than 0.05, suggesting that the squared residuals are uncorrelated over time, and hence the standardized residuals may be independent. The standardized residuals is plotted in Figure 9, its corresponding QQ plot in Figure 10 and the ACF of the squared residuals in Figure 11.

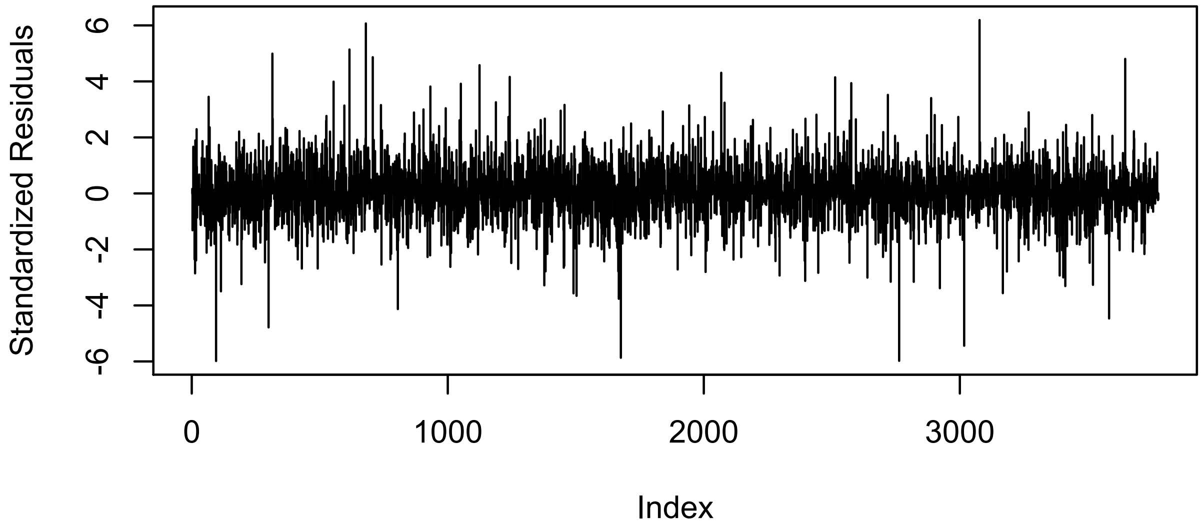

Figure 9

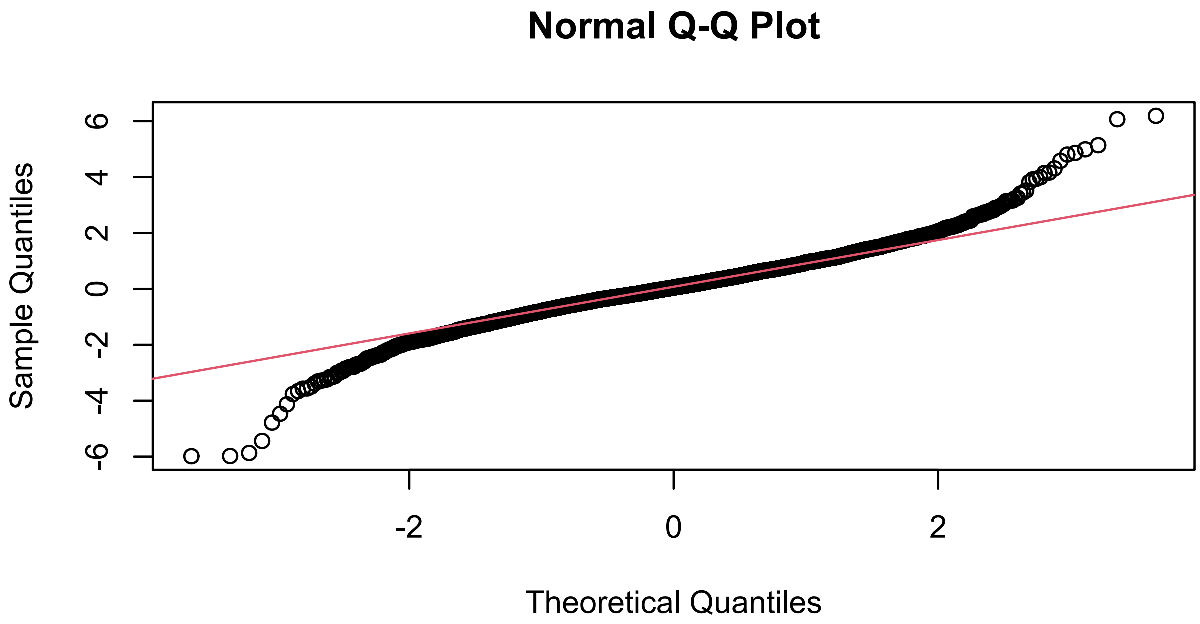

Figure 10

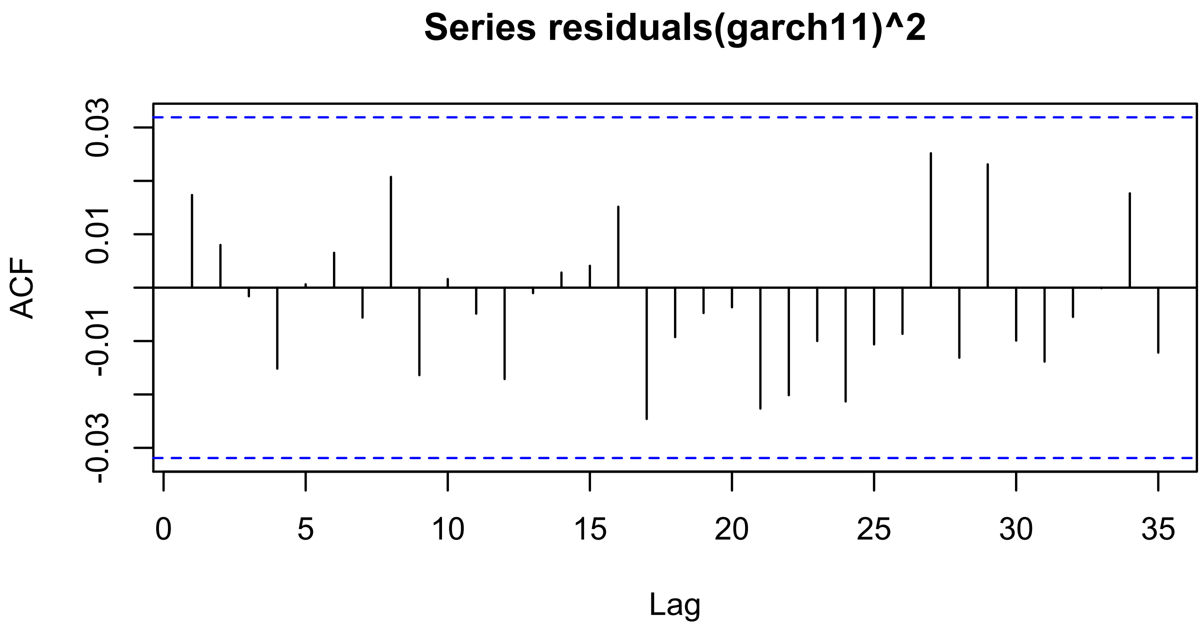

Figure 11

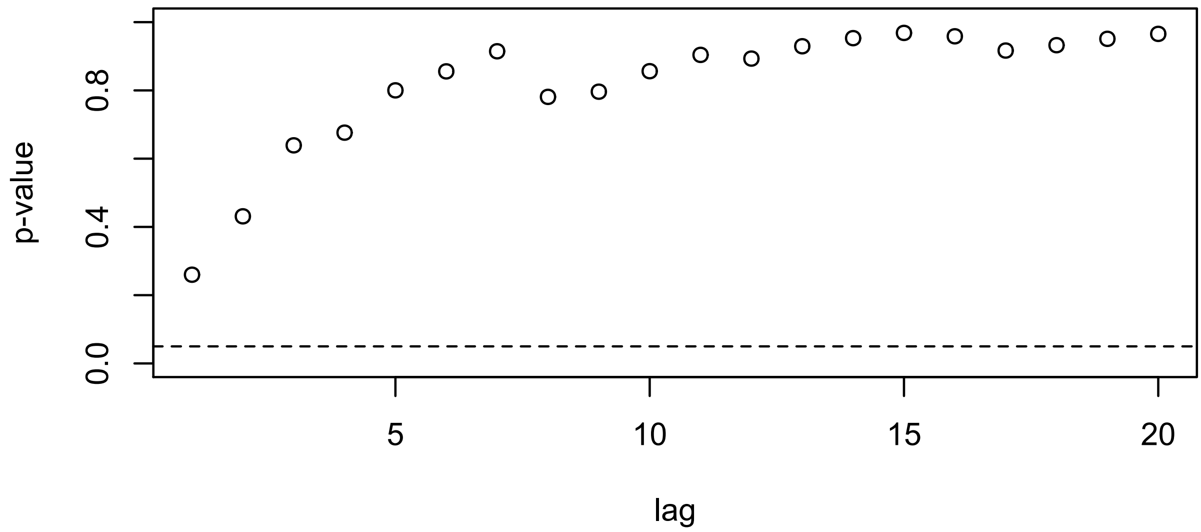

Figure 12

We also use the *rugarch* package to fit several standard $\mathrm{GARCH}(1,1)$ models with varying distributions for its innovations. Their corresponding AIC values are listed in the table below:

Distribution	**AIC**
Normal Distribution	4.2865
Skew Normal Distribution	4.2870
T-Distribution	4.1933
Skew T-Distribution	4.1936
Generalized Error Distribution	4.2056
Skew Generalized Error Distribution	4.2054
Normal Inverse Gaussian Distribution	4.1959
Generalized Hyperbolic Distribution	4.1942
