Econometrics and Time Series Applied to Finance

Group Assignment

CONTRIBUTION

We confirm that each of us has contributed equally to this project, whether in the code or in the report

HANDAR Saad | MASSE Josselin | BOUHLASSA Adil

Table des matière

Task 1 :

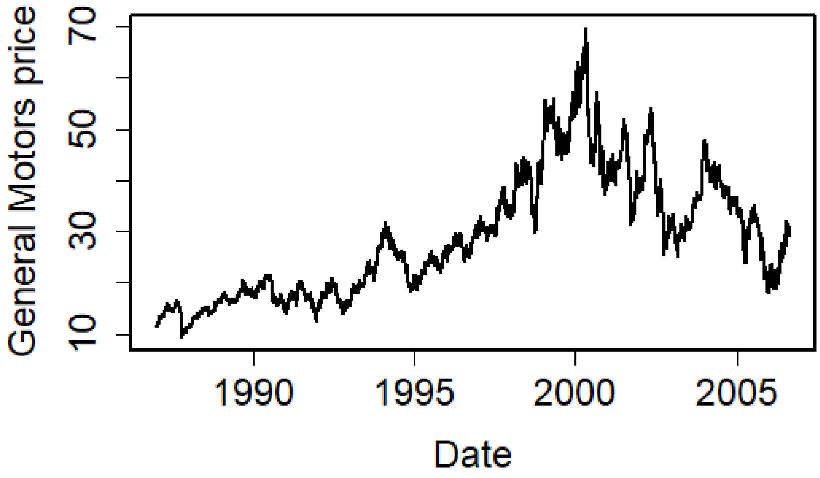
1.1 Names of the variables and adjusted closing price of General Motor and Ford

the names of the variables in the data set are :

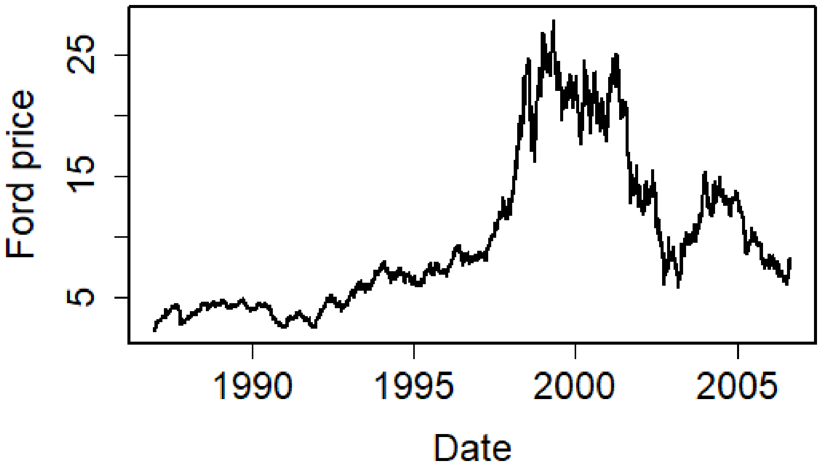
```
> print(cnames)
 [1] "Date"     "GM_AC"    "F_AC"     "MRK_AC"   "PFE_AC"   "MSFT_AC"  "X"        "X.1"
 [9] "X.2"      "X.3"      "X.4"      "X.5"
```

Screen Shot 1: The name of the variables in the Data set

General Motors and Ford Plots



Graphic 1: Price of General Motors price from 1987 to 2006



Graphic 2: Price of ford price from 1987 to 2006

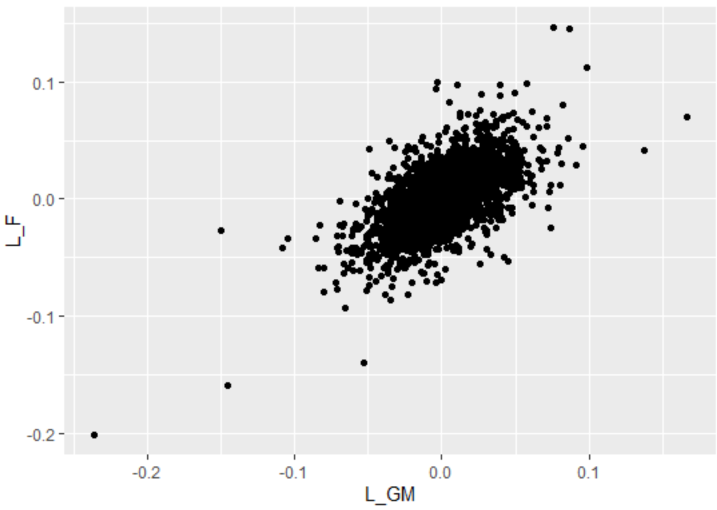
The two curves seems to fluctuate in same directions, whenever one stock price falls the other follows and so on. As we are talking about stock prices, it's hard to notice any stationarity in the series we notice as well that the mean isn't constant.

1.2 Comparison of Ford and General Motors stock returns

The sample size of the two series are the same :

```
> print(samplesize_F)
[1] 4963
> print(samplesize_GM)
[1] 4963
```

Figure 1 : Sample size of Ford and General Motors

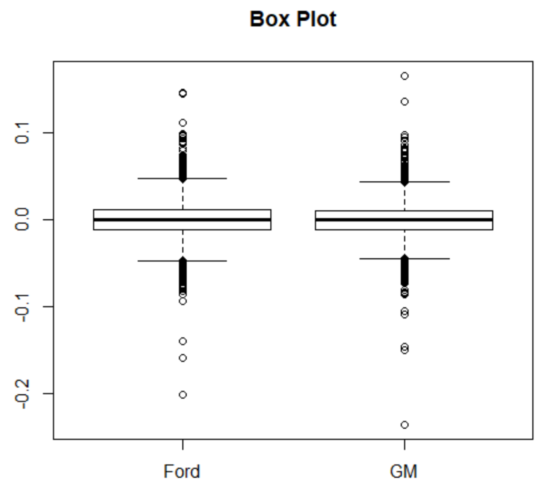


Graphic 3: General Motors returns against the Ford returns

Graphically both stock returns of Ford and General Motors seem correlated. If we have a look at graphic above we can easily grasp the positive correlation relationship between both stock returns.

We also notice that whenever an abnormal return occurred for General Motors or Ford, the other one followed almost simultaneously in the same direction.

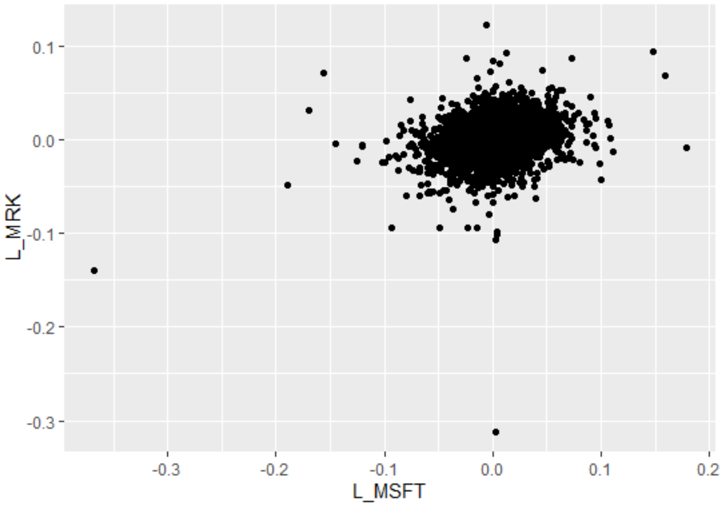
1 .3 Boxplot of the General Motors and Ford returns



Graphic 4: Ford and General Motors box plot

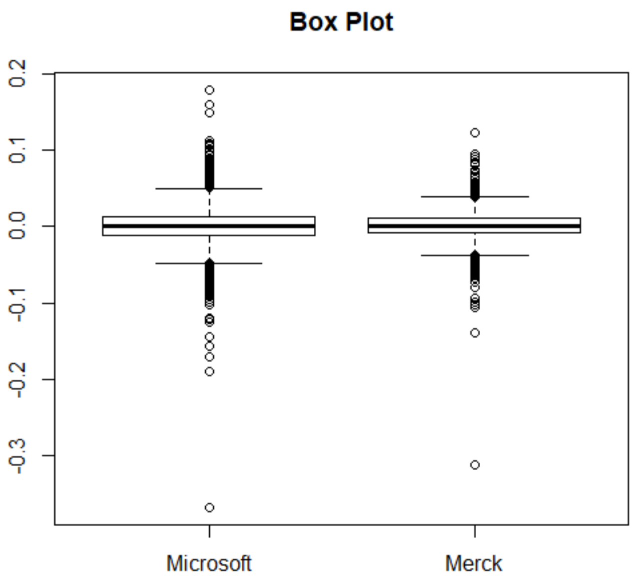
The boxplot indicates that both stocks have same log returns' mean which is approximately around 0. It also shows how scattered the returns are, for example we can easily notice that the dispersion around for the mean for log stock return of General Motors is more important than the one noticed for Ford, stating the existence of more extreme values

1.4 Comparison of Merck and Microsoft stock returns



Graphic 5: Microsoft returns against Merck returns

Again, following the same process as before, we can notice according to the first and second graph plotted that both stocks' return are positively correlated. Abnormal returns in this case are quite scattered, we can't therefore assume any correlation between these abnormalities.



Graphic 6: Microsoft and Merck box plots

The boxplot confirms the point regarding the abnormalities as it states that the log stock return of Merck are more centered around the mean which assumes less abnormalities and so correlation significance in abnormal returns between both stocks. we notice that Microsoft has more dispersed outliers.

1.5 Descriptive Statistics of stock returns and interpretations

```
[1] "For General Motors"
> summary(ret)
      Min.    1st Qu.     Median       Mean    3rd Qu.       Max.
-0.2359369 -0.0110811  0.0000000  0.0001959  0.0111987  0.1664248
> kurtosis(ret)
[1] 10.05246
> sd(ret)
[1] 0.02064997
> skewness(ret)
[1] -0.1724492
>
```

Screen Shot 2 : General Motors descriptive statistics

```
[1] "For Ford"
> summary(ret)
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
-0.200992 -0.011656  0.000000  0.000257  0.011972  0.146284
> kurtosis(ret)
[1] 8.031902
> sd(ret)
[1] 0.02104706
> skewness(ret)
[1] -0.0165494
```

Screen Shot 3: Ford descriptive statistics

```
[1] "For Microsoft"
> summary(ret)
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
-0.367725 -0.010987  0.000000  0.001052  0.013437  0.178821
> kurtosis(ret)
[1] 17.64122
> sd(ret)
[1] 0.02443024
> skewness(ret)
[1] -0.7395659
```

Screen Shot 4: Microsoft descriptive statistics

```
> summary(ret_mrk)
      Min.    1st Qu.     Median       Mean    3rd Qu.       Max.
-0.3116427 -0.0089822  0.0000000  0.0004678  0.0100636  0.1223275
> kurtosis(ret_mrk)
[1] 22.75197
attr(,"method")
[1] "excess"
> sd(ret_mrk)
[1] 0.01772822
> skewness(ret_mrk)
[1] -1.274747
```

Screen Shot 5: Merck descriptive statistics

A Normal distribution has standard attributes which are essentially symmetry (Skewness = 0) and a specific measure of tailedness (Kurtosis = 3).

From all the descriptive statistics that we got from our data set we notice that all stock's return have a skew different from 0 and a positive excess kurtosis which demonstrates that the stocks follow a leptokurtic distribution characterized by fatter tails and so, much more probability of abnormal return occurrence.

We confirmed this with Jarque-Bera test that gave us the same results for the 4 series:

```
> jarque.bera.test(ret_msft)

        Jarque Bera Test

data:  ret_msft
X-squared = 137053, df = 2, p-value < 2.2e-16
```

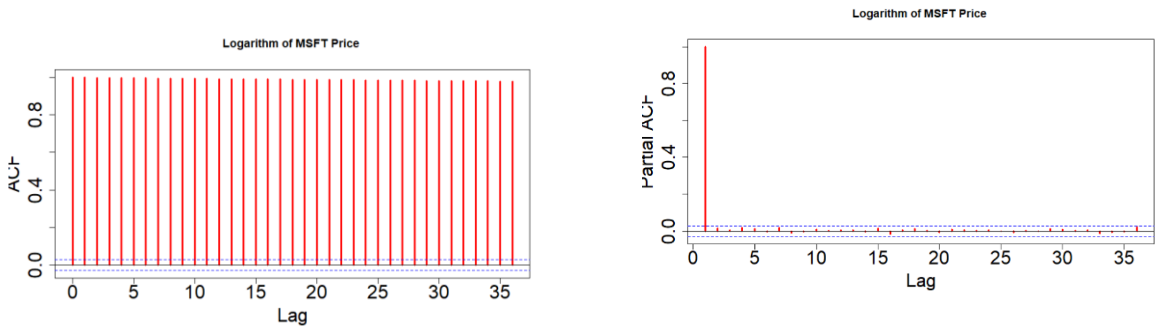
reminder: Jarque-Bera test H0: S=0 & K=3

As we can notice p-value≈0 <0.05 we need to reject the null hypothesis that the skewness is equal to 0 and the kurtosis is equal to 3 and from that we confirm that the time series is not normally distributed.

