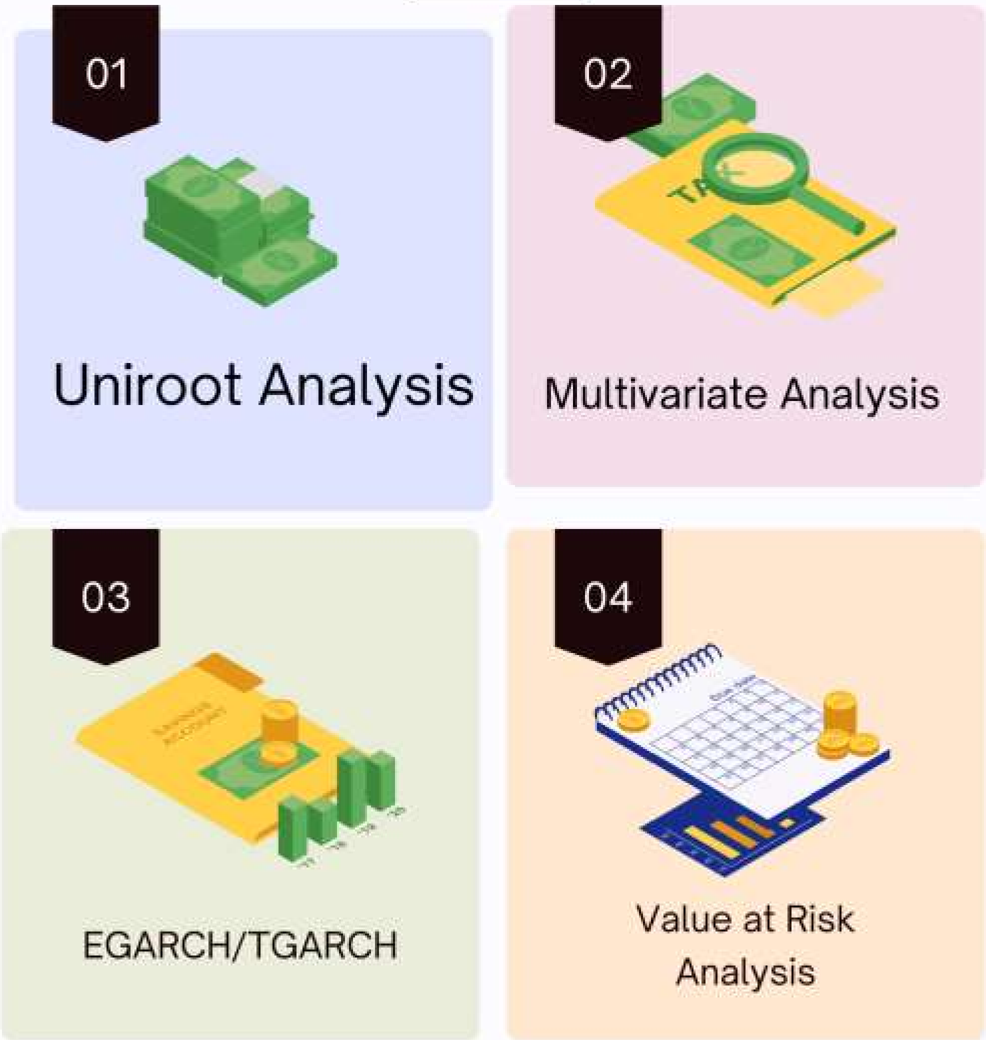
PROJECT REPORT

FINANCIAL ECONOMETRICS-FIN 620

APPLE

The cure for Apple is not cost-cutting. The cure for Apple is to innovate its way out of its current predicament.

Steve Jobs



Written by:
Abhimanyu Sheth, Aidana Bekboeva, Jayesh Kartik, Shivangi Sharma
Under the guidance of
Dr. Pallavi Pal

Contents

A) Overview of the asset and the market for the asset

Apple-An Introduction

Apple Inc. is an American multinational technology company that designs, develops, and sells consumer electronics, computer software, and online services. The company is headquartered in Cupertino, California and was founded in 1976 by Steve Jobs, Steve Wozniak, and Ronald Wayne. Apple is known for its innovative products such as the iPhone, iPad, MacBook, and Apple Watch, which have transformed the way people interact with technology. In addition to its hardware products, Apple also provides a range of online services such as the App Store, Apple Music, and iCloud. Apple is one of the world's largest technology companies with a market capitalization of over $2 trillion as of April 2023. The company has a global presence and operates in over 20 countries, employing over 150,000 people. It has consistently been ranked among the most valuable brands in the world and has been recognized for its design, innovation, and user experience. The company's success has been attributed to its focus on creating premium products and its ability to anticipate and meet the changing needs of consumers.

B) Properties of the time-series

i) Descriptive Statistics-Basic statistical examination of daily and monthly prices, returns, and Trade Volume

The analysis performed on the daily and monthly adjusted price of AAPL stock provides a summary of the key statistical measures for the dataset. The basicStats command reports the number of observations, minimum and maximum values, quartiles, mean and median, sum, standard error of the mean, confidence interval for the mean, variance, standard deviation, skewness, and kurtosis.

1. Analysis of Daily Prices-

- Based on the summary statistics of the daily data, we can see that the dataset is large (3,851 observations) and has no missing values.
- The minimum price is 2.38, and the maximum price is 180.68, indicating a wide range of prices over the period of the data. The median price is 25.10, which suggests that half of the data falls below this value, and half of the data falls above it.
- The mean price is 45.43, which indicates the average price over the period of the data. The quartiles provide information about the spread of the data. The first quartile (25% of the data) is at 11.94, while the third quartile (75% of the data) is at 51.69. This suggests that the data is moderately spread out with a wide range of values between the first and third quartiles.
- The variance and standard deviation provide information about the variability of the data. The dataset has a high variance of 2,385.04, which indicates that the data is spread out and has a high degree of variability. The standard deviation of 48.84 indicates that the data is spread out around the mean by about 49 units on average.
- The skewness value of 1.37 indicates that the distribution of the data is skewed to the right. This means that there are more observations with high values than low values, and the tail of the distribution extends towards higher prices. The kurtosis value of 0.49 indicates that the distribution is relatively flat compared to a normal distribution, which means that there are fewer extreme values in the data than we would expect in a normal distribution.

2. Analysis of Monthly Prices-

- The dataset consists of 184 observations with no missing values. The minimum price is 2.59, and the maximum price is 176.28. The median price is 25.07, and the mean price is 46.02.
- The first quartile is at 12.20, and the third quartile is at 52.96. These statistics indicate that the data is moderately spread out with a wide range of values between the first and third quartiles.
- The variance of the data is 2,461.41, and the standard deviation is 49.61, which indicate a high degree of variability in the monthly prices of AAPL stock.

- The skewness value of 1.35 indicates that the distribution is skewed to the right, which means that there are more observations with high values than low values, and the tail of the distribution extends towards higher prices.
- The kurtosis value of 0.41 indicates that the distribution is relatively flat compared to a normal distribution, which means that there are fewer extreme values in the data than we would expect in a normal distribution.

3. Analysis of Daily Returns-

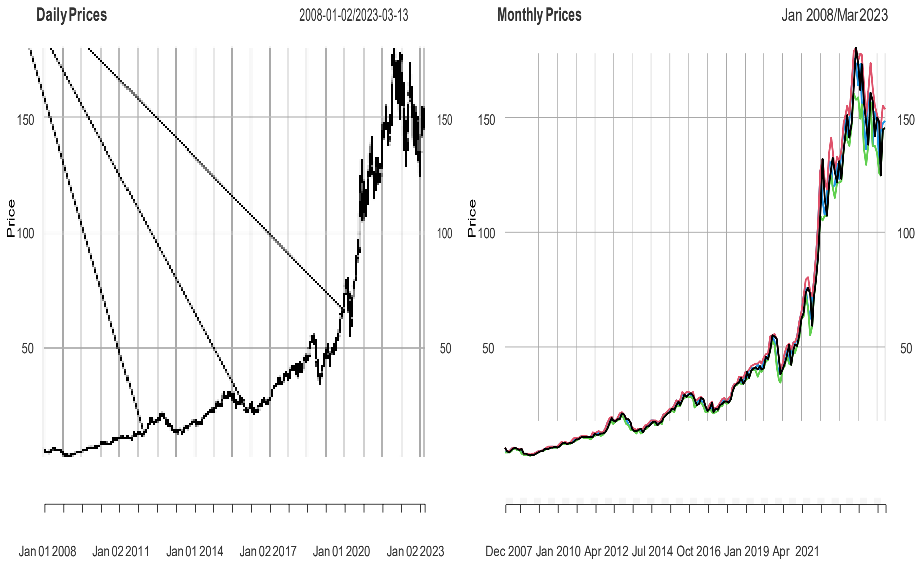
- The dataset consists of 3,850 observations with no missing values. The minimum log return is -0.197, and the maximum log return is 0.130. The median log return is 0.001, and the mean log return is 0.0009, which indicates that the overall trend in the data is relatively flat.
- The first quartile is at -0.008, and the third quartile is at 0.011, which indicates that most of the log returns are within a relatively narrow range. The variance of the log returns is 0.0004, and the standard deviation is 0.020, which indicates a low degree of variability in the daily log returns of AAPL stock.
- The skewness value of -0.416 indicates that the distribution is slightly skewed to the left, which means that there are more negative log returns than positive log returns. The kurtosis value of 6.907 indicates that the distribution has a high peak and heavy tails, which means that there are a few extreme log returns in the dataset.

4. Analysis of Monthly Returns-

- The dataset consists of 183 observations with no missing values. The minimum monthly return is -0.399, and the maximum monthly return is 0.194. The median monthly return is 0.028, and the mean monthly return is 0.020, which indicates that the overall trend in the data is positive, and the stock has increased in value on average over the time covered by the data.
- The first quartile is at -0.031, and the third quartile is at 0.075, which indicates that there is a relatively wide range of monthly returns. The variance of the monthly returns is 0.0074, and the standard deviation is 0.086, which indicates a moderate degree of variability in the monthly returns of AAPL stock.
- The skewness value of -0.747 indicates that the distribution is slightly skewed to the left, which means that there are more negative monthly returns than positive monthly returns. The kurtosis value of 2.030 indicates that the distribution has a moderate peak and relatively heavy tails, which means that there are some extreme returns in the dataset, but not as many as in the daily log returns.

ii)Visual behavior of daily and monthly prices, returns, and trade volume

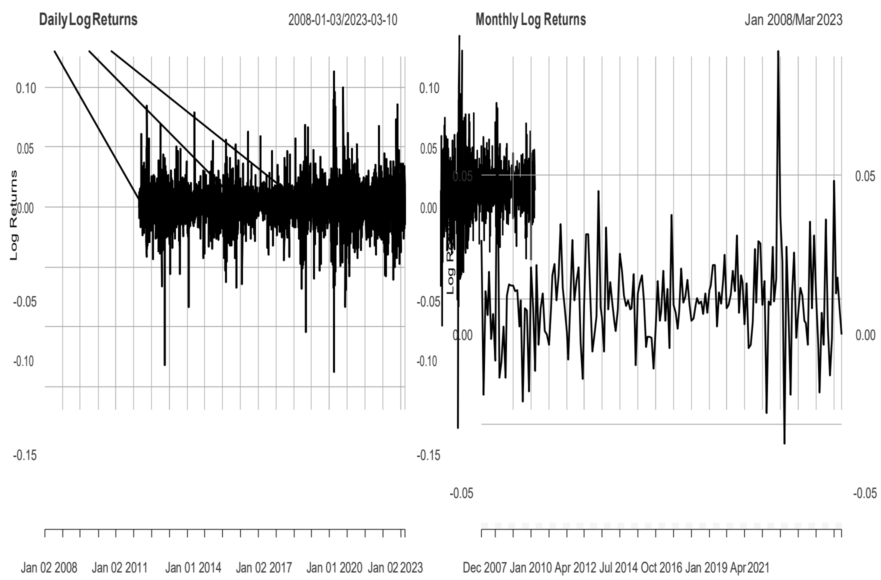
1. Visual behavior of daily and monthly prices-



Analysis-

- In general, we can observe that there is an upward trend in the price of Apple, particularly in the long-term monthly data. There are some periods of volatility where the price either increased or decreased rapidly, particularly in the daily data.
- Comparing the daily and monthly data, we can see that the daily data appears more volatile and has more noise compared to the monthly data. This makes sense, as daily data is influenced by a larger number of factors including news, investor sentiment, and overall market conditions, while monthly data may be influenced by broader economic trends and the company's financial performance.
- Overall, the plot provides a useful visualization of the trends and patterns in Apple's prices over time and allows us to easily compare the daily and monthly data.

