ETTALI Achraf
TRAN Francis
Joseph WILLSON

2023

Econometrics and Time Series Applied to Finance

SUMMARY

TASK 1

P3 Task 1.1

P4 Task 1.2

P5 Task 1.3

P5 Task 1.4

P7 Task 1.5

TASK 2

P9 Task 2.1

P13 Task 2.2

P15 Task 2.3

P17 Task 2.4

P19 Task 2.5

TASK 3

P22 Task 3.1

P22 Task 3.2

P24 Task 3.3

P25 Task 3.4

TASK 4

P26 Task 4.1

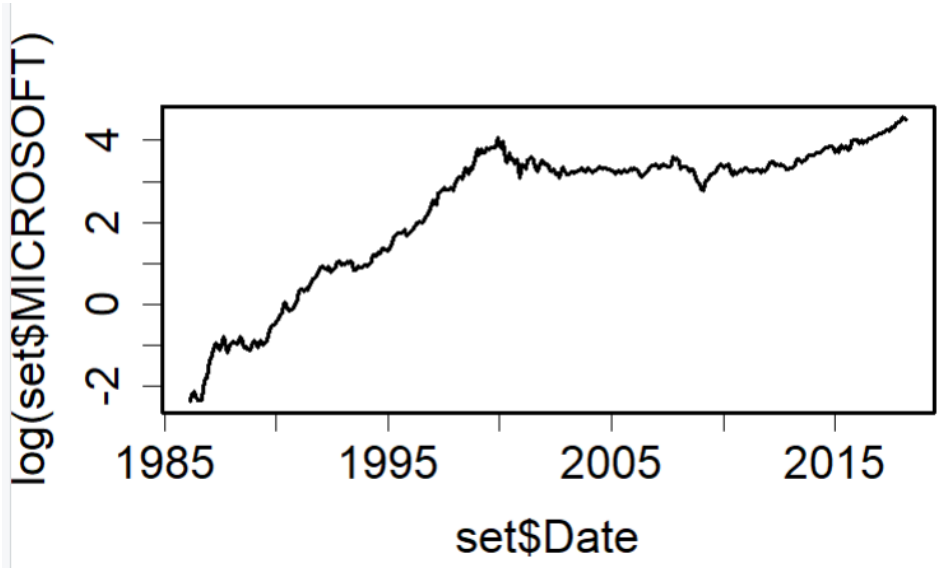
P28 Task 4.2

P29 Task 4.3

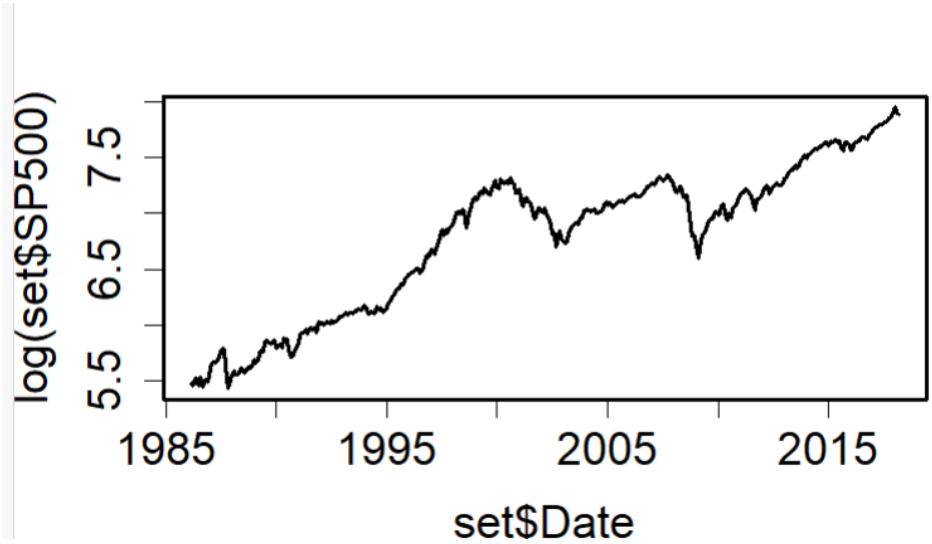
P30 Task 4.4

Task 1

1.1



Plot of log Microsoft time series

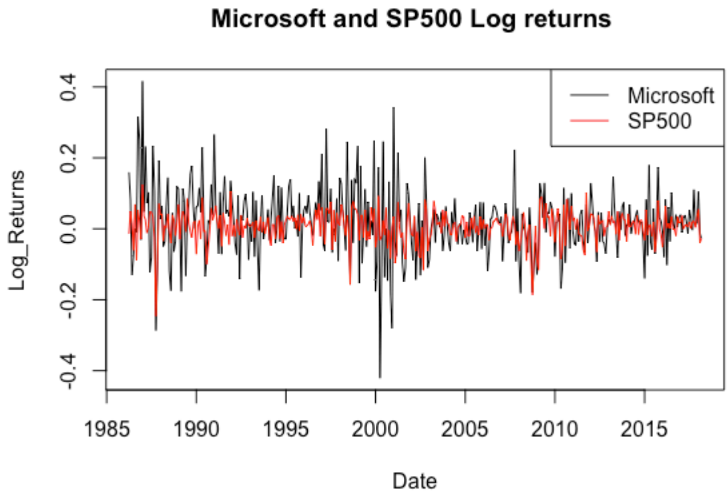


Plot of log SP500 time series

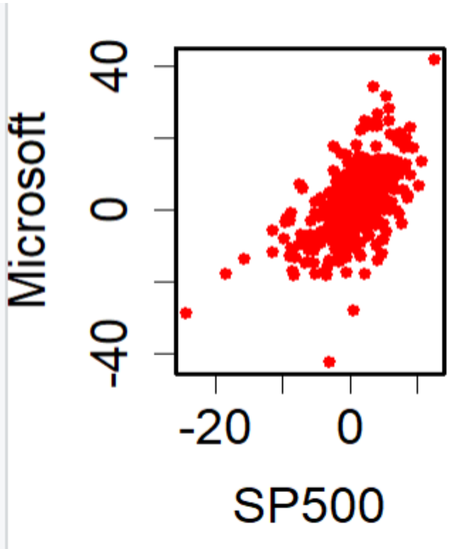
Globally and for most of the year, the two plots follow an increasing trend. For the first plot we can see a quick incresing from 1986 to 2000, but then the series seem to be constant and incresing slowly between 2015 and 2018.

