Forecasting Time Series Data **STAT-GB 6018**

Forecasting the Stock Price of Apple Inc.

Future Liang Fall 2023

Contents

1 Introduction • **2** ARIMA: Preliminary Analysis • **3** ARIMA: Model Selection • **4** Evidence of Conditional Heteroscedasticity • **5** GARCH: Model Selection • **6** ARIMA-GARCH: Forecast Interval • **7** Examination of Conditional Variance • **8** ARIMA-GARCH: Forecast Interval Graph • **9** ARIMA-GARCH: Residual Analysis • **10** Forecast Accuracy

1 Introduction

As Jeremy Siegel said, "volatility scares enough people out of the market to generate superior returns for those who stay in." As a result, in this project 2, I will delve into the volatile stock market and try to forecast the stock price of Apple Inc. (henceforth Apple).

The detailed information of this project is summarized in Table 1 below:

Table 1: Project Key Information

Series	Apple's Stock Price (Adjusted Close)
Time	November 30, 2018 - November 30, 2023
Frequency	Daily
Observations	1257 Trading Days
Method	ARIMA-GARCH
Model Selection Criteria	Minimization of AIC_c
Forecast Lead	1 Day
Data Source	Yahoo Finance

The project will start by fitting the best ARIMA model with some graphical analysis and the minimization of AIC_c.

Then, I will use the residuals from the ARIMA model to select the best model between ARCH and GARCH(1,1) models with some similar graphical analysis and the minimization of AIC_c.

Afterwards, I will provide the one step ahead forecast based on the ARIMA-GARCH model and compare it with the ARIMA-only model.

Finally, I will analyze the adequacy of the ARIMA-GARCH model through a series of graphical analysis and statistical hypothesis tests.

2 ARIMA: Preliminary Analysis

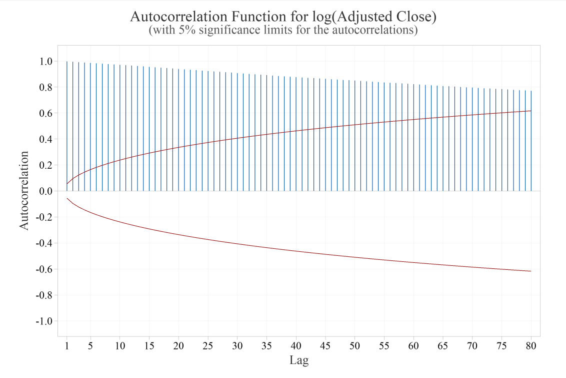
The time series of Adjusted Close and log(Adjusted Close) as well as their first differences of Apple are plotted below:

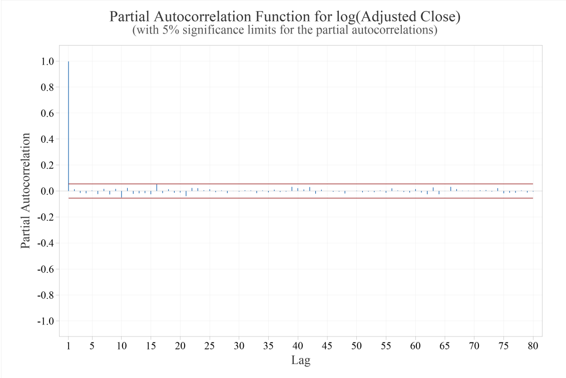
The graph of Adjusted Close exhibits a certain degree of level-dependent volatility. This is obvious by checking the graph of First Difference of Adjusted Close.

To mitigate this level-dependent volatility, I decide to take log(Adjusted Close). By checking the graphs of log(Adjusted Close) and First Difference of log(Adjusted Close), such level-dependent volatility indeed seems to be mitigated by the logarithmic transformation.

CONTINUE ON THE NEXT PAGE.