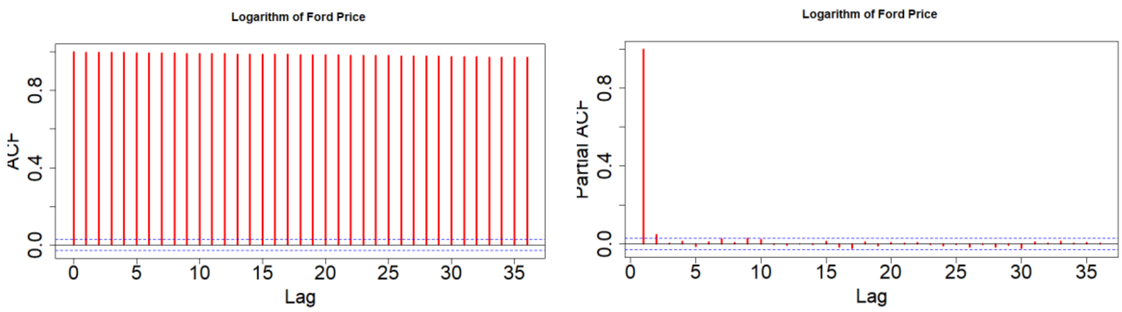
Task 2 :

2.1 Plot the ACF and PACF of the four price series (General Motors, Ford, Microsoft and Merck), and the ACF and PACF of the return series.

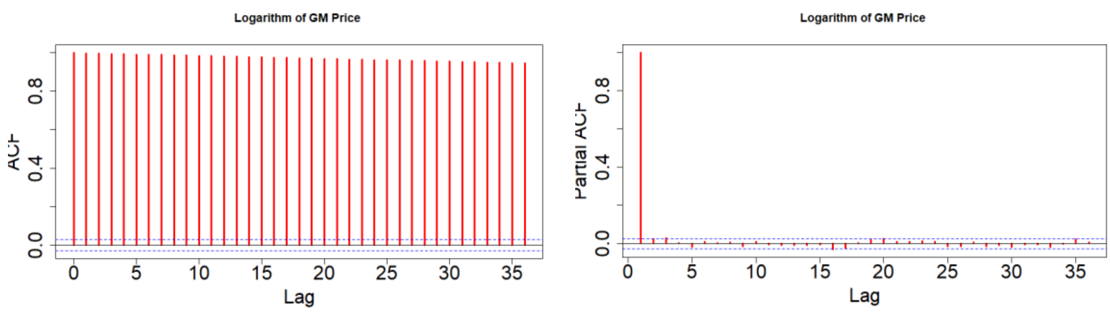
ACF and PACF of Price Series:



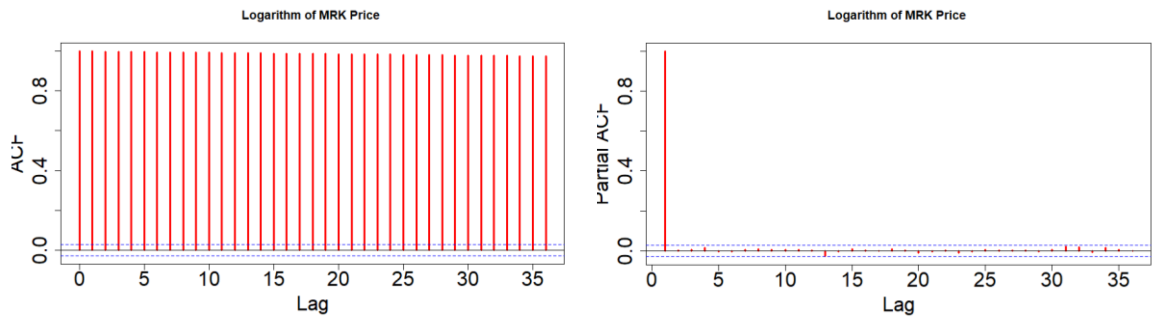
Graphic 7: Microsoft price ACF and PACF



Graphic 8: Ford price ACF and PACF



Graphic 9: General Motors price ACF and PACF



Graphic 10: Merck price ACF and PACF

Here, we notice a common pattern in the autocorrelation function and the partial autocorrelation function among the price series of the different firms. Therefore, what we will describe is valid for each price series.

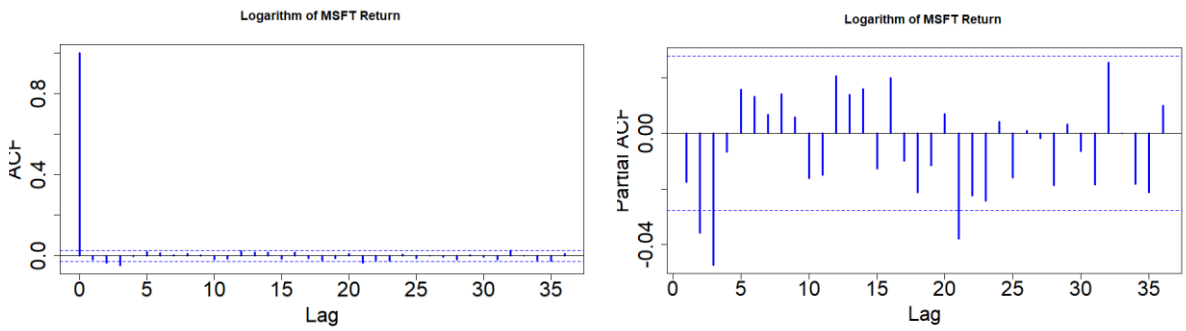
Initially, one may wonder whether it is not the first-order autocorrelation that is reflected in the other lags.

Looking at the PACF, we can rightly see that the ACF whose order is greater than 1 is ultimately explained by the first-order autocorrelation.

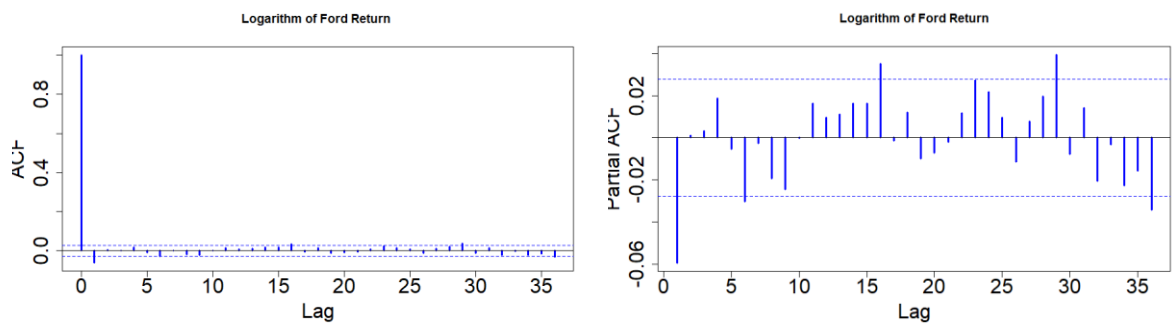
Moreover, we notice that the PACF at lag 1 is predominant and shows that an AR(1) model should be used in the case where we do not differentiate the series.

Finally, given that the PACF value at lag 1 is almost equivalent to 1, this means that we are dealing with a kind of AR model where the Beta1 coefficient is equal to 1 (unit root), which means that the series is not stationary and is similar to a random walk as seen in chapter 5.

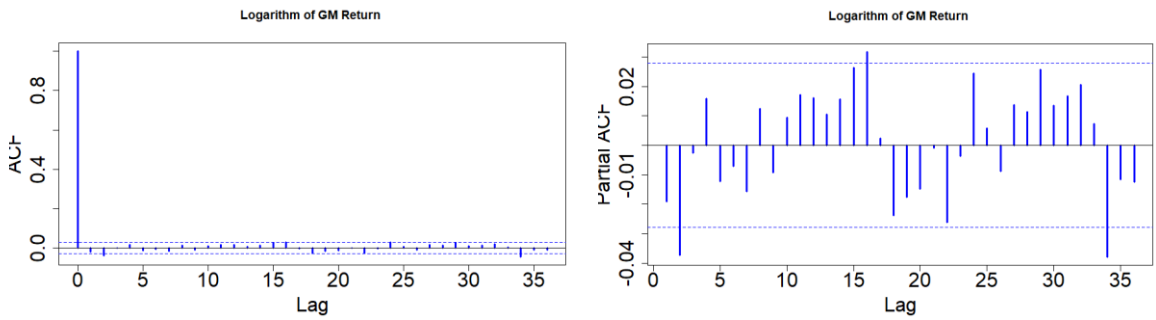
ACF and PACF of Price Series:



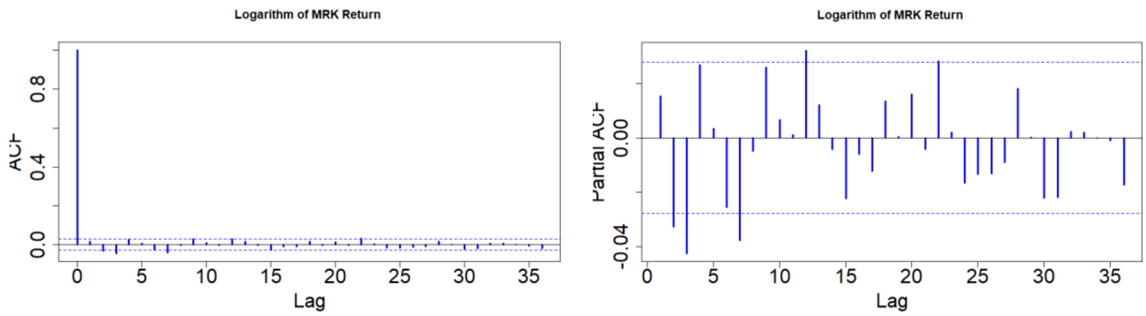
Graphic 11: Micosoft return ACF and PACF



Graphic 12: Ford return ACF and PACF



Graphic 13: General Motors return ACF and PACF



Graphic 14: Merckreturn ACF and PACF

Here, we notice a common pattern in the autocorrelation function and the partial autocorrelation function among the return series of the different firms. Therefore, what we will describe is valid for each return series.

Concerning the returns series, the reasoning is different.

First, we notice that there are few significant lags in the ACF .The other orders are approximately equal to 0. This is why we can deduce that an MA type model seems appropriate. However, there remains the PACF which does not converge to 0 and therefore does not correspond to the usual pattern of the MA model. In order to precisely detect if the series can be adapted to a MA model, the PACF should decay geometrically towards 0.

On the other hand, we can conclude that the series are stationary because of the ACF pattern.

2.2 Compute the appropriate tests to check if the price series and the return series are: stationary, serially correlated, homoscedastic, and normally distributed

As a preamble to exercise 2.2, it should be mentioned that in order to answer the question posed in its entirety, we had to create models that follow the reasoning we had in Part 2.1. In other words, it was not possible to perform a serial correlation test on anything other than errors. That is why we used the models we mentioned previously to describe the price and return series.

The price series thus followed an AR(1) model while the return series followed an MA(0) model.

Stationarity test for price and return series

```
> adf.test(LP_MSFT)

        Augmented Dickey-Fuller Test

data:  LP_MSFT
Dickey-Fuller = -0.8138, Lag order = 17, p-value = 0.9604
alternative hypothesis: stationary

> adf.test(LR_MSFT)

        Augmented Dickey-Fuller Test

data:  LR_MSFT
Dickey-Fuller = -16.94, Lag order = 17, p-value = 0.01
alternative hypothesis: stationary
```

Screen Shot 6: ADF test for Microsoft price and return series

```
> adf.test(LP_F)

        Augmented Dickey-Fuller Test

data:  LP_F
Dickey-Fuller = -1.2124, Lag order = 17, p-value = 0.9045
alternative hypothesis: stationary

> adf.test(LR_F)

        Augmented Dickey-Fuller Test

data:  LR_F
Dickey-Fuller = -15.74, Lag order = 17, p-value = 0.01
alternative hypothesis: stationary
```

Screen Shot 7: : ADF test for Ford price and return series

```
> adf.test(LP_GM)

        Augmented Dickey-Fuller Test

data:  LP_GM
Dickey-Fuller = -2.6814, Lag order = 17, p-value = 0.2898
alternative hypothesis: stationary

> adf.test(LR_GM)

        Augmented Dickey-Fuller Test

data:  LR_GM
Dickey-Fuller = -15.556, Lag order = 17, p-value = 0.01
alternative hypothesis: stationary
```

Screen Shot 8: : ADF test for GM price and return series

```
> adf.test(LP_MRK)

        Augmented Dickey-Fuller Test

data:  LP_MRK
Dickey-Fuller = -1.3019, Lag order = 17, p-value = 0.874
alternative hypothesis: stationary

> adf.test(LR_MRK)

        Augmented Dickey-Fuller Test

data:  LR_MRK
Dickey-Fuller = -16.712, Lag order = 17, p-value = 0.01
alternative hypothesis: stationary
```

Screen Shot 9: : ADF test for Merck price and return series

First of all, recalling the hypothesis of the Augmented Dickey Fuller test:

- H0 : there is a unit root in the time series, which means that the series is non-stationary
- H1: there is not unit root in the time series, which means that the series is stationary.

For the sake of this exercise, we fixed the confidence interval at 95%.

For each price series, we notice that the p-value is greater than 5% which means that we fail to reject H0. Therefore, price series seem to be non-stationary (as we mentioned in the 2.1 exercise).

On the other hand, we note for each return series that the p-value is less than 5% which means that we reject the null hypothesis. Therefore, return series seem to be stationary (as we mentioned in the 2.2 exercise).