The second plot is more volatile. We can see an increasing trend from 1986 to 2000 but then it's a variation of small increasing and decreasing trend from 2000 to 2018.

1.2



Plot of Microsoft and SP500 log-returns

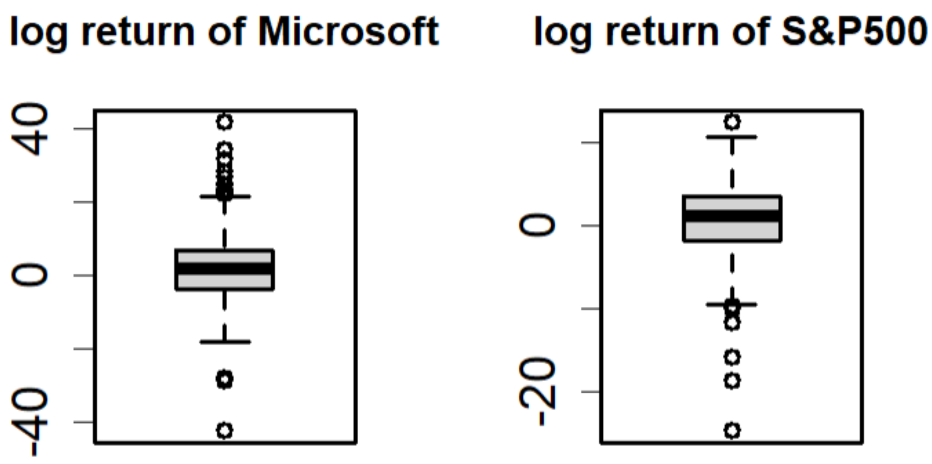


Plot correlation between our 2 log-returns series

First of all, we plotted the 2 log-returns and checked their evolution through time. Microsoft log-return has a higher volatility but it follows the same trend as SP500 log-return so we can suppose that there is a correlation.

Secondly, we have plotted the log return of Microsoft in function of S&P500 log return, and as we can look on the graphic above, the two log-returns describe an increasing curve and seems to be correlated positively. There are just two abnormal errors that we see on the graphic, two points on the bottom of the graphic. They are no correlation between those two points, on possible explanation is: they were a special event for Microsoft.

1.3



Boxplot of the Microsoft and SP500 log-returns

Let's describe those two Boxplots.

Microsoft log return:

We can see that the most of our data are present between the third quantile and the max which is between 5 and 20

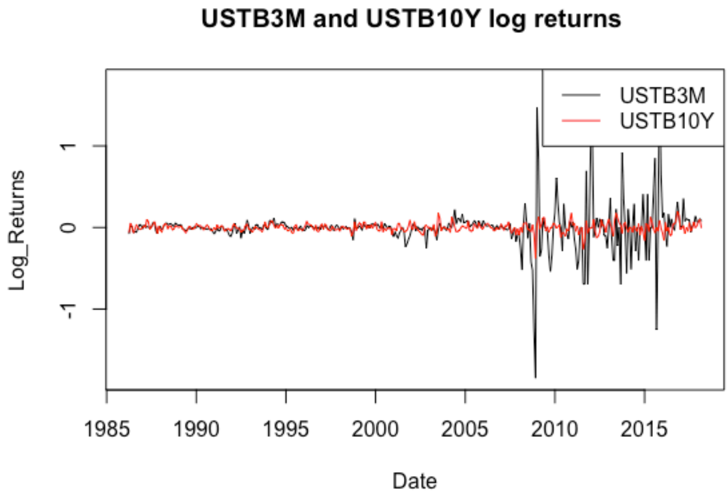
We have also, a lot of data in the first quantile which is –20 and –7, and there are outliers above 20

S&P 500 log return:

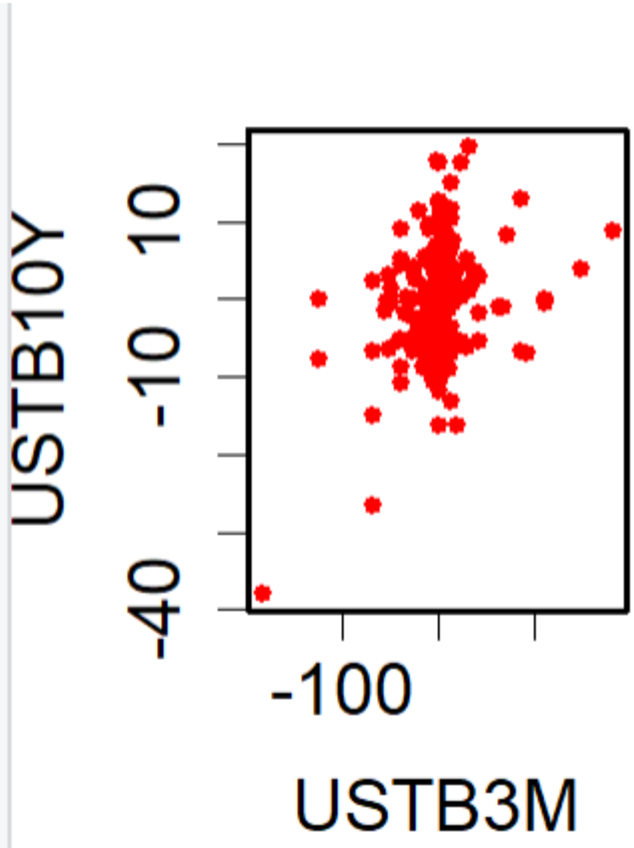
Most of our data are between the third quantile and the max, but they still a lot of outliers in the two boxplots.