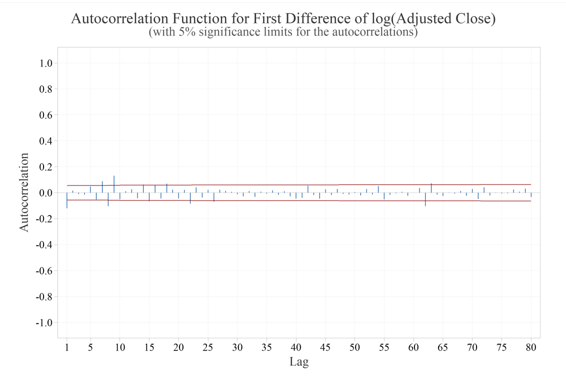
To fit a potential ARIMA model, I will start by plotting the ACF and PACF.
The ACF and PACF of log(Adjusted Close) of Apple are plotted below:

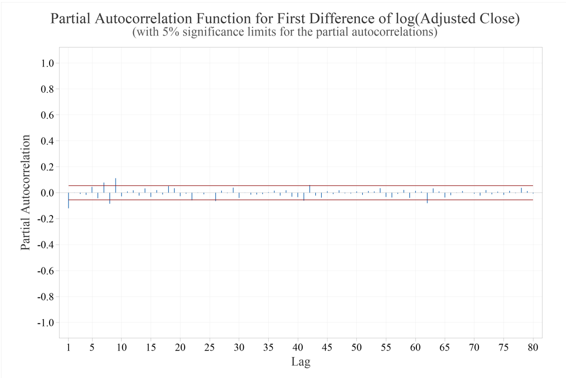




The ACF dies down very slowly, and the PACF has a statistically significant cutoff at lag 1. This indicates that the log(Adjusted Close) is non-stationary, so differencing is needed.

The ACF and PACF of the first difference of log(Adjusted Close) of Apple are plotted below:





Note that this is the first difference of log(Adjusted Close) of Apple, so the d in ARIMA(p,d,q) should be 1.
The ACF suggests the ARIMA(0,1,1) since the ACF has a statistically significant cutoff at lag 1.
The PACF suggests the ARIMA(1,1,0) since the PACF has a statistically significant cutoff at lag 1.
Therefore, based on the ACF and PACF, picking either ARIMA(0,1,1) or ARIMA(1,1,0) might be reasonable, with consideration of parsimony.

3 ARIMA: Model Selection

> Normally, while doing the model selection, it is essential to minimize the AIC_c between the model without constant and the model with constant.
> But for the purpose of this project, I will only fit the model with constant.

The Table 2 below provides the AIC_c of ARIMA(p,d,q) for log(Adjusted Close) of Apple **with constant** based on the formula:

$$AIC_c = N\log(\frac{SS}{N}) + 2(p+q+2)\frac{N}{N-p-q-3}$$

Table 2: ARIMA Model Trial With Constant

Model	p	d	q	N	SS	AIC_c
1	0	1	0	1256	0.5308872	-9753.720100
2	0	1	1	1256	0.5233028	-9769.783389
3	0	1	2	1256	0.5232217	-9767.965236
4	0	1	3	1256	0.5231105	-9766.216161
5	0	1	4	1256	0.5231091	-9764.20035
6	0	1	5	1256	0.5222338	-9764.281191
7	1	1	0	1256	0.5232010	-9770.027861
8	1	1	1	1256	0.5232006	-9768.016089
9	1	1	2	1256	0.5231911	-9766.022777
10	1	1	3	1256	0.5231103	-9764.197438
11	1	1	4	1256	0.5176389	-9775.38114
12	1	1	5	1256	*	*
13	2	1	0	1256	0.5232006	-9768.015919
14	2	1	1	1256	*	*
15	2	1	2	1256	0.5231602	-9764.077612
16	2	1	3	1256	*	*
17	2	1	4	1256	0.5074216	-9798.394537
18	2	1	5	1256	*	*
19	3	1	0	1256	0.5231510	-9766.119072
20	3	1	1	1256	0.5231343	-9764.139904
21	3	1	2	1256	0.5187934	-9772.582881

Continued on next page

Table 2: ARIMA Model Trial With Constant (Continued)