Johnson's S_U Distribution	4.1942

Using a T-distribution for modelling the innovations in the GARCH model yielded the best lowest AIC. We thus also fit several types of GARCH models using T-distribution for the conditional density of the innovations as listed below:

Distribution	**AIC**
fGARCH, GARCH	4.1933
fGARCH, TGARCH	4.1821
fGARCH, AVGARCH	4.1823
fGARCH, NGARCH	4.1887
fGARCH, NAGARCH	4.1858
fGARCH, APARCH	4.1826
fGARCH, GJRGARCH	4.1885
fGARCH, ALLGARCH	4.1825
eGARCH	4.1812
gjrGARCH	4.1885
apARCH	4.1826
iGARCH	4.1929
csGARCH	4.1871

The eGARCH model yielded the lowest AIC, thus our best model is $\mathrm{eGARCH}(1,1)$ with a T-distribution for the conditional density of the innovations, which we will use for forecasting. The model also passes diagnostic testing as shown below:

```
*---------------------------------*
*          GARCH Model Fit        *
*---------------------------------*

Conditional Variance Dynamics
-----------------------------------
GARCH Model    : eGARCH(1,1)
Mean Model     : ARFIMA(0,0,0)
Distribution   : std

Optimal Parameters
------------------------------------
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.125716    0.025566   4.9173  1.0e-06
omega   0.021964    0.001620  13.5546  0.0e+00
alpha1 -0.051308    0.007111  -7.2151  0.0e+00
beta1   0.984111    0.002073 474.8254  0.0e+00
gamma1  0.155904    0.038824   4.0156  5.9e-05
shape   5.173837    0.410101  12.6160  0.0e+00

Weighted Ljung-Box Test on Standardized Residuals
------------------------------------
                        statistic p-value
Lag[1]                      3.490 0.06173
Lag[2*(p+q)+(p+q)-1][2]     3.703 0.09062
Lag[4*(p+q)+(p+q)-1][5]     6.191 0.08107
d.o.f=0
H0 : No serial correlation

Weighted Ljung-Box Test on Standardized Squared Residuals
------------------------------------
                        statistic p-value
Lag[1]                     0.3689  0.5436
Lag[2*(p+q)+(p+q)-1][5]    2.1800  0.5765
Lag[4*(p+q)+(p+q)-1][9]    2.8247  0.7875
d.o.f=2
```