1.4

We have repeated the process but with the US tbill at 3-month maturity and US tbill at 10 years maturity.

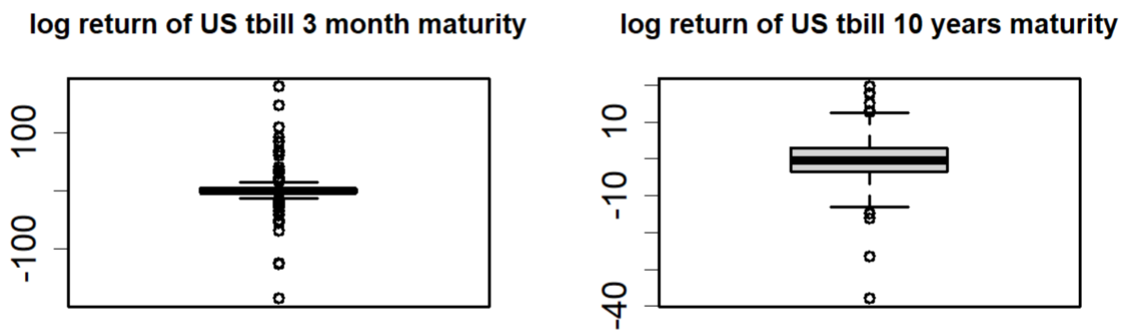


Plot of USTB3M and USTB10Y log-returns



Plot correlation between our 2 log-returns series

On the graph, we can see that the 2 log-returns series follow the same trend from 1986 to 2008 but then the USTB3M log-return series has a higher volatility. We can imagine that these 2 log-returns are correlated which is confirmed on the plot correlation. They seem to be correlated positively but not a lot as we can see in the graphic above.



<u>Boxplot of the USTB3M and USTB10Y log-returns</u>

On the first boxplot (log returns of US tbill 3-month maturity) there is a lot of outliers and in the second boxplot (log returns of US tbill 10 years maturity) our data are especially concentrated between the third quantile and max and there are outliers but not as much as for the USTB3M boxplot.

1.5

SP500

- **mean(log_returns_SP500) = 0.00625734**
- **sd(log_returns_SP500) = 0.0434562**
- **kurtosis(log_returns_SP500) = 6.797415**
- **skewness(log_returns_SP500) = -1.086611**
- **median(log_returns_SP500) = 0.01101732**
- **quantile(log_returns_SP500,0.25)** = **-0.01758331**
- **quantile(log_returns_SP500_1,0.75)** = **0.03412809**
- **max(log_returns_SP500) = 0.12378**
- **min(log_returns_SP500) = -0.245428**

For the log return of S&P 500: the series can't be normally distributed because the kurtosis is not equal to 3 and the skewness is not equal to 0.

Microsoft

- **mean(log_returns_Microsoft) = 0.01787135**
- **sd(log_returns_Microsoft) = 0.09639607**
- **kurtosis(log_returns_Microsoft) = 5.112244**
- **skewness(log_returns_Microsoft) = 0.1244276**
- **median(log_returns_Microsoft) = 0.01976367**
- **quantile(log_returns_Microsoft,0.25) = -0.03839747**
- **quantile(log_returns_Microsoft,0.75) = 0.06552241**
- **max(log_returns_Microsoft) = 0.4157718**
- **min(log_returns_Microsoft) = -0.4208774**

For the log return of microsoft : the series can't be normally distributed because the kurtosis is not equal to 3 and the skewness is not equal to 0.

USTB3M

- **mean(log_returns_USTB3M) = -0.003516569**
- **sd(log_returns_USTB3M) = 0.2613333**
- **kurtosis(log_returns_USTB3M) = 21.65039**
- **skewness(log_returns_USTB3M) = 0.2747321**
- **median(log_returns_USTB3M) = 0**
- **quantile(log_returns_USTB3M,0.25) = -0.03285758**
- **quantile(log_returns_USTB3M,0.75) = 0.03653661**
- **max(log_returns_USTB3M) = 1.791759**
- **min(log_returns_USTB3M) = -1.845827**

For the log return of us tbill 3-month maturity: the series can't be normally distributed because the kurtosis is not equal to 3 and the skewness is not equal to 0.

USTB10Y

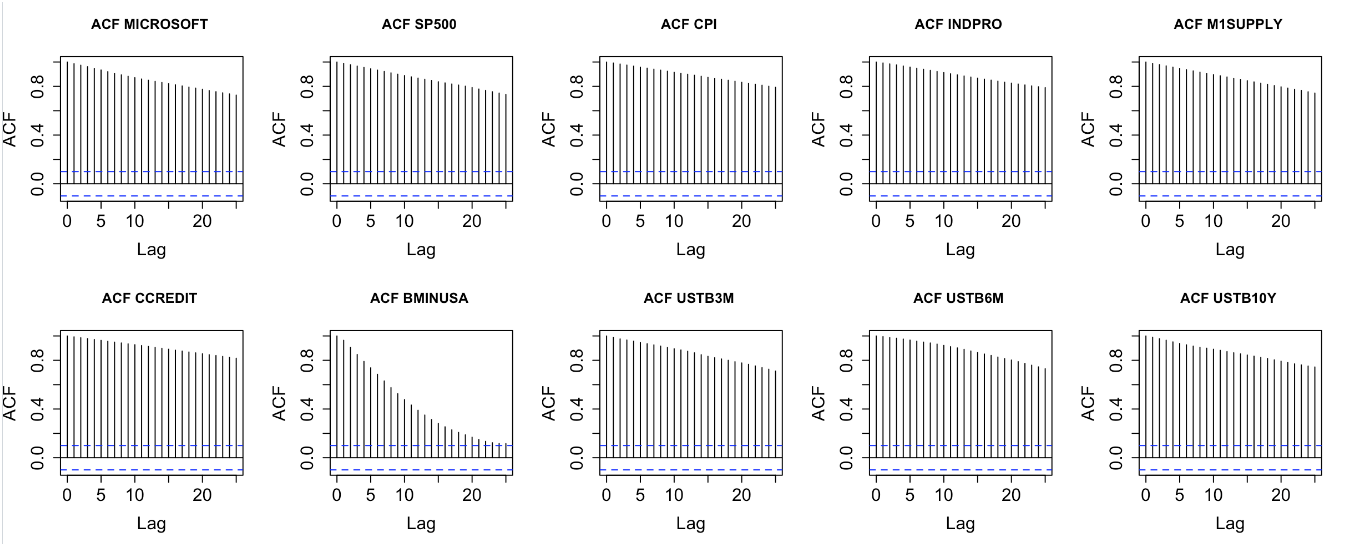
- **mean(log_returns_USTB10Y) = -0.002624355**
- **sd(log_returns_USTB10Y) = 0.0575709**
- **kurtosis(log_returns_USTB10Y) = 8.847052**
- **skewness(log_returns_USTB10Y) = -0.5909696**
- **median(log_returns_USTB10Y) = -0.005006904**
- **quantile(log_returns_USTB10Y,0.25) = -0.03473474**
- **quantile(log_returns_USTB10Y,0.75) = 0.02972636**
- **max(log_returns_USTB10Y) = 0.195492**
- **min(log_returns_USTB10Y) = -0.3775303**