Model	p	d	q	N	SS	AIC_c
22	3	1	3	1256	*	*
23	3	1	4	1256	0.5071732	-9796.980659
24	3	1	5	1256	0.5094079	-9789.426448
25	4	1	0	1256	0.5230078	-9764.443525
26	4	1	1	1256	*	*
27	4	1	2	1256	*	*
28	4	1	3	1256	0.5073661	-9796.503037
29	4	1	4	1256	*	*
30	4	1	5	1256	0.5181242	-9766.081728
31	5	1	0	1256	0.5219654	-9764.927022
32	5	1	1	1256	0.5146412	-9780.6502
33	5	1	2	1256	0.5153319	-9776.936597
34	5	1	3	1256	0.5069833	-9795.418779
35	5	1	4	1256	0.5065559	-9794.442659
36	5	1	5	1256	*	*
					Lowest	-9798.394537

Note: `Minitab` refuses to fit the model with asterisk (*).

Based on the table with constant above, ARIMA(2,1,4) has the lowest AIC_c value, so it is reasonable to choose ARIMA(2,1,4) with constant in this case.

CONTINUE ON THE NEXT PAGE.

The estimates of the parameters of ARIMA(2,1,4) with constant are provided by `Minitab` in Table 3 as follow:

Table 3: Parameter Estimates of ARIMA(2,1,4) With Constant

Type	**Coef**	**SE Coef**	***t*-Value**	***p*-Value**
AR 1	-1.7581	0.0297	-59.19	0.000
AR 2	-0.8995	0.0358	-25.09	0.000
MA 1	-1.6619	0.0155	-107.49	0.000
MA 2	-0.7210	0.0411	-17.52	0.000
MA 3	0.1152	0.0548	2.10	0.036
MA 4	0.0511	0.0316	1.62	0.106
Constant	0.00425	0.00181	2.35	0.019

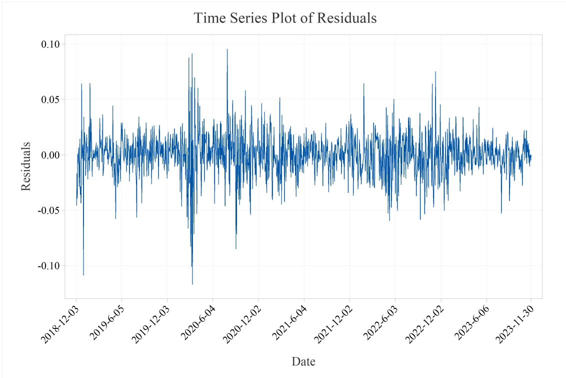
Assume x_t represents log(Adjusted Close) of Apple.
Set $y_t = x_t - x_{t-1}$.
Then, the complete form of the ARIMA(2,1,4) model with constant can be written as:

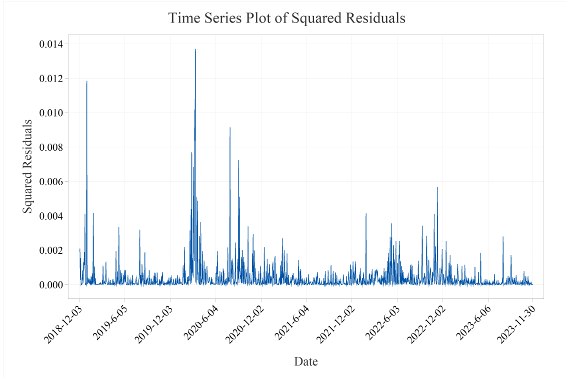
$$y_t = 0.00425 - 1.7581 y_{t-1} - 0.8995 y_{t-2} + \varepsilon_t + 1.6619 \varepsilon_{t-1} + 0.7210 \varepsilon_{t-2} - 0.1152 \varepsilon_{t-3} - 0.0511 \varepsilon_{t-4}$$

From `Minitab`, the ARIMA one step ahead forecast is 5.24711.
The ARIMA one step ahead 95% forecast interval is $[5.20759, 5.28662]$.