Normality test for price and return series

```
> #Normality test on the series
> jarque.bera.test(LP_MSFT)

        Jarque Bera Test

data:  LP_MSFT
X-squared = 490.9, df = 2, p-value < 2.2e-16

> jarque.bera.test(LR_MSFT)

        Jarque Bera Test

data:  LR_MSFT
X-squared = 44772, df = 2, p-value < 2.2e-16

> jarque.bera.test(LP_MRK)

        Jarque Bera Test

data:  LP_MRK
X-squared = 413.31, df = 2, p-value < 2.2e-16

> jarque.bera.test(LR_MRK)

        Jarque Bera Test

data:  LR_MRK
X-squared = 108467, df = 2, p-value < 2.2e-16

> jarque.bera.test(LP_F)

        Jarque Bera Test

data:  LP_F
X-squared = 237.71, df = 2, p-value < 2.2e-16

> jarque.bera.test(LR_F)

        Jarque Bera Test

data:  LR_F
X-squared = 5235.1, df = 2, p-value < 2.2e-16


> jarque.bera.test(LP_GM)

        Jarque Bera Test

data:  LP_GM
X-squared = 202.74, df = 2, p-value < 2.2e-16

> jarque.bera.test(LR_GM)

        Jarque Bera Test

data:  LR_GM
X-squared = 10308, df = 2, p-value < 2.2e-16
```

Screen Shot 10:

Recalling the hypothesis of the Jarque Bera Test :

- H0 : the series follow a normal distribution.
- H1: the series does not follow a normal distribution.

For the sake of this exercise, we fixed the confidence interval at 95%.

According to Jarque Bera Test, all financial series do not follow a normal distribution since the p-value is less than 5% which means that we have to reject the null hypothesis of normal distribution.

From now on, before making any model, based on the fact that price series are not stationary, we should differentiate for non-stationary series in order to fit them in an ARMA (p,q) model. Not doing so would make the model irrelevant. However, differentiating on prices would be the same as working on return series since the price series have been transformed in log series for clarity. This is why the rest of the process will be done on the return series (the price series after having been differentiated once).

Serially Correlation test for error terms return series

```
> Box.test(MA_ret_GM$residuals, lag=5)

        Box-Pierce test

data:  MA_ret_GM$residuals
X-squared = 10.874, df = 5, p-value = 0.05394

> Box.test(MA_ret_F$residuals, lag=5)

        Box-Pierce test

data:  MA_ret_F$residuals
X-squared = 19.511, df = 5, p-value = 0.001543

> Box.test(MA_ret_MSFT$residuals, lag=5)

        Box-Pierce test

data:  MA_ret_MSFT$residuals
X-squared = 20.171, df = 5, p-value = 0.00116

> Box.test(MA_ret_MERCK$residuals, lag=5)

        Box-Pierce test

data:  MA_ret_MERCK$residuals
X-squared = 19.527, df = 5, p-value = 0.001532
```

Screen Shot 11:BOX test for return series

As we did for ADF test, recalling the hypothesis of the Box-Pierce test:

- H0 : there is not autocorrelation between errors, which means there is not serial correlation.
- H1: there is autocorrelation between errors, which means there is serial correlation.

For the sake of this exercise, we fixed the confidence interval at 95%.

Thus, according to the Box – Pierce test, since the p-value is still less than 5% for all return series, it means there is not correlation for them. For instance, it could mean that there is seasonality of the error terms. Sometimes, there is seasonality in the error term since the error term reflect another variable that has been omitted in our model. In this case, the right method would be to add it in the model in order to make the error term stationary.

From an economic outlook, it means that for MERCK and MSFT, the past values of the time series have an effect on the future values of the time series. The pattern is not random. It could reflect the idea of market inefficiency according to the definition of E. Fama.

Homoskedasticity test for error terms in return series

```
> ArchTest(MA_ret_GM$residuals, lags=60, demean = TRUE)

        ARCH LM-test; Null hypothesis: no ARCH effects

data:  MA_ret_GM$residuals
Chi-squared = 373.86, df = 60, p-value < 2.2e-16

> ArchTest(MA_ret_F$residuals, lags=60, demean = TRUE)

        ARCH LM-test; Null hypothesis: no ARCH effects

data:  MA_ret_F$residuals
Chi-squared = 382.68, df = 60, p-value < 2.2e-16

> ArchTest(MA_ret_MSFT$residuals, lags=60, demean = TRUE)

        ARCH LM-test; Null hypothesis: no ARCH effects

data:  MA_ret_MSFT$residuals
Chi-squared = 437.39, df = 60, p-value < 2.2e-16

> ArchTest(MA_ret_MERCK$residuals, lags=60, demean = TRUE)

        ARCH LM-test; Null hypothesis: no ARCH effects

data:  MA_ret_MERCK$residuals
Chi-squared = 58.092, df = 60, p-value = 0.5458
```

Screen Shot 12:ARCH test for returns series

As we did for Box-Pierce test, recalling the hypothesis of the ArchTest:

- H0 : there is not ARCH effects, which means that the series is homoscedastic.
- H1: there is ARCH effects, which means that the series is heteroskedastic.

For the sake of this exercise, we fixed the confidence interval at 95%.

According to ArchTest, since the p-value is less than 5%, all series have heteroskedasticity except MERCK series.

Normality test for error terms return series

```
> jarque.bera.test(MA_ret_GM$residuals)

        Jarque Bera Test

data:  MA_ret_GM$residuals
X-squared = 10308, df = 2, p-value < 2.2e-16

> jarque.bera.test(MA_ret_F$residuals)

        Jarque Bera Test

data:  MA_ret_F$residuals
X-squared = 5235.1, df = 2, p-value < 2.2e-16

> jarque.bera.test(MA_ret_MSFT$residuals)

        Jarque Bera Test

data:  MA_ret_MSFT$residuals
X-squared = 44772, df = 2, p-value < 2.2e-16

> jarque.bera.test(MA_ret_MERCK$residuals)

        Jarque Bera Test

data:  MA_ret_MERCK$residuals
X-squared = 108467, df = 2, p-value < 2.2e-16
```

Screen Shot 13: Jarque Bera test return series

As we did for ArchTest , recalling the hypothesis of the Jarque Bera Test :

- H0 : the series follow a normal distribution.
- H1: the series does not follow a normal distribution.

For the sake of this exercise, we fixed the confidence interval at 95%.

According to Jarque Bera Test, all error terms from the series do not follow a normal distribution since the p-value is less than 5% which means that we have to reject the null hypothesis.

If the error terms are not normally distributed, it means that there is some skewness. When there is skewness in the error terms, this means that the mean of the error is not equal to 0. In other words, the model taken for the time series does not correctly describe the time series.

One more thing to conclude the exercise 2.2.
We are aware that the models that have been using in this section have not allowed us to properly capture and describe the financial series. However, they were necessary for us to carry out tests of serial correlation for example. According to the results, there seems to be no series suitable for an ARMA-type model. The ARMA model does not consider the non-constancy of the error variance. However, if one were to choose at least one, we believe that the MERCK return series ticks the most boxes.
Thus, the results lead us to believe that models such as ARCH/ GARCH are more relevant when we find that the errors of the models present heteroskedasticity.

2.3 Based on your answers to the task 2.2, fit an AR(1) model on the General Motors series

```
arima(x = LR_GM, order = c(1, 0, 0))

Coefficients:
          ar1  intercept
      -0.0191      2e-04
s.e.   0.0142      3e-04

sigma^2 estimated as 0.0004262:  log likelihood = 12213.4,  aic = -24420.8
```

Screen Shot 14: AR(1) Model for GM

Here are the main characteristics of our AR(1) model.
Before any economic comments, it should be said that ar1 is the coefficient showing what is the impact of the variable Xt-1 over Xt.
The intercept indicates the value that should be taken by Xt when Xt-1 is equal to 0.
From an economic perspective, it means that Xt-1 (the value at a day before) is not relevant to predict the day after. The advantage of that is that economics shocks should be disappeared rapidly.

```
> Box.test(AR_ret_GM$residuals, lag=5)

        Box-Pierce test

data:  AR_ret_GM$residuals
X-squared = 9.1549, df = 5, p-value = 0.103

> ArchTest(AR_ret_GM$residuals, lags=60, demean = TRUE)

        ARCH LM-test; Null hypothesis: no ARCH effects

data:  AR_ret_GM$residuals
Chi-squared = 353.39, df = 60, p-value < 2.2e-16

> jarque.bera.test(AR_ret_GM$residuals)

        Jarque Bera Test

data:  AR_ret_GM$residuals
X-squared = 10486, df = 2, p-value < 2.2e-16
```

According to the above test, the errors are not serially correlated.
However, the errors do not meet some CLRM criteria since they do not follow a normal distribution and they are heteroskedastic.
The implication to those result for the OLS model is the fact that it makes it impossible to use since there are two main assumptions which are not met.

In other words, the heteroskedasticity contradicts the fact that the estimator is the best (BLUE) since the it would not display the lowest variance. One could also say that the probability that the fitted parameter is equal to the real parameter is lower when there is strong heteroscedasticity.
A possible solution to analyze this time series is to use an ARCH/GARCH model since this is a type of model which considers the conditional heteroskedasticity.

2.4 Compute the appropriate tests to check if the price series and the return series are: stationary, serially correlated, homoscedastic, and normally distributed

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

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

Optimal Parameters
------------------------------------
        Estimate  Std. Error   t value Pr(>|t|)
mu      0.000605    0.000256   2.36124 0.018214
ar1     0.008495    0.015322   0.55442 0.579288
omega   0.000012    0.000001  21.70848 0.000000
alpha1  0.077620    0.003439  22.56846 0.000000
beta1   0.896147    0.005449 164.47354 0.000000

Robust Standard Errors:
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.000605    0.000251  2.40699 0.016085
ar1     0.008495    0.015979  0.53162 0.594988
omega   0.000012    0.000001  9.03787 0.000000
alpha1  0.077620    0.006941 11.18272 0.000000
beta1   0.896147    0.010118 88.56895 0.000000

LogLikelihood : 12517.96

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

Akaike       -5.0435
Bayes        -5.0370
Shibata      -5.0435
Hannan-Quinn -5.0412

Weighted Ljung-Box Test on Standardized Residuals
------------------------------------
                        statistic p-value
Lag[1]                     0.3252  0.5685
Lag[2*(p+q)+(p+q)-1][2]    1.9253  0.2398
Lag[4*(p+q)+(p+q)-1][5]    3.4365  0.3299
d.o.f=1
H0 : No serial correlation
```

```
Weighted Ljung-Box Test on Standardized Squared Residuals
------------------------------------
                        statistic p-value
Lag[1]                      3.549 0.05957
Lag[2*(p+q)+(p+q)-1][5]     4.122 0.23928
Lag[4*(p+q)+(p+q)-1][9]     4.836 0.45392
d.o.f=2

Weighted ARCH LM Tests
------------------------------------
            Statistic Shape Scale P-Value
ARCH Lag[3]   0.05775 0.500 2.000  0.8101
ARCH Lag[5]   0.62350 1.440 1.667  0.8467
ARCH Lag[7]   1.04414 2.315 1.543  0.9063

Nyblom stability test
------------------------------------
Joint Statistic:  23.6967
Individual Statistics:
mu     0.05229
ar1    0.16202
omega  2.77593
alpha1 0.35449
beta1  0.67135

Asymptotic Critical Values (10% 5% 1%)
Joint Statistic:         1.28 1.47 1.88
Individual Statistic:    0.35 0.47 0.75

Sign Bias Test
------------------------------------
                   t-value     prob sig
Sign Bias           1.5035 0.132775
Negative Sign Bias  2.5860 0.009739 ***
Positive Sign Bias  0.4175 0.676367
Joint Effect        6.8656 0.076309   *


Adjusted Pearson Goodness-of-Fit Test:
------------------------------------
  group statistic p-value(g-1)
1    20     151.9    9.394e-23
2    30     269.1    9.932e-41
3    40     378.7    3.095e-57
4    50     448.3    6.955e-66
```

First and foremost, the model that was chosen is a GARCH(1,1) model.
Initially, in order to take into account the conditional variance, we could have simply chosen an ARCH(1) model. This could have been relevant since the criticism of our model in section 2.3 was that it did not consider the conditional variance. However, we chose a GARCH(1,1). This is because of what the GARCH model has in addition to an ARCH(1) model. It is richer than the ARCH(p) model since it incorporates a measure of conditional variance at time t-1 in its model (if it is a GARCH(1,1)).

From the GARCH model we decided to implement, we can make several conclusions:
Optimal parameters window:

> All parameters are not statistically significant. Ar1 parameter has a p-value above 5% which means that the parameter is not significant. We could remove it from our model in order to obtain a more precise GARCH model. The other parameters are statistically significant.
> Alpha1 + beta 1 < which shows that our model is stationary.

Information criteria:

> It helps to choose the simplest model. The simplest model is the one which has the lowest information criteria. In our case, this is Aikake and Shibata.

Ljung Box Test:

It specifies if there is any form of serial correlation in our model. When the p-value is above 0.05, we fail to reject the null hypothesis and we conclude to the non-serial correlation of the residuals. This is the case in our testing model.

Goodness of fit :

Since the p-value is by far lower than the significance level, it means that the distribution we use to model our errors is not the right one.

Diagnosis testing on our GARCH(1,1) model:

```
> Box.test(StandRes)

        Box-Pierce test

data:  StandRes
X-squared = 0.32499, df = 1, p-value = 0.5686

> ArchTest(StandRes)

        ARCH LM-test; Null hypothesis: no ARCH effects

data:  StandRes
Chi-squared = 9.0763, df = 12, p-value = 0.6964

> jarque.bera.test(StandRes)

        Jarque Bera Test

data:  StandRes
X-squared = 5052.9, df = 2, p-value < 2.2e-16
```

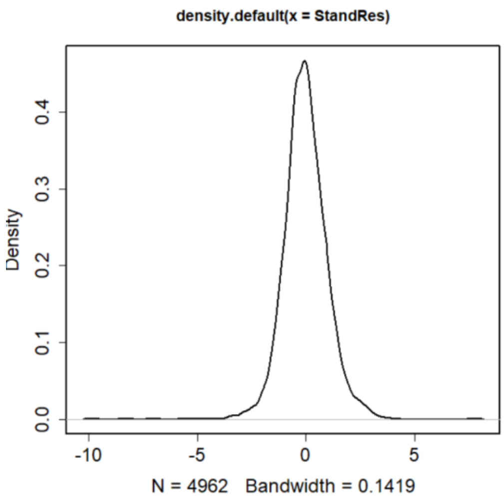
According to the above test, the errors are not serially correlated (Boxtest, p-value>0.05), and the errors are not heteroskedastic anymore(ArchTest, p-value>0.05).

This means that our model was able to capture the effect of the variability of the conditional variance of our errors. This was not the case when dealing with part 2.3.

Regarding our errors, they are still non-normal according to the Jarque Bera Test.

However, to see how our errors behave, let's do a quick inspection to see how non-normal our error distribution is.

First, we plot our StandRes:



At first sight, our distribution seems to be highly concentrated around the mean which makes it sharp.

So, let’s compute the metrics used in Jarque Bera Test in order to see what are the reasons why our distribution is non-normal.

