ThermoFisher
SCIENTIFIC

The world leader in serving science

FIN 620:
PROJECT REPORT
Group 4

UNDER THE GUIDANCE OF: PROF. PALLAVI PAL

WRITTEN BY: PRIYANKA DUBEY

Table of Contents

A. OVERVIEW OF THE ASSET AND THE MARKET FOR THE ASSET

Thermo Fisher Scientific Inc. offers life sciences solutions, analytical instruments, specialty diagnostics, and laboratory products and service worldwide. The company's Life Sciences Solutions segment offers reagents, instruments, and consumables for biological and medical research, discovery, and production of drugs and vaccines, as well as diagnosis of infections and diseases to pharmaceutical, biotechnology, agricultural, clinical, healthcare, academic, and government markets. Its Analytical Instruments segment provides instruments, consumables, software, and services for use in laboratory, on production line, and in field for pharmaceutical, biotechnology, academic, government, environmental, and other research and industrial markets, as well as clinical laboratories. The company offers products and services through a direct sales force, customer-service professionals, electronic commerce, third-party distributors, and catalogs. It has a strategic alliance with the University of California, San Francisco. The company was incorporated in 1956 and is based in Waltham, Massachusetts.

Market: As of May 2023, Thermo Fisher Scientific has a market cap of $206.68 Billion. This makes Thermo Fisher Scientific the world's 40th most valuable company by market cap. The Stock is currently trading at $535.92. Thermo Fisher Scientific competitors include Agilent Technologies, Beckman Coulter, Bruker Corporation and Bio-Rad Laboratories.

B. PROPERTIES OF THE TIME-SERIES

i. Descriptive statistics

All data is being taken directly from RStudio via the **getsymbol** command using TMO, which directly populates and sorts the data for the **Thermo Fisher Scientific Inc.** from *Yahoo Finance* in real time. Cleaning of said data is not necessary for this analysis as it is already cleaned and compiled into R. Data that is being taken and sorted by R are: Close price, Adjusted close, Daily log of returns (via log function applied to Adjusted price), High/Lows of TMO, and Trading Volume (all over time)

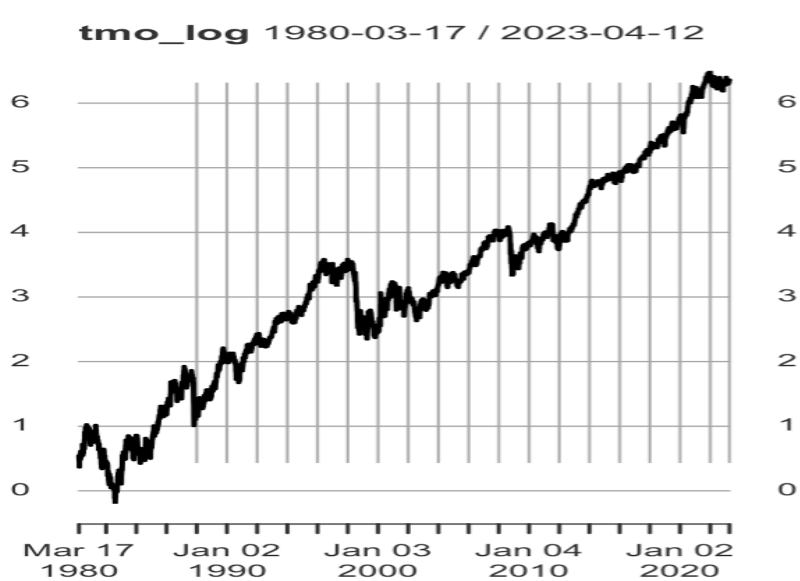
```
> basicStats(tmo_d)
                TMO.Adjusted
nobs            1.086000e+04
NAs             0.000000e+00
Minimum         8.181920e-01
Maximum         6.653384e+02
1. Quartile     8.033847e+00
3. Quartile     5.752713e+01
Mean            7.938293e+01
Median          2.536309e+01
Sum             8.620986e+05
SE Mean         1.299885e+00
LCL Mean        7.683491e+01
UCL Mean        8.193094e+01
Variance        1.835015e+04
Stdev           1.354627e+02
Skewness        2.526987e+00
Kurtosis        5.722497e+00
```

```
> basicStats(tmo_d_lr)
                TMO.Adjusted
nobs            10859.000000
NAs                 0.000000
Minimum            -0.206337
Maximum             0.163039
1. Quartile        -0.009390
3. Quartile         0.010126
Mean                0.000538
Median              0.000000
Sum                 5.836982
SE Mean             0.000194
LCL Mean            0.000158
UCL Mean            0.000917
Variance            0.000407
Stdev               0.020173
Skewness            0.107565
Kurtosis            6.838843
```

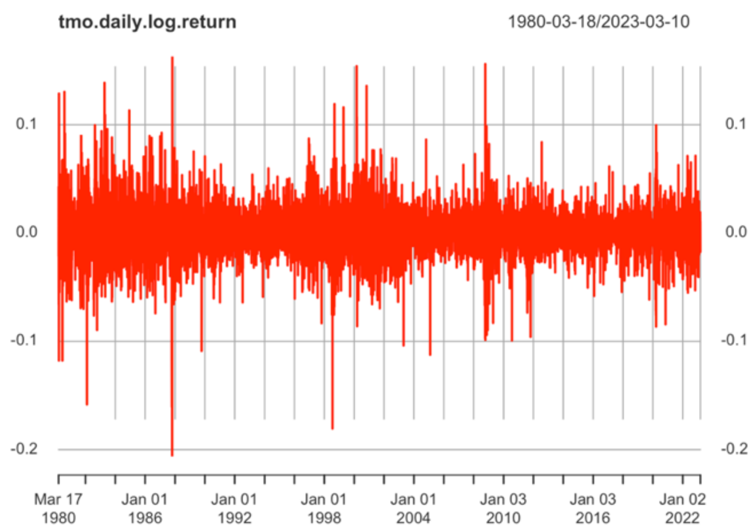
ii. Visualization

Analysis of Log Returns

Graph of daily log prices



Graph of daily log returns



For our further analysis and model building we will be using the log return series for our stock Thermo Fisher Scientific. As we can observe in the above graph for the log returns we can observe that there is stationarity in data and there is no clear trend, but there may be seasonality. We will be conducting all the necessary test for testing the normality and stationarity of our data and build and test different econometric models as per our analysis going ahead.

iii. Unit-Root and Seasonality Test

Augmented Dickey-fuller Test

```
> adf.test(ret_d)

        Augmented Dickey-Fuller Test

data:  ret_d
Dickey-Fuller = -22.159, Lag order = 22, p-value = 0.01
alternative hypothesis: stationary
```

Here we are rejecting the null hypothesis that the data has a unit root and conclude that log returns are stationary since the p-value is low and this validates our observation from the log return series graph that the log return series is stationary.

T-test for Mean log return

```
> t.test(ret_d)

        One Sample t-test

data:  ret_d
t = 2.7767, df = 10858, p-value = 0.005502
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 0.0001580588 0.0009169909
sample estimates:
   mean of x
0.0005375249
```

We performed a t-test to test for the mean log return to be zero, and received a p- value of 0.005502, which is quite low and tells us that we reject the null hypothesis and can conclude that the true mean for the log return is not equal to zero; the true mean is positive since the 95% confidence interval range is positive from 0.0001526457 to 0.0009127904.

Jarque Bera test – Log Returns

```
> normalTest(ret_d, method="jb")

Title:
 Jarque - Bera Normalality Test

Test Results:
  STATISTIC:
    X-squared: 21193.5394
  P VALUE:
    Asymptotic p Value: < 2.2e-16
```

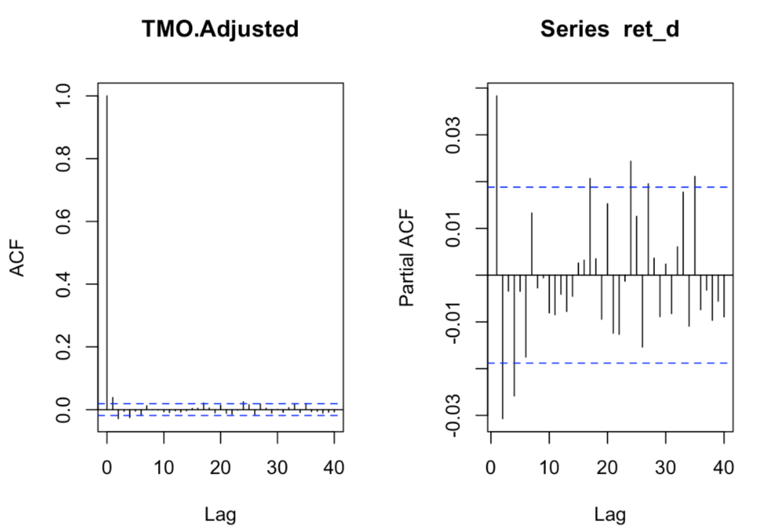
The p-value is less than the significance level (usually 0.05), thus we reject the null hypothesis that the data is normally distributed. The distribution of log returns is not normal.