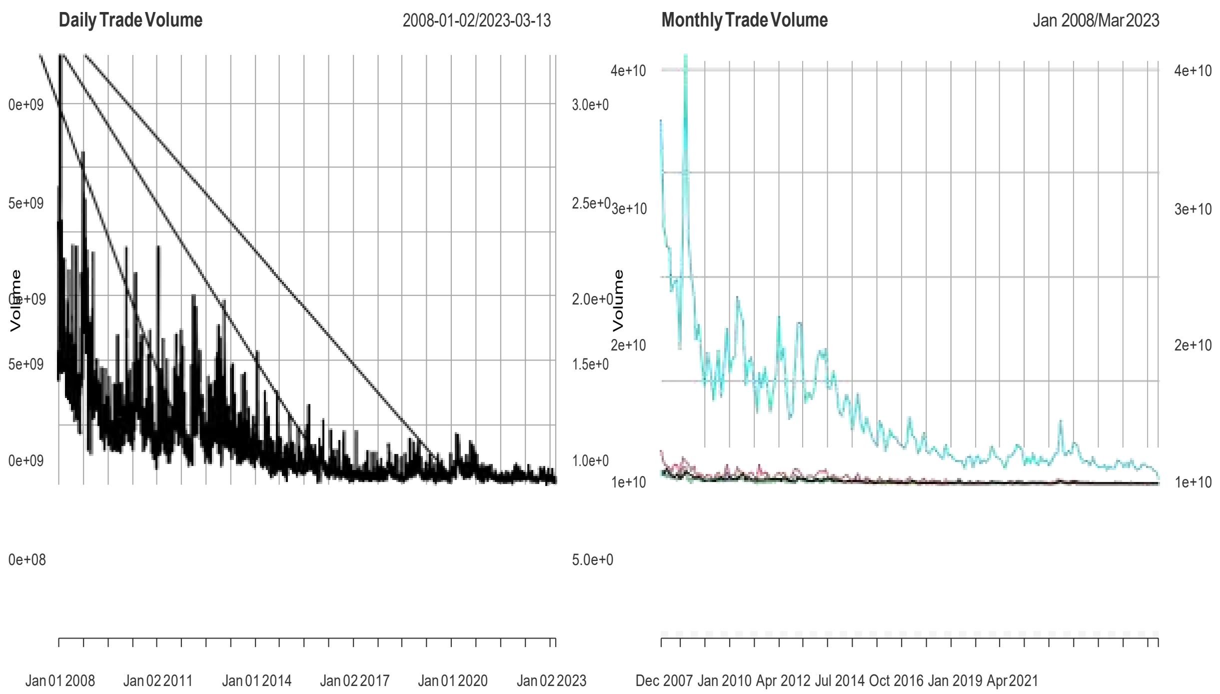
2. Visual behavior of daily and monthly log returns-



Analysis-

- Comparing the daily and monthly log returns, we can see that the daily log returns are more volatile and have more noise compared to the monthly log returns. This is consistent with our understanding of financial markets, where short-term fluctuations are more unpredictable and influenced by a larger number of factors compared to longer-term trends.
- Overall, the plot provides a useful visualization of the trends and patterns in Apple's log returns over time and allows us to easily compare the daily and monthly data.
- It can also be used to identify specific periods of high or low volatility in Apple's log returns, which could be useful for investors or analysts seeking to understand the performance of the stock over time.

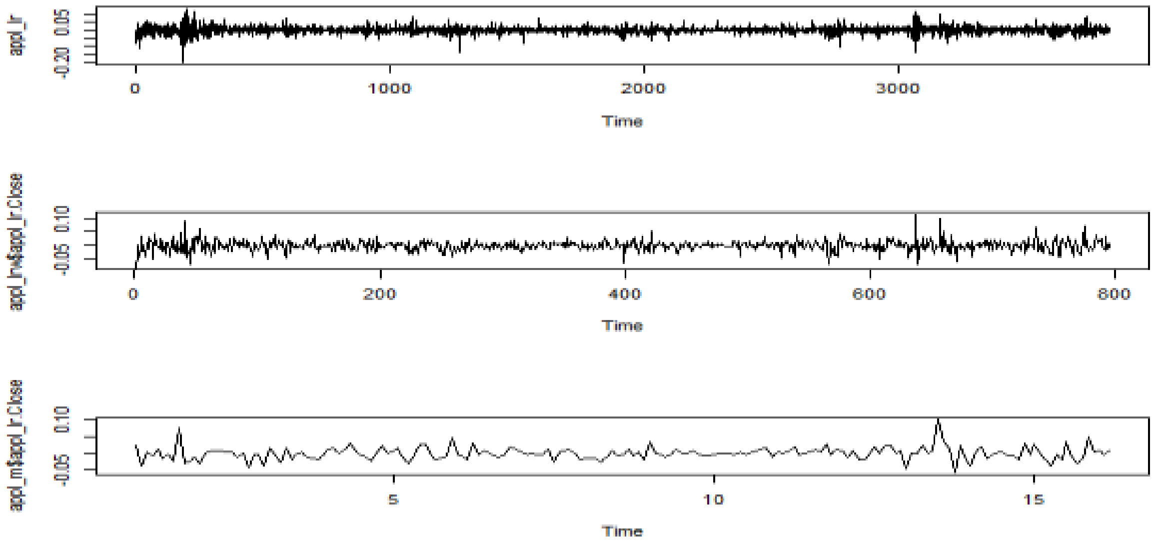
3. Visual behavior of daily and monthly volumes-



Analysis-

- Looking at the daily traded volumes, we can see that there are a few notable spikes in volume that occur periodically over the time period depicted. These spikes in volume typically occur during times of increased market volatility or during key events affecting the company or broader market. In general, daily traded volumes tend to be highly variable, reflecting the short-term fluctuations in market activity.
- On the other hand, the monthly traded volumes tend to be more stable and less volatile. This is consistent with the fact that monthly data aggregates daily data over a longer time, smoothing out some of the short-term fluctuations that can occur in the daily data.

Relationship between returns over different frequencies (daily, weekly, monthly)



Analysis-

Daily Returns:

- The average daily return for Apple from January 2008 until yesterday is 0.033%
- It has a standard deviation of 1.74%.
- Apple's daily returns are relatively volatile, with a wide range of positive and negative returns.

Weekly Returns:

- The average weekly return for Apple from January 2008 until yesterday is 0.22%,
- It has a standard deviation of 3.81%.
- Lesser volatile than daily returns but still have a relatively wide range of positive and negative returns.

Monthly Returns:

- The average monthly return for Apple from January 2008 until yesterday is 1.68%
- It has a standard deviation of 7.22%.
- Lesser volatile than both daily and weekly returns and tend to be more positive.

iii)UNIT ROOT TEST ON PRICES

Result ofT-test

```
        One Sample t-test

data:  appl
t = 57.438, df = 3823, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 43.09332 46.13917
sample estimates:
mean of x
 44.61624
```

Analysis:

Clearly p-value is less than alpha at 5% hence we can reject the null hypothesis and can accept the alternative hypothesis that suggests that the true mean is not equal to 0.

Result of Jarque-Bera Normality Test

```
Title:
 Jarque - Bera Normalality Test

Test Results:
  STATISTIC:
    X-squared: 1325.1148
  P VALUE:
    Asymptotic p Value: < 2.2e-16
```

Analysis:

When we used Jarque-Bera test, again the p-value is less than alpha at 5% hence we can reject the null hypothesis and can accept the alternative hypothesis that suggests that true mean is not equal to 0 and normality does not exist.

Result of Skewness and Kurtosis Test-

```
sk_p = skewness(appl);
T <- length(appl);
tst = abs(sk_p/sqrt(6/T))
pv  <-  2*(1-pnorm(tst))
pv

## [1] 0
## attr(,"method")
## [1] "moment"

kt_p  <-  kurtosis(appl)
tst <-abs(kt_p/sqrt(24/T))
```

```
pv <- 2*(1-pnorm(tst))
pv

## [1] 7.771561e-15
## attr(,"method")
## [1] "excess"
```

Analysis:

- We can see that skewness is equal to 0 which means that it is symmetrical i.e., its left and right sides are mirror images.
- It can be inferred that here it's excess kurtosis and therefore there are higher chances of outliers.

RESULT OF LJUNG BOX TEST-

```
        Box-Ljung test

data:  price_m
X-squared = 827.69, df = 5, p-value < 2.2e-16
```

Analysis:

Clearly p-value is less than alpha at 5% hence we can reject the null hypothesis and can accept the alternative hypothesis that suggests that ACF is not equal to 0. This also means that the data is not independently distributed and exhibits serial correlation.

UNIT ROOT TEST ON RETURNS

T-test on Returns

(From here on whenever I am writing returns, I am taking it as daily)

```
        One Sample t-test

data:  appl_lr
t = 2.5888, df = 3822, p-value = 0.009668
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 0.0002045012 0.0014809893
sample estimates:
   mean of x 
0.0008427453 
```

Analysis:

Clearly p-value is less than alpha at 5% hence we can reject the null hypothesis and can accept the alternative hypothesis that suggests that the true mean is not equal to 0.

Jarque-Bera Test on Returns

```
Title:
 Jarque - Bera Normalality Test

Test Results:
  STATISTIC:
    X-squared: 7664.3261
  P VALUE:
    Asymptotic p Value: < 2.2e-16
```

Analysis:

When we used Jarque-Bera test, again the p-value is less than alpha at 5% hence we can reject the null hypothesis and can accept the alternative hypothesis that suggests that true mean is not equal to 0 and normality does not exist.

Skewness and Kurtosis Test on Returns

```
> #Perform skewness and kurtosis test on returns
> sk_r = skewness(appl_lr);
> T <- length(appl_lr);
> tst = abs(sk_r/sqrt(6/T))
> pv <- 2*(1-pnorm(tst))
> pv
[1] 0
attr(,"method")
[1] "moment"
>
> kt_r <- kurtosis(appl_lr)
> tst <-abs(kt_r/sqrt(24/T))
> pv <- 2*(1-pnorm(tst))
> pv
[1] 0
attr(,"method")
[1] "excess"
```

Analysis:

a) We can see that skewness is equal to 0 which means that it is symmetrical i.e., its left and right sides are mirror images.

b)It can be inferred that there is an excess kurtosis and therefore there are higher chances of outliers.

Ljung Box Test on Returns

```
        Box-Ljung test

data:  appl_lr
X-squared = 30.521, df = 8, p-value = 0.0001709
```

Analysis: Clearly p-value is less than alpha at 5% hence we can reject the null hypothesis and can accept the alternative hypothesis that suggests that ACF is not equal to 0. This also means that the data is not independently distributed and exhibits serial correlation.

C) ARIMA Model on Prices and Returns

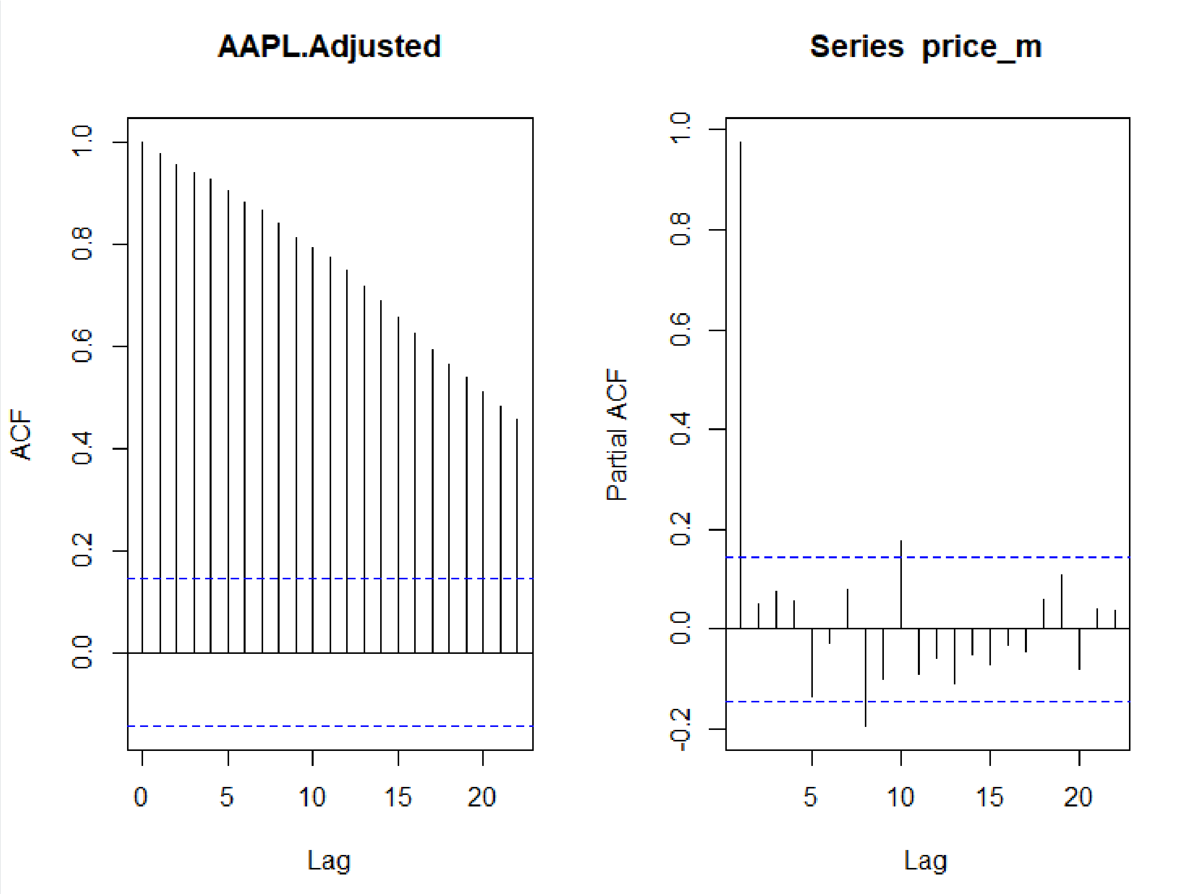
ADF TEST ON PRICES AND RETURNS

```
        Augmented Dickey-Fuller Test

data:  na.omit(appl$AAPL.Adjusted)
Dickey-Fuller = -1.4614, Lag order = 15, p-value = 0.8064
alternative hypothesis: stationary

> adf.test(appl_lr)

        Augmented Dickey-Fuller Test

data:  appl_lr
Dickey-Fuller = -14.648, Lag order = 15, p-value = 0.01
alternative hypothesis: stationary
```

Analysis:

a) For Apple prices, the p -value is greater than alpha at 5%, hence there is no reason to reject the null hypothesis. So, the time series for Apple prices is in fact non-stationary.

b) For Apple log returns, the p-value is less than alpha at 5%, hence we can reject the null hypothesis and take that the series is stationary. ACF,PACF,EACF OF PRICES –



Analysis

From the ACF of monthly prices, there is trend in the closing price of Apple.
We can remove the trend from the graph by taking the first difference.
The PACF of the monthly prices cuts off at lag 8 and lag 10.