```
> kurtosis(StandRes)
[1] 4.922546
attr(,"method")
[1] "excess"
> skewness(StandRes)
[1] -0.2102759
attr(,"method")
[1] "moment"
```

We understand now that our error distribution is non-normal for the Jarque Bera Test however it comes very close to it. The skewness which indicates the symmetry of the distribution is very low in absolute values.

The issue seems to come from the kurtosis, since it has an excess of 1.9 compared to the normal distribution kurtosis.

To conclude, it seems that our ARMA GARCH model is not so bad when it comes to model our General Motors return series. Nevertheless, this does not mean that the model is perfect. It has a better describing effect than the previous ones in part 2.2 and 2.3.

2.5 Fit an EGARCH model for the GM series you used in 2.4

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

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

Optimal Parameters
------------------------------------
        Estimate  Std. Error   t value Pr(>|t|)
mu      0.000182    0.000410   0.44466  0.65657
ar1    -0.006066    0.026043  -0.23291  0.81583
omega  -0.141580    0.005086 -27.83464  0.00000
alpha1 -0.061254    0.004852 -12.62331  0.00000
beta1   0.981411    0.000547 1795.68293 0.00000
gamma1  0.106955    0.014272   7.49411  0.00000

Robust Standard Errors:
        Estimate  Std. Error   t value Pr(>|t|)
mu      0.000182    0.003469  0.052545 0.958094
ar1    -0.006066    0.223983 -0.027081 0.978395
omega  -0.141580    0.036406 -3.888881 0.000101
alpha1 -0.061254    0.055676 -1.100181 0.271253
beta1   0.981411    0.004489 218.605336 0.000000
gamma1  0.106955    0.050262  2.127946 0.033342

LogLikelihood : 12549.91

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

Akaike       -5.0560
Bayes        -5.0481
Shibata      -5.0560
Hannan-Quinn -5.0532

Weighted Ljung-Box Test on Standardized Residuals
------------------------------------
                        statistic p-value
Lag[1]                     0.1875  0.6650
Lag[2*(p+q)+(p+q)-1][2]    2.0644  0.1924
Lag[4*(p+q)+(p+q)-1][5]    3.8399  0.2521
d.o.f=1
H0 : No serial correlation
```

```
Weighted Ljung-Box Test on Standardized Squared Residuals
------------------------------------
                        statistic  p-value
Lag[1]                      9.166 0.002466
Lag[2*(p+q)+(p+q)-1][5]     9.751 0.010662
Lag[4*(p+q)+(p+q)-1][9]    10.219 0.045034
d.o.f=2

Weighted ARCH LM Tests
------------------------------------
            Statistic Shape Scale P-Value
ARCH Lag[3]    0.6825 0.500 2.000  0.4087
ARCH Lag[5]    0.6854 1.440 1.667  0.8279
ARCH Lag[7]    1.0218 2.315 1.543  0.9100

Nyblom stability test
------------------------------------
Joint Statistic:  1.1284
Individual Statistics:
mu     0.1347
ar1    0.1048
omega  0.4405
alpha1 0.1420
beta1  0.4155
gamma1 0.2846

Asymptotic Critical Values (10% 5% 1%)
Joint Statistic:         1.49 1.68 2.12
Individual Statistic:    0.35 0.47 0.75

Sign Bias Test
------------------------------------
                   t-value    prob sig
Sign Bias            1.601 0.10936
Negative Sign Bias   2.558 0.01057  **
Positive Sign Bias   1.104 0.26969
Joint Effect         9.953 0.01897  **

Adjusted Pearson Goodness-of-Fit Test:
------------------------------------
  group statistic p-value(g-1)
1    20     176.1    1.772e-27
2    30     248.7    9.246e-37
3    40     339.1    1.623e-49
4    50     442.5    9.424e-65
```

Screen Shot 15: Egarch results for General Motors

The interest of using such GARCH is to consider leverage effect. So, we should wonder if our GARCH model has some leverage effects. To do so, we have to look at the value of gamma. Here the gamma, aka gamma1 is positive and statistically significant(p-value < 0.05). A non-zero value of gamma1 means that there is more volatility in time trouble than other situation. From an economic outlook, we could say that a negative shock has more effect on the variability of the data than a positive one. Let's make a quick analysis of the model.

Optimal parameters window:

Some parameters seem to be non-significant, as it is the case for mu and ar1.

Alpha1 and beta1, when added are less than 1 which confirms the stationarity.

Gamma1 as we mentioned above, it is statistically significant, and the value is positive.

Information criteria:

It helps to choose the simplest model. The simplest model is the one which has the lowest information criteria. In our case, this is Aikake and Shibata.

Ljung Box Test:

It specifies if there is any form of serial correlation in our model. When the p-value is above 0.05, we fail to reject the null hypothesis and we conclude to the non-serial correlation of the residuals. This is the case in our testing model.

Goodness of fit :

Since the p-value is by far lower than the significance level, it means that the distribution we use to model our errors is not the right one.

Micorosoft

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

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

Optimal Parameters
------------------------------------
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.001165    0.000482   2.4194 0.015548
ar1    -0.021365    0.014347  -1.4891 0.136459
omega  -0.103244    0.017157  -6.0176 0.000000
alpha1 -0.020288    0.005335  -3.8028 0.000143
beta1   0.985455    0.001924 512.3161 0.000000
gamma1  0.131462    0.016038   8.1969 0.000000

Robust Standard Errors:
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.001165    0.006846  0.17017  0.86487
ar1    -0.021365    0.017137 -1.24673  0.21250
omega  -0.103244    0.214171 -0.48207  0.62976
alpha1 -0.020288    0.103987 -0.19510  0.84532
beta1   0.985455    0.024093 40.90191  0.00000
gamma1  0.131462    0.197086  0.66703  0.50475

LogLikelihood : 11935.16

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

Akaike       -4.8082
Bayes        -4.8003
Shibata      -4.8082
Hannan-Quinn -4.8054

Weighted Ljung-Box Test on Standardized Residuals
------------------------------------
                        statistic p-value
Lag[1]                     0.3196  0.5718
Lag[2*(p+q)+(p+q)-1][2]    0.6012  0.9353
Lag[4*(p+q)+(p+q)-1][5]    1.0624  0.9371
d.o.f=1
H0 : No serial correlation

Weighted Ljung-Box Test on Standardized Squared Residuals
------------------------------------
                        statistic p-value
Lag[1]                     0.5291  0.4670
Lag[2*(p+q)+(p+q)-1][5]    1.3125  0.7858
Lag[4*(p+q)+(p+q)-1][9]    1.7044  0.9372
d.o.f=2

Weighted ARCH LM Tests
------------------------------------
            Statistic Shape Scale P-Value
ARCH Lag[3]    0.1499 0.500 2.000  0.6987
ARCH Lag[5]    0.2929 1.440 1.667  0.9422
ARCH Lag[7]    0.4713 2.315 1.543  0.9809

Nyblom stability test
------------------------------------
Joint Statistic:  4.3722
Individual Statistics:
mu     1.44941
ar1    0.11676
omega  0.45259
alpha1 0.12716
beta1  0.37073
gamma1 0.05091

Asymptotic Critical Values (10% 5% 1%)
Joint Statistic:         1.49 1.68 2.12
Individual Statistic:    0.35 0.47 0.75

Sign Bias Test
------------------------------------
                   t-value    prob sig
Sign Bias           1.5248 0.12737
Negative Sign Bias  2.2395 0.02517  **
Positive Sign Bias  0.4215 0.67341
Joint Effect        5.5041 0.13839


Adjusted Pearson Goodness-of-Fit Test:
------------------------------------
  group statistic p-value(g-1)
1    20     900.4    9.096e-179
2    30    1273.7    8.335e-250
3    40    1067.1    2.060e-198
4    50    1160.7    6.654e-211


Elapsed time : 0.500694
```

Screen Shot 16: Egarch results for Microsoft

We will now analyse Microsoft under the EGarch model in comparison with the General Motors financial series.

Optimal parameters window:

Some parameters seem to be non-significant, as it is the case for mu and ar1.

Alpha1 and beta1, when added are less than 1 which confirms the stationarity.

Gamma1 as we mentioned above, it is statistically significant, and the value is positive. From that respect, the value is higher than the one from General Motors. It means that the leverage effect in Microsoft return series is more important.

Information criteria:

In general, the information criteria is less important for Microsoft return series than for General Motors. This would mean that the model of General Motors series seems to be a bit better in terms of describing its series.

The simplest model is the one which has the lowest information criteria. In our case, this is Aikake and Shibata.

Ljung Box Test:

It specifies if there is any form of serial correlation in our model. When the p-value is above 0.05, we fail to reject the null hypothesis and we conclude to the non-serial correlation of the residuals. This is the case in our testing model.

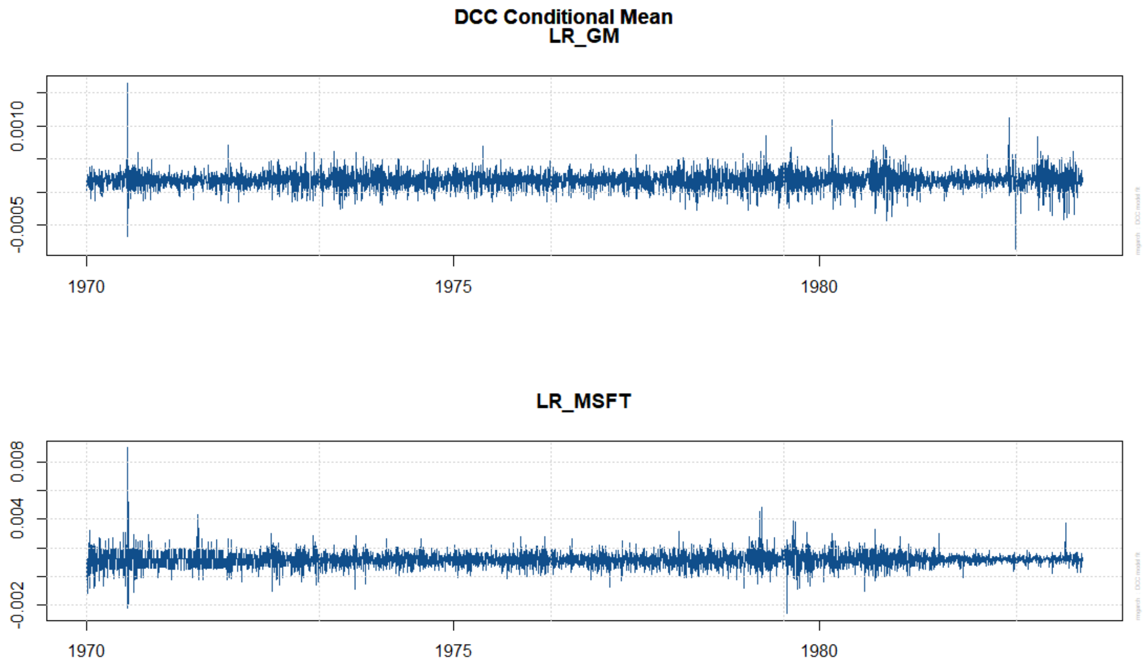
Goodness of fit :

Here we note a big gap in the GOF between Microsoft and General Motors. The GOF of Microsoft is far lower than the General Motors' one. It means that the distribution used to model our errors seems to be more appropriate for GM than Microsoft.

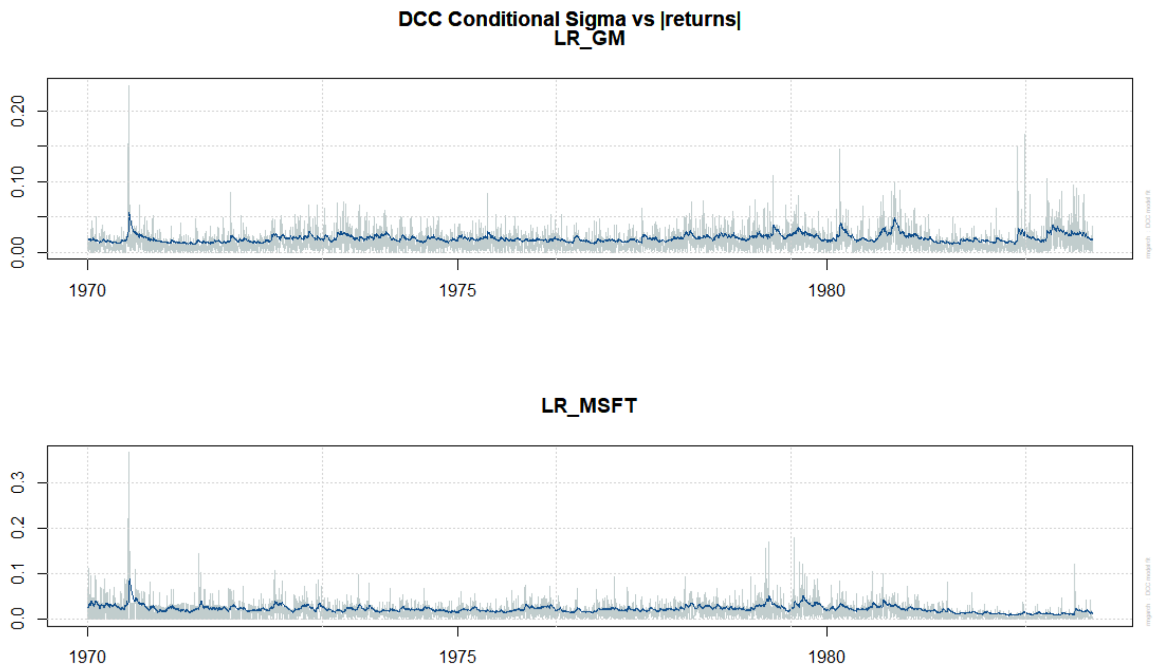
2.6 Fit a GARCH-DCC model with the General Motors and Microsoft series