Forecasting

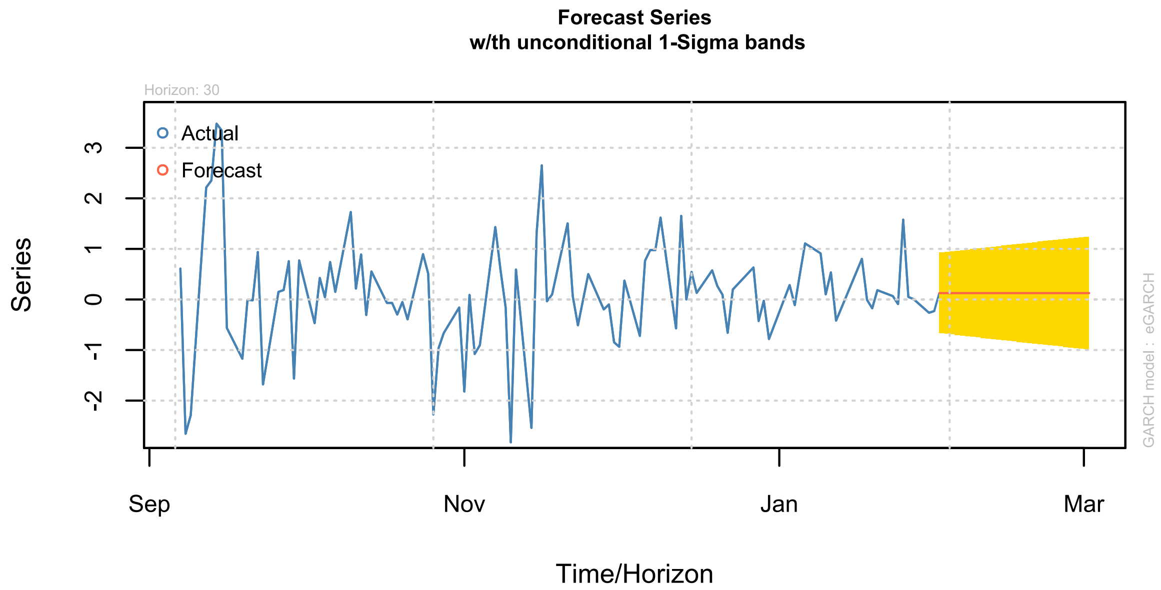

Figure 13

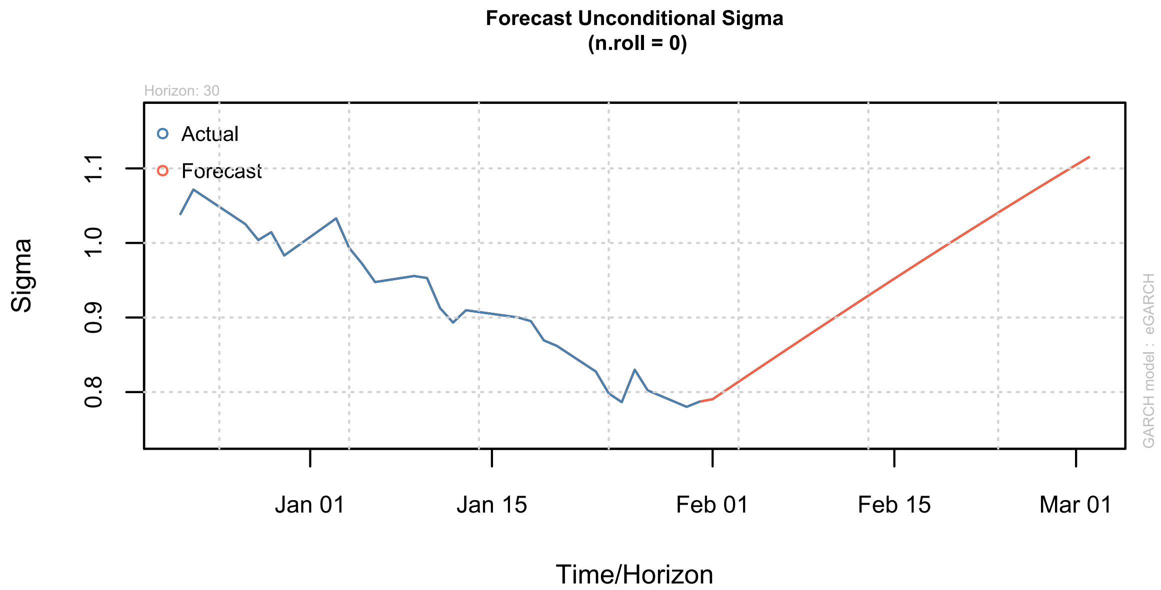

Figure 14

Figure 13 and Figure 14 shows the forecasted volatility for 30 days, using the entire series as training data for fitting the model. The red line being the forecast and the yellow shaded area the 95% upper and lower bounds.