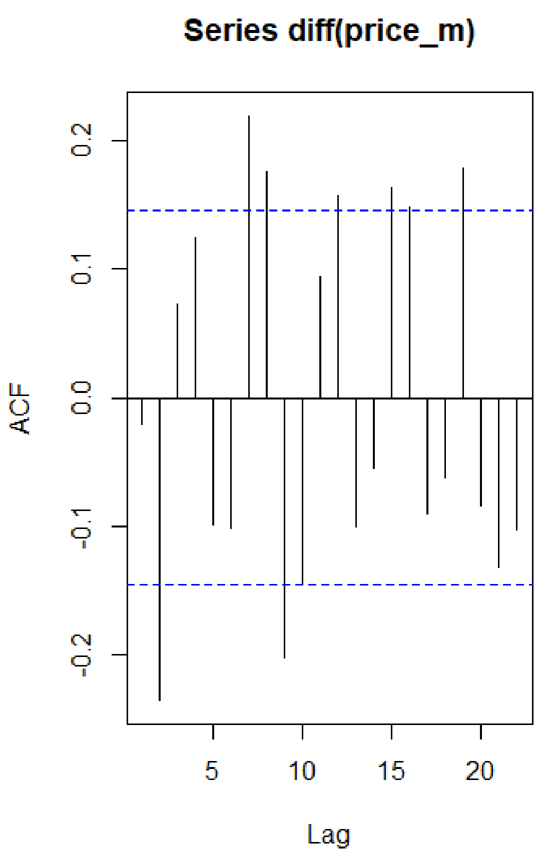
We have plotted the EACF of the price-

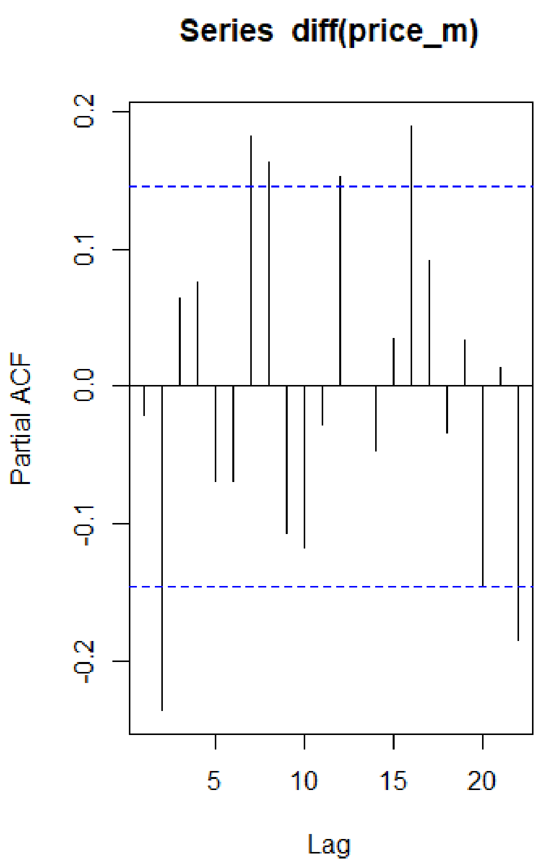
```
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 x x x x x x x x x x x  x  x  x
1 o x o o o o x x x x o  x  o  o
2 o x x o o o x x x x o  x  o  o
3 x x o o o o o o o o o  x  o  o
4 x x o o o o o o o o o  o  o  o
5 x x o o o o o o o o o  o  o  o
6 x o x o o o x o o o o  o  o  o
7 x o x x o o x o o o o  o  o  o
>
```

Analysis

From the EACF of the price, we chose these models to fit the model – (1,2), (1,3), (1,0), (2,0), (2,3), (2,4), (3,2), (3,3)

But since, there was trend in the price graph, we differenced the graph one time and repeat the above analysis once again-





Analysis

From the above ACF and PACF plot, after one differencing, lag1, lag6, lag7 are significant and hence we can easily fit AR of these orders or MA of these orders too. Again we performed the dickey fuller test, and it is very evident from the result that p value is less than the significant level, hence, we can reject the null hypothesis concluding the price graph to be stationary.

```
        Augmented Dickey-Fuller Test

data:  diff(price_m)
Dickey-Fuller = -6.4661, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary
```

Again, we plotted the EACF of the price-series:

```
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 o x o o o o x x x x o  x  o  o
1 o x o o o o x x x x o  o  o  o
2 x x o o o o o o o o o  o  o  o
3 x x o 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 o o o o o o o o o  o  o  o
6 x x o x o x o o o o o  o  o  o
7 x x x x x x x o o o o  o  o  o
```

Order-(0,2), (0,3), (1,3), (1,4), (2,2) (3,5)

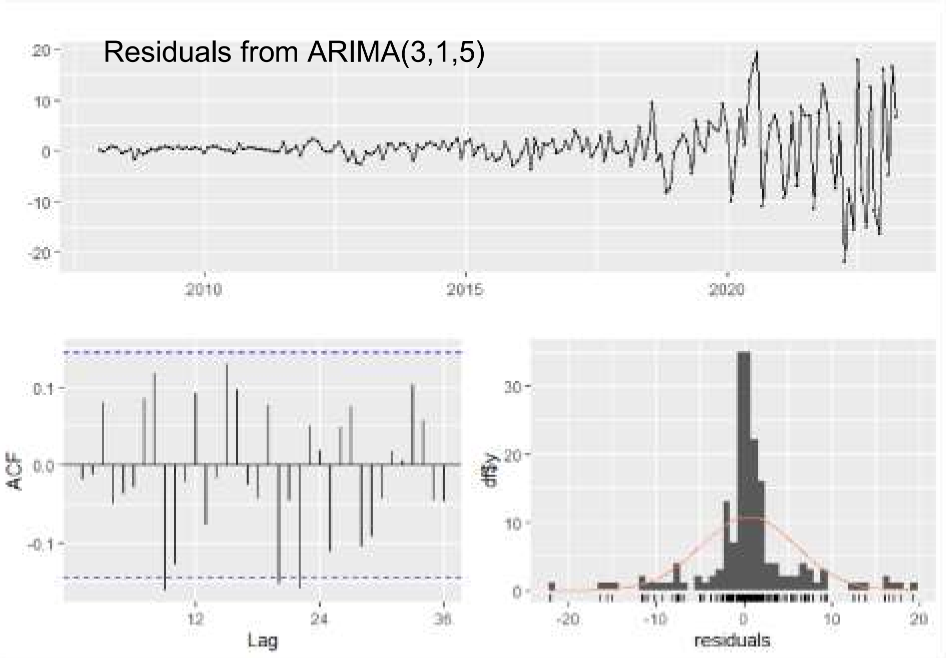
We made models of these orders with a difference of 1 in the middle.

FINALISING THE ORDER OF THE MODEL

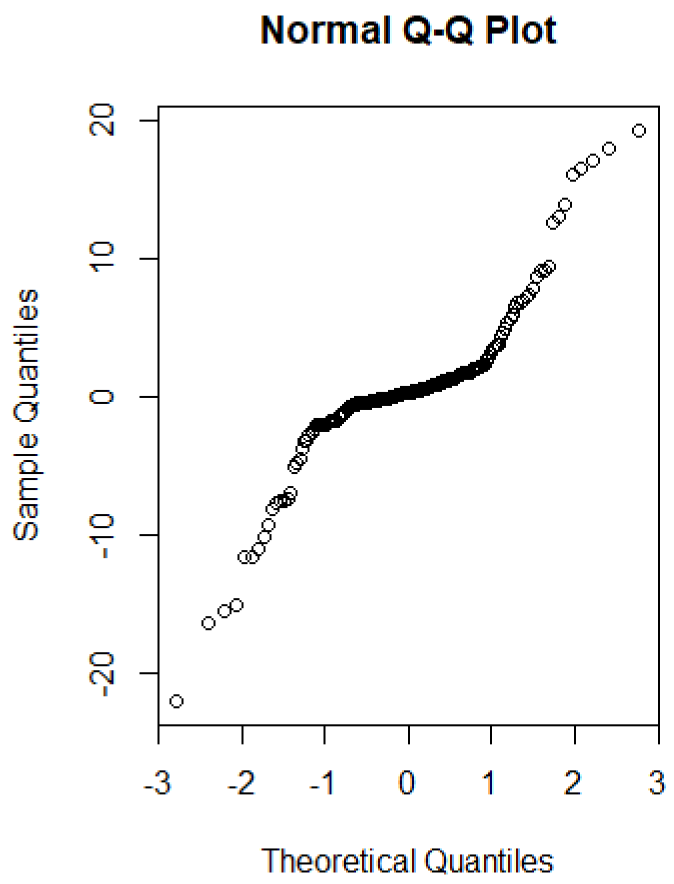
Here is the link of the all the models-
https://docs.google.com/document/d/1RmJFU4oa3AoCcZijXywwBED3JoXjlEHg/edit?usp=sharing&ouid=101276239542398720241&rtpof=true&sd=true

M5

```
> m5

Call:
arima(x = price_m, order = c(3, 1, 5))

Coefficients:
         ar1      ar2     ar3      ma1     ma2      ma3     ma4
      0.3466  -0.2811  0.9002  -0.3195  0.1371  -0.9581  0.0826
s.e.  0.0560   0.0702  0.0562   0.0979  0.1131   0.0631  0.0961
         ma5
      0.1518
s.e.  0.0945

sigma^2 estimated as 29.2:  log likelihood = -574.06,  aic = 1164.12
>
```




```
> Box.test(m5$residuals, lag = 10, type = "Ljung")

        Box-Ljung test

data:  m5$residuals
X-squared = 14.469, df = 10, p-value = 0.1527
```

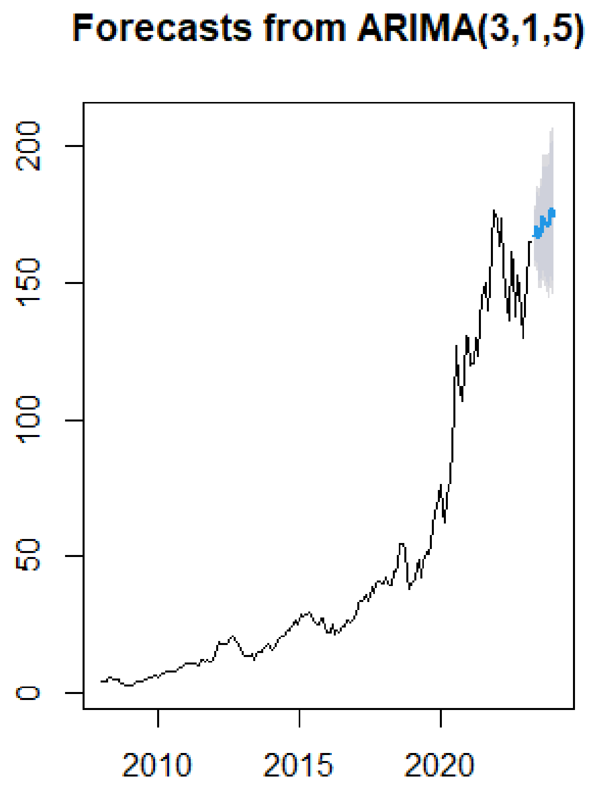



```
> print(aic_p)
[[1]]
[1] 1167.228

[[2]]
[1] 1168.46

[[3]]
[1] 1147.569

[[4]]
[1] 1149.552

[[5]]
[1] 1164.115
```

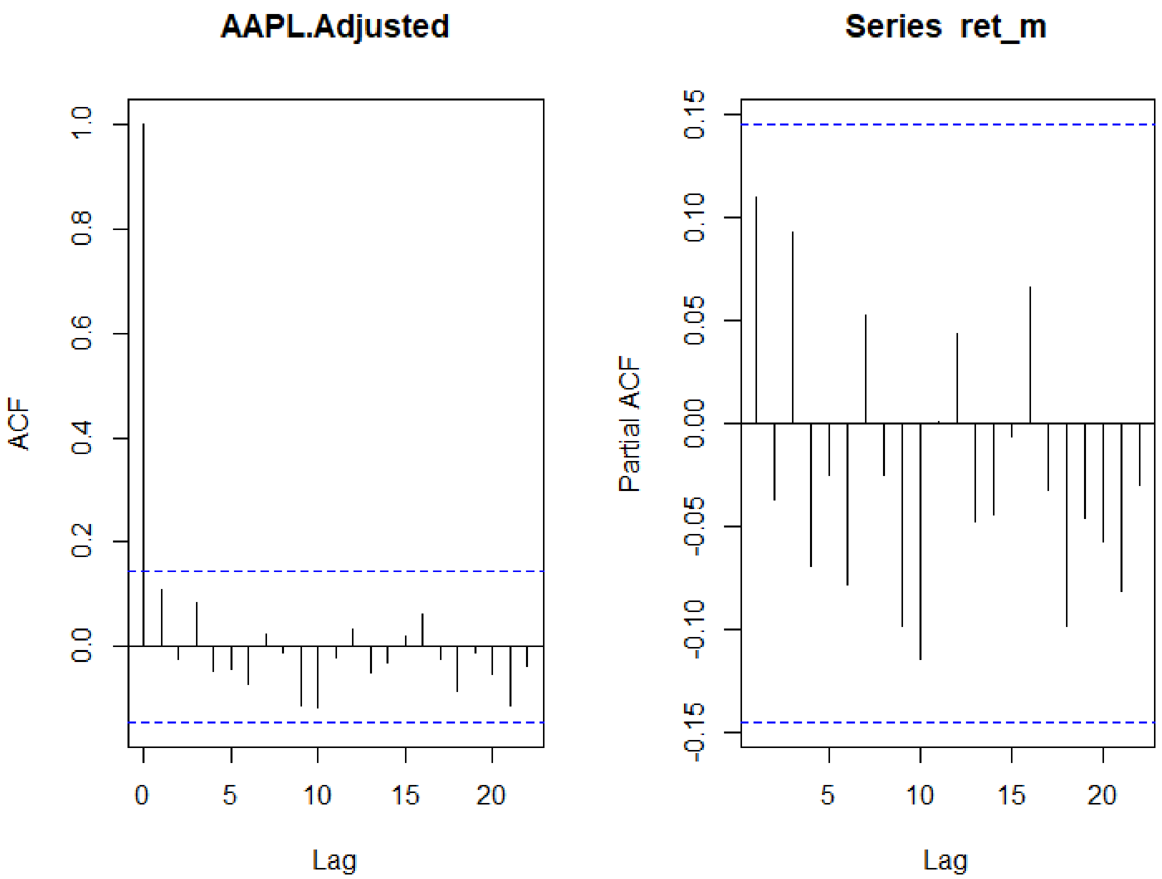
Based on the above chart of AIC, we chose the 5th model as our preferred model for prices.

It has the lowest AIC, and the residual of this model is also not significant.