```
*         DCC GARCH Fit          *
*---------------------------------*

Distribution         :  mvnorm
Model                :  DCC(1,1)
No. Parameters       :  15
[VAR GARCH DCC UncQ] : [0+12+2+1]
No. Series           :  2
No. Obs.             :  4962
Log-Likelihood       :  24722.16
Av.Log-Likelihood    :  4.98

Optimal Parameters
-----------------------------------
                  Estimate  Std. Error   t value Pr(>|t|)
[LR_GM].mu        0.000180    0.000891   0.20239 0.839613
[LR_GM].ar1      -0.006250    0.048649  -0.12847 0.897780
[LR_GM].omega    -0.141589    0.174968  -0.80922 0.418386
[LR_GM].alpha1   -0.061257    0.274737  -0.22297 0.823562
[LR_GM].beta1     0.981409    0.024407  40.20988 0.000000
[LR_GM].gamma1    0.106935    0.155215   0.68895 0.490856
[LR_MSFT].mu      0.001165    0.006463   0.18022 0.856976
[LR_MSFT].ar1    -0.021374    0.020824  -1.02643 0.304689
[LR_MSFT].omega  -0.103232    0.201918  -0.51126 0.609170
[LR_MSFT].alpha1 -0.020283    0.098866  -0.20516 0.837449
[LR_MSFT].beta1   0.985457    0.022678  43.45485 0.000000
[LR_MSFT].gamma1  0.131455    0.185551   0.70846 0.478660
[Joint]dcca1      0.012716    0.004069   3.12541 0.001776
[Joint]dccb1      0.980855    0.008286 118.37079 0.000000

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

Akaike       -9.9585
Bayes        -9.9389
Shibata      -9.9586
Hannan-Quinn -9.9516