For the log return of us tbill 3-month maturity: the series can't be normally distributed because the kurtosis is not equal to 3 and the skewness is not equal to 0.

Task 2

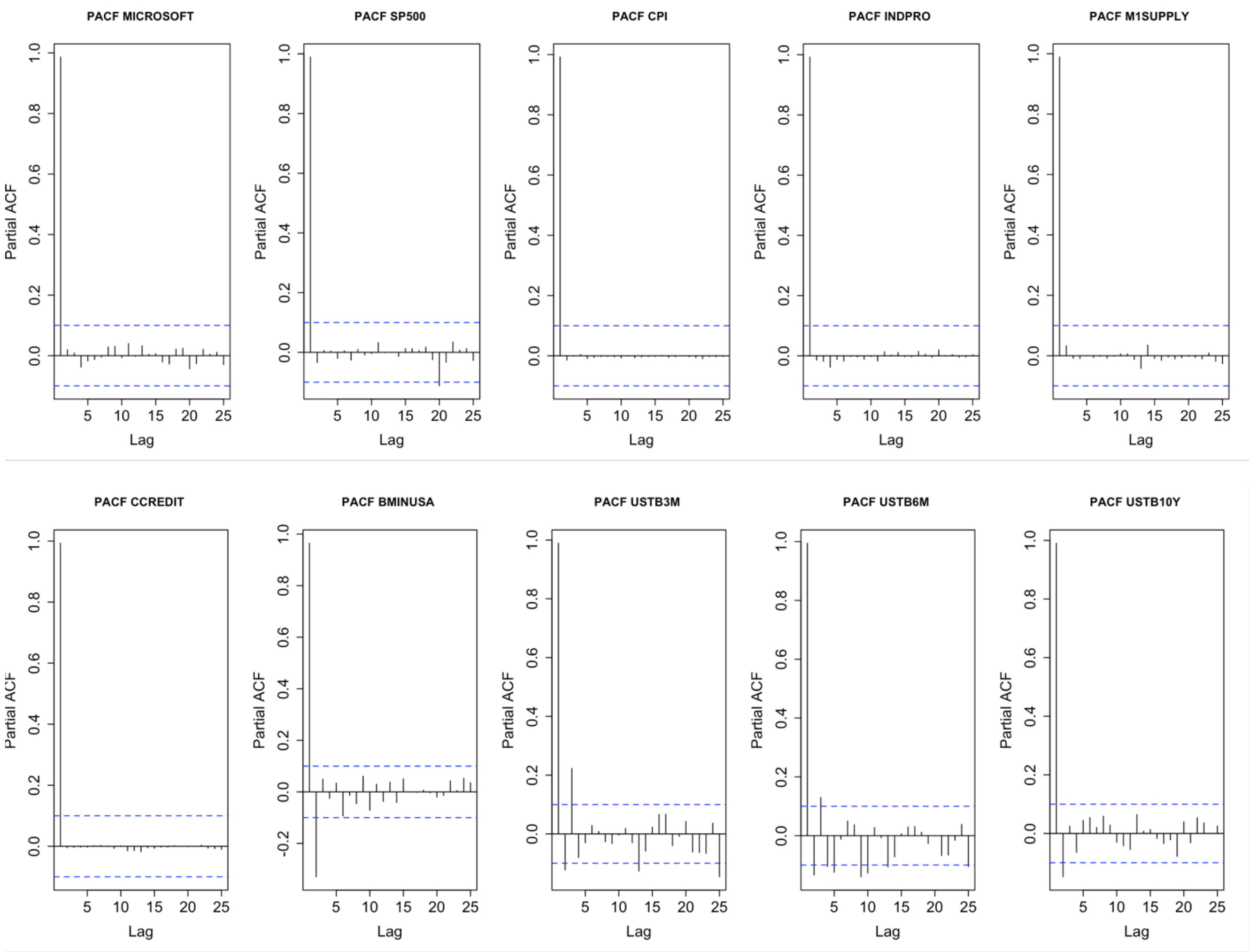
2.1 (0.5 Points)

Log Series



<u>ACF of Log series</u>

For the ACF on our log series, we know that ACF represents the correlation between the time series and its lagged values. We can see that trends are similar for all of our time series. The others 10 series have a decreasing trend as the lags progress. We can see that for the lag zero, ACF is equal to one which is logical because our first value of our time series is always perfectly correlated with itself. The small decreasing trend means that the values are always strongly correlated but the more we progress on the lag, the little bit less correlated it becomes. For the BMINUSA log series, the correlation between a current value and its previous one fades away faster than for the other times series. Above the 20th lag of this time serie, we can see that our value are on our error bands which means that they are not statistically significant anymore.