FORECASTING OF PRICES-



ACF, PACF, EACF of RETURNS-



Analysis

From the ACF and PACF of the return graph, we can clearly see that all the lags in both the plots are insignificant. Hence, its difficult to choose which AR model should be considered. But its clear from the graph that there is no trend or seasonality present-

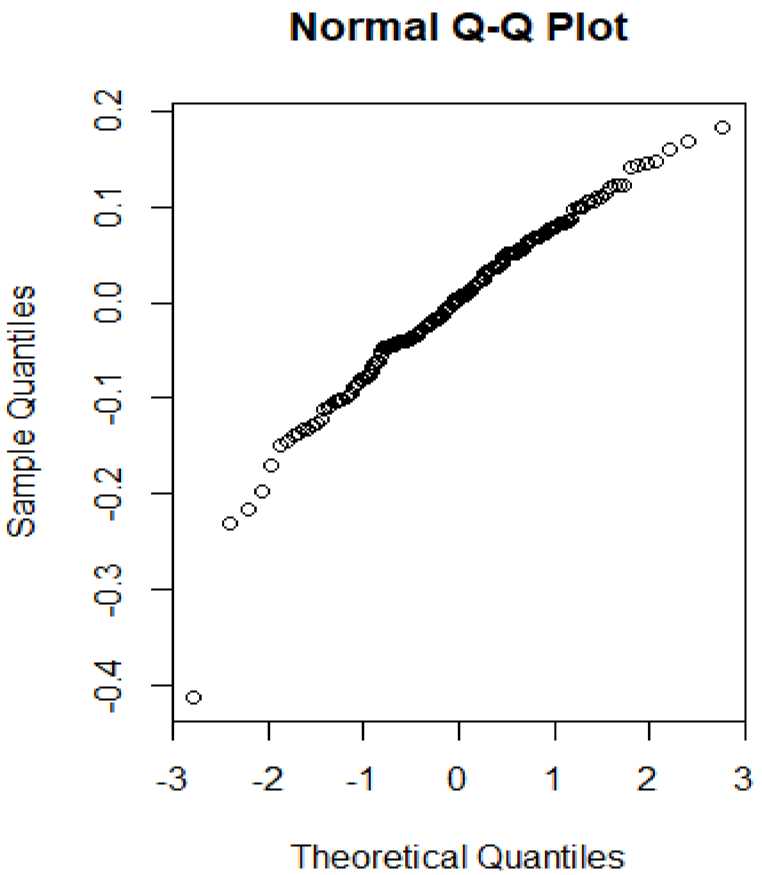
```
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 o o o o o o o o o o o  o  o  o
1 x o o o o o o o o o o  o  o  o
2 x x o o o o o o o o o  o  o  o
3 x x x o o o o o o o o  o  o  o
4 x x x x o o o o o o o  o  o  o
5 x x x o o o o o o o o  o  o  o
6 x o x o o o o o o o o  o  o  o
7 x o x x o o o o o o o  o  o  o
```

Analysis

The chosen models from the EACF returns are- (0,0), (0,1), (0,2), (1,1), (1,2), (0,3), (1,3), (2,2)

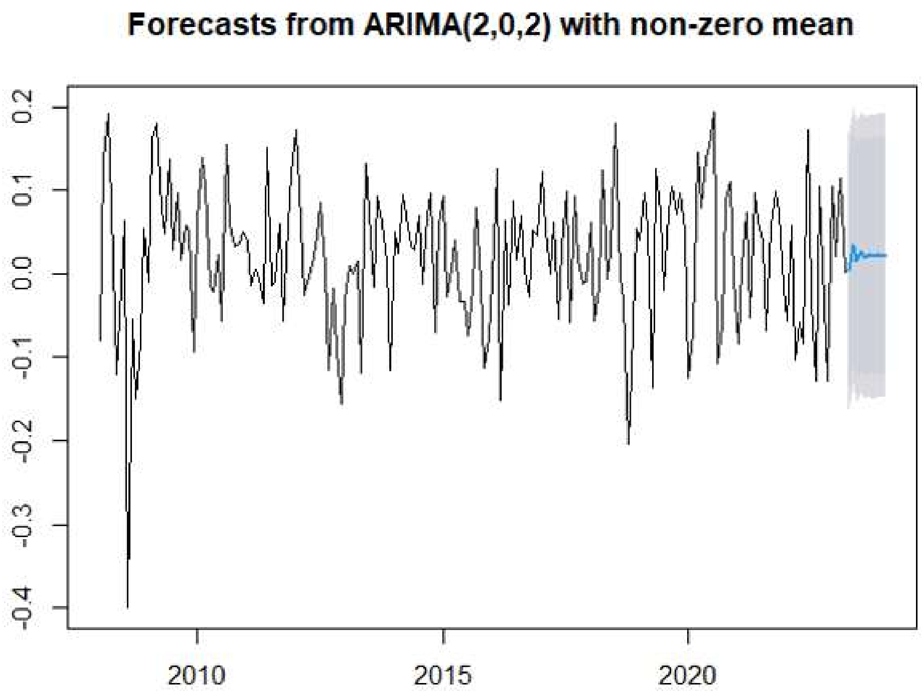
We made ARIMA models with the above orders without differencing. The following data frame shows the AICs of these models. Next, we will be seeing the plots of the residuals and then will perform a Ljung Box test on residuals to see if the residuals are not that significant.

```
> print(aic_r)
[[1]]
[1] -374.9119

[[2]]
[1] -375.2308

[[3]]
[1] -373.634

[[4]]
[1] -375.1882

[[5]]
[1] -373.2187

[[6]]
[1] -373.3828

[[7]]
[1] -371.7839

[[8]]
[1] -376.2861
```



Link to file containing all the plot of the return series-
https://docs.google.com/document/d/1FHxdQ7p2X8nZeftNqDi3WB3UVUCvin0o/edit?usp=sharing&ouid=101276239542398720241&rtpof=true&sd=true
Based on the AIC of different models, we chose the 13th model as our preferred model. We did residual testing and performed Ljung box test on the residuals. Finally, we plot the QQ plot of the residuals to check if the residuals are not significant. Hence, we are going to fit m13 to our data.

Forecasting based on the selected model



D)Multivariate Analysis: VAR

Comparison with S&P Market Data-

To create a VAR model, we have considered the daily price data of apple stock and compared its movement with S&P 500.

Reasons for considering the apple stock price are-

- Apple stock prices and the S&P 500 move in the same direction.
- By analyzing both together, you can identify patterns and trends in the broader market, and how they relate to Apple specifically.
- S&P 500 is used as a benchmark for performance of the stock market, so analyzing Apple's performance relative to it can give you a sense of how well the company is doing compared to its peers.

Why Prices?

Prices are not stationary; hence we cannot apply the VAR method directly. So, we took Log of Prices, and the movement happens to be the same direction again.

- It means that the prices of assets, in their original form, are not stationary, and it would not be appropriate to use VAR models directly to analyze their relationships.
- However, by taking the logarithm of prices, we can transform the non-stationary series into a stationary series, which allows us to apply VAR models to analyze the relationships between the assets. Logarithmic transformation is often used in finance because it has the effect of stabilizing the variance of the series over time.
- By taking the logarithm of prices, we can obtain a stationary series, and the movement of the series in its transformed form would have the same direction as the original non-stationary series. This means that the logarithmic transformation preserves the direction of movement of the original series while making it stationary, which makes it suitable for the application of VAR models to analyze the relationships between the assets.

3.Comparing Log Prices of Apple and S&P 500

- Looking at the graph of the log prices of Apple and S&P 500 from 2008-01-01 to 2023-03-31, we can see that both Apple and S&P 500 have had significant growth over the years, but with some fluctuations and periods of decline.
- Starting with Apple, we can see that the log prices of the company's stocks have steadily increased over time. However, there are some noticeable periods of decline, such as in 2008-2009 during the financial crisis, and again in 2015-2016. In both cases, the log prices eventually recovered and continued to grow.
- Looking at the S&P 500, we can also see that the log prices have grown over the years, but with some significant fluctuations. There were notable dips in 2008-2009 during the financial crisis, and again in 2020 during the COVID-19 pandemic.
- However, in both cases, the index eventually recovered and continued to grow. Comparing the two graphs, we can see that there are some periods where Apple outperforms the S&P 500, and other periods where the opposite is true. For example, from 2008-2012, Apple's log prices grew at a much faster rate than the S&P 500, while from 2013-2016, the S&P 500 outperformed Apple. More recently, from 2017-2023, Apple has once again outperformed the S&P 500. Overall, the graphs suggest that both Apple and the S&P 500 have experienced significant growth over the past decade and a half, despite some periods of decline and volatility. Investors in both Apple and the broader market have seen positive returns over this time, but with some variations in performance over different periods.

4.Cross correlation matrices

Fig 1. – Sample ACF of Apple log prices and log

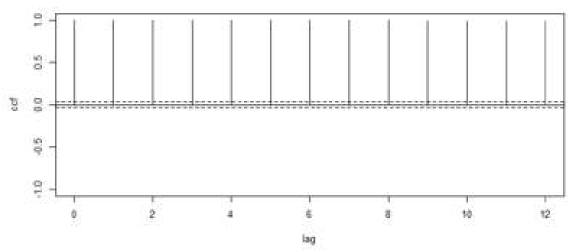


Fig 2. - Cross correlation between apple price price of S&P 500

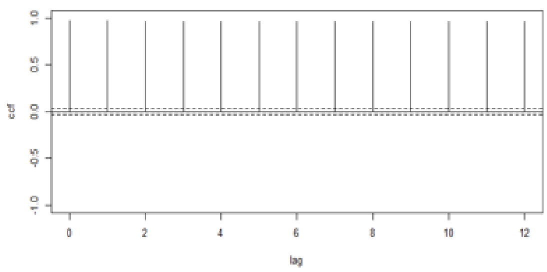
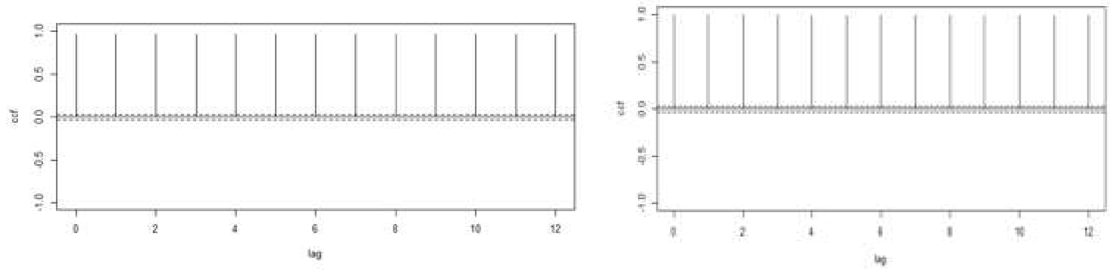


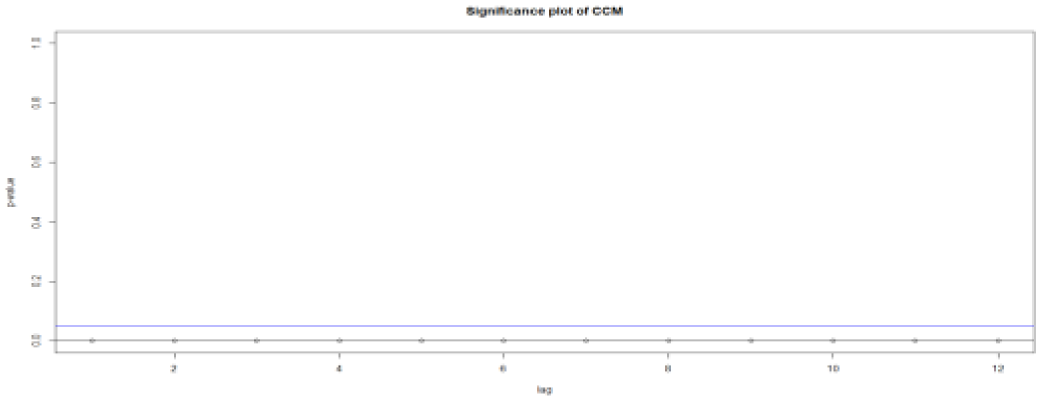
Fig 3. - Cross correlation between apple price and log price of S&P 500

Fig 4. – Sample ACF of S&P500 log prices



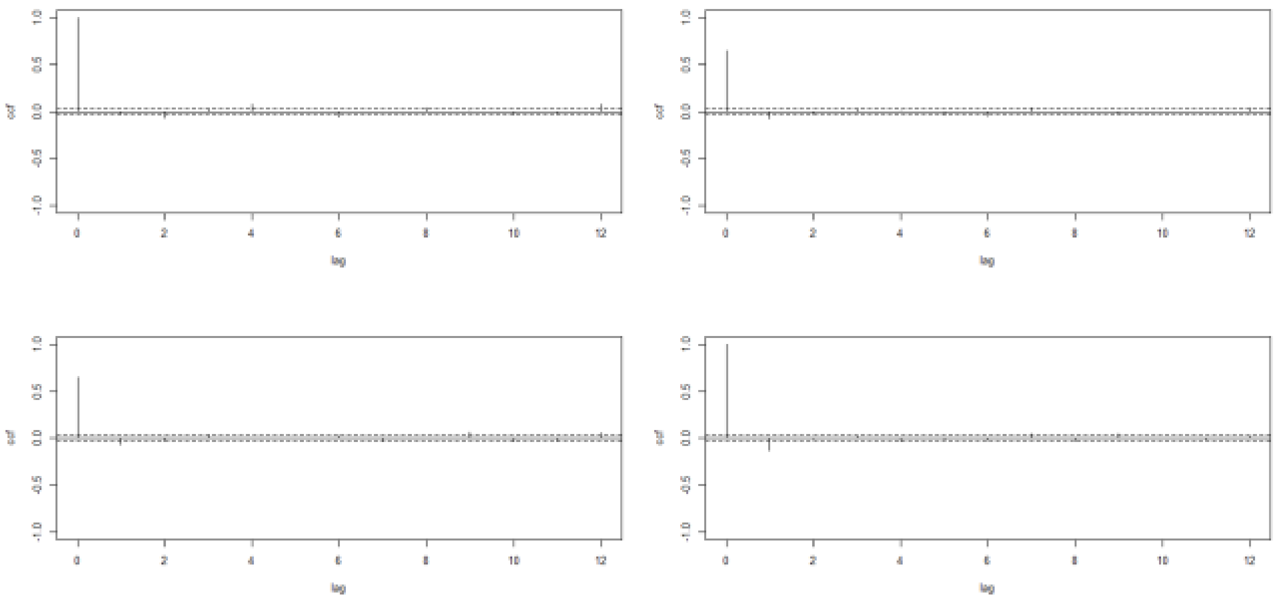
Analysis
All the lags look significant. Hence, we will take the first difference.