Elapsed time : 7.036962
```

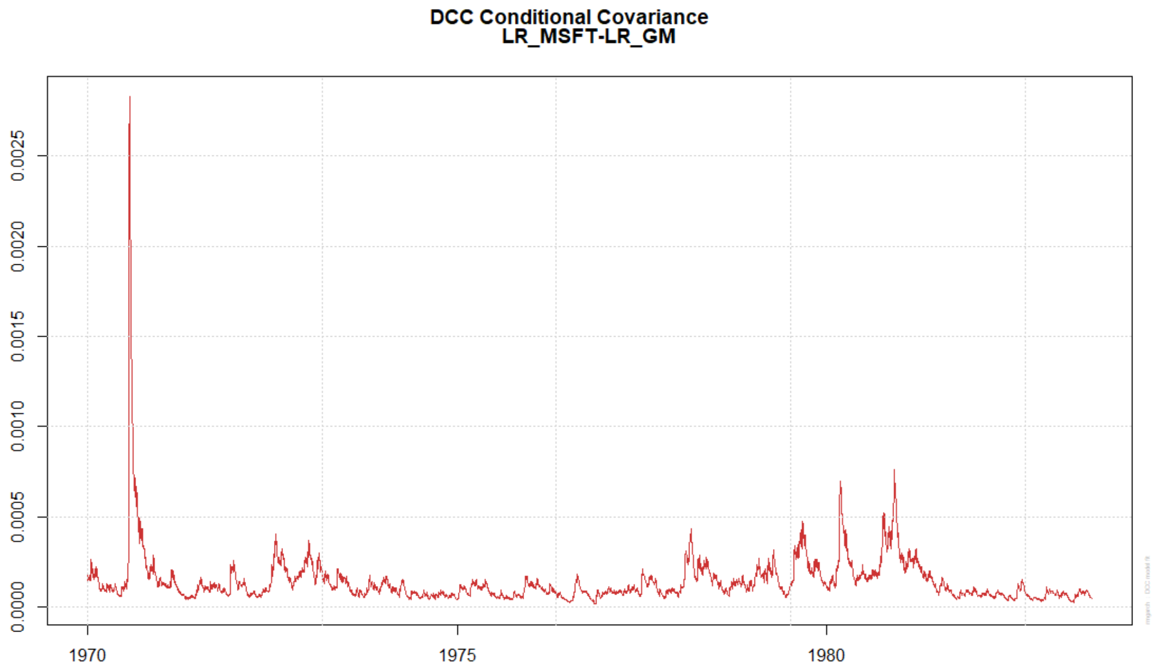
Screen Shot 22: DCC GARCH results



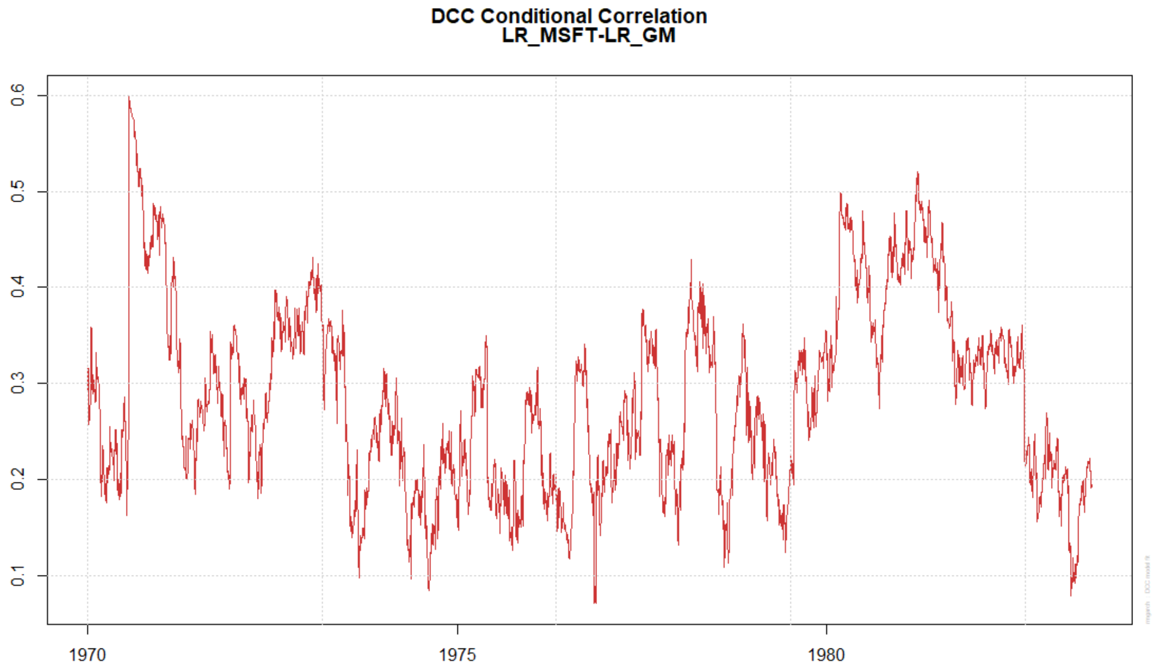
Graphic 13:DCC Conditional Mean



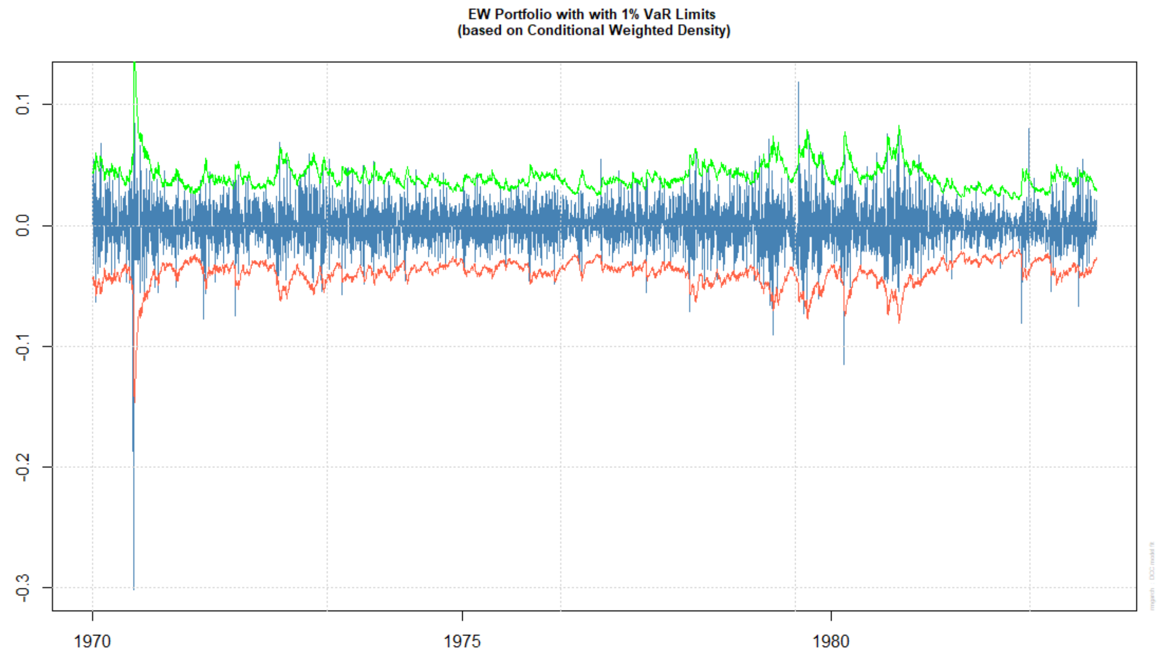
Graphic 7: DCC conditional Sigma



Graphic 8: DCC conditional Covariance



Graphic 9: DCC conditional Correlation



Graphic 10: EW Portfolio with 1% VAR Limits

DCC model is a multivariate Garch models which allows to estimate and forecast the conditional covariances and correlations.

Optimal parameters window:

Several parameters seem to be non-significant.
However, dcca1 and dccb1 are jointly significant and positive.
Moreover, when the correlation is expected to be declining over time, you would expect dcaa1 to be 0 and dccb1 between 0 and 1. In our test, this is not exactly the case.
Gamma1 is positive and not statistically significant. Therefore, we cannot conclude anything from that.

Information criteria:

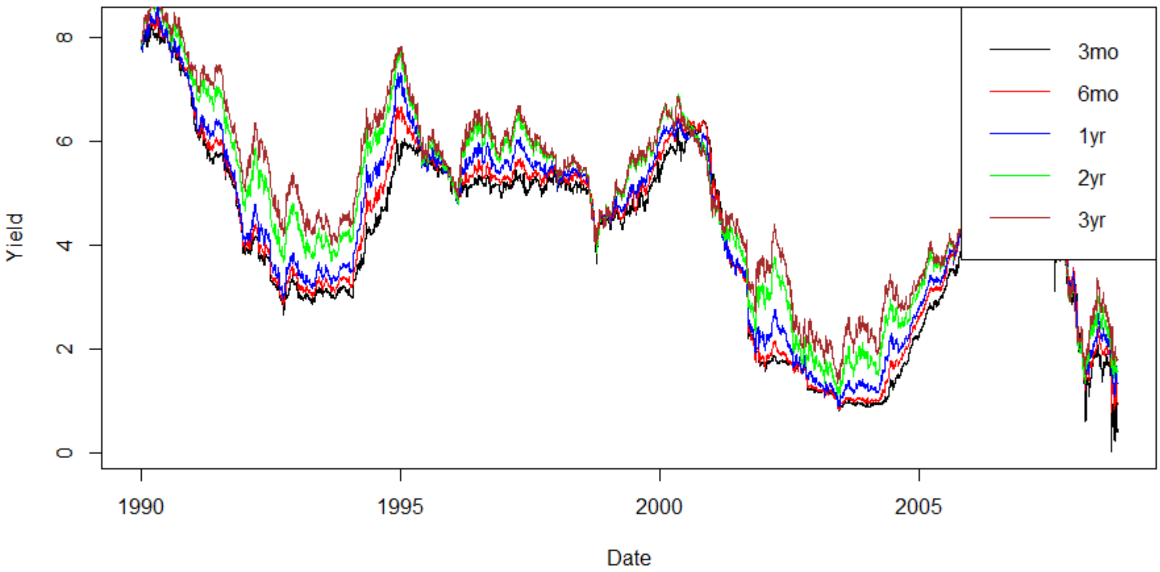
The simplest model is the one which has the lowest information criteria. In our case, this is and Shibata.

Graphs corroborate the idea behind DCC conditional correlation and DCC conditional covariance. We see a declining curve in those graphs as we expected.

Task 3

3.1 Plot of the five yields

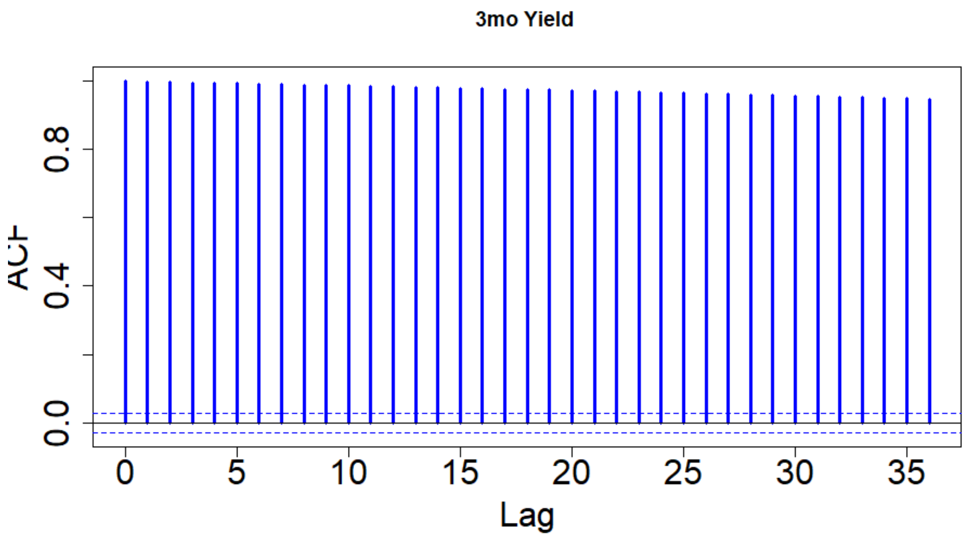
The idea behind this question is to analyze visually if a cointegration is detected. Through visualization we try to identify scenarios where two or more non-stationary time series are integrated in such a way that they cannot deviate from equilibrium in the long term.



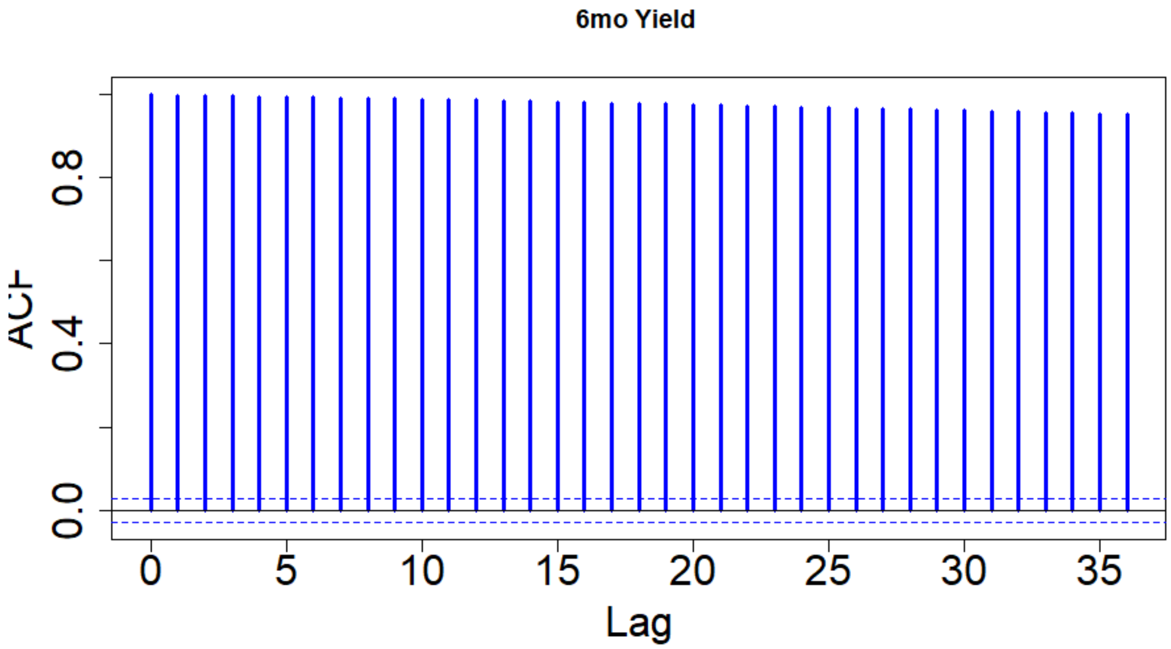
Graphic 11: five yield series

Comment: We can see from the plot that time series have the same trend (when the first goes up the second follows…) and we can also notice that the gap between those series is more or less constant over the time, except some periods. But overall, we can say that the time series are cointegrated.

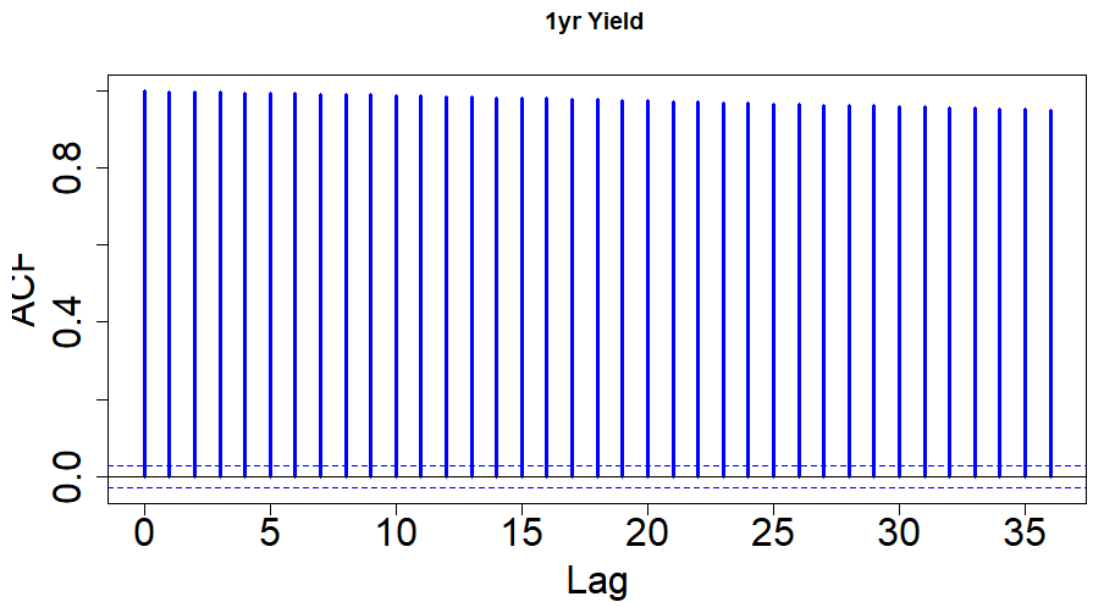
3.2 ACF and ADF test of the five yields



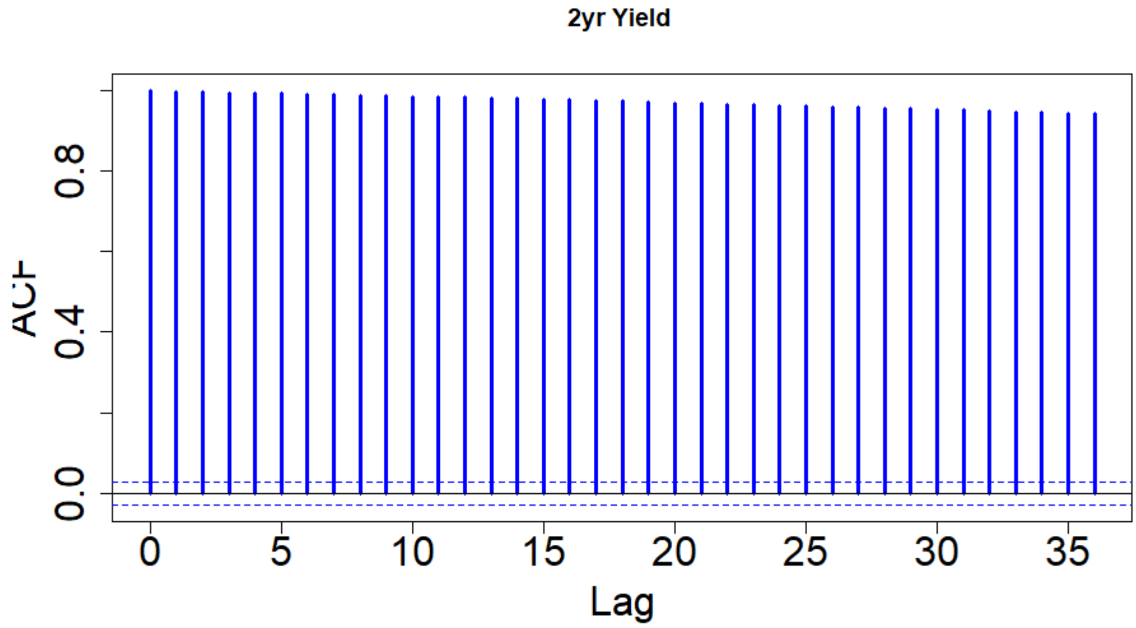
Graphic 8: 3mo ACF



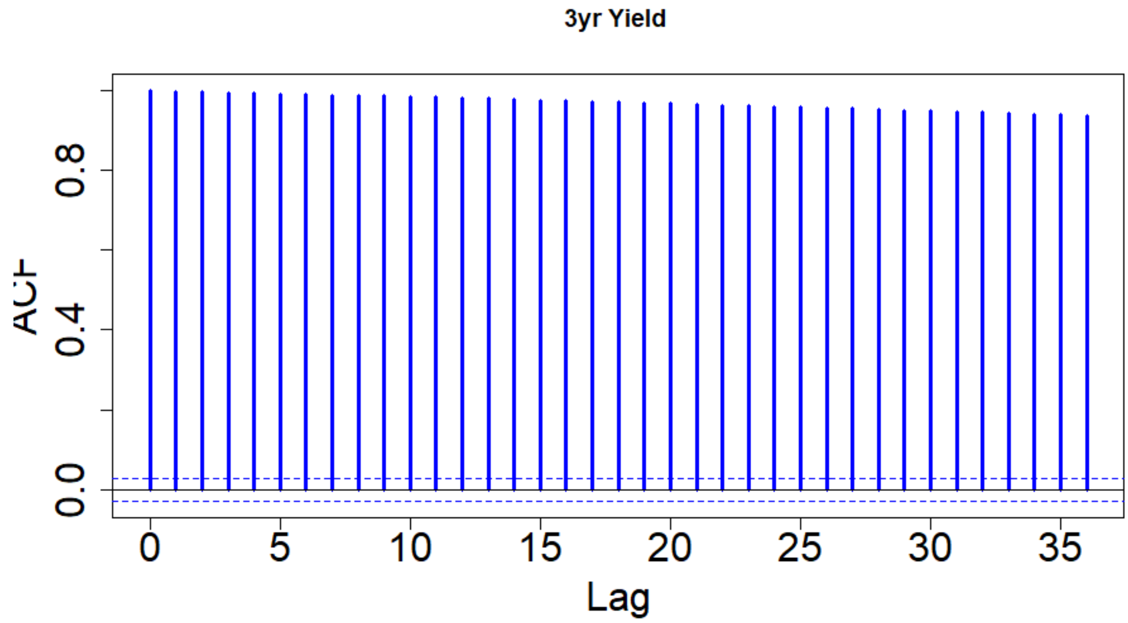
Graphic 9: 6mo ACF



Graphic 9: 1yr ACF



Graphic 10: 2yr ACF



Graphic 9: 3yr ACF

Comment: From all the ACF we highlight the non-stationarity of the time series because of its small decay.

Stationarity test of 3mo Yield

```
> adf.test(`3mo`)

        Augmented Dickey-Fuller Test

data:  3mo
Dickey-Fuller = -1.4174, Lag order = 16, p-value = 0.8251
alternative hypothesis: stationary
```

Screen Shot 17: ADF test on 3mo yield

Comment: After performing the ADF test we see that the p-value=0.825>0.05 (our significance level) which means that we can accept the null hypothesis. So we can consider our 3mo series as Non-Stationary confirming what we said earlier.

Stationarity test of 6mo Yield

```
> adf.test(`6mo`)

        Augmented Dickey-Fuller Test

data:  6mo
Dickey-Fuller = -1.5646, Lag order = 16, p-value = 0.7627
alternative hypothesis: stationary
```

Screen Shot 18:ADF test on 6mo yield

Comment: After performing the ADF test we can see that the p-value=0.7627>0.05 (our significance level) that means that we can accept the null hypothesis so we can consider our 6mo series as Non-Stationary confirming what we said earlier.

Stationarity test of 1yr Yield

```
> adf.test(`1yr`)

        Augmented Dickey-Fuller Test

data:  1yr
Dickey-Fuller = -1.68, Lag order = 16, p-value = 0.7138
alternative hypothesis: stationary
```

Screen Shot 19:ADF test on 1yr yield

Comment: After performing the ADF test we can see that the p-value=0.7138>0.05 (our significance level) that means that we can accept the null hypothesis so we can consider our 1yr series as Non-Stationary confirming what we said earlier.

Stationarity test of 2yr Yield

```
> adf.test(`2yr`)

        Augmented Dickey-Fuller Test

data:  2yr
Dickey-Fuller = -1.9582, Lag order = 16, p-value = 0.596
alternative hypothesis: stationary
```

Screen Shot 20:ADF test on 2yr yield

Comment: After performing the ADF test we can see that the p-value=0.596>0.05 (our significance level) that means that we can accept the null hypothesis so we can consider our 2yr series as Non-Stationary confirming what we said earlier.

Stationarity test of 3yr Yield

```
> adf.test(`3yr`)