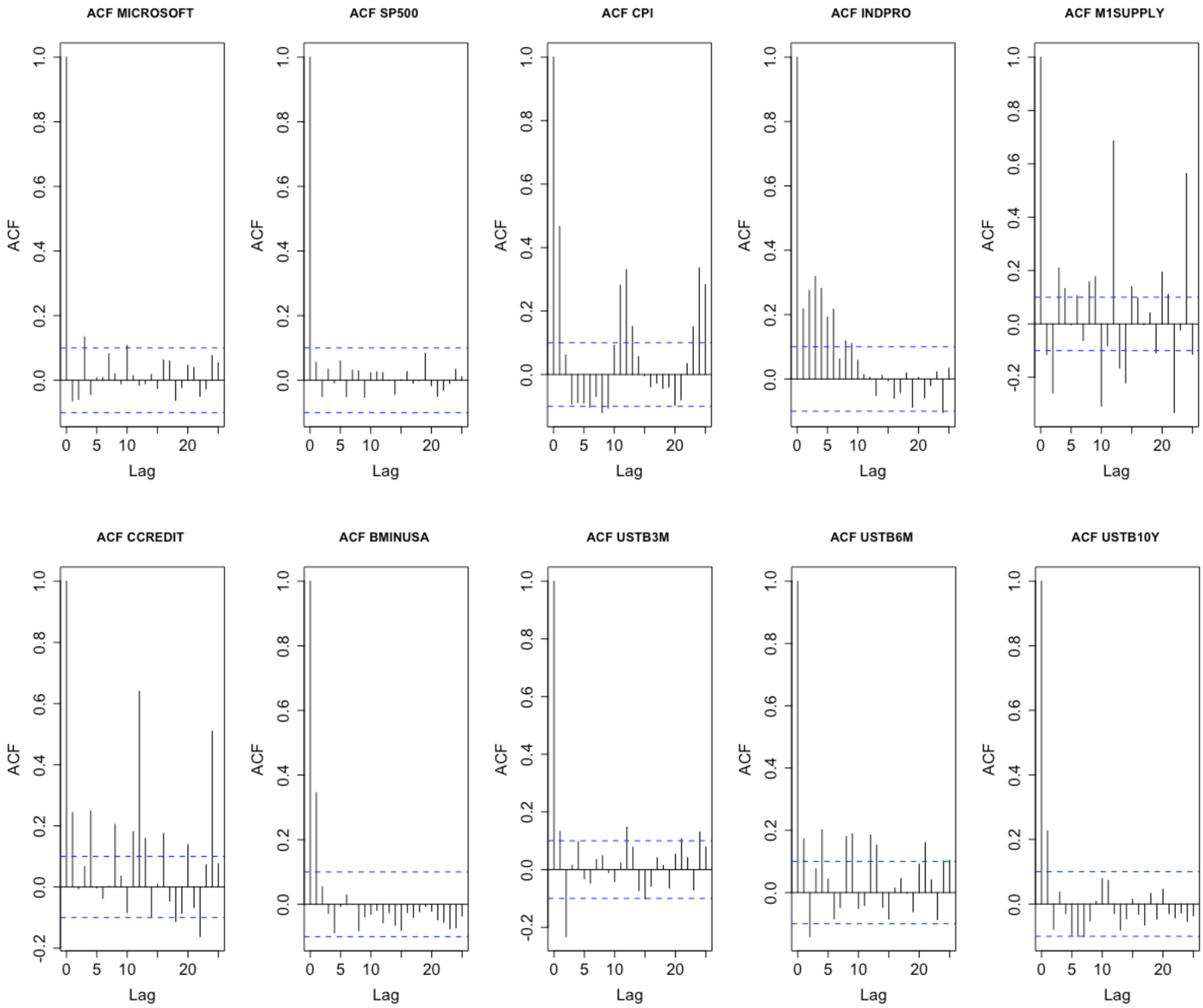


<u>PACF of Log series</u>

Concerning the PACF, for our 10 times series, there is a significant spike on the first lag which mean that the higher order autocorrelations are explained by the first lag autocorrelation. There is always a significant correlation at lag 1 followed by non-significant correlations which indicates an autoregressive term of order 1 in our 10-time series.