5.Cross correlation matrices after differencing the prices



Analysis
After we difference the matrices of log of price, we see that the CCM looks better with less significant lags, meaning less dependence. CCM is still not the most informative method for determining the lag.

6.Plotting the cross-correlation matrices again



1.Analysis
After plotting the cross-correlation matrices, we saw that they are all significant. Difference the matrices of log of price, we see that the CCM looks better with less significant lags, meaning less dependence. Reasons why differencing has improved the results is-

- It might have removed the trend component of the data.
- It can reduce the impact of autocorrelation on the results.

2.Order specification

Order determination is the process of selecting the appropriate number of lags to include in a VAR model. Determined the order by examining the autocorrelation and partial autocorrelation functions of the variables in the system, and selecting the lag order that best captures the relationship between the variables while minimizing any remaining autocorrelation in the model residuals. Decision criteria for VAR was based on Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC).

RESULTS OF VAR ORDER DETERMINATION

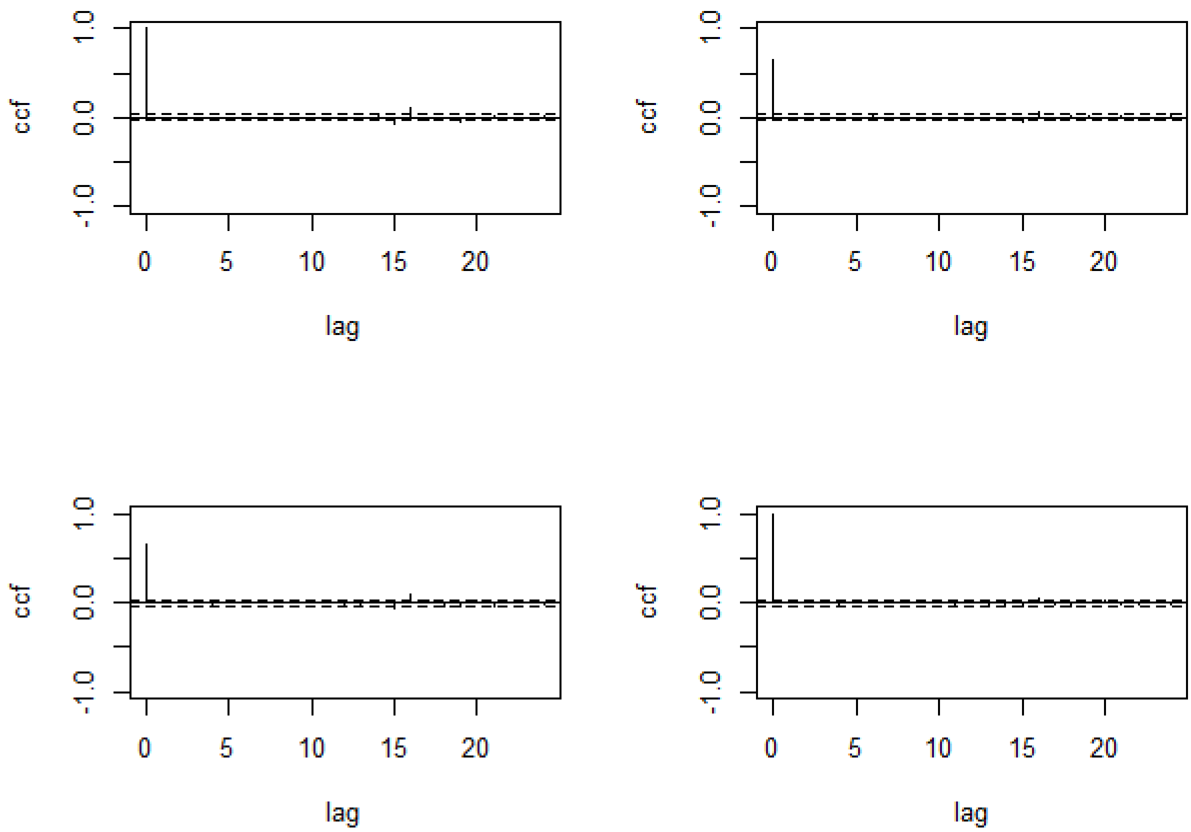
```
selected order: aic =  13
selected order: bic =  1
selected order: hq =  9
Summary table:
        p      AIC      BIC       HQ    M(p) p-value
 [1,]   0 -22.5776 -22.5776 -22.5776  0.0000  0.0000
 [2,]   1 -22.5931 -22.5866 -22.5908 67.4231  0.0000
 [3,]   2 -22.5958 -22.5828 -22.5912 18.1793  0.0011
 [4,]   3 -22.5952 -22.5757 -22.5883  5.7878  0.2156
 [5,]   4 -22.6071 -22.5812 -22.5979 53.6549  0.0000
 [6,]   5 -22.6071 -22.5747 -22.5956  8.0893  0.0884
 [7,]   6 -22.6214 -22.5825 -22.6076 62.5918  0.0000
 [8,]   7 -22.6247 -22.5793 -22.6086 20.5165  0.0004
 [9,]   8 -22.6249 -22.5730 -22.6065  8.8920  0.0639
[10,]   9 -22.6298 -22.5714 -22.6090 26.3894  0.0000
[11,]  10 -22.6295 -22.5646 -22.6065  6.9787  0.1370
[12,]  11 -22.6283 -22.5570 -22.6030  3.3846  0.4956
[13,]  12 -22.6306 -22.5528 -22.6030 16.7276  0.0022
[14,]  13 -22.6315 -22.5471 -22.6015 11.2121  0.0243
```

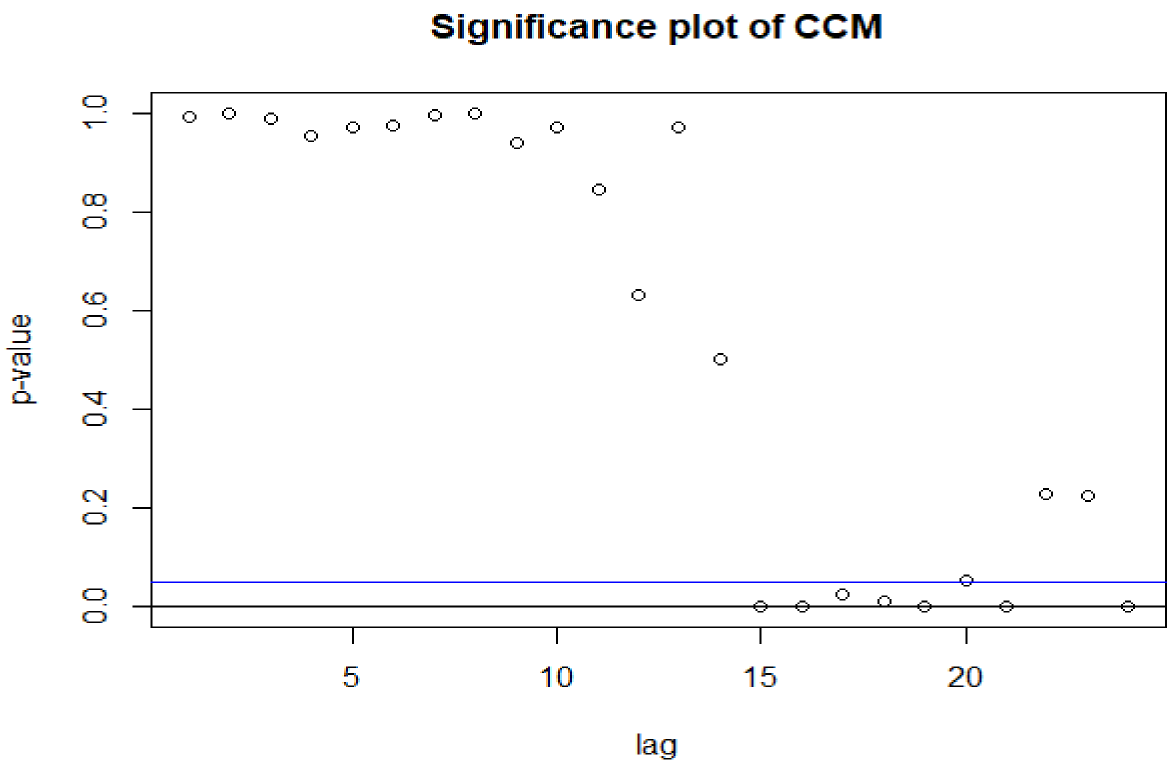
Based on the VAR order determination, we decided to make three models with p=13(selected based on AIC), p=1(selected based on BIC) and p=9(selected because of HQ). After building the three models m1, m2 and m3, we have summarized below the coefficients of all the three models.

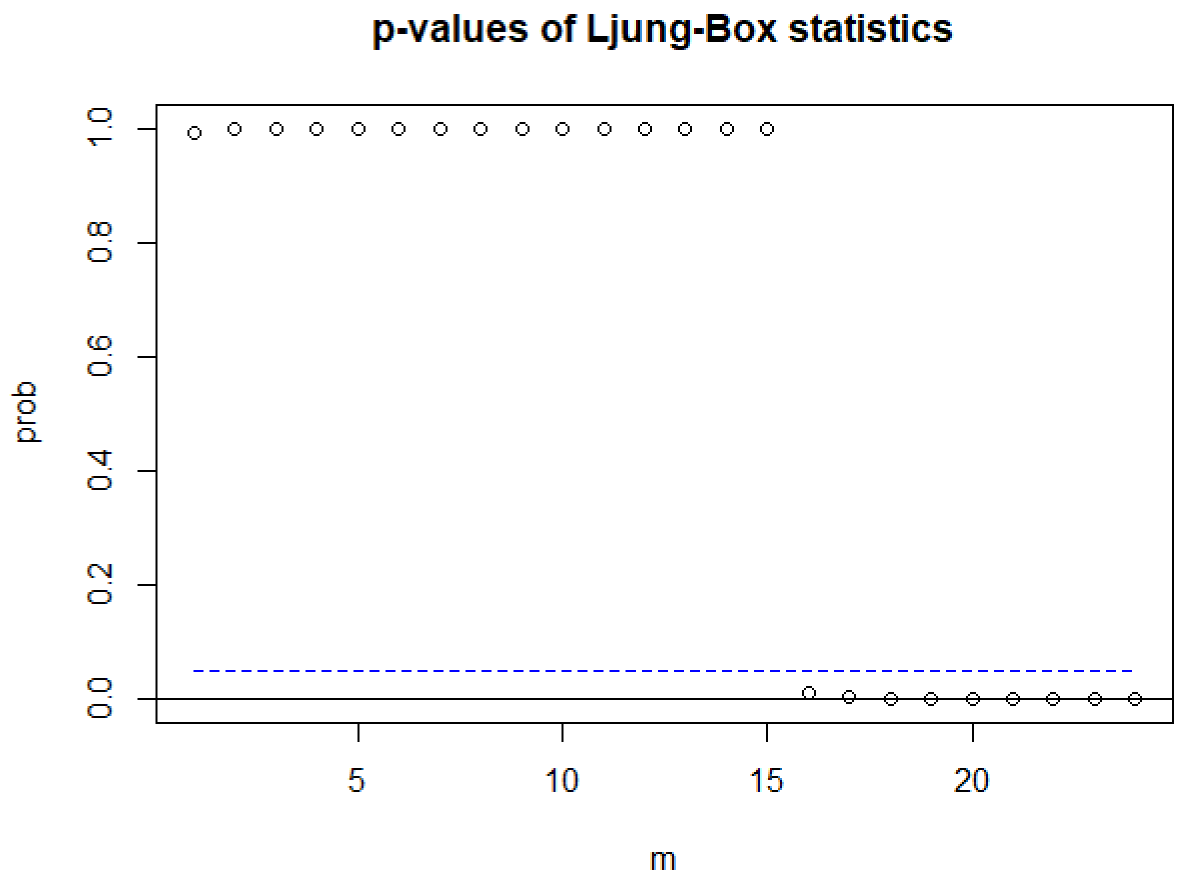
```
> m1$coef
                [,1]          [,2]
 [1,]  0.0002833919  3.993523e-05
 [2,]  0.0330411192  3.668810e-03
 [3,] -0.5223167477 -1.370609e-01
 [4,] -0.0500137574 -8.228773e-03
 [5,]  0.0109792153  6.907681e-03
 [6,]  0.0125470732  9.990121e-03
 [7,]  0.0657445528 -2.436036e-02
 [8,]  0.1123285924  7.431175e-03
 [9,] -0.3688890241 -5.754151e-02
[10,]  0.0245236115  1.638021e-03
[11,] -0.3194889912 -3.094869e-02
[12,] -0.0274995966  1.937562e-02
[13,] -0.1136900544 -9.085946e-02
[14,] -0.0476410080 -1.453052e-02
[15,]  0.3349576244  8.118265e-02
[16,]  0.0385007063  1.884796e-03
[17,] -0.1622148121 -2.546402e-02
[18,]  0.0278499700  1.115596e-02
[19,] -0.2104962617 -1.858560e-05
[20,] -0.0195642366 -9.328210e-03
[21,]  0.0610909596  4.044211e-02
[22,] -0.0226220895 -7.544973e-03
[23,]  0.0668431493  1.748475e-02
[24,]  0.0683876477  1.237444e-02
[25,] -0.0631314720 -2.081783e-02
[26,]  0.0558146689  1.868692e-03
[27,] -0.2464371017 -2.649418e-02
> 
> m2$coef
               [,1]          [,2]
[1,]  0.0002600392  3.817714e-05
[2,]  0.0358655708  5.366993e-03
[3,] -0.5144687573 -1.468194e-01
> 
> m3$coef
                [,1]          [,2]
 [1,]  0.0002803935  3.715653e-05
 [2,]  0.0369960855  4.131350e-03
 [3,] -0.5240525371 -1.395833e-01
 [4,] -0.0548005961 -7.759928e-03
 [5,]  0.0096529986  2.797233e-03
 [6,]  0.0151853509  9.489651e-03
 [7,]  0.0776581012 -1.773465e-02
 [8,]  0.1207637584  9.341278e-03
 [9,] -0.3868692806 -6.436775e-02
[10,]  0.0240898192  2.294750e-03
[11,] -0.3323693090 -3.779755e-02
[12,] -0.0332495932  1.678389e-02
[13,] -0.0912695302 -8.374346e-02
[14,] -0.0535370506 -1.482264e-02
[15,]  0.3623805835  8.270506e-02
[16,]  0.0513225222  4.044392e-03
[17,] -0.1905811916 -3.145586e-02
[18,]  0.0326970446  1.187518e-02
[19,] -0.1975838157  1.162125e-03
```