Ljung-Box test for log-returns

```
> Box.test(ret_d, lag = 10, type = "Ljung")

        Box-Ljung test

data:  ret_d
X-squared = 37.907, df = 10, p-value = 3.939e-05
```

Since p-value is very low we reject the null hypothesis that there is zero autocorrelation. Hence, we conclude that there is autocorrelation.

Graph of ACF & PACF for TMO Daily Log Return



As per the ACF graph we can use lag (1), lag(2) and lag(4) and some later lags as those lags are significant but we do not consider them and in the Partial ACF graph lag (1), lag(3) and later lags are also significant which suggests that we can try AR(1), AR(3) model

C. VARIOUS ARIMA MODELS FITTED TO PRICE AND RETURNS

EACF MATRIX

```
> eacf(ret_d, ar.max = 7, ma.max = 13)
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 x x o x o o o o o o o  o  o  o
1 x x o x o o o o o o o  o  o  o
2 x x x o o o o o o o o  o  o  o
3 x x x o o x 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 x x x 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
```

To test for more models apart from the ones observed in the ACF and PACF graph we run the EACF function. As we can observe in the EACF matrix it suggests that we can choose ARIMA (0,2),(0,4), (1,4),(1,2), (1,3), (2,3) for our further analysis and decide which model suits the best.

Auto ARIMA Function

```
> #Checking for various ARIMA model log return series
> m1 = auto.arima(ret_d)
> m1
Series: ret_d
ARIMA(1,0,1) with non-zero mean

Coefficients:
          ar1     ma1   mean
      -0.6866  0.7233  6e-04
s.e.   0.1066  0.1009  2e-04

sigma^2 = 0.000406:  log likelihood = 26993.31
AIC=-53978.62   AICc=-53978.61   BIC=-53949.45
```

Auto ARIMA function gives a ARIMA(1,0,1) series

AR & ARIMA TABLE

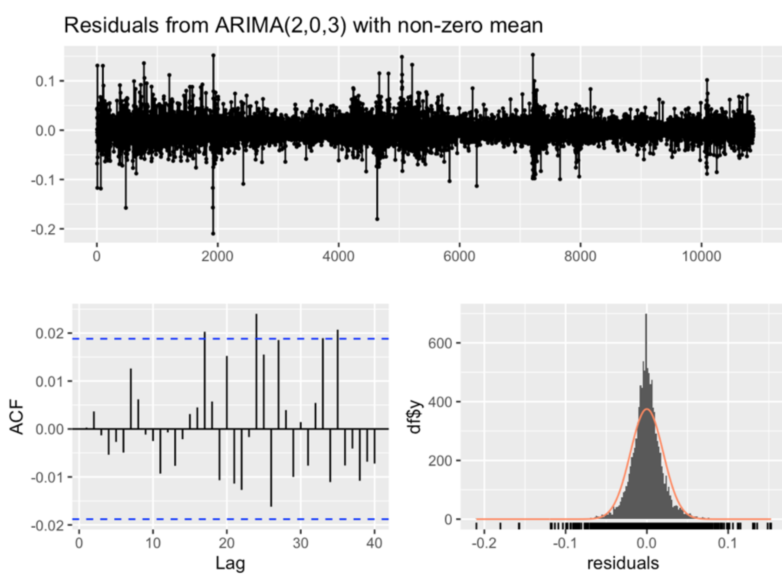
MODEL/ORDER	AIC	MODEL/ORDER	AIC
AR(1)	-53849.11	ARMA(1,0,1)	-53980.62
AR(2)	-53857.44	ARMA (1,0,2)	-53977.32
AR(3)	-53855.56	ARMA (1,0,3)	-53975.49
AR(4)	-53860.85	ARMA (1,0,4)	-53980.81
MA(1)	-53850.15	ARMA (2,0,3)	-53981.18
MA(2)	-53857.78	ARMA(0,0,4)	-53980.78
MA(3)	-53855.95		
MA(4)			
MA(5)			

To choose from all the model suggestions as per different function we further go on to check the AIC and BIC for all the models selected and choose the one with the lowest AIC. The above table represents different AR/MA/ARIMA models and their corresponding AIC. The model with the lowest AIC is ARIMA (2,0,3).

RESIDUAL CHECKING

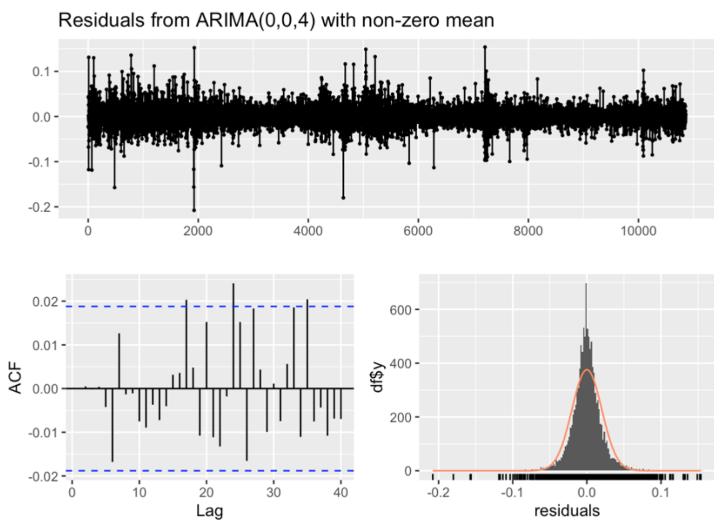
#CASE 1

As per the above table we are selecting the ARIMA(2,0,3) model. To choose the best model we also check for residuals and we have plotted residuals for all the models and then choose the model with the best fit for the residuals.
Here we are plotting the residuals for the Arima model [ARIMA(2,0,3)] by using the check residual function and we can observe that there is no significant residuals.



#CASE 2

Here we have plotted the residuals for the Arima model [ARIMA(0,0,4)] because we also had non-significant residual plot for the ARIMA(0,0,4).



After we observe both the graphs we conclude that ARIMA(2,0,3) has the best fitted residuals and so we choose ARIMA(2,0,3).