        Augmented Dickey-Fuller Test

data:  3yr
Dickey-Fuller = -2.2534, Lag order = 16, p-value = 0.4711
alternative hypothesis: stationary
```

Screen Shot 21: ADF test on 3yr yield

Comment: After performing the ADF test we can see that the p-value=0.4711>0.05 (our significance level) that means that we can accept the null hypothesis so we can consider our 3yr series as Non-Stationary confirming what we said earlier.

3.3 Correlation matrix and multicollinearity

	3mo	6mo	1yr	2yr	3yr
3mo	1.0000000	0.9968180	0.9869009	0.9554240	0.9249541
6mo	0.9968180	1.0000000	0.9949354	0.9673946	0.9381265
1yr	0.9869009	0.9949354	1.0000000	0.9852847	0.9629922
2yr	0.9554240	0.9673946	0.9852847	1.0000000	0.9942084
3yr	0.9249541	0.9381265	0.9629922	0.9942084	1.0000000

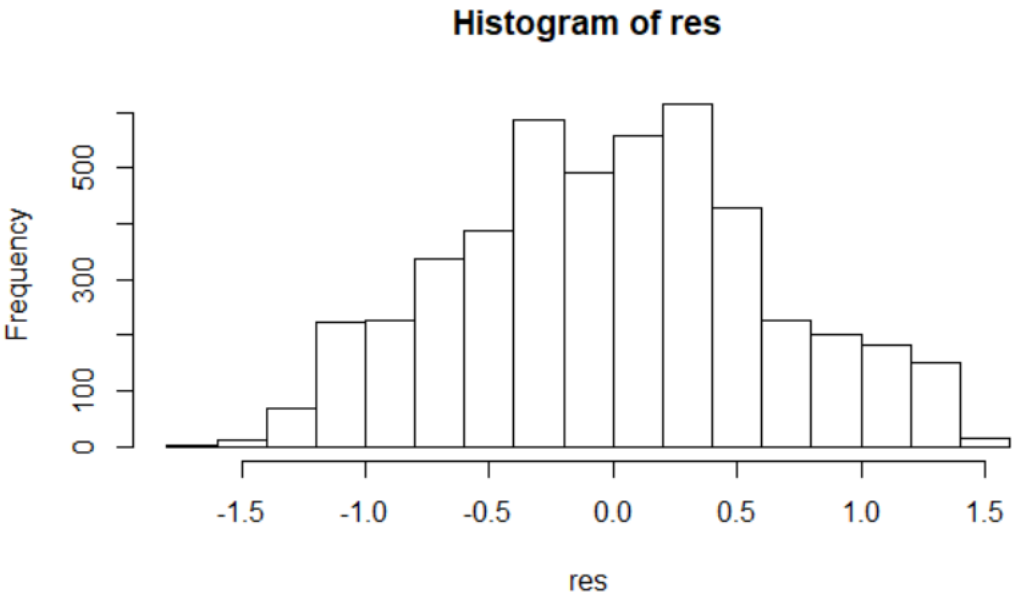
Screen Shot 22: Correlation Matrix

The series are very correlated with a minimum correlation of 0.92 between 3yr and 3mo series

```
> regression <- lm(`3yr`~`3mo`)
> summ(regression)
MODEL INFO:
Observations: 4714
Dependent Variable: 3yr
Type: OLS linear regression

MODEL FIT:
F(1,4712) = 27906.03, p = 0.00
R² = 0.86
Adj. R² = 0.86

Standard errors: OLS
-----------------------------------------------
                     Est.   S.E.   t val.     p
----------------- ------ ------ -------- ------
(Intercept)          1.46   0.02    63.77   0.00
3mo                  0.85   0.01   167.05   0.00
-----------------------------------------------
```

Screen Shot 23: Summary of the regression

The multicollinearity can have some consequences on R^2, it might cause a high R^2 even if the model is not good and the inference might be wrong.
Having all the explanatory variables in the model would mean that strongly correlated variables would be at the right-hand side of the equation model.
But how to measure the impact of one of the variables in the right-hand side of the model if they are mutually correlated ? That's why we cannot have all this explanatory variables at the same time in the model.



Graphic 12: Residuals Histogram

3.4 Test of cointegration

In this question we want to test the presence of a cointegrating relationship between the two-yield series for this we need to use Engle-Granger Approach following the steps :

- We already checked that all the individual variables are I(1). using ADF test we found that they are non-stationarity and the only method to make them stationary is by diff them 1 time "I(1)"
- Then estimate the regression using OLS on the previous question
- Below we saved the OLS residuals
- And we perform ADF test on the residuals to ensure that they are stationary and I(0)

```
> adf.test(Resids)

        Augmented Dickey-Fuller Test

data:  Resids
Dickey-Fuller = -3.7881, Lag order = 16, p-value = 0.01955
alternative hypothesis: stationary
```

Screen Shot 24 ADF test on residuals

Comment: as the p-value of the ADF test is p-value<0.05 we reject the null hypothesis that mean we consider out residuals as stationary. With this we can finally conclude that the 2 series are cointegrated

3.5 ECM

Error correction model is useful for estimating both short-term and long-term effects of one time series on another it also estimates the speed at which a dependent variable returns to equilibrium after a change in other variables. In our case we want to estimate the speed of which 3mo returns to it normal equilibrium with 3yr series after a shock on the first one

```
MODEL INFO:
Observations: 4713
Dependent Variable: DL_3yr
Type: OLS linear regression

MODEL FIT:
F(2,4710) = 495.90, p = 0.00
R² = 0.17
Adj. R² = 0.17

Standard errors: OLS
-------------------------------------------------
                     Est.   S.E.   t val.      p
----------------- -------- ------ -------- ------
(Intercept)         -0.00   0.00    -0.66   0.51
DL_3mo               0.46   0.01    31.49   0.00
ResidsAdj           -0.00   0.00    -1.86   0.06
-------------------------------------------------
```

Screen Shot 25: ECM results

Comment: we notice that R^2 is low but both the speed of adjustment back to the long term equilibrium and and the coefficient of the coefficient of the short term relationship are significant.

3.6 Regression of 3mo on 3yr yield

```
> regression2 <- lm(`3mo`~`3yr`)
> summ(regression2)
MODEL INFO:
Observations: 4714
Dependent Variable: 3mo
Type: OLS linear regression

MODEL FIT:
F(1,4712) = 27906.03, p = 0.00
R² = 0.86
Adj. R² = 0.86

Standard errors: OLS
-------------------------------------------------
                     Est.   S.E.   t val.      p
----------------- -------- ------ -------- ------
(Intercept)         -0.87   0.03   -27.70   0.00
3yr                  1.00   0.01   167.05   0.00
-------------------------------------------------
```

Comment : the OLS results are the same as the regression of 3yr ~ 3mo with R squared of 0.86

```
> Resids2=regression2$residuals
> adf.test(Resids2)

        Augmented Dickey-Fuller Test

data:  Resids2
Dickey-Fuller = -3.0275, Lag order = 16, p-value = 0.1432
alternative hypothesis: stationary
```

Comment: we notice that the residuals are non-stationary and that confirms that our series are not cointegrated.

According to our results it should be impossible to run ECM model since the series aren't cointegrated

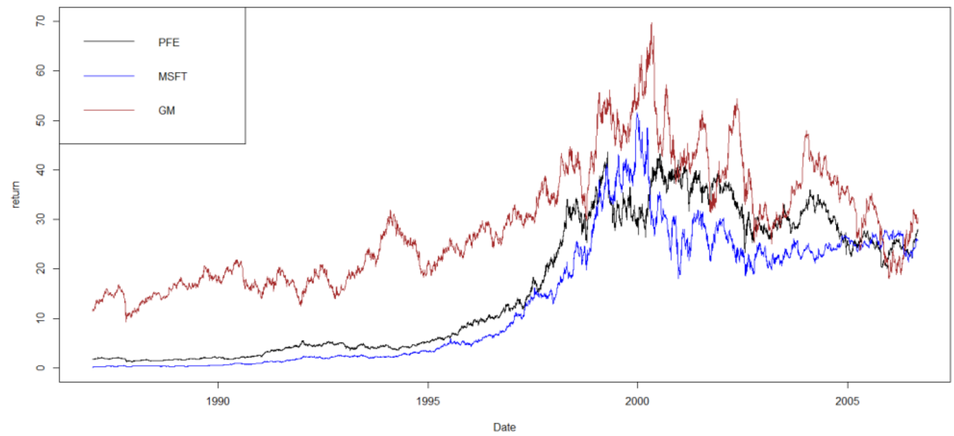
Whereas the series in the first model (3yr~3mo) were cointegrated and allowed us to run an ECM model, it's no more the case in the second model(3mo~3yr)

Task 4

For this task we use again the data set Stocks AC.csv the main objective is to study the VAR model on General Motor and Microsoft and Pfizer

4.1 Price and log return of Microsoft, General Motors and Pfizer

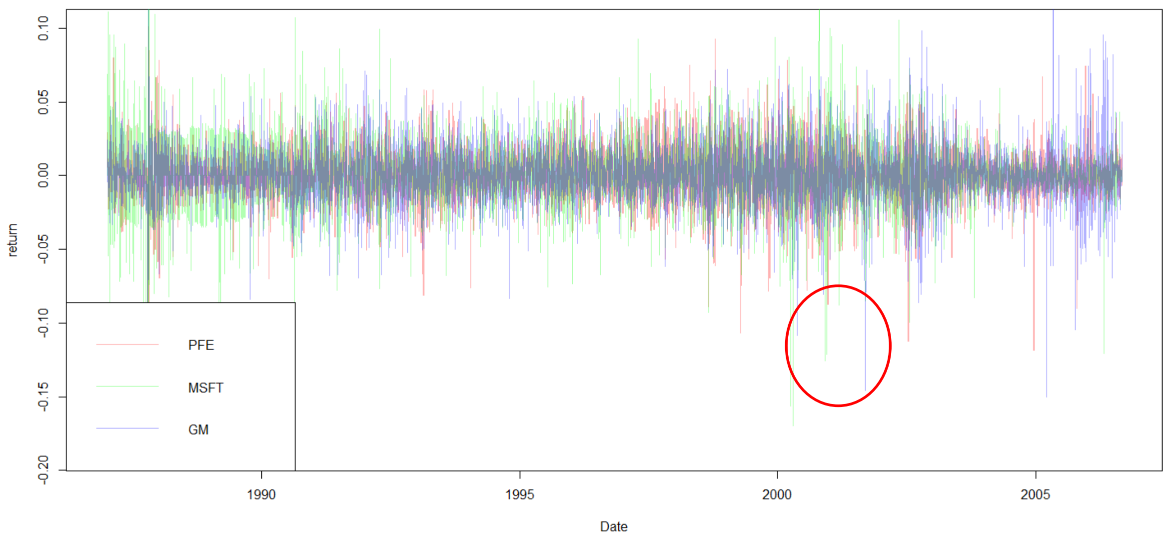
We plot the prices of the 3 series :



Graphic 15: GM, PFE, MSFT Prices plot from 1987 to 2006

Comment : we notice that the three series are not stationary they seems to have a positive trend until the year 2000 and then a negative trend until 2006. We can notice that the 3 series are quite correlated.

We perform the logarithm return on the prices of the 3 series to get a stationary



Graphic 12: PFE,GM, MSFT Return plot from 1987 to 2006

Comment: As expected the return time series are stationary and we can notice that there is some independent shocks(outliers) in the 3 return series

i.e : in the circle we have 2 dependent shocks the first one MSFT but GM series get the shock months after MSFT shock

4.2 VAR model

For this question we should estimate VAR model in the following order (GM, MSFT, PFE) and determinate appropriate length lag the purpose behind this is to reduce residual correlation.

```
$selection
AIC(n)  HQ(n)  SC(n) FPE(n)
     3      2      1      3

$criteria
                   1             2             3             4             5
AIC(n) -2.335321e+01 -2.335939e+01 -2.336054e+01 -2.335905e+01 -2.335908e+01
HQ(n)  -2.334769e+01 -2.334972e+01 -2.334672e+01 -2.334110e+01 -2.333698e+01
SC(n)  -2.333746e+01 -2.333182e+01 -2.332115e+01 -2.330784e+01 -2.329605e+01
FPE(n)  7.208227e-11  7.163835e-11  7.155635e-11  7.166262e-11  7.166068e-11
```

Screen Shot 26: Var selection results

Comment: Using the information criteria we have 4 models each model gave us a result AIC and FPE 3, HQ 2 and SC 1. We choose SC because with less parameters we obtain more precise estimates. We may need more lags in specific cases like when generating hump-shaped impulse response or to simulate cyclical pattern…, and this is not what we're trying to do here.