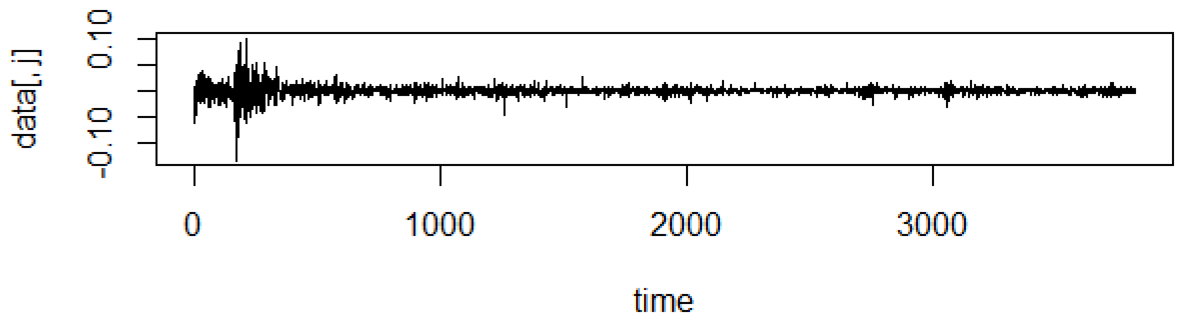
MODEL CHECKING

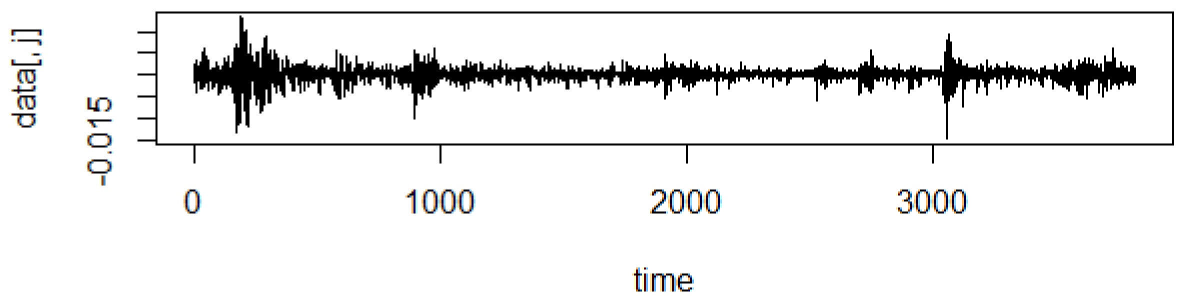
The three models will now be checked by running a code of MTSdiag on each of these models separately m1, m2 and m3 respectively. Below is the output of m1 after running MTSdiag-











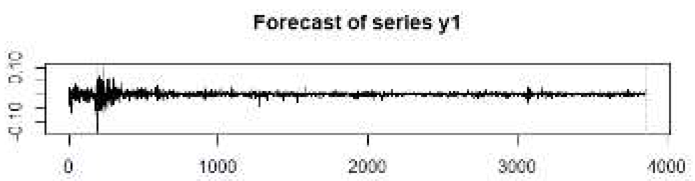
Link to other model MTSdiag results-https://docs.google.com/document/d/1f5rCaZzqp47SXT_rp8P3AaXEJTOidmZQ/edit?usp=sharing&ouid=101276239542398720241&rtpof=true&sd=true

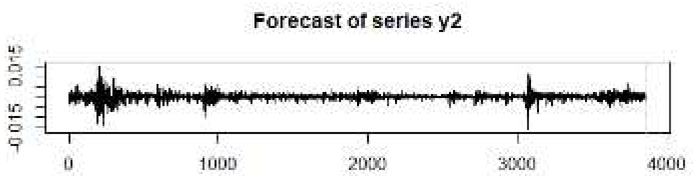
Comparing all the above models, we chose the model m1 to be the best fitted.

MODEL SELECTION

Based on the outputs given by the MTSDiag, I have finalized my m1 model to be the selected model. The m3 model can also be considered, but since this is more based on how we look to see the diagrams, we can select the m1 model to be the model selected for forecasting.

MODEL FORECASTING (using M1 Model)





E) Conditional variance analysis: Various types of GARCH models: GARCH Models

GARCH Model fitting on returns

The code is estimating GARCH models in R with different p, q (autoregressive and moving average lags).

```
#GARCH Model fitting on returns
library(rugarch)

#Selecting GARCH order
g1 = garch(ret_r, order = c(1,1))
g2 = garch(ret_r, order = c(1,2)) #Chosen model based on lowest AIC
g3 = garch(ret_r, order = c(2,2))
g4 = garch(ret_r, order = c(2,3))
g5 = garch(ret_r, order = c(3,2))
g6 = garch(ret_r, order = c(3,3))

aic_res <- list(AIC(g1), AIC(g2), AIC(g3), AIC(g4), AIC(g5), AIC(g6))
print(aic_res)

summary(g2)
#Ljung box test p>0.05, suggests no autocorrelation
```

AIC (Akaike Information Criterion)

We use AIC (Akaike Information Criterion) to choose a model because it helps to select the model that best balances the fit to the data and complexity of the model. AIC is a statistical measure of the quality of a model relative to the amount of information it uses, and it considers both the goodness of fit of the model and the number of parameters used in the model. The lower the AIC value, the more accurate the model is thought to be. As a result, the model with the lowest AIC value was picked as the optimal model in this circumstance.

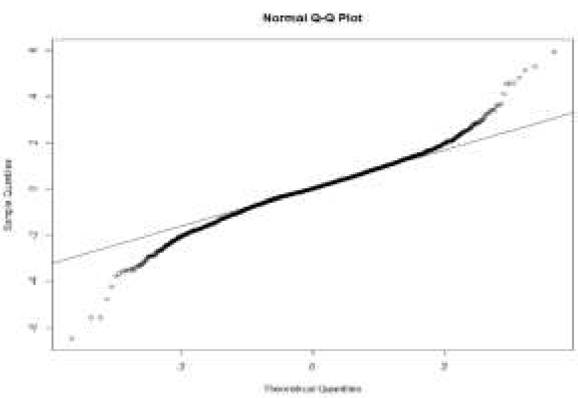
```
> print(aic_res)
[[1]]
[1] -20005.6

[[2]]
[1] -20015.11

[[3]]
[1] -20013.71

[[4]]
[1] -20004.72

[[5]]
[1] -20004.78

[[6]]
[1] -19996.08
```

Chosen model based on lowest AIC equals to -20015.11 is g2. After choosing the GARCH(1,2) model, running its summary provides us with important information about the model's fit and performance. The summary includes various statistical measures such as the coefficients of the model, their standard errors, t-values, p-values, and the likelihood ratio test results.

```
> summary(g2)

Call:
garch(x = ret_r, order = c(1, 2))

Model:
GARCH(1,2)

Residuals:
      Min        1Q    Median        3Q       Max 
-6.477265 -0.523972  0.003551  0.587940  5.941934 

Coefficient(s):
    Estimate  Std. Error  t value Pr(>|t|)    
a0 1.485e-05   1.687e-06    8.802  < 2e-16 ***
a1 1.114e-01   1.863e-02    5.980 2.23e-09 ***
a2 3.258e-10   1.928e-02    0.000        1    
b1 8.507e-01   1.349e-02   63.063  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Diagnostic Tests:
	Jarque Bera Test

data:  Residuals
X-squared = 1284.6, df = 2, p-value < 2.2e-16


	Box-Ljung test

data:  Squared.Residuals
X-squared = 0.26449, df = 1, p-value = 0.6071

> |
```

As we can see in the Ljung box test, p-value > 0.05 (0.6071), which suggests that there is no autocorrelation.

Checking normality assumption



In the context of GARCH modeling, the residuals are assumed to follow a normal distribution. However, if the QQ plot shows a deviation from normality, it may indicate that the normal distribution is not an appropriate assumption for the residuals. In this case, the residuals can be modeled using a different distribution, such as the student’s t-distribution, which is more robust to outliers and heavy tails. Changing the distribution can improve the model's fit and accuracy, which is why it may be necessary to modify the distribution to student t in this

case.

Fitting our chosen GARCH (1,2) model with chosen ARMA (0,4) mean model

```
> ##Fitting our chosen GARCH(1,2) model with chosen ARMA(0,4) mean model
> #Specify the mean and garch models
> spec <- ugarchspec(mean.model = list(armaorder = c(0,4)),
+                    variance.model = list(model = "sGARCH", garchOrder = c(1,2)))
> def.fit = ugarchfit(spec = spec, data = ret_d)
> print(def.fit)

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

Conditional Variance Dynamics
-----------------------------------
GARCH Model     : sGARCH(1,2)
Mean Model      : ARFIMA(0,0,4)
Distribution    : norm

Optimal Parameters
------------------------------------
        Estimate  Std. Error   t value Pr(>|t|)
mu      0.001775    0.000264  6.727732 0.000000
ma1     0.004021    0.017819  0.225655 0.821470
ma2     0.000584    0.017354  0.033632 0.973170
ma3    -0.041357    0.017602 -2.349530 0.018797
ma4     0.036719    0.018111  2.027458 0.042616
omega   0.000015    0.000003  6.160289 0.000000
alpha1  0.124421    0.017870  6.962425 0.000000
beta1   0.731517    0.156153  4.684630 0.000003
beta2   0.105139    0.148889  0.706153 0.480093

Robust Standard Errors:
        Estimate  Std. Error   t value Pr(>|t|)
mu      0.001775    0.000309  5.731384 0.000000
ma1     0.004021    0.016965  0.237015 0.812645
ma2     0.000584    0.016375  0.035644 0.971566
ma3    -0.041357    0.017100 -2.418347 0.015583
ma4     0.036719    0.018908  1.941917 0.052147
omega   0.000015    0.000008  2.016372 0.043761
alpha1  0.124421    0.050769  2.450743 0.014256
beta1   0.731517    0.303182  2.412801 0.015830
beta2   0.105139    0.296162  0.355004 0.722587

LogLikelihood : 10019.79

Information Criteria
------------------------------------

Akaike       -5.2180
Bayes        -5.2034
Shibata      -5.2180
Hannan-Quinn -5.2128

Weighted Ljung-Box Test on Standardized Residuals
------------------------------------
                        statistic p-value
Lag[1]                     0.3988  0.6157
Lag[2*(p+q)+(p+q)-1][11]   5.2855  0.9167
Lag[4*(p+q)+(p+q)-1][19]  10.0900  0.4350
d.o.f=4
H0 : No serial correlation

Weighted Ljung-Box Test on Standardized Squared Residuals
------------------------------------
                        statistic p-value
Lag[1]                     0.6705  0.4129
Lag[2*(p+q)+(p+q)-1][8]    2.5331  0.7780
Lag[4*(p+q)+(p+q)-1][14]   4.0604  0.8743
d.o.f=2

Weighted ARCH LM Tests
------------------------------------
            Statistic Shape Scale P-Value
ARCH Lag[4]     1.283 0.500 2.000  0.2573
ARCH Lag[6]     1.494 1.461 1.711  0.6121
ARCH Lag[8]     2.716 2.368 1.583  0.5977

Nyblom stability test
------------------------------------
Joint Statistic:  26.1438
Individual Statistics:
mu     0.06298
ma1    0.29046
ma2    0.56603
ma3    0.04888
ma4    0.08148
omega  1.72122
alpha1 0.29084
beta1  0.28520
beta2  0.27170

Asymptotic Critical Values (10% 5% 1%)
Joint Statistic:         2.1 2.32 2.82
Individual Statistic:    0.35 0.47 0.75

Sign Bias Test
------------------------------------
                   t-value    prob sig
Sign Bias           1.5916 0.11169
Negative Sign Bias  0.4724 0.63664
Positive Sign Bias  0.5101 0.60999
Joint Effect        8.9079 0.03054  **


Adjusted Pearson Goodness-of-Fit Test:
------------------------------------
  group statistic p-value(g-1)
1    20     134.6    1.973e-19
2    30     147.8    7.228e-18
3    40     156.1    1.306e-15
4    50     173.4    7.755e-16


Elapsed time : 0.6581878

>
```

After examining the statistics of the fitted model, the conclusion was drawn that there is no sign of bias (values are above the level of significance 0.05). But, as we can see, there is a joint effect, as the value is 0.03054 (< 0.05). Therefore, we decided to move on to fitting an EGARCH or TGARCH model with a student's t-distribution to remove the joint effect.

Moving to EGARCH and TGARCH with student t-distribution

EGARCH with t-distribution

```
Weighted Ljung-Box Test on Standardized Residuals
------------------------------------
                         statistic p-value
Lag[1]                      0.5148  0.4731
Lag[2*(p+q)+(p+q)-1][11]    6.0894  0.4282
Lag[4*(p+q)+(p+q)-1][19]   11.0822  0.3024
d.o.f=4
H0 : No serial correlation

Weighted Ljung-Box Test on Standardized Squared Residuals
------------------------------------
                         statistic p-value
Lag[1]                      0.1234  0.7254
Lag[2*(p+q)+(p+q)-1][8]     0.4923  0.9964
Lag[4*(p+q)+(p+q)-1][14]    1.2903  0.9988
d.o.f=3
```

```
Sign Bias Test
------------------------------------
                   t-value   prob sig
Sign Bias           1.1024 0.2703
Negative Sign Bias  0.7667 0.4433
Positive Sign Bias  0.1255 0.9001
Joint Effect        1.4974 0.6829
```

The Weighted Ljung-Box Test

The Weighted Ljung-Box Test used to test for the presence of serial correlation in time series data. We test whether the residuals are independently and identically distributed. The p-values indicate the significance of the test.