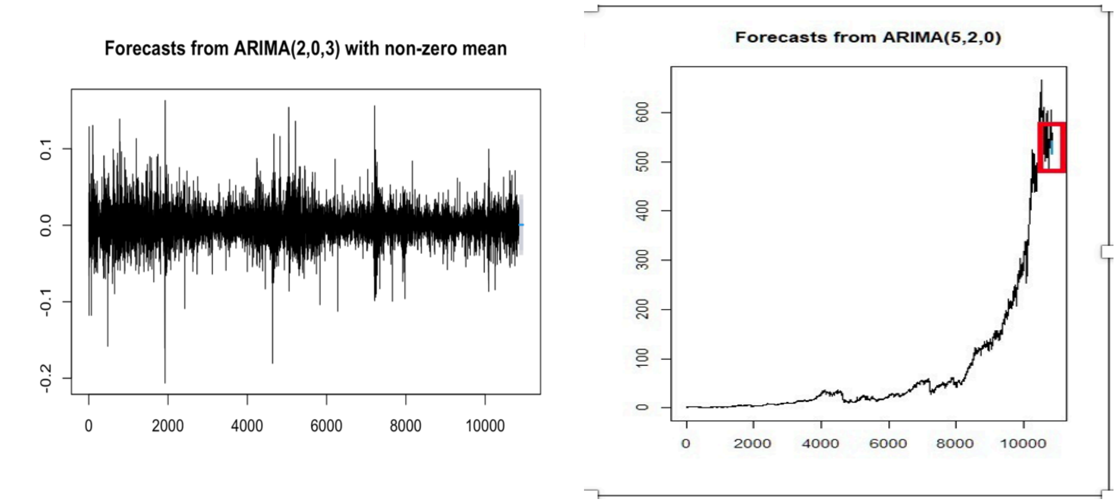
Ljung- Box Test for Residuals

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

        Box-Ljung test

data:  m5$residuals
X-squared = 3.0519, df = 10, p-value = 0.9802
```

To validate our results for the chosen model ARIMA(2,0,3) we conduct the Ljung-Box test. As we can observe we fail to reject the null hypothesis as the p-value is high and conclude that there is no autocorrelation between the residuals.

i. FORECASTING POWER OF THE ARIMA MODEL

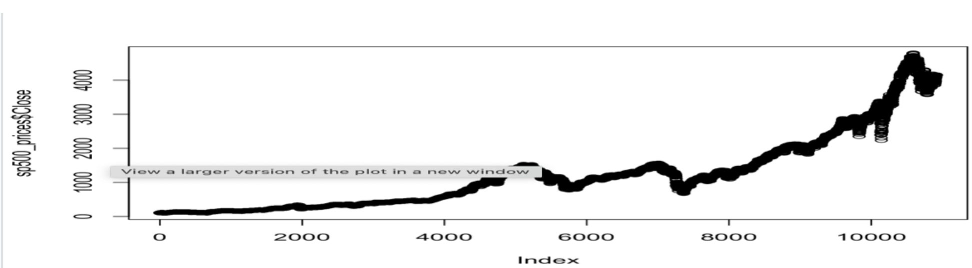
We have forecasted for the future returns and price series using the Arima model

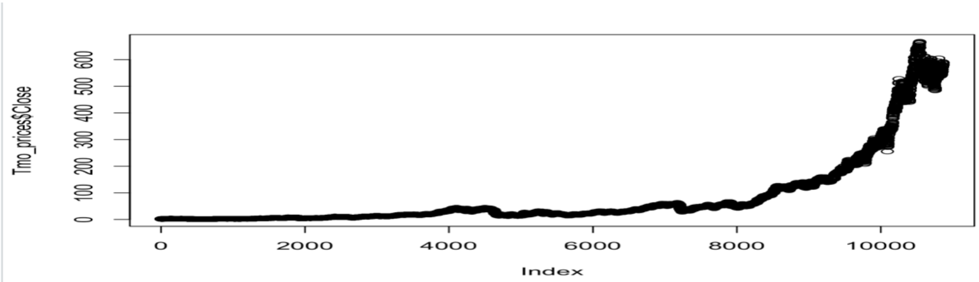


D. MULTI-VARIATE ANALYSIS: VAR

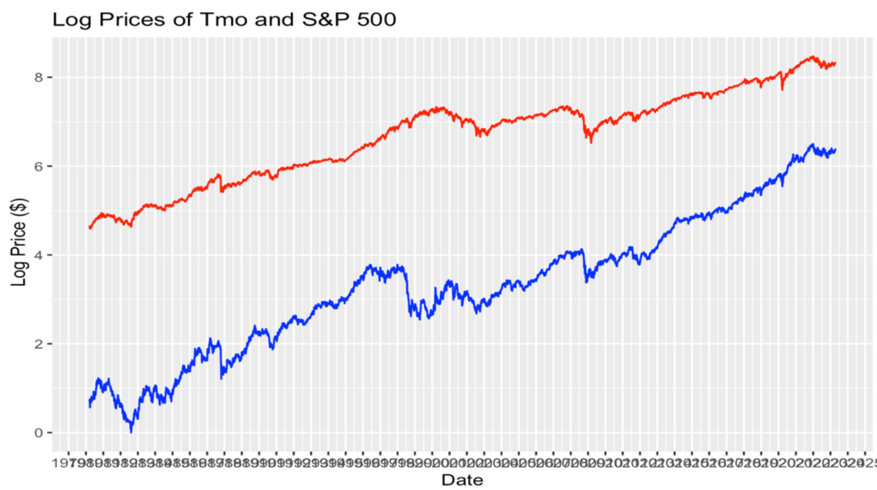
As part of our analysis, we considered the cointegration between TMO and the S&P500 (The Standard and Poor's 500).

The first figure below shows that the prices of TMO and S&P500 generally move in the same direction for a similar magnitude. A similar conclusion can be made about the second figure, which shows the log of prices. We will use the log of prices for further analysis given that prices are non-stationary, so we cannot apply the VAR method directly.



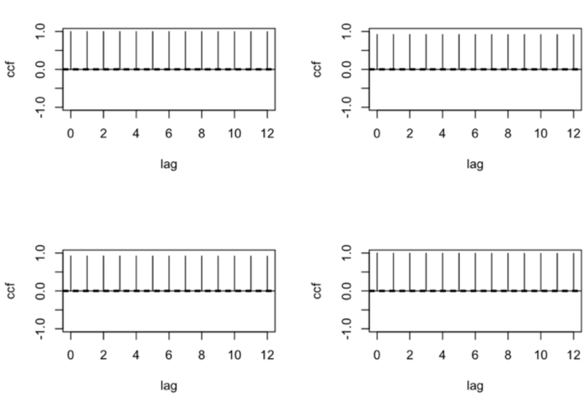


GRAPH FOR LOG-PRICES

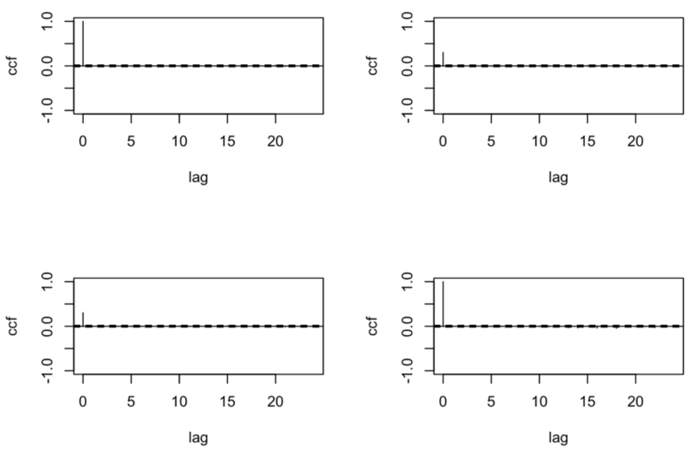


Furthermore, when we plot the cross-correlation matrices below, we see that they are all significant. However, after we difference the matrices of log of price, we see that the CCM looks better with less significant lags, meaning less dependence. CCM is still not the most informative method for determining the lag.

CCM for Log Prices



CCM for Difference of Log Prices



ORDER SELECTION

```
> VARorder(diff_prices)
selected order: aic =  13
selected order: bic =  6
selected order: hq =  9
Summary table:
       p      AIC      BIC       HQ      M(p) p-value
 [1,]  0 -13.8550 -13.8550 -13.8550    0.0000  0.0000
 [2,]  1 -14.2814 -14.2787 -14.2804 4633.3287  0.0000
 [3,]  2 -14.3797 -14.3743 -14.3779 1074.6153  0.0000
 [4,]  3 -14.4286 -14.4206 -14.4259  538.7265  0.0000
 [5,]  4 -14.4338 -14.4231 -14.4302   64.5541  0.0000
 [6,]  5 -14.4333 -14.4199 -14.4288    2.0919  0.7189
 [7,]  6 -14.4410 -14.4249 -14.4356   91.5356  0.0000
 [8,]  7 -14.4430 -14.4242 -14.4366   28.9790  0.0000
 [9,]  8 -14.4432 -14.4217 -14.4359   10.5686  0.0319
[10,]  9 -14.4461 -14.4219 -14.4380   39.4807  0.0000
[11,] 10 -14.4459 -14.4191 -14.4369    5.9570  0.2024
[12,] 11 -14.4461 -14.4166 -14.4362   10.5051  0.0327
[13,] 12 -14.4462 -14.4140 -14.4354    8.7805  0.0668
[14,] 13 -14.4481 -14.4132 -14.4363   28.2312  0.0000
```

When selecting a VAR order, the order with the lowest AIC is 13 and the order with the lowest BIC is 6 and hq is 9. This is a large difference in order, but we will look into all the models.

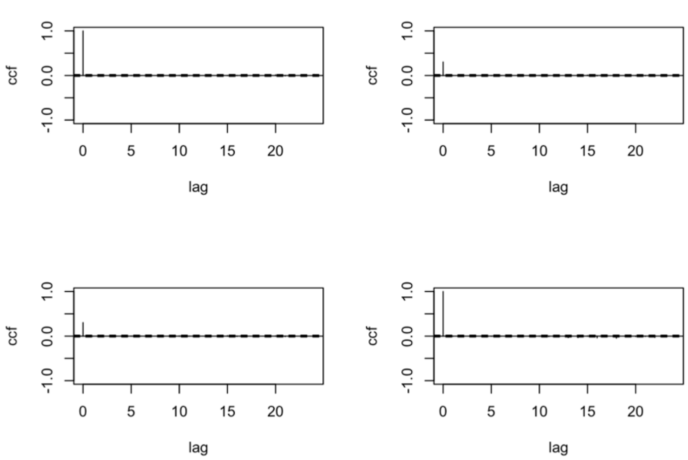
VAR ORDER SELECTION OUTPUT