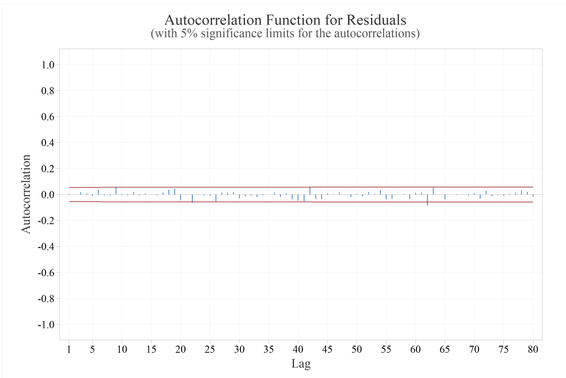
4 Evidence of Conditional Heteroscedasticity

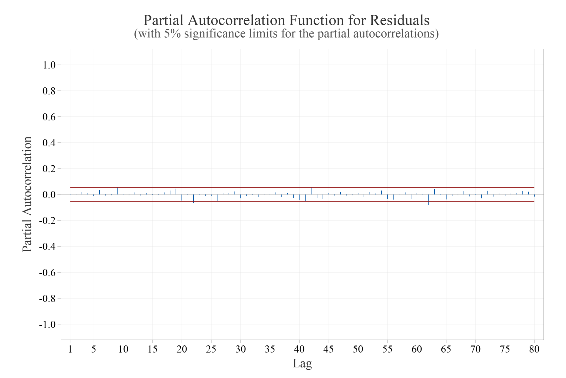
The residuals and squared residuals of log(Adjusted Close) of Apple are plotted below:



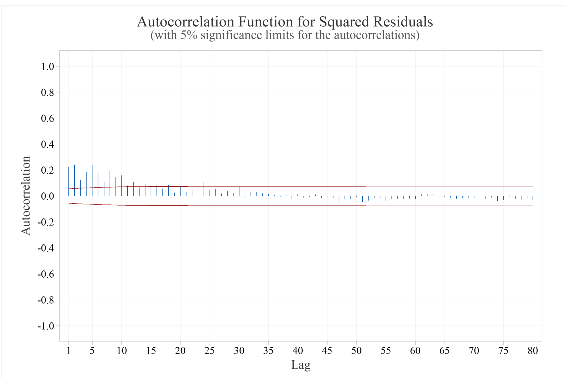


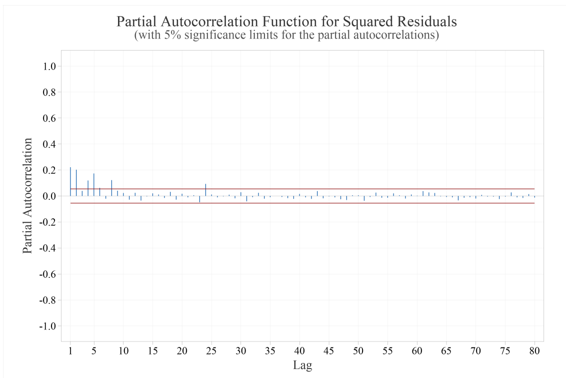
The ACF and PACF of the residuals of log(Adjusted Close) of Apple are plotted below:





The ACF and PACF of the squared residuals of log(Adjusted Close) of Apple are plotted below:





CONTINUE ON THE NEXT PAGE.

The ACF and PACF of the residuals show that there are few statistically significant cutoffs. This suggests that the residuals are approximately uncorrelated.

However, they are not independent because for the squared residuals, the ACF has a statistically significant cutoff at lag 10, and the PACF has a statistically significant cutoff at lag 5. This suggests that large residuals tend to be followed by large residuals (of either sign), and small residuals by small residuals, which is also manifested in the plots of residuals and squared residuals. This is a classic sign of volatility clustering, which is a form of conditional heteroscedasticity.