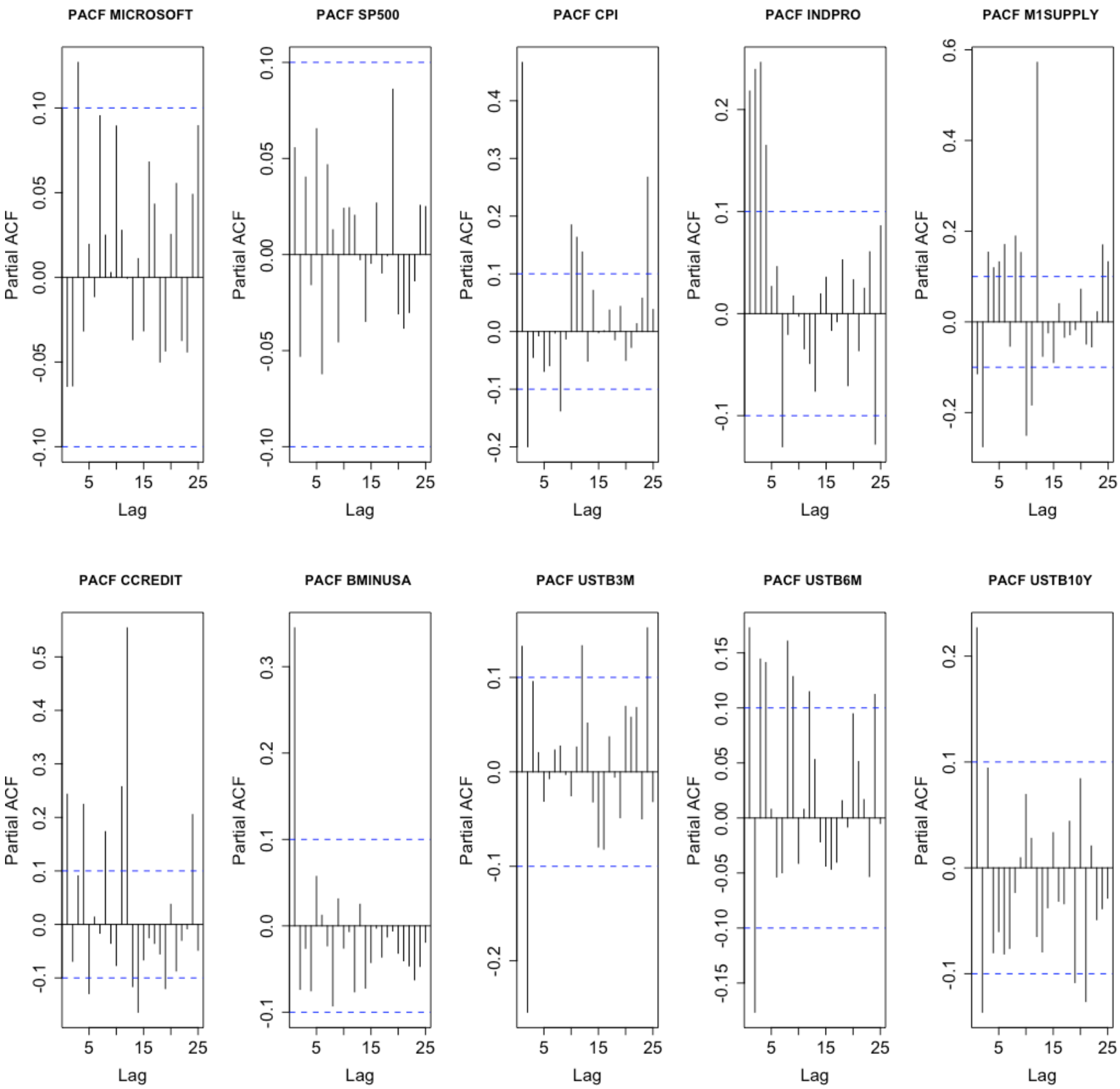
As we have a declining ACF and PACF which is significant only for a few lags, we can deduce that it’s an AR process more precisely an AR (1) process as the lags of the process are determined by the number of significant lags of the PACF.

Logs returns



<u>ACF of log-returns series</u>

For the INDPRO time series, we can see that the values of the first lag to the 9^{h} lags are correlated but from the 10^{th} lags the values are not statistically significant anymore, for the CPI time series there are some period cycles on which value are autocorrelated (lags from 10 to 15 and 22 to 25 for example). For the BMINUSA and USTB10Y we can see 2 significant spikes on the 2 first lags but then values become statistically significant. For the M1SUPPLY we can see a significant spike at lag 0, 12, and 24. For all others time series it’s difficult to identify a common decreasing trend such as for the log computed previously, we can deduce a huge correlation between a value and its previous one for a certain lag.



<u>PACF of log-returns series</u>

For the PACF of our log-returns, we can see that for the MICROSOFT and SP500 our values for all lags are not significant. For the INDPRO times series, we can see a large spike at lag 1 that decreases after a few lags which indicates a moving average term in the serie. For all other times series, there is a large spike at lag 1 and then some alternative values between positive and negative correlations. It indicates a higher order moving average term in the times series.

2.2

Log series

We are going to sum up all the p-values of our series to determine if it's higher or not than 0.05 and to see if we reject or not the null hypothesis.

Test	***Null hypothesis***	***Microsoft***	***SP500***	***CPI***	***INDPRO***	***M1Supply***	***CCREDIT***	***BMINUSA***	***USTB3M***	***USTB6M***	***USTB10Y***
Stationarity (ADF)	***H0: Unit Root non-stationarity***	***p-value = 0.3458***	***p-value = 0.567***	***p-value = 0.5466***	p-value = 0.6388	***p-value =*** 0.9897	***p-value =*** 0.9595	***p-value =*** 0.02304	***p-value =*** 0.7132	***p-value =*** 0.8951	***p-value =*** 0.03796
Normality Test Shapiro_Wilk's	***H0: No serial correlations***	***p-value < 2.2e-16***	***p-value = 9.438e-13***	***p-value = 7.145e-11***	***p-value < 2.2e-16***	***p-value = 2.582e-13***	***p-value = 1.117e-13***	***p-value = 1.158e-11***	***p-value < 2.2e-16***	***p-value < 2.2e-16***	***p-value = 8.979e-11***
Serial-correlation Ljung-Box	***H0: Normal distribution***	***p-value < 2.2e-16***	***p-value < 2.2e-16***	***p-value < 2.2e-16***	***p-value < 2.2e-16***	***p-value < 2.2e-16***	***p-value < 2.2e-16***	***p-value < 2.2e-16***	***p-value < 2.2e-16***	***p-value < 2.2e-16***	***p-value < 2.2e-16***
Homoskedasticity : Arch Test	HO : Homoscedastic series	***p-value < 2.2e-16***	***p-value < 2.2e-16***	***p-value < 2.2e-16***	***p-value < 2.2e-16***	***p-value < 2.2e-16***	***p-value < 2.2e-16***	***p-value < 2.2e-16***	***p-value < 2.2e-16***	***p-value < 2.2e-16***	***p-value < 2.2e-16***

Recap table for the log series

For the ADF test, we reject the null hypothesis for the BMINUSA and USTB10Y times series which means that the data are stationary, for all of the other times series, we don't reject the null hypothesis which means that the data are not stationary.

For the Shapiro test, we reject the null hypothesis for all of our times series which means that the data don't follow a normal distribution law.

For the Ljung box test, we reject the null hypothesis for all of our times series which means that the residuals of our times series are not independently distributed so there is a serial correlation.

For the ArchTest, same we reject the null hypothesis for all of our times series which means that the residuals of our time series exhibit heteroscedasticity.