```
              [,1]          [,2]
 [1,]  0.0005520069  6.272761e-05
 [2,]  0.0377924549  5.446676e-01
 [3,]  0.0081203250 -8.488860e-01
 [4,] -0.0356637135  2.758177e-01
 [5,]  0.0098708983 -4.932157e-01
 [6,] -0.0050589276  1.543967e-01
 [7,]  0.0034332911 -2.375604e-01
 [8,] -0.0324471555 -1.869235e-02
 [9,]  0.0106299851  3.003738e-03
[10,] -0.0036841782 -5.991000e-02
[11,]  0.0010968359  1.043761e-01
[12,] -0.0161692401 -4.310176e-02
[13,] -0.0018098902  1.109163e-01
[14,]  0.0109359879  7.065772e-03
[15,]  0.0010108337  3.127172e-02
[16,] -0.0176468595 -3.299035e-03
[17,]  0.0242017285  3.466377e-05
[18,] -0.0083384482 -2.083829e-02
[19,]  0.0130608874 -2.944726e-02
[20,] -0.0078494659 -3.053117e-02
[21,] -0.0006791567  2.025225e-02
[22,] -0.0002032507  2.217608e-03
[23,] -0.0168915043  1.294848e-02
[24,]  0.0007174248  5.785674e-03
[25,] -0.0091854863 -1.440507e-02
[26,] -0.0032784593 -2.218785e-02
[27,] -0.0056859435 -4.584308e-02
> m2=VAR(diff_prices, p = 6) ### Fit a
```

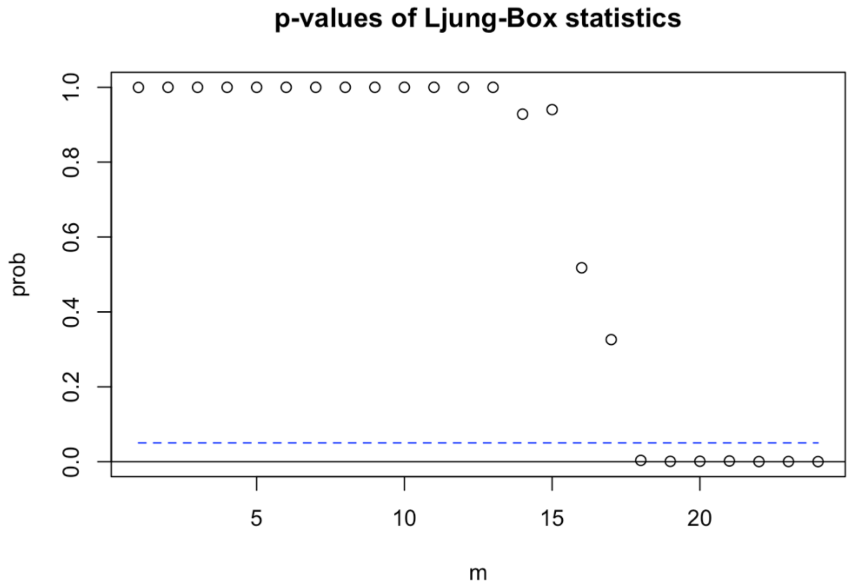
As part of model checking, we check the CCF for VAR(13), and they are mostly insignificant, which is what we want to see in a good model. However, since some of the lags are difficult to interpret in the CCF, we also looked at the Ljung-Box test, which shows no significance at small lags, and there is some significance between lags 13 and 15, but since those lags are high, we will move forward.

When it comes to VAR(6) and VAR(9), the model does not pass the Ljung-Box test as all lags are significant. We should not use this model going forward.

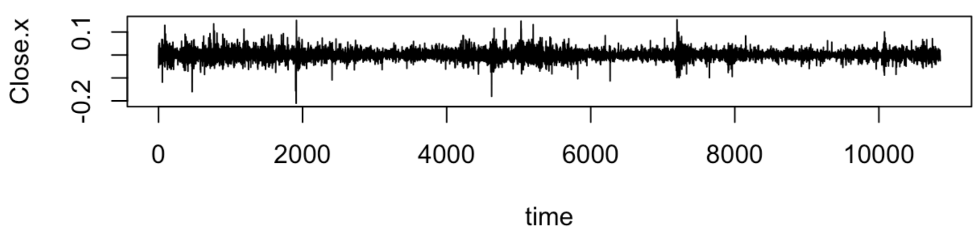
VAR(13): Model checking using CCF

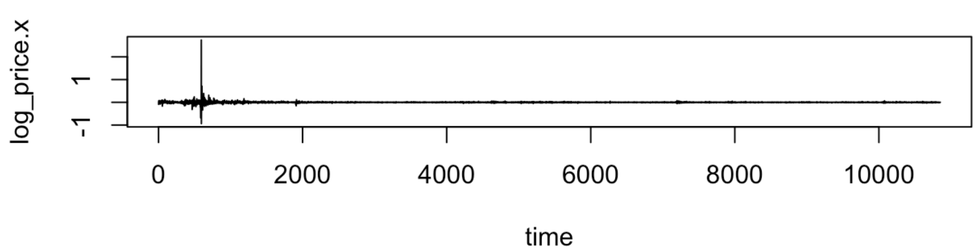


VAR(13): Model checking using Ljung-Box test



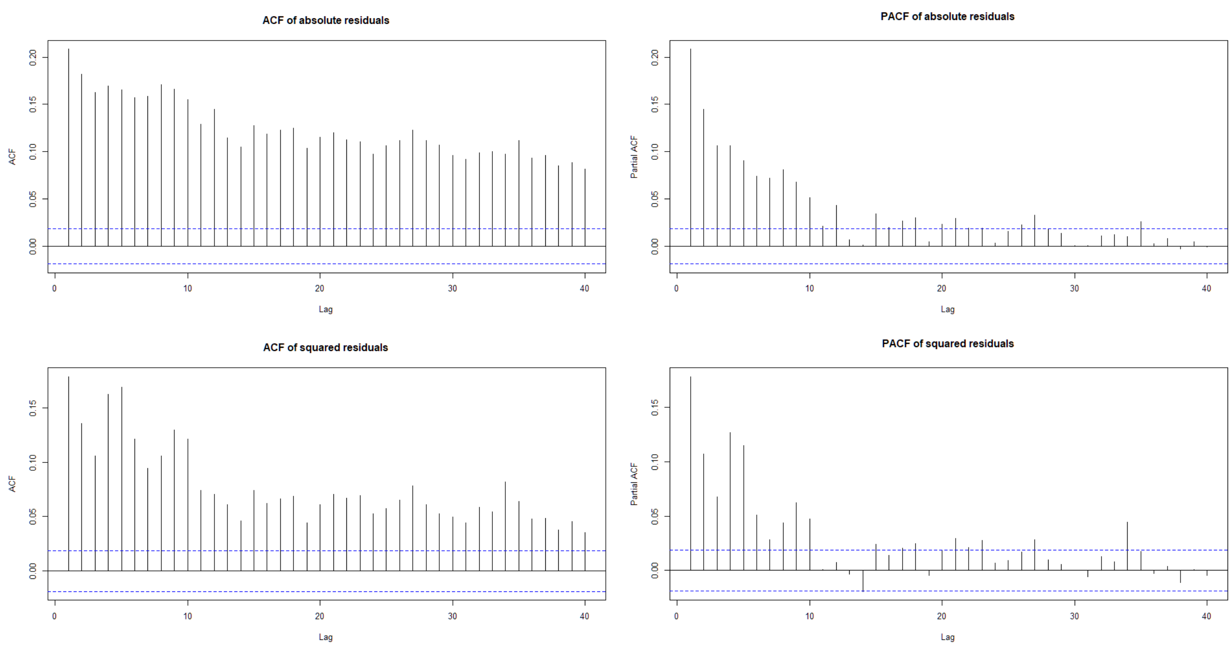
FORECAST





E. CONDITIONAL VARIANCE ANALYSIS: ARCH/GARCH MODELS

1. Based on the lowest AIC criteria, we have selected the ARIMA (2,0,3) model on the daily Log. Return series of TMO stock price timeseries spanning from 17-Mar-1980 to 12-Apr-2023.
2. We plotted the ACF and PACF of absolute and squared residuals of the ARIMA (2,0,3) model to test for heteroscedasticity problem. ACF and PACF plots of absolute and squared residuals show significant spikes i.e. significant lag correlation, which indicates we can improve this model by going for ARCH/GARCH methods.



3. Next, we did tests for presence of ARCH effect.
    a. First, we defined “at” as the mean differenced series of residuals of the ARIMA (2,0,3) model.
    b. Then, we used “Generalized Portmanteau Test for the ARCH effect” which is the Ljung Box test on ‘at^2’ series. For the best power performance, we used lags as the natural log of the number of data points in the series which comes out to be 9.29 ~ 9.
    c. H0 of the Generalized Portmanteau Test: No ARCH effect.