5 GARCH: Model Selection

The Table 4 below provides the AIC_c of ARCH(q) for the residuals from ARIMA(2,1,4) model **with constant** for log(Adjusted Close) of Apple based on the formula:

$$AIC_c = -2 \cdot \log(\text{Likelihood}) + 2(q+1)\frac{N}{N-q-2}$$

Table 4: ARCH Model Trial

Model	q	N	log(Likelihood)	$AICc$
0	0	1256	3125.068	-6248.132810
1	1	1256	3155.825	-6307.640423
2	2	1256	3176.464	-6346.908831
3	3	1256	3181.830	-6355.628026
4	4	1256	3188.830	-6367.612000
5	5	1256	3214.890	-6417.712746
6	6	1256	3222.657	-6431.224256
7	7	1256	3220.617	-6425.118523
8	8	1256	3225.544	-6432.943538
9	9	1256	3223.180	-6426.183293
10	10	1256	3219.618	-6417.023781
			Lowest	-6432.943538

Using the same formula above, with $q = 2$ and $\log(\text{Likelihood}) = 3240.729$, AIC_c can be computed for GARCH(1,1) as:

$$AIC_c = -2 \cdot 3240.729 + 2(2+1)\frac{1256}{1256-2-2} = \boxed{-6475.439}$$

Lowest Overall

Among the ARCH(q) where $q = 1, 2, \cdots, 10$ and GARCH(1,1) models, AIC_c is the lowest for GARCH(1,1) model.
As a result, GARCH(1,1) model is selected.

CONTINUE ON THE NEXT PAGE.

The output of GARCH(1,1) model for the residuals from ARIMA(2,1,4) model **with constant** for log(Adjusted Close) of Apple is provided by `R` below:

CODE Output for GARCH(1,1) Model

```
Call:
garch(x = res, order = c(1, 1), trace = F)

Model:
GARCH(1,1)

Residuals:
     Min       1Q   Median       3Q      Max
-4.29817 -0.56482  0.02478  0.61084  5.08107

Coefficient(s):
    Estimate  Std. Error  t value Pr(>|t|)
a0 1.146e-05   2.590e-06    4.426 9.59e-06 ***
a1 9.626e-02   1.397e-02    6.891 5.54e-12 ***
b1 8.749e-01   1.704e-02   51.337  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Diagnostic Tests:
        Jarque Bera Test

data:  Residuals
X-squared = 141.24, df = 2, p-value < 2.2e-16


        Box-Ljung test

data:  Squared.Residuals
X-squared = 0.33074, df = 1, p-value = 0.5652
```

Based on the output from `R` above, all of the coefficients are statistically significant (after dividing the p-value by 2).

The complete form of the GARCH(1,1) model can be written as:

$$h_t = 1.146 \times 10^{-5} + 9.626 \times 10^{-2} \varepsilon_{t-1}^2 + 8.749 \times 10^{-1} h_{t-1}$$

The unconditional variance of the shocks can be calculated as:

$$\frac{1.146 \times 10^{-5}}{1 - 9.626 \times 10^{-2} - 8.749 \times 10^{-1}} = 3.974 \times 10^{-4}$$

Note that the unconditional variance is very close to 0.

6 ARIMA-GARCH: Forecast Interval

The one step ahead ARIMA-GARCH 95% forecast interval will be constructed through:

$$\begin{cases} f_{n,1} \pm 1.96\sqrt{h_{n+1}} \\ h_{n+1} = 1.146 \times 10^{-5} + 9.626 \times 10^{-2}\varepsilon_n^2 + 8.749 \times 10^{-1} h_n \end{cases}$$

From R, I code the ARIMA-GARCH one step ahead 95% forecast interval as follow:

CODE ARIMA-GARCH One Step Ahead 95% Forecast Interval

```
> arimamean <- 5.24710517991707 80 # From Minitab
> h_t <- model$fitted.values[,1]^2
> h_tplus1 <- coef(model)[1]+coef(model)[2]*res[1256]^2+coef(model)[3]*h_t[1256]
> lower <- arimamean-qnorm(p = 1-0.05/2)*sqrt(h_tplus1)
> upper <- arimamean+qnorm(p = 1-0.05/2)*sqrt(h_tplus1)
> lower
       a0
5.223818
> upper
       a0
5.270392
> qnorm(p = 0.05,mean = arimamean,sd = sqrt(h_tplus1))
[1] 5.227562
```