Log returns

Test	***Null hypothesis***	***Microsoft***	***SP500***	***CPI***	***INDPRO***	***M1Supply***	***CCREDIT***	***BMINUSA***	***USTB3M***	***USTB6M***	***USTB10Y***
Stationarity (ADF)	***H0: Unit Root non-stationarity***	***p-value = 0.01***	***p-value = 0.01***	***p-value = 0.01***	p-value = 0.01	***p-value = 0.01***	***p-value = 0.01***	***p-value = 0.01***	***p-value = 0.01***	***p-value = 0.01***	***p-value = 0.01***
Normality Test Shapiro_Wilk's	***H0: No serial correlations***	***p-value = 2.474e-06***	***p-value = 1.674e-10***	***p-value = 6.443e-11***	***p-value = 1.514e-14***	***p-value = 0.005904***	***p-value = 8.206e-10***	***p-value = 1.535e-10***	***p-value < 2.2e-16***	***p-value < 2.2e-16***	***p-value = 2.301e-11***
Serial-correlation Ljung-Box	***H0: Normal distribution***	***p-value = 0.2054***	***p-value = 0.2743***	***p-value < 2.2e-16***	***p-value = 1.775e-05***	***p-value = 0.02445***	***p-value = 1.688e-06***	***p-value = 1.157e-11***	***p-value = 0.008944***	***p-value = 0.006788***	***p-value = 8.201e-06***
Homoskedasticity : Arch Test	HO : Homoscedastic series	***p-value =*** 2.911e-10	***p-value =*** 0.09188	***p-value =*** 2.058e-09	***p-value =*** 1.356e-08	***p-value =*** 3.71e-06	***p-value =*** 6.579e-05	***p-value =*** 0.003481	***p-value =*** 3.472e-16	***p-value =*** 7.457e-06	***p-value =*** 0.446

Recap table of the log-returns series

For the ADF test, we reject the null hypothesis for all of our times series which means that the data are stationary.

For the Shapiro test, we reject our null hypothesis for all of our times series which means that the data.

For the Ljung-box test, we reject our null hypothesis for the CPI, INDPRO, M1Supply, CCREDIT, BMINUSA, USTB3M, USTB6M, USTB10Y times series which means that the residuals of our times series are not independently distributed. For the Microsoft and SP500 times series we don't reject it which means that the residuals of our time series are independently distributed.

For the Arch Test, we reject the null hypothesis for all of our times series which means that the residuals of our time series exhibit heteroscedasticity.   expect for the SP500 and USTB10Y times series where we don't reject it which means that the residuals don't exhibits heteroscedasticity.

2.3

```
  log_returns_SP500       log_returns_CPI    log_returns_INDPRO log_returns_M1SUPPLY  log_returns_CCREDIT
           1.038419              1.191539              1.069194             1.202582             1.117692
log_returns_BMINUSA    log_returns_USTB3M    log_returns_USTB6M  log_returns_USTB10Y
           1.140201              3.743047              3.781975             1.220608
```

<u>Summary of our linear model</u>

We check if there are some coefficients above 10 but it's not the case here so there is no a multicollinearity in our model. It seems logical because the Microsoft times series cannot be perfectly predicted other times series which are not linked to Microsoft (except for the SP500).

- Check CLRM assumptions :

1) E(Res)=0 The mean of our ErrorTerms is equal to 5.140714e-18 -16 and the standard deviation to 0.07861163.

The first assumption is not violated as the mean is equal to 0.

2) Var(Res)=sigma^2 (constant & finite)

For the BPtest with have p-value = 0.4535 and for the ARCH test p-value = 0.05823.

We don't reject the null hypothesis of our ARCH test which means that there is a homoscedasticity trend. The second assumption is not violated.

3) Residuals no autocorrelated

For the bgtest p-value = 0.781 and for the box test, p-value = 0.7775.

For the Breusch-Godfrey test, we don't reject the null hypothesis which means that a first order autocorrelation exists in our model and there is patter in the errors. The third assumption is violated which means that OLS estimates are still unbiased and inefficient (not BLUE). Our standard errors might be inappropriate too. Also, R^2 might be inflated relative to its correct data for positively autocorrelated residuals. The possible solutions is to add to our linear regression model some previous values of the dependent variable or previous values of the independent variables to have a dynamic models.

4) X(t) and e(t) non-correlated

```
> Check <- lm(log_returns_SP500~ErrorTerms)
> summary(Check)

Call:
lm(formula = log_returns_SP500 ~ ErrorTerms)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.25168 -0.02384  0.00476  0.02787  0.11752 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)   
(Intercept)  6.257e-03  2.221e-03   2.818  0.00508 **
ErrorTerms  -6.765e-17  2.828e-02   0.000  1.00000   
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.04351 on 382 degrees of freedom
Multiple R-squared:  3.121e-32,	Adjusted R-squared:  -0.002618 
F-statistic: 1.192e-29 on 1 and 382 DF,  p-value: 1

> cor(ErrorTerms,log_returns_SP500)
[1] 8.303731e-20
```

Summary of our linear model between the SP500 log-return and error Terms

We can see that there is no correlation between our ErrorTerms and Log return of SP500. So, the assumption is not violated.

5) Residuals are normally distributed

For the Jarque test p-value <2.2e-16 so we reject the null hypothesis that our errors are normally distributed, so the assumption is not violated.

- Structural breaks:

For the sctest, p-value = 0.06634.

Our p-value is above 0.05 which mean we don’t reject the null hypothesis of our sctest that there is no structural break in the coefficients of our APT model.