```
##  Box-Ljung test
##
## data:  at^2
## X-squared = 1821.5, df = 9, p-value < 2.2e-16
```

    As $p<0.05$: There is ARCH effect.

    d. We also did ‘Arch Test(r)’: $p<0.05$, i.e. There is ARCH effect.

4. Now, as we cannot determine the GARCH model from the above ACF and PACF graphs, we plotted the sample EACF for the absolute and squared residual series.
    a. EACF of absolute residual series suggested the following orders for the GARCH model : (1,1), (1,2), (2,3), (3,3)

```
## 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 o o o x  x  o  o
## 2 x x x o o o o o o o o  x  o  x
## 3 x x x o o o o o o o o  o  x  x
## 4 x x x x o o o o o o o  o  o  x
## 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 o o o o o  x  o  x
```

    b. EACF of the squared residual series was inconclusive.

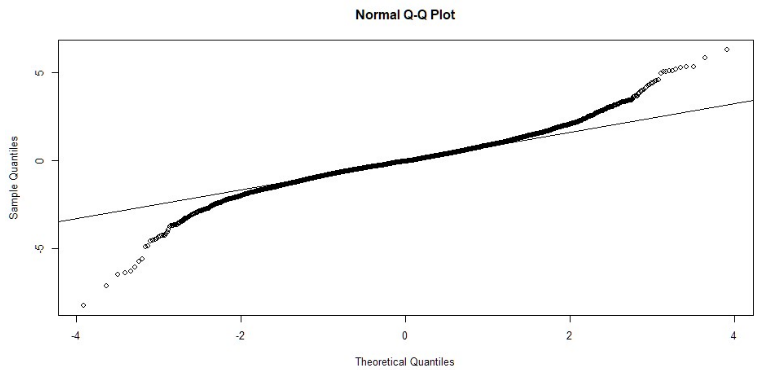
```
## AR/MA
##   0 1 2 3 4 5 6 7 8 9 10 11 12 13
## 0 x x x x x x x x x x x  x  x  x
## 1 x o x x x o x o x x x  o  o  o
## 2 x o x x x o o o o x x  o  o  x
## 3 x x x x o o o o o x o  o  o  x
## 4 x x x o o x o o o x o  x  o  x
## 5 x x x o x o o o o o o  o  o  o
## 6 x x x x x x o o o x o  x  o  o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

    c. We ran the 4 GARCH models selected from the EACF of the absolute residual series and selected the one with the lowest AIC : GARCH (1,1). Summary of the GARCH (1,1) showed that all the coefficient estimates are significant.

```
## Coefficient(s):
##     Estimate  Std. Error  t value Pr(>|t|)
## a0 5.723e-06   4.118e-07     13.9   <2e-16 ***
## a1 7.199e-02   2.891e-03     24.9   <2e-16 ***
## b1 9.155e-01   3.244e-03    282.2   <2e-16 ***
```

Jarque-Bera Test on Residuals: $p<0.05$ i.e. Normality.

Qqplot of residuals suggests student distribution.



Ljung-Box test on Residuals: p>0.05 i.e. No Autocorrelation in residuals.

d. Hence, we select GARCH (1,1) model as the best model at this stage.

5. Next, we fit the GARCH (1,1) model with Std distribution on the Log.Return series using rugarch package changing the ARFIMA order of the mean model to (2,0,3) from the default (1,0,1).
   a. We find that there is significant 'sign bias' meaning that there is leverage effect i.e. Volatility reacts differently for positive and negative jumps. Hence, to counter the leverage effect, we go for EGarch and TGarch models.

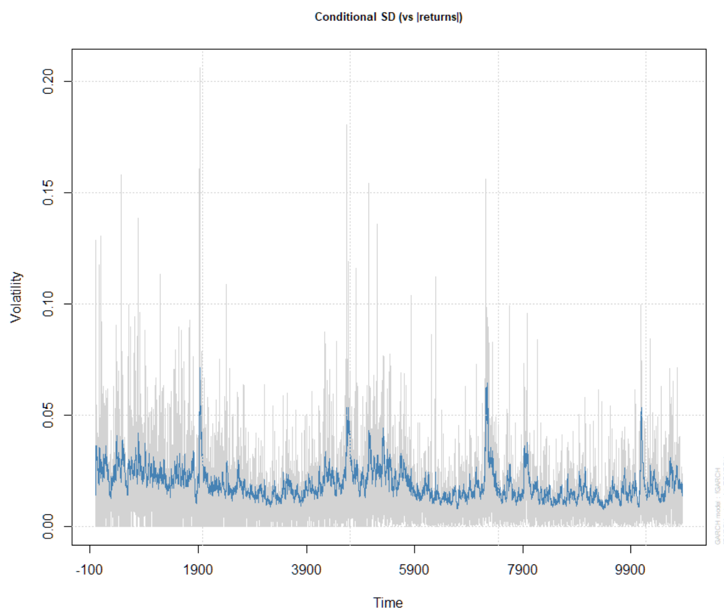
```
## Sign Bias Test
## ------------------------------------
##                    t-value     prob sig
## Sign Bias           0.3295 0.741818
## Negative Sign Bias  2.8256 0.004729 ***
## Positive Sign Bias  0.1814 0.856086
## Joint Effect       10.3570 0.015763  **
##
```

6. Next, we fit the EGarch (1,1) and TGarch (1,1) models using "Student-t distribution" and plotted the different graphs.
   a. Threshold GARCH - TGarch (1,1) model:

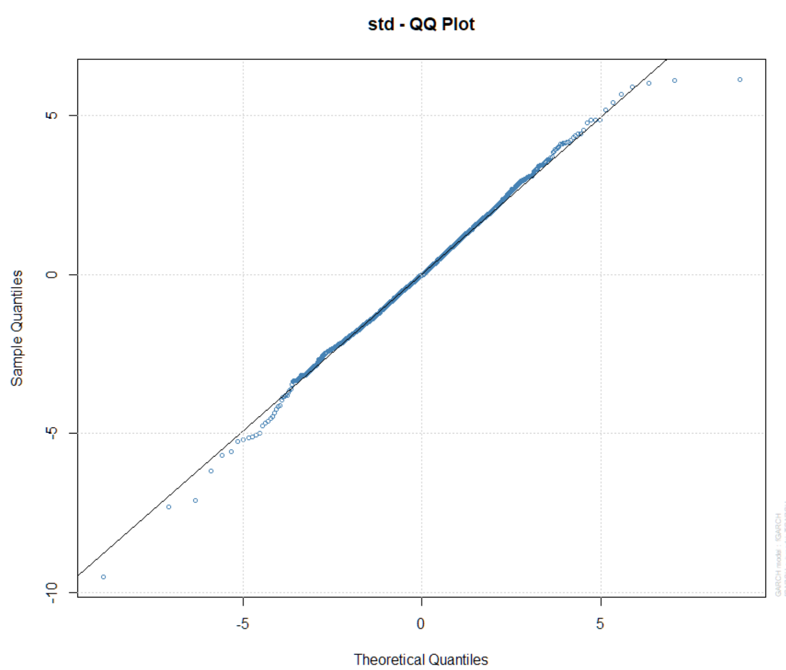
Sign Bias are removed.

```
## Sign Bias Test
## ------------------------------------
##                    t-value   prob sig
## Sign Bias            0.856 0.3920
## Negative Sign Bias   1.535 0.1247
## Positive Sign Bias   1.298 0.1943
## Joint Effect         5.194 0.1581
```

Grey line is the plot of the series. Blue line represents the volatility of the series: Seems like a good fit.