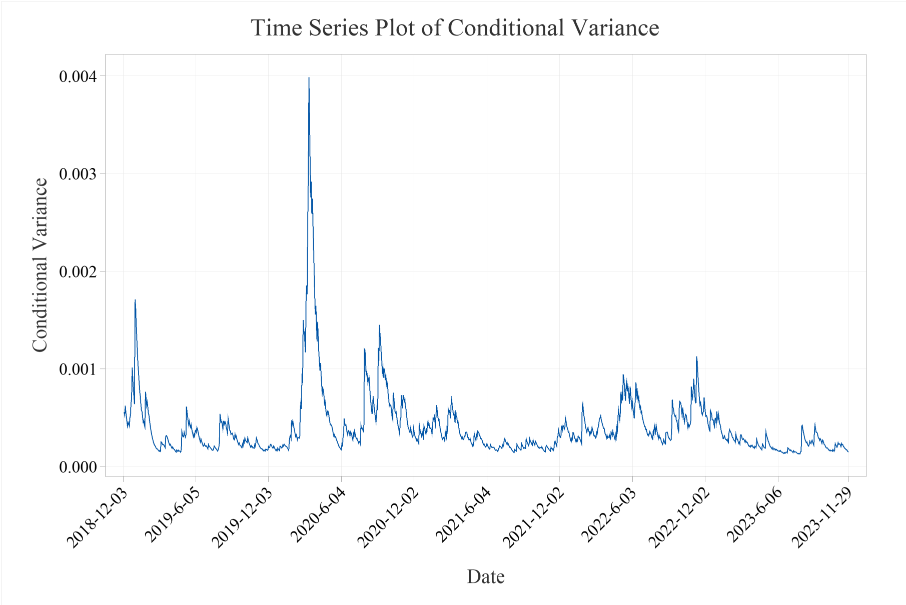
From R, the ARIMA-GARCH one step ahead 95% forecast interval is $[5.223818, 5.270392]$.

In this case, the 95% forecast interval of the ARIMA-GARCH model is slightly narrower than that of the ARIMA-only model.

The 5^{th} percentile of the conditional distribution of the next period's log(Adjusted Close) of Apple is 5.227562, as computed in R.

7 Examination of Conditional Variance

The conditional variance from the ARIMA-GARCH model for log(Adjusted Close) of Apple is plotted below:

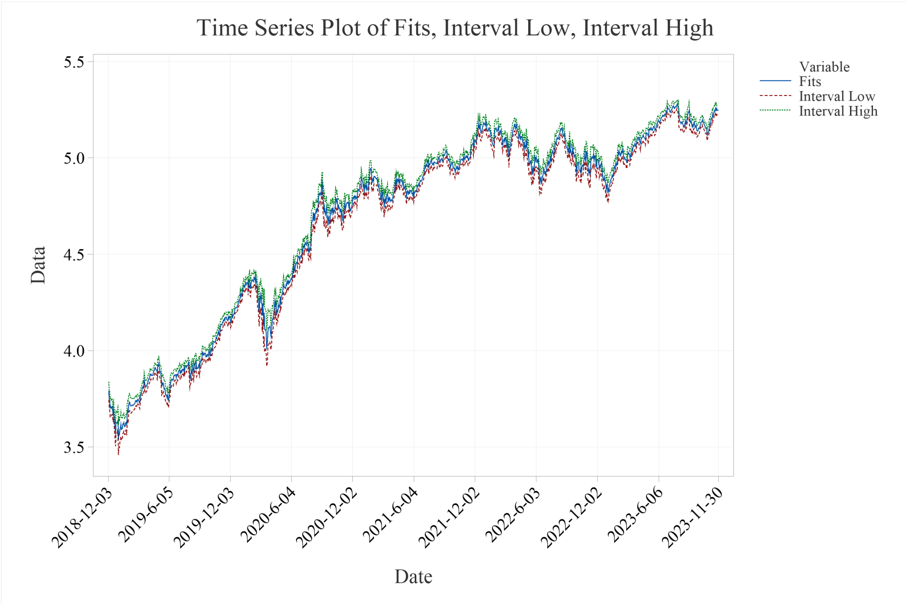


The location of bursts of high volatility are around dates 01/04/2019, 03/16/2020, 07/31/2020, and 09/08/2020.