```
VAR Estimation Results:
=========================
Endogenous variables: LR_GM, LR_MSFT, LR_PFE
Deterministic variables: const
Sample size: 4961
Log Likelihood: 36819.041
Roots of the characteristic polynomial:
0.02992 0.02992 0.02853
Call:
VAR(y = VARData, p = 1)


Estimation results for equation LR_GM:
======================================
LR_GM = LR_GM.l1 + LR_MSFT.l1 + LR_PFE.l1 + const

            Estimate Std. Error t value Pr(>|t|)
LR_GM.l1   -0.0261371  0.0151015  -1.731   0.0836 .
LR_MSFT.l1  0.0250617  0.0129918   1.929   0.0538 .
LR_PFE.l1  -0.0105561  0.0165566  -0.638   0.5238
const       0.0001757  0.0002934   0.599   0.5494
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Residual standard error: 0.02064 on 4957 degrees of freedom
Multiple R-Squared: 0.001123,   Adjusted R-squared: 0.000518
F-statistic: 1.857 on 3 and 4957 DF,  p-value: 0.1346
```

```
Estimation results for equation LR_MSFT:
========================================
LR_MSFT = LR_GM.l1 + LR_MSFT.l1 + LR_PFE.l1 + const

            Estimate Std. Error t value Pr(>|t|)
LR_GM.l1   -0.0242124  0.0178584  -1.356  0.17523
LR_MSFT.l1 -0.0114572  0.0153636  -0.746  0.45586
LR_PFE.l1   0.0007033  0.0195792   0.036  0.97135
const       0.0010541  0.0003470   3.038  0.00239 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Residual standard error: 0.02441 on 4957 degrees of freedom
Multiple R-Squared: 0.0006876,   Adjusted R-squared: 8.282e-05
F-statistic: 1.137 on 3 and 4957 DF,  p-value: 0.3326


Estimation results for equation LR_PFE:
=======================================
LR_PFE = LR_GM.l1 + LR_MSFT.l1 + LR_PFE.l1 + const

            Estimate Std. Error t value Pr(>|t|)
LR_GM.l1   -0.0064167  0.0136909  -0.469   0.6393
LR_MSFT.l1 -0.0007831  0.0117783  -0.066   0.9470
LR_PFE.l1   0.0276786  0.0150101   1.844   0.0652 .
const       0.0005485  0.0002660   2.062   0.0393 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Residual standard error: 0.01872 on 4957 degrees of freedom
Multiple R-Squared: 0.0007166,   Adjusted R-squared: 0.0001118
F-statistic: 1.185 on 3 and 4957 DF,  p-value: 0.3138
```

```
Covariance matrix of residuals:
            LR_GM   LR_MSFT    LR_PFE
LR_GM   4.262e-04 0.0001557 8.772e-05
LR_MSFT 1.557e-04 0.0005960 1.330e-04
LR_PFE  8.772e-05 0.0001330 3.503e-04

Correlation matrix of residuals:
        LR_GM LR_MSFT LR_PFE
LR_GM   1.000   0.309  0.227
LR_MSFT 0.309   1.000  0.291
LR_PFE  0.227   0.291  1.000
```

Screen Shot 27: Var model summary

Comment:

- For General Motors we can notice that none of the coefficient are significant if we consider a significance level of 5% but if we considered a significance level of 10% only coefficients of the GM and MSFT return are significant.
- For Microsoft we can notice that none of the return coefficients is significant for a 5% and 10% significance levels. However the constant is significant in the case of 5% significance level.
- For Pfizer we can notice that none of the return coefficients is significant for a 5%, however the constant is significant for this level but only PFE coefficient of return is significant in the case of 10% confidence level.
- We can also conclude that our 3 models aren't significant because of the low Adjusted R Squared

4.3 Causality test

In this question we want to see test causality the purpose is to determine whether one time series is useful for forecasting another

```
> causality(VAR_Model, cause = "LR_GM")
$Granger

        Granger causality H0: LR_GM do not Granger-cause LR_MSFT LR_PFE

data:  VAR object VAR_Model
F-Test = 0.92209, df1 = 2, df2 = 14871, p-value = 0.3977


$Instant

        H0: No instantaneous causality between: LR_GM and LR_MSFT LR_PFE

data:  VAR object VAR_Model
Chi-squared = 515.71, df = 2, p-value < 2.2e-16
```

Screen Shot 28: GM Causality results

Comment: The P-value=0.3977 > 0.05 we can't reject the null hypothesis so the return of GM don't cause the returns of MSFT and PFE in other terms MSFT and PFE aren't useful for forecasting GM return

```
> causality(VAR_Model, cause = "LR_MSFT")
$Granger

        Granger causality H0: LR_MSFT do not Granger-cause LR_GM LR_PFE

data:  VAR object VAR_Model
F-Test = 1.9947, df1 = 2, df2 = 14871, p-value = 0.1361


$Instant

        H0: No instantaneous causality between: LR_MSFT and LR_GM LR_PFE

data:  VAR object VAR_Model
Chi-squared = 635.51, df = 2, p-value < 2.2e-16
```

Screen Shot 29:MSFT Causality results

Comment: The P-value=0.1361 > 0.05 we can't reject the null hypothesis so the return of MSFT don't cause the returns of GM and PFE in other terms GM and PFE aren't useful for forecasting MSFT return

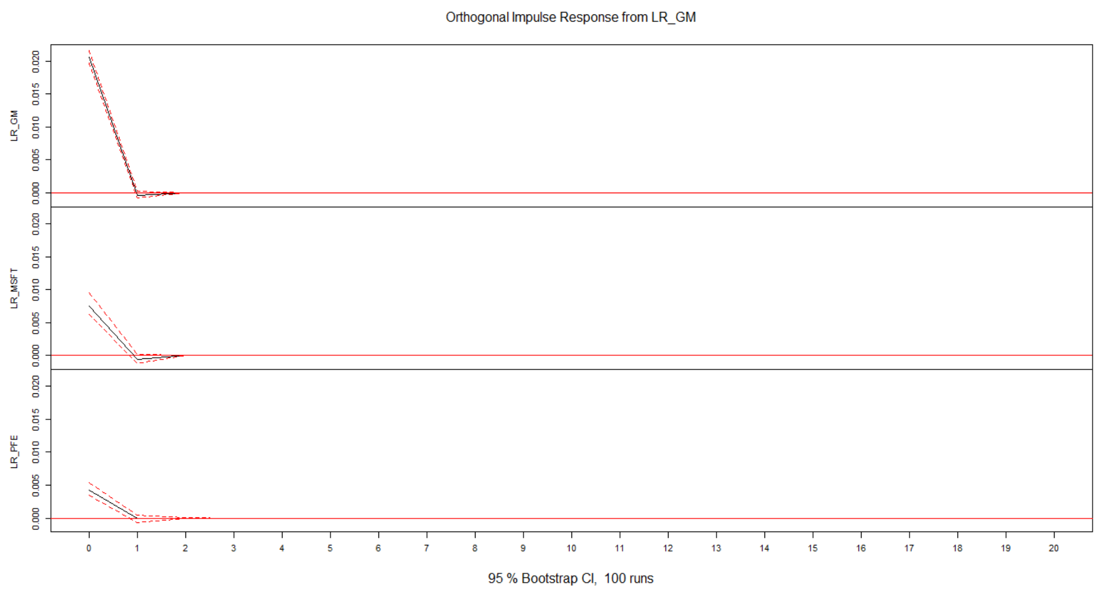
```
> causality(VAR_Model, cause = "LR_PFE")
$Granger

        Granger causality H0: LR_PFE do not Granger-cause LR_GM LR_MSFT

data:  VAR object VAR_Model
F-Test = 0.23324, df1 = 2, df2 = 14871, p-value = 0.792


$Instant

        H0: No instantaneous causality between: LR_PFE and LR_GM LR_MSFT

data:  VAR object VAR_Model
Chi-squared = 473.4, df = 2, p-value < 2.2e-16
```

Screen Shot 30: PFE Causality results

Comment: The P-value=0.792 > 0.05 we can't reject the null hypothesis so the return of PFE don't cause the returns of GM and MSFT in other terms GM and MSFT aren't useful for forecasting PFE return

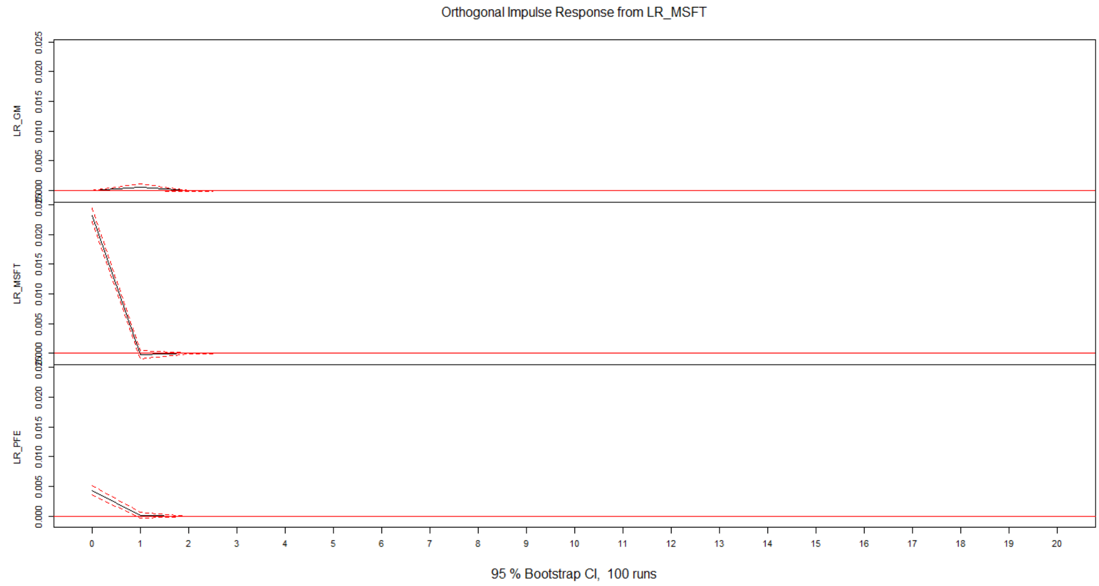
4.4 impulse responses for estimated VAR

The purpose of the impulse response is to describe the evolution of a model's variables in reaction to a shock in one or more variables. This feature allows to trace the transmission of a single shock within an otherwise noisy system of equations.



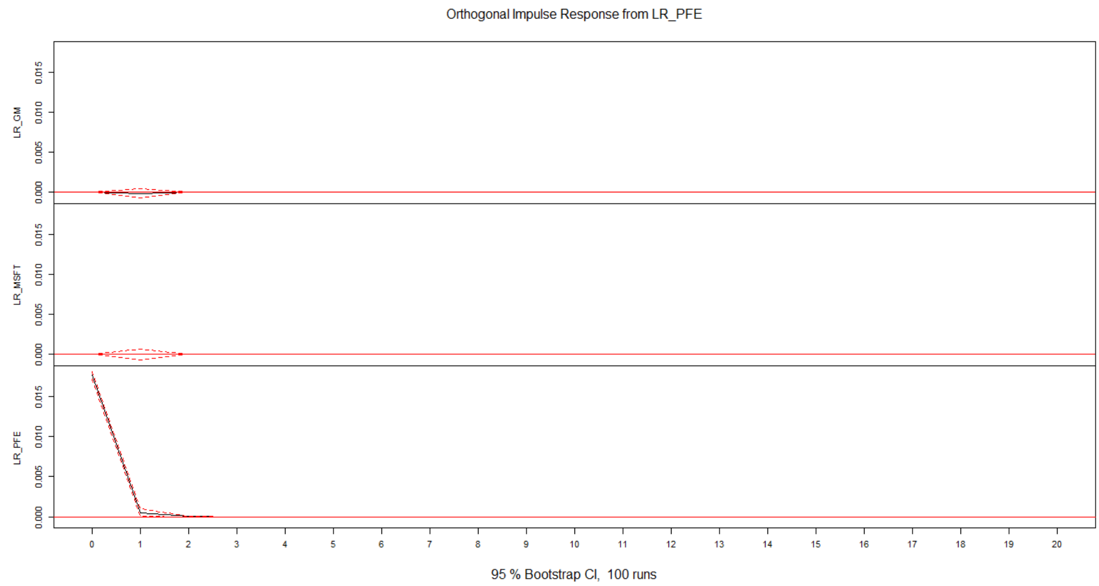
Graphic 13:GM impulse response

Comment: When we have a shock on GM returns It last only 1 period and after that it become stable, we can also notice that a shock on GM create a shock on MSFT which is greater that the one of PFE



Graphic 14: MSFT impulse response

Comment: When we have a shock on MSFT returns It last only 1 period and after that it become stable, we can also notice that a shock on MSFT create a small shock on PFE and don't affect GM

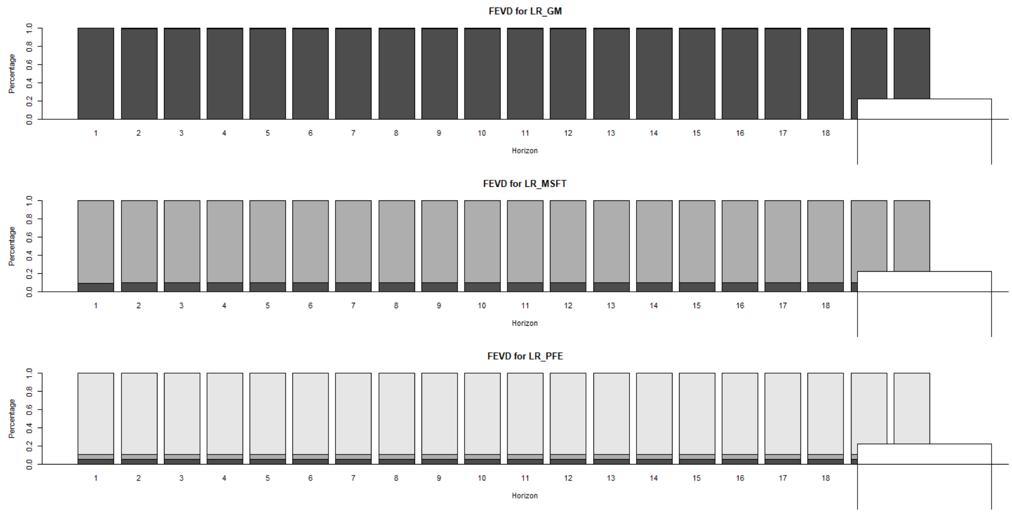


Graphic 15: PFE impulse response

Comment: When we have a shock on PFE returns It also last only 1 period and after that it stabilizes, we can also notice that PFE shock don't create a shock on GM and MSFT

4.5 Forecast error variance decomposition (FEVD)

Forecast error variance decomposition (FEVD) is a part of structural analysis which decomposes the variance of the forecast error into the contributions from specific exogenous shocks. This is useful because it demonstrates how important a shock is in explaining the variations of the variables in the model and also shows how that importance changes over time. For example, some shocks may not be responsible for variations in the short-run but may cause longer-term fluctuations.



Graphic 16: Forecast error variance decomposition

Comment:

- For GM returns, we notice that the movement comes 100% due to it's own shock.
- For MSFT returns, we notice that 10% came from GM shocks and 90% of the movement cames from its own shocks
- For PFE returns, 90% of the movement came from its own shock, 5% from GM and 5% from