For the first test on the standardized residuals, the null hypothesis (H0) is *that there is no serial correlation*. The p-values suggest that there is no evidence to reject the null hypothesis.

For the second test on the standardized squared residuals, the null hypothesis is again that there is no serial correlation. The p-values for each lag suggest that there is no evidence to reject the null hypothesis. That is, there is no significant serial correlation in the standardized squared residuals.

Therefore, based on these results, we can conclude that the EGARCH model adequately captures the serial correlation in the data.

<u>The Sign Bias Test</u>

The Sign Bias Test is used to check if there is a bias in the signs of the standardized residuals. If the p-value of the Sign Bias test is less than the significance level (usually 0.05), it indicates that there is a sign bias in the residuals, which means that the model tends to overestimate or underestimate the actual values in a consistent direction.

In this case, the p-values of all three Sign Bias tests are greater than 0.05, which suggests that there is no significant sign bias in the standardized residuals. The Joint Effect test combines the Sign Bias, Negative Sign Bias, and Positive Sign Bias tests, and again the p-value is greater than 0.05, indicating that there is no evidence of sign bias in the model. Therefore, we can conclude that the model does not suffer from significant sign bias.

TGARCH with t-distribution

```
Weighted Ljung-Box Test on Standardized Squared Residuals
------------------------------------
                        statistic p-value
Lag[1]                     0.1255  0.7231
Lag[2*(p+q)+(p+q)-1][8]    0.4870  0.9965
Lag[4*(p+q)+(p+q)-1][14]   1.2263  0.9990
d.o.f=3

Weighted ARCH LM Tests
------------------------------------
            Statistic Shape Scale P-Value
ARCH Lag[4]   0.08615 0.500 2.000  0.7691
ARCH Lag[6]   0.09208 1.461 1.711  0.9895
ARCH Lag[8]   0.60255 2.368 1.583  0.9730

Sign Bias Test
------------------------------------
                   t-value   prob sig
Sign Bias           0.8496 0.3956
Negative Sign Bias  0.9714 0.3314
Positive Sign Bias  0.1510 0.8800
Joint Effect        1.3951 0.7067
```

The Weighted Ljung-Box test

For the standardized residuals, the p-values for all lags are greater than the significance level of 0.05, indicating that we fail to reject the null hypothesis of no serial correlation.

Similarly, for the standardized squared residuals, the p-values for all lags are greater than the significance level of 0.05, suggesting that we fail to reject the null hypothesis of no serial correlation.

Therefore, based on these results, it appears that the TGARCH with t-distribution model adequately captures the serial correlation present in the data.

In the Sign Bias Test

The "t-value" indicates the test statistic value for each of the three tests: "Sign Bias", "Negative Sign Bias" and "Positive Sign Bias".

The p-value for "Sign Bias" is 0.3956, which means that there is no significant sign bias in the model. The p-value for "Negative Sign Bias" is 0.3314 and for "Positive Sign Bias" it is 0.8800, indicating no significant negative or positive sign bias in the model.

If the p-value is less than 0.05, then we reject the null hypothesis and conclude that there is significant sign bias. In this case, all the p-values are greater than 0.05, indicating that we cannot reject the null hypothesis and therefore there is no significant sign bias in the model.

The "Joint Effect" test examines the joint significance of all three tests. The "t-value" for the Joint Effect is 1.3951 and the "prob" is 0.7067, indicating that there is no significant sign bias in the model.

GARCH-M model

```
Weighted Ljung-Box Test on Standardized Residuals
------------------------------------
                        statistic p-value
Lag[1]                     0.4179  0.5180
Lag[2*(p+q)+(p+q)-1][11]   6.0519  0.4525
Lag[4*(p+q)+(p+q)-1][19]  10.9603  0.3193
d.o.f=4
H0 : No serial correlation

Weighted Ljung-Box Test on Standardized Squared Residuals
------------------------------------
                        statistic p-value
Lag[1]                     0.1152  0.7343
Lag[2*(p+q)+(p+q)-1][8]    0.4758  0.9968
Lag[4*(p+q)+(p+q)-1][14]   1.2811  0.9988
d.o.f=3
```

```
Sign Bias Test
------------------------------------
                   t-value   prob sig
Sign Bias          1.02105 0.3073
Negative Sign Bias 0.71519 0.4745
Positive Sign Bias 0.07768 0.9381
Joint Effect       1.32442 0.7233
```

There is no evidence of serial correlation in the residuals. Similarly, for the standardized squared residuals, the p-values for all lags are greater than 0.05, which indicates that there is no evidence of residual autocorrelation squared.

The Sign Bias Test examines whether the standardized residuals tend to be positive or negative. The null hypothesis is that there is no such tendency. The t-values for the Sign Bias Test are all less than 2 in absolute value, which indicates that there is no significant sign bias in the standardized residuals.

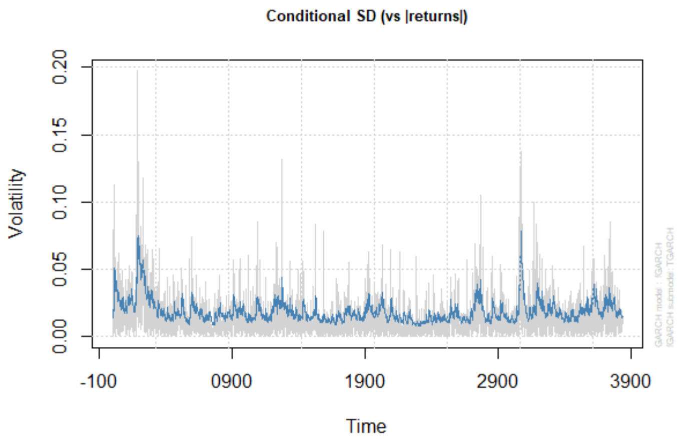
Additionally, the p-values for all three tests are greater than 0.05, which suggests that there is insufficient evidence to reject the null hypothesis of no sign bias in the standardized residuals. Therefore, the model has no sign bias.

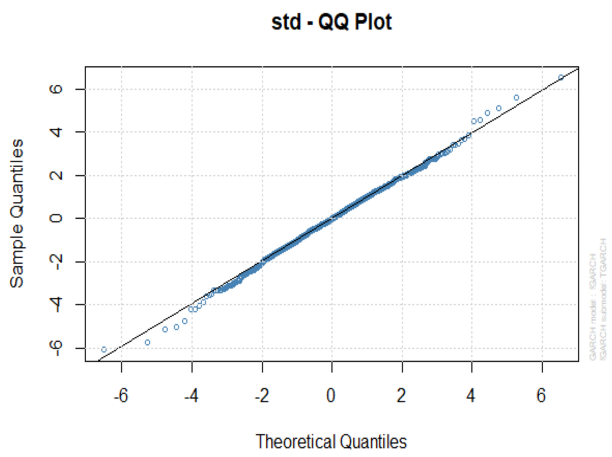
However, as we can see, here the archm coefficient is not significant. If the archm coefficient is not significant in a GARCH-M model, it means that the inclusion of the additional predictor(s) (which are assumed to influence volatility) does not provide a significant improvement in the model's ability to predict volatility.

MODEL SELECTION

1. **<u>To proceed we want to choose one of the three GARCH models. We do so by picking the one with the lowest AIC. As we can see the TGARCH model with t-distribution has the lowest AIC of -5.3170.</u>**

We proceed with this model and plot the series. On this graph the grey line is the series themselves and the blue line represents the volatility of the series. By looking at this plot, we can see which were the periods of high volatility.

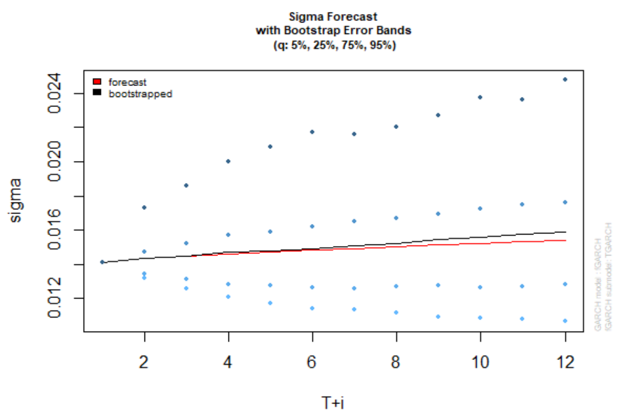


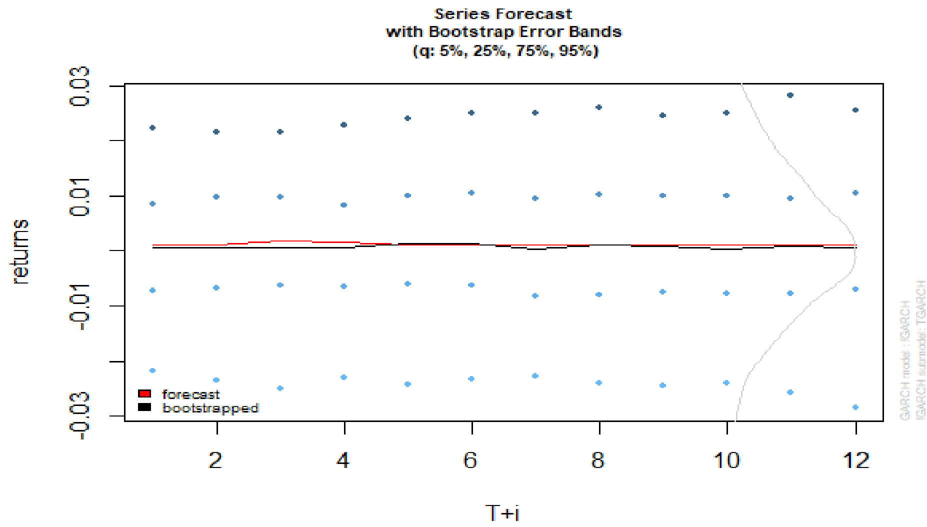


<u>QQ plot looks normal, which is an indication that the data is likely normally distributed</u>

Volatility and series forecast:

The forecast line closely tracks the actual data, this suggests that the model is well-specified and is capturing the underlying patterns in the data. Model is predicting a positive return over the forecast period, with some level of uncertainty in the later part.





F) Value at Risk

Risk Metrics

In general, the risk metric method used in VaR calculations involves estimating the potential distribution of portfolio returns, selecting an appropriate level of confidence, and identifying the potential loss level at that confidence level. The method used will depend on the characteristics of the portfolio and the data available. It is important to note that VaR is just one measure of risk and should be used in conjunction with other risk measures to fully assess the potential risks of an investment portfolio.

Risk Metric results for AAPL stock

When Beta is estimated

```
Coefficient(s):
      Estimate  Std. Error  t value   Pr(>|t|)
beta 0.96039934  0.00345551  277.933 < 2.22e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

 Volatility prediction:
     Orig       Vpred
[1,] 3837 0.01511316

 Risk measure based on RiskMetrics:
      prob        VaR         ES
[1,] 0.950 0.02485893 0.03117411
[2,] 0.990 0.03515846 0.04027981
[3,] 0.999 0.04670317 0.05088737
```

When Beta is fixed

```
Default beta = 0.96 is used.

 Volatility prediction:
     Orig       Vpred
[1,] 3837 0.01507747

 Risk measure based on RiskMetrics:
      prob        VaR         ES
[1,] 0.950 0.02480023 0.03110048
[2,] 0.990 0.03507543 0.04018468
[3,] 0.999 0.04659288 0.05076719
```

The given results show the impact of estimating the beta coefficient on volatility prediction and risk measures based on Risk Metrics.

a) When beta is estimated:

The estimated beta is 0.96039934, and it is statistically significant (p-value < 2.22e-16). The volatility prediction is 0.01511316, and the risk measures based on Risk Metrics show that the value at risk (VaR) and expected shortfall (ES) increase with increasing probability level. The VaR at 0.99 probability level is 0.03515846, while the ES is 0.04027981.

b) When beta is fixed:

In this case, the default beta value of 0.96 is used. The volatility prediction is slightly lower than the previous case, at 0.01507747. The risk measures based on Risk Metrics are also slightly lower than the previous case. For example, the VaR at 0.99 probability level is 0.03507543, and the ES is 0.04018468.
Overall, the difference between the two cases is not significant. This implies that the choice of estimating beta or fixing it does not have a substantial impact on the volatility prediction and risk measures based on Risk Metrics. However, it is worth noting that estimating beta could be more appropriate in situations where the underlying asset's risk characteristics are expected to change over time. In contrast, fixing beta could be more appropriate when the asset's risk characteristics are stable and well-known.

Econometric Modelling

Econometric modeling involves the use of statistical and mathematical methods to analyze economic and financial data. In the context of VaR estimation, econometric modeling involves using historical data on asset returns to estimate the asset's probability distribution and then using that distribution to compute VaR.

Here are the general steps involved in using econometric modeling to compute VaR:

- Data collection and preparation: The first step is to collect historical data on the asset or portfolio being analyzed. This data should be properly cleaned, normalized, and organized for analysis.
- Model specification: Next, an appropriate econometric model is selected that can capture the asset's return distribution. Some commonly used models include the GARCH, EVT, and Cornish-Fisher VaR models.
- Estimation: The model's parameters are then estimated using historical data. This involves estimating the model's coefficients, standard errors, and other relevant statistical measures.
- VaR computation: Once the model is estimated, VaR can be computed using a variety of methods. One common approach is Monte Carlo simulation, which involves simulating future asset returns using the estimated model and then calculating the potential loss at the desired probability level.
- Validation: Finally, the VaR estimates are validated to ensure their accuracy and reliability. This involves comparing the predicted VaR with actual losses to assess the model's effectiveness and identify any shortcomings that may need to be addressed.

In summary, econometric modeling involves using statistical and mathematical methods to estimate an asset's return distribution and compute VaR. The process involves data collection and preparation, model specification, parameter estimation, VaR computation, and validation.