These highly volatile periods agree with those found from the time series plot of log(Adjusted Close) of Apple itself.

8 ARIMA-GARCH: Forecast Interval Graph

The time series of the fits of log(Adjusted Close) and its ARIMA-GARCH 95% one step ahead forecast interval are plotted below:



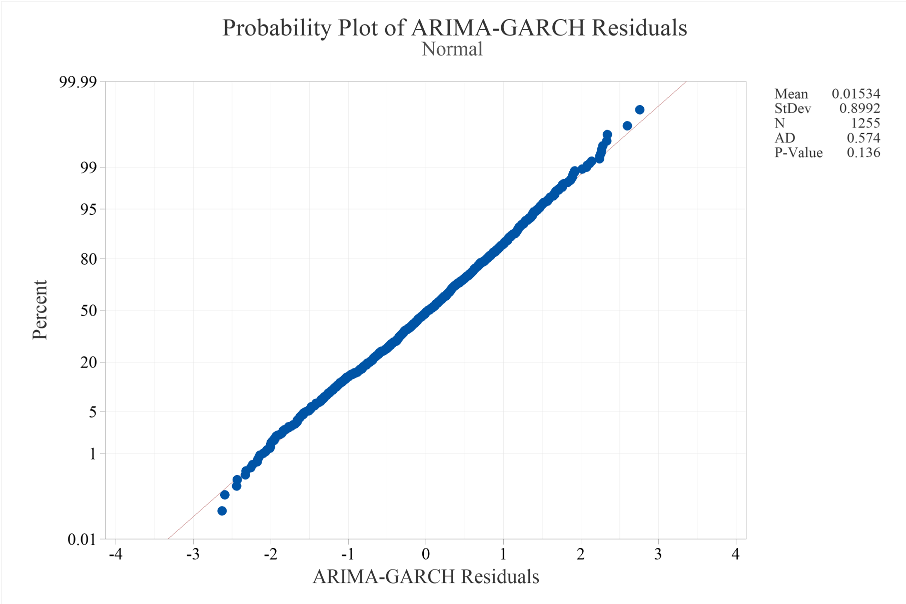
The ARIMA-GARCH one step ahead 95% forecast interval is very close to the true value of log(Adjusted Close) of Apple, which suggests that the forecast intervals are accurate.

The narrow width of the forecast intervals implies that the model's predictions are precise and provide a high level of confidence in the estimated values.

Note that the performance may be somewhat better here than in an actual forecasting context, since the ARIMA-GARCH parameters are estimated from the entire data set, not just the observations up to the time at which the forecast is to be constructed.

9 ARIMA-GARCH: Residual Analysis

The normal probability plot of the residuals from the ARIMA-GARCH model of log(Adjusted Close) of Apple is plotted below:



The hypothesis of the Anderson-Darling Normality Test is:

$$\begin{cases} H_0 : \text{The data are normally distributed.} \\ H_a : \text{The data are not normally distributed.} \end{cases}$$

Although some dots fall away from the line, the p-value of the Anderson-Darling Normality Test is larger than 0.05, which means we fail to reject H_0. So we have enough evidence to conclude that the residuals from the ARIMA-GARCH model is normally distributed.

Therefore, the model seems to have adequately described the leptokurtosis in the data.

10 Forecast Accuracy

The `Minitab` reports that there are 35 observations out of the ARIMA-GARCH 95% forecast interval.
Therefore, the failure rate of the ARIMA-GARCH 95% forecast interval is about 2.787%.

The one step ahead true value is 5.25353, which is captured by the ARIMA-GARCH 95% forecast interval.
As a result, the ARIMA-GARCH 95% forecast interval seems just right, while the ARIMA-only 95% forecast interval seems to be a little bit wide.