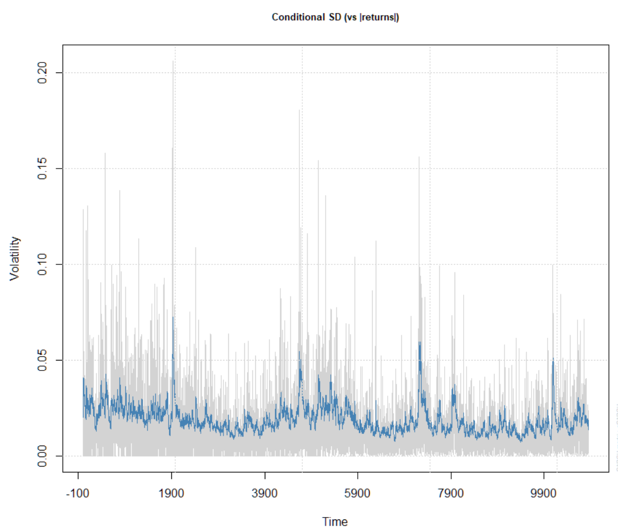
Qqplot also seems like a good fit.



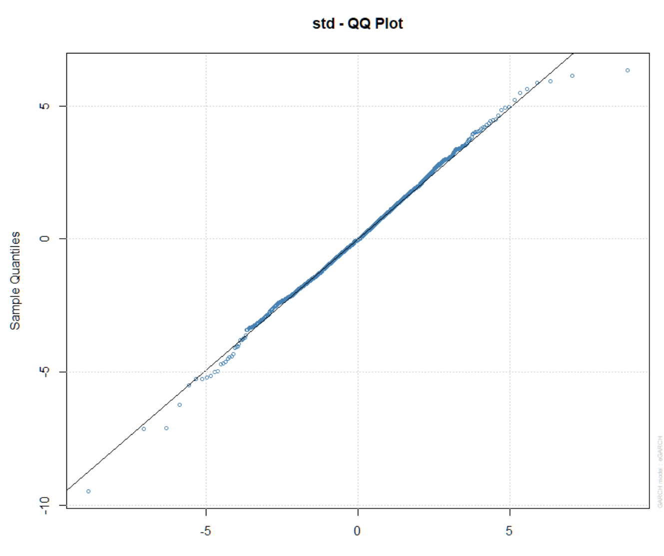
b. Exponential GARCH – EGarch (1,1) model:
Sign Bias are removed.

```
## Sign Bias Test
## ------------------------------------
##                    t-value   prob sig
## Sign Bias           0.775 0.4383
## Negative Sign Bias  1.603 0.1089
## Positive Sign Bias  1.477 0.1396
## Joint Effect        5.841 0.1196
```

Grey line is the plot of the series. Blue line represents the volatility of the series: Seems like a good fit.



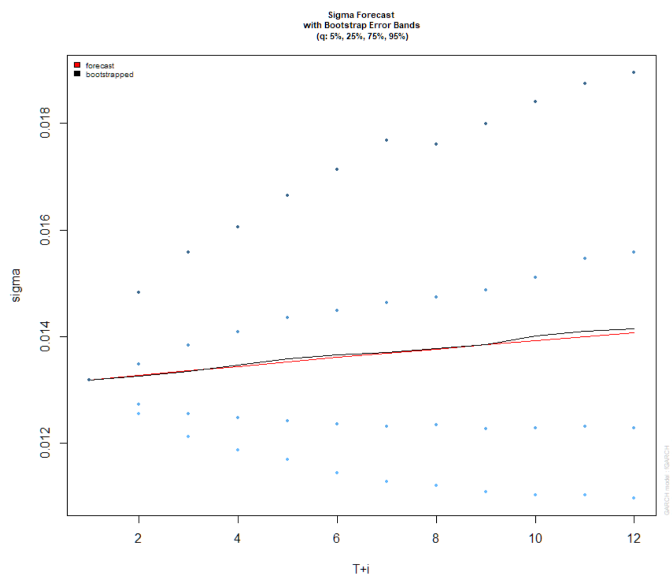
Qqplot also seems like a good fit.



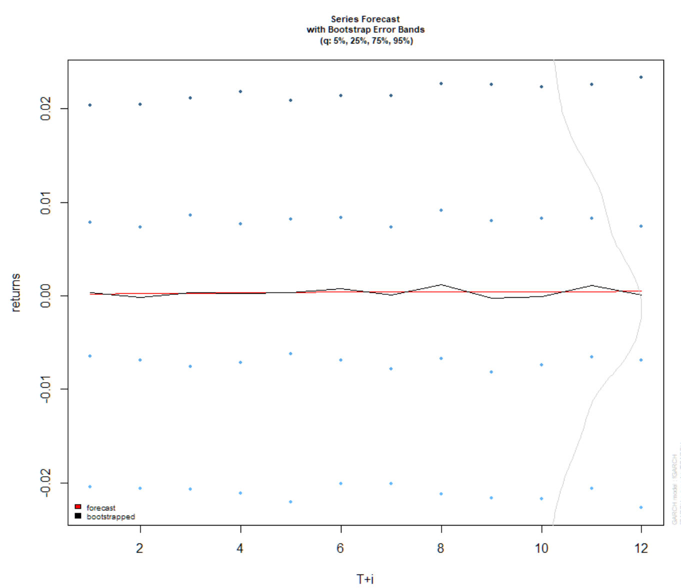
7. We also went for GARCH-in-mean (GARCHM) model, and found that p>0.05 for archm parameter, which means archm is not significant and GARCHM model is not needed.

```
## Optimal Parameters
## ------------------------------------
##          Estimate  Std. Error   t value Pr(>|t|)
## archm   -0.040539    0.025218  -1.60752 0.107941
```

8. Next, we forecasted next 12 datapoints using the EGarch (1,1) and TGarch (1,1) models with Standard-t distributions and ARIMA (2,0,3) model on the Log.Return series.
   a. TGarch (1,1):
   Density plot for sigma forecast using the Full Bootstrap method:

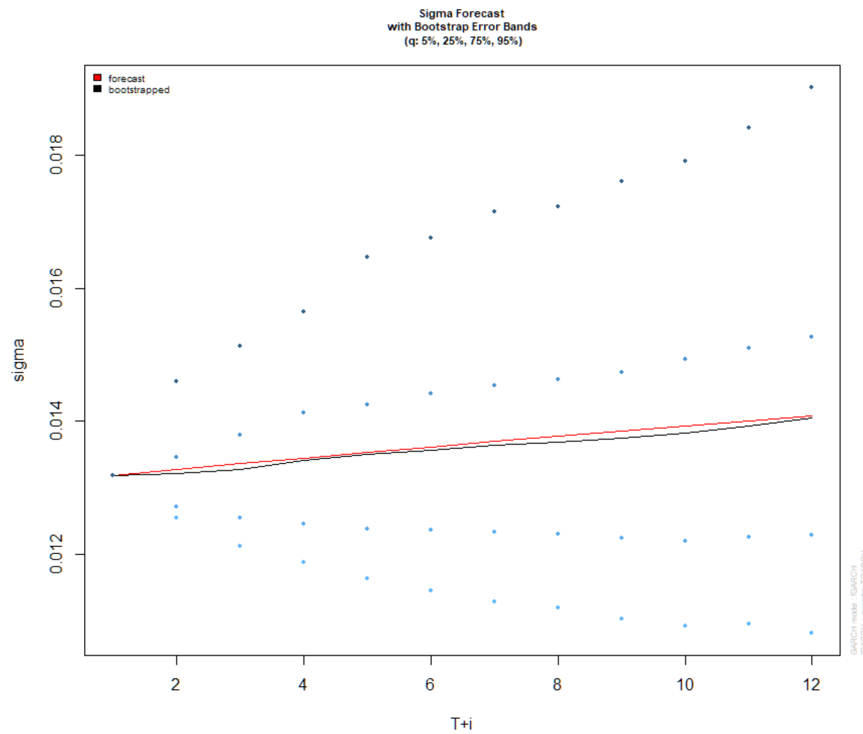


Density plot for Log.Returns forecast using the Full Bootstrap method:

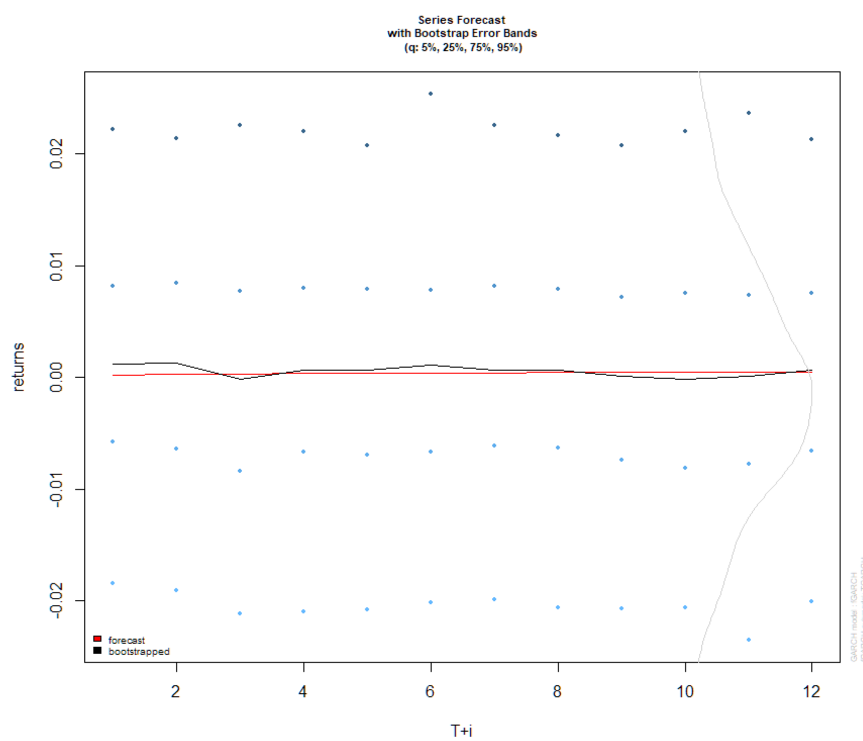


b. E-Garch (1,1):

Density plot for sigma forecast using the Full Bootstrap method:



Density plot for Log.Returns forecast using the Full Bootstrap method:



F. VALUE AT RISK ANALYSIS

a. Risk Metrics

Risk Metrics assumes that the continuously compounded daily return of a portfolio follows a conditional normal distribution. The method assumes that the logarithm of the daily price, pt = ln(Pt), of the portfolio satisfies the difference equation pt − pt−1 = at , where at = σt €t is an IGARCH(1,1) process without drift. Expected Shortfall (ES) is the expected value of the loss function if the VaR is exceeded.

RESULT FOR TMO STOCK

1) Beta is estimated.

```
> ### RiskMetrics #########
> source("RMfit.R")
> RMfit(ntmo)

Coefficient(s):
       Estimate  Std. Error  t value   Pr(>|t|)
beta 0.96161062  0.00248339  387.218 < 2.22e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

 Volatility prediction:
      Orig      Vpred
[1,] 10859 0.01338383

 Risk measure based on RiskMetrics:
      prob        VaR         ES
[1,] 0.950 0.02201444 0.02760699
[2,] 0.990 0.03113544 0.03567077
[3,] 0.999 0.04135914 0.04506455
```

2) Beta is fixed.

```
> RMfit(ntmo,estim=F)
Default beta = 0.96 is used.

 Volatility prediction:
      Orig      Vpred
[1,] 10859 0.01326526

 Risk measure based on RiskMetrics:
      prob        VaR         ES
[1,] 0.950 0.02181941 0.02736242
[2,] 0.990 0.03085961 0.03535476
[3,] 0.999 0.04099274 0.04466533
```

ANALYSIS

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 0.96161062, and it is statistically significant (p-value < 2.22e-16). The volatility prediction is 0.01338383, 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 0.03113544, while the ES is 0.03567077.

At 99.9% confidence level, VaR = 0.04135 and ES = 0.04506

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 0.01326526. 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 0.03085961, and the ES is 0.03535476

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.

b. Econometric Modelling

Econometric modeling uses a time-series model to predict the mean return (e.g. AR model) and it uses a volatility model to predict the volatility (e.g. GARCH model with innovations). The estimates will give us the forecasted one day ahead mean and variance. These estimates can then be used to calculate the VaR and ES.

Summary of the model

```
Title:
 GARCH Modelling

Call:
 garchFit(formula = ~garch(1, 1), data = tmo, cond.dist = "std",
    trace = F)

Mean and Variance Equation:
 data ~ garch(1, 1)
<environment: 0x7ff1d32efc10>
 [data = tmo]

Conditional Distribution:
 std

Coefficient(s):
        mu       omega      alpha1       beta1       shape
5.8732e-04  4.3969e-06  7.0154e-02  9.2239e-01  4.8484e+00

Std. Errors:
 based on Hessian

Error Analysis:
        Estimate  Std. Error  t value Pr(>|t|)
mu     5.873e-04   1.390e-04    4.224 2.40e-05 ***
omega  4.397e-06   8.479e-07    5.186 2.15e-07 ***
alpha1 7.015e-02   7.214e-03    9.724  < 2e-16 ***
beta1  9.224e-01   7.737e-03  119.225  < 2e-16 ***
shape  4.848e+00   2.292e-01   21.150  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Log Likelihood:
 28663.24    normalized:  2.639584

Description:
 Wed May 10 16:38:35 2023 by user:


Standardised Residuals Tests:
                                Statistic p-Value
 Jarque-Bera Test   R    Chi^2  6507.178  0
 Shapiro-Wilk Test  R    W      NA        NA
 Ljung-Box Test     R    Q(10)  26.46881  0.003158286
 Ljung-Box Test     R    Q(15)  27.46501  0.02516817
 Ljung-Box Test     R    Q(20)  33.04184  0.03338577
 Ljung-Box Test     R^2  Q(10)  8.152222  0.613971
 Ljung-Box Test     R^2  Q(15)  16.23474  0.3666247
 Ljung-Box Test     R^2  Q(20)  22.25272  0.3269562
 LM Arch Test       R    TR^2   8.951292  0.707085

Information Criterion Statistics:
      AIC       BIC       SIC      HQIC
-5.278247 -5.274889 -5.278248 -5.277115
```

ANALYSIS

- 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 4.808, which suggests that the distribution has heavier tails than a normal distribution.
- The log-likelihood value for the model is 28663.24, and the normalized log-likelihood is 2.6395. 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 (AIC), (BIC), (SIC), and (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

```
> pm1=predict(m1,10)
> pm1
   meanForecast  meanError standardDeviation
1  0.0007564997 0.01320674        0.01320674
2  0.0007564997 0.01334519        0.01334519
3  0.0007564997 0.01348050        0.01348050
4  0.0007564997 0.01361279        0.01361279
5  0.0007564997 0.01374217        0.01374217
6  0.0007564997 0.01386874        0.01386874
7  0.0007564997 0.01399260        0.01399260
8  0.0007564997 0.01411383        0.01411383
9  0.0007564997 0.01423252        0.01423252
10 0.0007564997 0.01434876        0.01434876
```

ANALYSIS

- The output shows the predicted values for the next 10 periods using the econometric model 'm1'.
- The column 'meanForecast' displays the predicted mean value of the dependent variable, while the column 'standardDeviation' shows the standard deviation of the prediction error, also known as the mean error or the forecast error.