Econometric Modelling results for AAPL stock (Using student-t distribution)

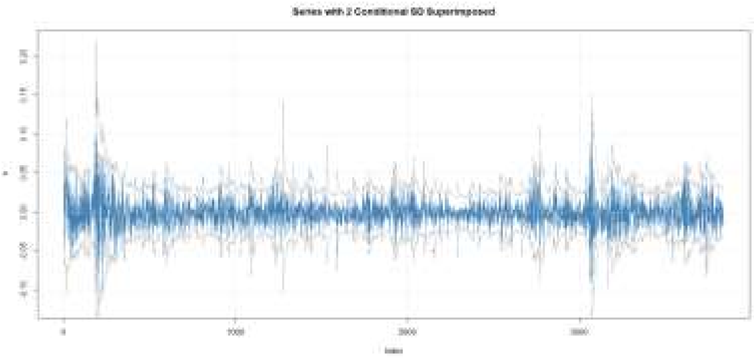
Summary of the model

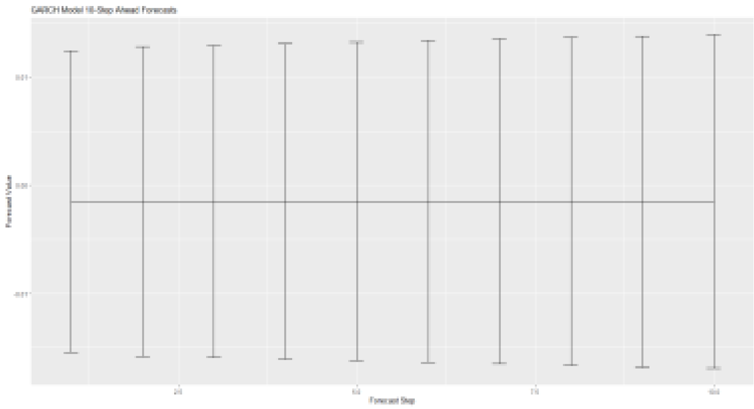
```
Conditional Distribution:
 std

Coefficient(s):
         mu        omega       alpha1        beta1        beta2        shape
-1.4913e-03   8.9603e-06   1.1853e-01   6.1149e-01   2.5295e-01   5.0659e+00

Std. Errors:
 based on Hessian

Error Analysis:
         Estimate  Std. Error  t value Pr(>|t|)
mu     -1.491e-03   2.332e-04   -6.395 1.60e-10 ***
omega   8.960e-06   2.602e-06    3.443 0.000575 ***
alpha1  1.185e-01   2.331e-02    5.085 3.68e-07 ***
beta1   6.115e-01   2.241e-01    2.729 0.006359 **
beta2   2.529e-01   2.061e-01    1.227 0.219651
shape   5.066e+00   4.104e-01   12.343  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log Likelihood:
 10171.5    normalized:  2.6509

Description:
 Thu May  4 20:20:52 2023 by user: abhim


Standardised Residuals Tests:
                                Statistic p-Value
 Jarque-Bera Test   R    Chi^2  1733.298  0
 Shapiro-Wilk Test  R    W      0.9698666 0
 Ljung-Box Test     R    Q(10)  16.14899  0.09544598
 Ljung-Box Test     R    Q(15)  20.00802  0.1716253
 Ljung-Box Test     R    Q(20)  28.28805  0.1027541
 Ljung-Box Test     R^2  Q(10)  5.283643  0.871443
 Ljung-Box Test     R^2  Q(15)  8.41726   0.9059675
 Ljung-Box Test     R^2  Q(20)  11.61067  0.9288249
 LM Arch Test       R    TR^2   5.843159  0.9237808

Information Criterion Statistics:
      AIC       BIC       SIC      HQIC
-5.298672 -5.288895 -5.298677 -5.295199
```

- The p-values for all the parameters except beta2 and the intercept term are significant at the 1% level, indicating that they are likely different from zero. The shape parameter of the student-t distribution is 5.066, which suggests that the distribution has heavier tails than a normal distribution.
- The log-likelihood value for the model is 10171.5, and the normalized log-likelihood is 2.6509. The normalized log-likelihood can be used to compare models with different sample sizes, and higher values indicate better fit.
- The standardized residuals tests indicate that the residuals are not normally distributed, based on the Jarque-Bera and Shapiro-Wilk tests. The Ljung-Box tests show that there is some evidence of residual autocorrelation at lags 10, 15, and 20 for the squared residuals, but not for the residuals themselves. The LM Arch test shows that there is no significant evidence of residual ARCH effects.
- The information criterion statistics include the Akaike information criterion (AIC), Bayesian information criterion (BIC), Schwarz information criterion (SIC), and Hannan-Quinn information criterion (HQIC). Lower values of these criteria indicate better fit, and they can be used to compare different models.

Predicted Mean Forecast, Mean Error and Standard Deviation for next 10 periods.

```
   meanForecast  meanError standardDeviation
1  -0.001491259 0.01394908        0.01394908
2  -0.001491259 0.01431274        0.01431274
3  -0.001491259 0.01441275        0.01441275
4  -0.001491259 0.01457478        0.01457478
5  -0.001491259 0.01471666        0.01471666
6  -0.001491259 0.01485940        0.01485940
7  -0.001491259 0.01499784        0.01499784
8  -0.001491259 0.01513342        0.01513342
9  -0.001491259 0.01526592        0.01526592
10 -0.001491259 0.01539552        0.01539552
```





The output shows the predicted values for the next 10 periods using the econometric model 'm1'.

- The column 'mean Forecast' displays the predicted mean value of the dependent variable, while the column 'standard Deviation' shows the standard deviation of the prediction error, also known as the mean error or the forecast error. The 'mean Error' column displays the same values as the 'standard Deviation' column.
- The predicted mean values are all the same, which is the estimated value for the conditional mean of the dependent variable. However, the predicted standard deviations (or forecast errors) increase with each prediction, indicating higher uncertainty in the predictions for future periods. Overall, the model appears to be producing reasonable predictions based on the available data.
- However, it's important to note that the accuracy of the predictions may decrease as we move further away from the last observation used to estimate the model. Therefore, it's important to evaluate the model's performance using additional metrics and by comparing the predicted values to the actual values observed in the future.

Risk Measures (VaR and ES)

1-Day VaR

```
> RMeasure(-.001491,.0139,cond.dist="std",df=5.0659)

 Risk Measures for selected probabilities:
       prob        VaR         ES
[1,] 0.9500 0.02023623 0.02960157
[2,] 0.9900 0.03469852 0.04627479
[3,] 0.9990 0.06161141 0.07874256
[4,] 0.9999 0.10160336 0.12796192
```

10 Day VaR

```
> RMeasure(-.001491,v1,cond.dist="std",df=5.0659)

 Risk Measures for selected probabilities:
       prob        VaR        ES
[1,] 0.9500 0.07150791 0.1029735
[2,] 0.9900 0.12009818 0.1589920
[3,] 0.9990 0.21051981 0.2680769
[4,] 0.9999 0.34488433 0.4334436
```

- The 1-day VaR and ES estimates indicate the potential loss for a single day holding period at different probabilities. For example, at the 99% confidence level, the 1-day VaR is 0.0347 and the ES is 0.0463. This means that with a 99% probability, the maximum loss on any given day is not expected to exceed 3.47% and the expected loss is 4.63%.
- On the other hand, the 10-day VaR and ES estimates suggest the potential loss over a 10-day holding period at different probabilities. For instance, at the 99% confidence level, the 10-day VaR is 0.1201 and the ES is 0.1590. This indicates that with a 99% probability, the maximum loss over a 10-day period is not expected to exceed 12.01% and the expected loss is 15.90%.
- Comparing the results for 1-day and 10-day holding periods, we can observe that the VaR and ES estimates increase with the increase in holding period. This is because as the holding period increases, the potential for fluctuations in the underlying asset or security increases, resulting in a higher risk of losses. Therefore, investors may want to consider longer holding periods for risk management and take appropriate measures to reduce potential losses.

Empirical Quantile & Quantile Regression

Quantile Regression results for AAPL stock

At 95%

```
Call: rq(formula = naapl ~ vol + GSPC, tau = 0.95, data = df)

tau: [1] 0.95

Coefficients:
            Value    Std. Error t value  Pr(>|t|)
(Intercept) -0.00321  0.00313   -1.02741  0.30429
vol          0.00000  0.00000   11.68710  0.00000
GSPC         0.00001  0.00000    8.74800  0.00000

> VaR_quant
[1] 0.0370681
```

At 99%

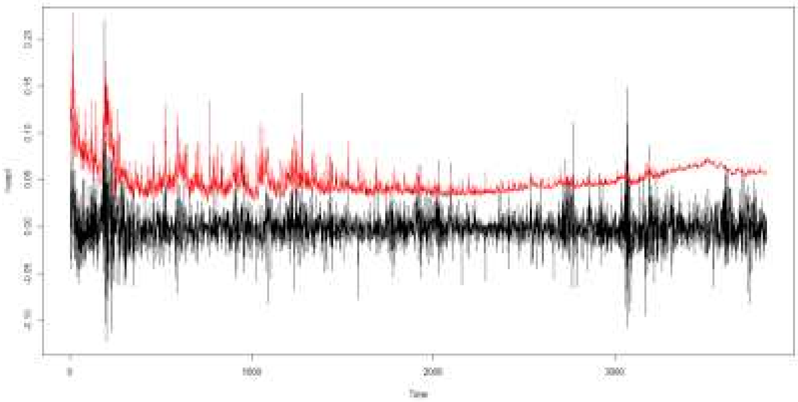
```
Call: rq(formula = naapl ~ vol + GSPC, tau = 0.99, data = df)

tau: [1] 0.99

Coefficients:
            Value    Std. Error t value  Pr(>|t|)
(Intercept) -0.00083  0.01097   -0.07544  0.93987
vol          0.00000  0.00000    3.42263  0.00063
GSPC         0.00001  0.00000    4.74295  0.00000

> VaR_quant
[1] 0.0394481
```

At 99.9%

```
Call: rq(formula = naapl ~ vol + GSPC, tau = 0.999, data = df)

tau: [1] 0.999

Coefficients:
            Value    Std. Error t value  Pr(>|t|)
(Intercept) -0.05413  0.02110   -2.56522  0.01035
vol          0.00000  0.00000   20.02756  0.00000
GSPC         0.00004  0.00001    2.78755  0.00534

> VaR_quant
[1] 0.1069824
```

- The quantile regression analysis was performed at three different confidence levels: 95%, 99%, and 99.9%.
- At 95% confidence level, the estimated VaR_quant is 0.0370681. This means that there is a 95% chance that the loss on the portfolio will not exceed 3.70681% in one day.
- At 99% confidence level, the estimated VaR_quant is 0.0394481. This means that there is a 99% chance that the loss on the portfolio will not exceed 3.94481% in one day.
- At 99.9% confidence level, the estimated VaR_quant is 0.1069824. This means that there is a 99.9% chance that the loss on the portfolio will not exceed 10.69824% in one day.
- Overall, as the confidence level increases, the estimated VaR_quant also increases, indicating that the risk of loss on the portfolio becomes greater. This is because as the confidence level increases, the probability of extreme events happening also increases, resulting in a larger estimated VaR_quant.

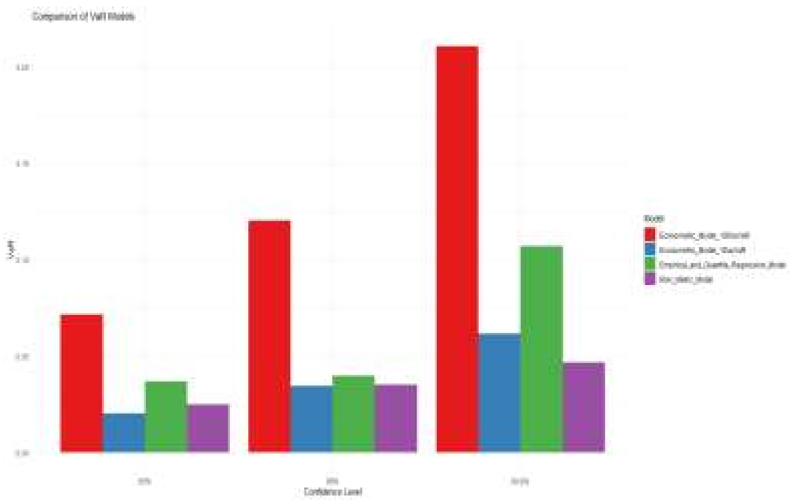
Comparing Loss with VaR at 99% confidence level



From the graph, we can say that at 99% confidence level, most of the periods are covered and there are very few periods when loss was greater than VaR.

Choosing the best Model

Confidence Level	Risk Metric	Econometric Model (1 Day VaR)	Econometric Model (10 Day VaR)	Empirical & Quantile Regression
95%	0.02485893	0.02023623	0.07150791	0.0370681
99%	0.03515846	0.03469852	0.12009818	0.0394481
99.9%	0.04670317	0.06161141	0.21051981	0.1069824



Analysis

- Based on the output provided, it appears that none of the models is suitable for every confidence level.

- From the graph, we can see that the "Risk Metric Model" has the lowest VaR at all confidence levels, followed by the "Econometric Model (1 Day VaR)" and the "Empirical & Quantile Regression Model." The "Econometric Model (10 Day VaR)" has the highest VaR at all confidence levels.

- Based on this graph, we can conclude that the "Risk Metric Model" is the best model for calculating VaR as it has the lowest VaR at all confidence levels. However, it's essential to consider other factors such as model assumptions and data quality before making a final decision.

G) Conclusion and Managerial Implications

- Based on our analysis, we found that the use of log returns provided a better model than price due to its stationarity. We applied various models to our data, including ARIMA, ARCH/GARCH, EGARCH, and TGARCH, to identify the best-fit model for our data.
- Through our analysis, we detected the presence of a leverage effect and sign bias in our data, which led us to select the EGARCH and TGARCH models to eliminate these effects.
- We then used the selected models to forecast the next 12 datapoints, and the density plots of volatility and log-return forecasts for both models looked favorable. Therefore, we can confidently use the EGARCH and TGARCH models with ARIMA (2,0,3) and Standard-t distributions for future predictions of the data.
- The implications of this analysis are significant for businesses that rely on financial data to make informed decisions. By using the appropriate models, businesses can accurately predict future market trends and adjust their strategies accordingly.
- Furthermore, understanding the presence of a leverage effect and sign bias in financial data can help businesses avoid making incorrect assumptions and ultimately improve their bottom line.