We can alternatively do a rolling forecast using the last quarter from the original data as out-of-sample data to test the performance of our model and also to forecast ahead of the original series for 30 days. This is correspondingly shown in all the following figures below:

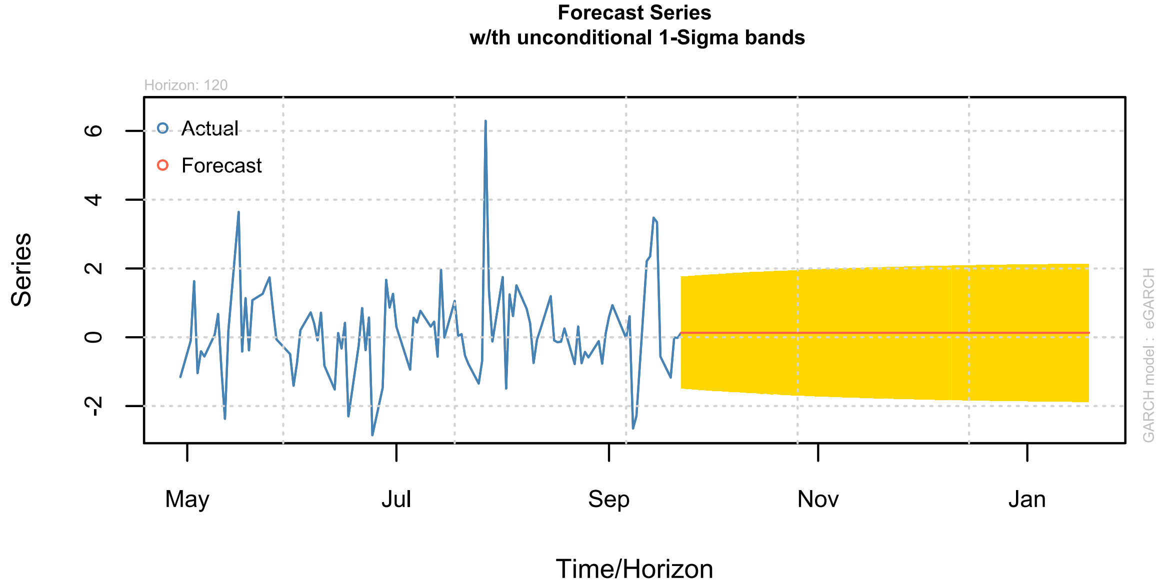

Figure 15

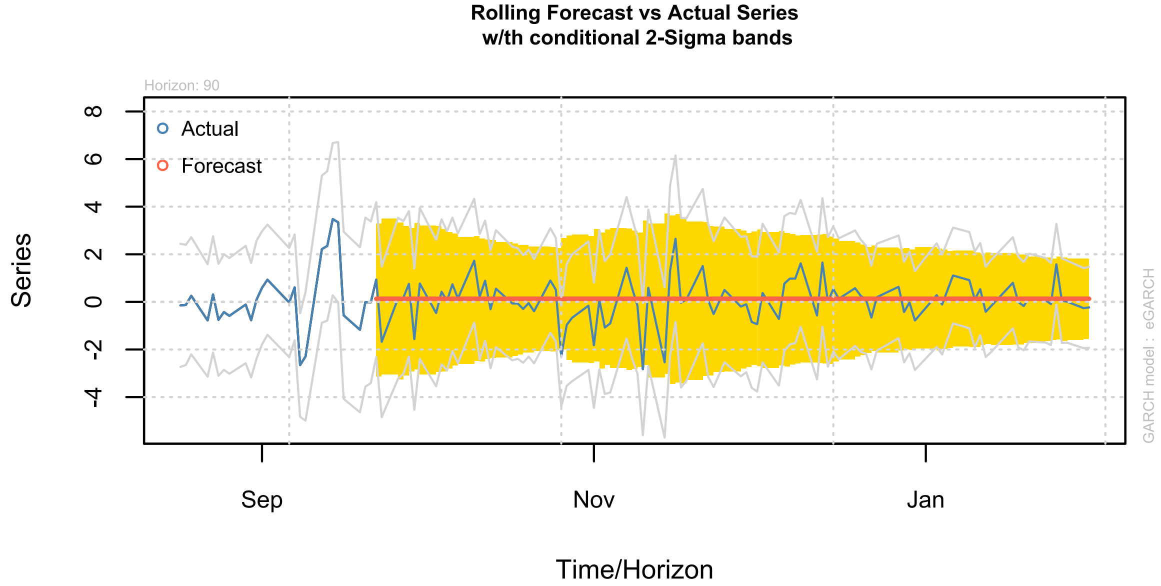

Figure 16

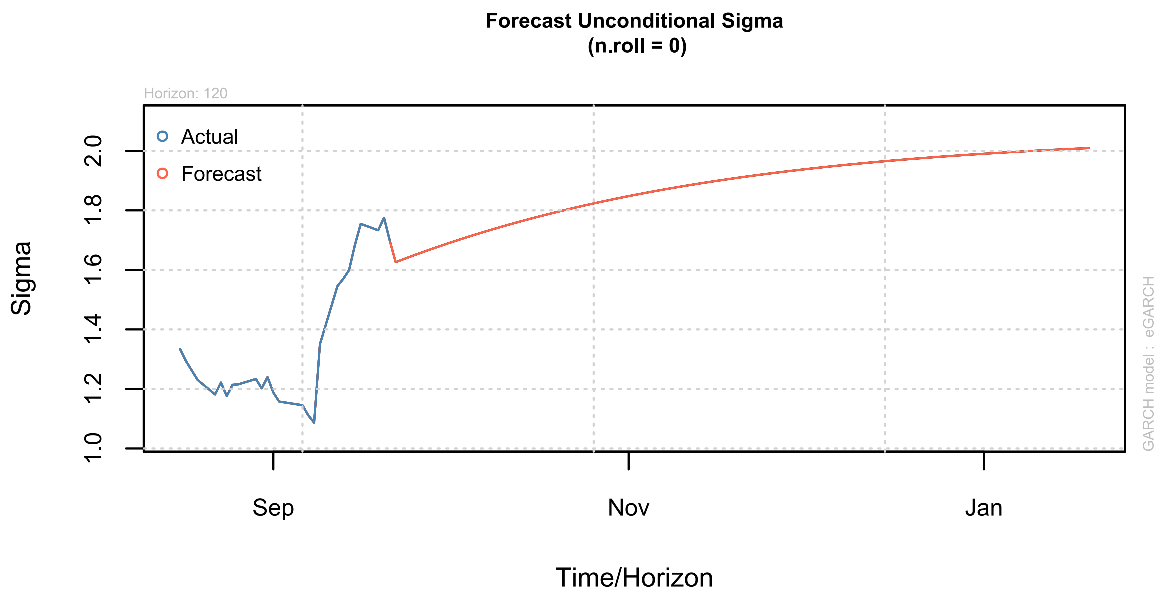

Figure 17

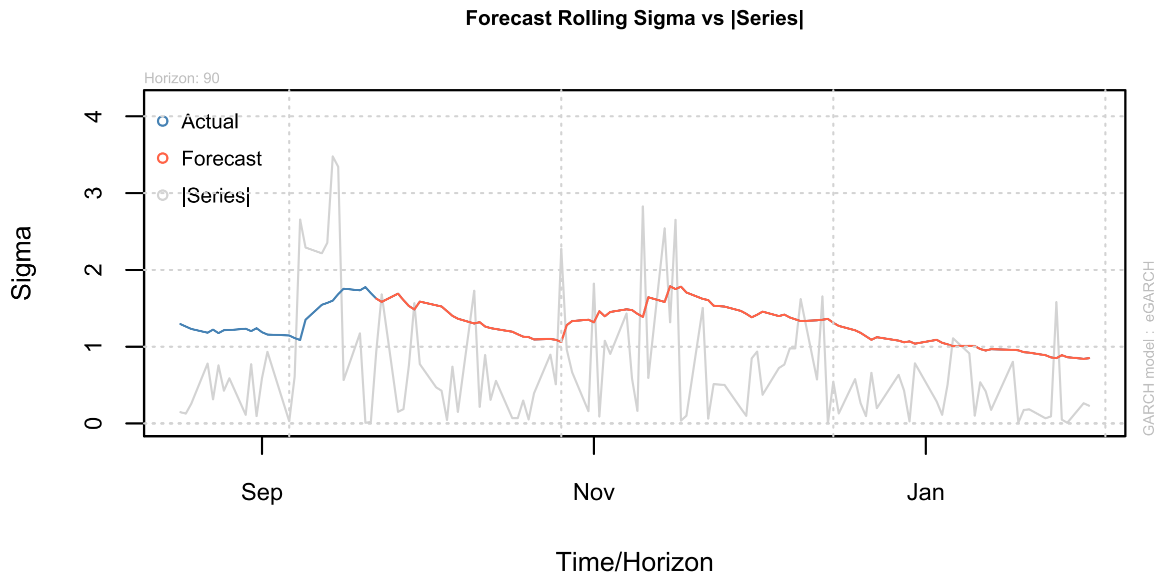

Figure 18