- 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=5.066)

 Risk Measures for selected probabilities:
       prob        VaR         ES
[1,] 0.9500 0.02321827 0.03258353
[2,] 0.9900 0.03768046 0.04925654
[3,] 0.9990 0.06459288 0.08172355
[4,] 0.9999 0.10458371 0.13094128
```

10-DAY VaR

```
 Risk Measures for selected probabilities:
       prob        VaR         ES
[1,] 0.9500 0.07015629 0.08839037
[2,] 0.9900 0.09989458 0.11468165
[3,] 0.9990 0.13322812 0.14530933
[4,] 0.9999 0.16066630 0.17111573
```

ANALYSIS

- 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 0.0376 and the ES is 0.04925. This means that with a 99% probability, the maximum loss on any given day is not expected to exceed 3.76% and the expected loss is 4.93%.

- 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 0.09989 and the ES is 0.1146. This indicates that with a 99% probability, the maximum loss over a 10-day period is not expected to exceed 9.98% and the expected loss is 11.46%.

- 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.

c. Empirical Quantile & Quantile Regression

Empirical quantile is a non-parametric approach that calculates the VaR by ordering the historical returns and finding the value that corresponds to a certain level of probability. Quantile regression, on the other hand, is a parametric approach that estimates the VaR using a linear regression model. It identifies the relationship between the independent variable (e.g., market returns) and the dependent variable (e.g., portfolio returns) at various quantiles of the distribution. In terms of comparing the two methods, empirical quantile is simpler and easier to implement, Quantile regression, on the other hand, is more complex and requires assumptions about the distribution and functional form of the relationship between the variables.

OUTPUT

1. **At 95%**

```
> # fit quantile regression model
> m3=rq(ntmo~vol+GSPC,data=df,tau=0.95)
> summary(m3)

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

tau: [1] 0.95

Coefficients:
            Value    Std. Error t value  Pr(>|t|)
(Intercept)  0.03188  0.00083   38.39640  0.00000
vol          0.00000  0.00000    1.62845  0.10346
GSPC         0.00000  0.00000   -4.41438  0.00001
```

```
> VaR_quant
[1] 0.03188
```

2. AT 99%

```
> # fit quantile regression model
> m3=rq(ntmo~vol+GSPC,data=df,tau=0.99)
> summary(m3)

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

tau: [1] 0.99

Coefficients:
            Value    Std. Error t value  Pr(>|t|)
(Intercept)  0.05498  0.00265   20.75054  0.00000
vol          0.00000  0.00000    2.82042  0.00480
GSPC        -0.00001  0.00000   -3.62065  0.00030
```

```
> VaR_quant
[1] 0.0148178
```

3. AT 99.99%

```
> # fit quantile regression model
> m3=rq(ntmo~vol+GSPC,data=df,tau=0.999)
> summary(m3)

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

tau: [1] 0.999

Coefficients:
            Value    Std. Error t value  Pr(>|t|)
(Intercept)  0.12691  0.02366    5.36407  0.00000
vol          0.00000  0.00000   -0.11477  0.90863
GSPC        -0.00001  0.00001   -2.68297  0.00731
```

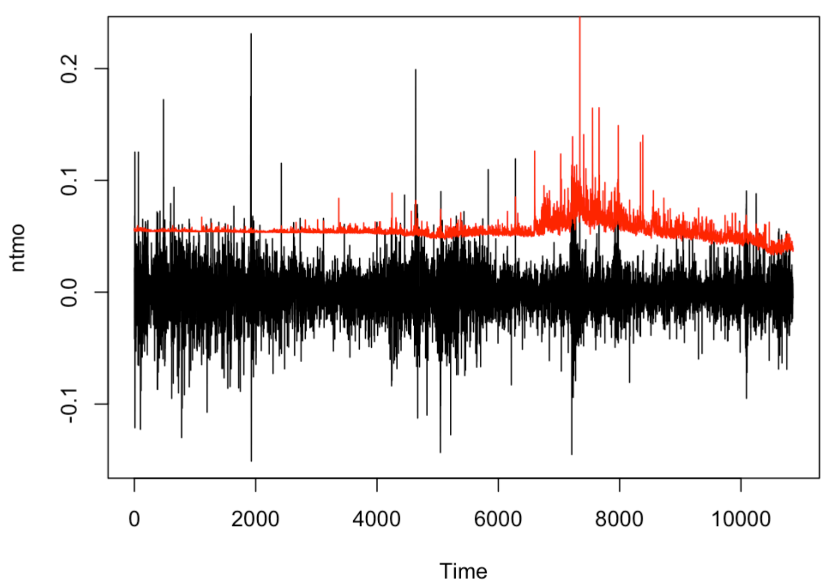
```
> VaR_quant
[1] 0.0867478
```

ANALYSIS

- 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 0.03188. This means that there is a 95% chance that the loss on the portfolio will not exceed 3.18% in one day.
- At 99% confidence level, the estimated VaR_quant is 0.01487. This means that there is a 99% chance that the loss on the portfolio will not exceed 1.48% in one day.
- At 99.9% confidence level, the estimated VaR_quant is 0.08674. This means that there is a 99.9% chance that the loss on the portfolio will not exceed 8.67% in one day.

- If the coefficient of the S&P 500 in the quantile regression is negative, it means that there is an inverse relationship between the S&P 500 and the negative log-returns of TMO stock. This may suggest that when the S&P 500 is performing well, TMO stock is likely to underperform, and vice versa.
- In terms of its impact on the VaR probability, a negative coefficient for the S&P 500 in the quantile regression implies that the inclusion of the S&P 500 as an explanatory variable in the model could potentially improve the accuracy of the VaR estimate.
- This is because the S&P 500 provides information about the broader market conditions that can affect the performance of TMO stock. By accounting for the relationship between the S&P 500 and TMO stock returns, the VaR estimate can be more informed and potentially more accurate.

Comparing Loss with VaR at 99% confidence level



ANALYSIS

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%	0.02201	0.02321	0.070156	0.03188
99%	0.03113	0.03768	0.09989	0.01487
99.9%	0.04135	0.06459	0.13329	0.08674

Analysis

- Based on the output provided, it appears that the Risk Metrics model performs the best across all confidence levels. This model consistently produces lower VaR estimates than the other models at the same confidence level.
- Therefore, based on the provided output and the reasoning above, the Risk Metrics model seems to be the best model for computing VaR. However, it is important to note that the selection of the best model may also depend on other factors such as the nature of the data and the specific goals of the analysis.

G. CONCLUSION AND MANAGERIAL IMPLICATIONS

- We first started our analysis with log returns because we had a preliminary understanding of stationarity and figured that log returns would give us a better model than price. It was clear from the start that price demonstrated some trend, and it was possible that there was seasonality too.
- After testing several ARMA models, we found that the best model was ARIMA(2,0,3), which we were satisfied with because it used a low order of lags and it passed the model-checking techniques.
- After finding ARIMA (2,0,3) as the best model at the previous stage on Log.Return series, we checked for leveraged effect by plotting acf and pacf of absolute and squared residuals of the ARIMA (2,0,3) model fitted on out Log.Returns data. We found that several lags are significant in all the acf and pacf plots which suggests we go for volatility models which might fit our data better.
- To analytically check for the leverage effect, we did the Generalized Portmanteau Test for the ARCH effect on the squared mean-differenced log-returns and found that there indeed is ARCH effect and we can move towards ARCH/GARCH models.
- As the above ACF and PACF models did not indicate which model to go for, we plotted the sample EACF on the absolute and squared residuals and found several possible models. We used the minimum AIC criteria to select GARCH(1,1) model.
- From the GARCH(1,1) model and plotting the qqplot, we can see that we need to change the distribution from normal to Student-t, and we can also see that all GARCH parameters are significant. We ran the GARCH(1,1) model with ARIMA(2,0,3) on log. Return series using Student-t distribution, and found that there is sign bias in our data (i.e. leverage effect : log. Returns reacts differently when there are similar positive and negative jumps), suggesting we need to go for EGARCH and TGARCH.
- We ran the EGARCH(1,1) and TGARCH(1,1) models with ARIMA(2,0,3) on log.returns with Student-t distribution and found that sign bias are removed, and respective plots also seemed good. We also ran the GARCHM(1,1) model and found that the archm parameter was not significant as the $p>0.05$.
- Next, we forecasted the next 12 datapoints using the EGarch (1,1) and TGarch (1,1) models with Standard-t distributions and ARIMA (2,0,3) model. Also, the density plots of volatility and log-return forecasts also seem good for both the selected models.
- Lastly, we concluded our report by performing Value at Risk Analysis by using three methods including Risk Metrics, Econometric Modelling and Quantile Regression, which suggested that the risk metric method gives the best result and can be used to analyze the value at risk in our portfolio and also the expected shortfall, which can be used to asses different business implication.