2.4 Fit an ARMA-EGARCH model for the Microsoft monthly returns

```
Optimal Parameters
------------------------------------
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.016413    0.003896   4.2124 0.000025
omega  -0.216979    0.091852  -2.3623 0.018163
alpha1  0.046731    0.032459   1.4397 0.149954
beta1   0.954898    0.019015  50.2192 0.000000
gamma1  0.205541    0.055714   3.6892 0.000225

Robust Standard Errors:
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.016413    0.003934   4.1716 0.000030
omega  -0.216979    0.087968  -2.4666 0.013641
alpha1  0.046731    0.031072   1.5040 0.132589
beta1   0.954898    0.018474  51.6887 0.000000
gamma1  0.205541    0.052868   3.8878 0.000101

LogLikelihood : 389.2892
```

Results of our ARMA-EGARCH model

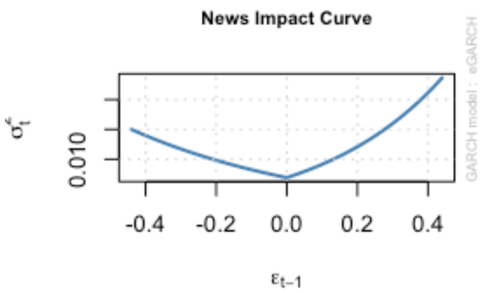
We can see that all of our coefficient except alpha1 are significant. Alpha1 seems not to be useful in our model setting.

```
Information Criteria
------------------------------------

Akaike       -2.0015
Bayes        -1.9501
Shibata      -2.0018
Hannan-Quinn -1.9811

Weighted Ljung-Box Test on Standardized Residuals
------------------------------------
                        statistic p-value
Lag[1]                     0.8407  0.3592
Lag[2*(p+q)+(p+q)-1][2]    0.9488  0.5151
Lag[4*(p+q)+(p+q)-1][5]    3.7258  0.2903
d.o.f=0
H0 : No serial correlation

Weighted Ljung-Box Test on Standardized Squared Residuals
------------------------------------
                        statistic p-value
Lag[1]                    0.01117  0.9158
Lag[2*(p+q)+(p+q)-1][5]   5.24255  0.1348
Lag[4*(p+q)+(p+q)-1][9]   7.76324  0.1432
d.o.f=2
```

Then on information criteria table, it displays some information about the model estimation. We know that the lower the Akakeb Bayes, Shibata, Hannan-Quinn values are which is the case here the better our model is in terms of fitting. Then there is the Ljung-Box test. The p-value is higher than 5% se we don't reject the null hypothesis. There is no serial correlation of the error term in our case.

```
Adjusted Pearson Goodness-of-Fit Test:
------------------------------------
  group statistic p-value(g-1)
1    20     12.35       0.8699
2    30     27.41       0.5498
3    40     31.00       0.8158
4    50     45.17       0.6293
```

Finally with the results of the Adjusted Pearson test, we can see that the p-value is greater than 0.05 so we don't reject the null hypothesis that the error term follows the normal distribution.



<u>News-Impact curve of our model</u>

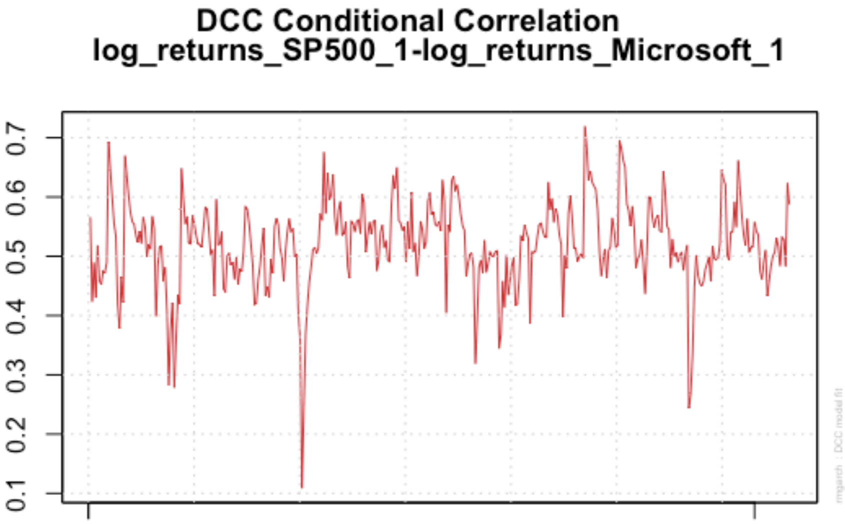
As we can see on the graph above, there is an asymmetric effect and it implies that the impact of positive news on our Microsoft EGARCH model is higher than the impact for a negative news.

2.5 Fit a GARCH-DCC model with the Microsoft and SP500 returns.

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

Distribution         :  mvnorm
Model                :  DCC(1,1)
No. Parameters       :  11
[VAR GARCH DCC UncQ] : [0+8+2+1]
No. Series           :  2
No. Obs.             :  384
Log-Likelihood       :  1138.933
Av.Log-Likelihood    :  2.97

Optimal Parameters
-----------------------------------
                                  Estimate  Std. Error  t value Pr(>|t|)
[log_returns_Microsoft_1].mu      0.016032    0.004184   3.8318 0.000127
[log_returns_Microsoft_1].omega   0.000303    0.000158   1.9180 0.055107
[log_returns_Microsoft_1].alpha1  0.111344    0.036725   3.0318 0.002431
[log_returns_Microsoft_1].beta1   0.851907    0.036024  23.6486 0.000000
[log_returns_SP500_1].mu          0.007083    0.001994   3.5513 0.000383
[log_returns_SP500_1].omega       0.000044    0.000032   1.3634 0.172751
[log_returns_SP500_1].alpha1      0.156839    0.047859   3.2771 0.001049
[log_returns_SP500_1].beta1       0.836376    0.043537  19.2109 0.000000
[Joint]dcca1                      0.093826    0.074651   1.2569 0.208803
[Joint]dccb1                      0.570799    0.435872   1.3096 0.190346

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

Akaike       -5.8747
Bayes        -5.7615
Shibata      -5.8762
Hannan-Quinn -5.8298


Elapsed time : 0.6546979
```

Results of GARCH-DCC model

We can see that almost of our values are significant except logReturns Microsoft omega, log return SP500 omega, dcca1 and dccb1. Alpha represent our ARCH model and Beta represents our GARCH model. When we create a DCC model we have to check that these values are significant which is the case here. DCCAlpha will tells us if there is a short-term persistence or no and DCC Beta will show us the long-term persistence, but these values are not significant, so we don't have any persistence between Log Microsoft and log SP500. Also, the sum of these both coefficients should be less than 1 which is the case here so there is a dynamic relationship and our model is dynamic conditional correlation between the 2 times series.

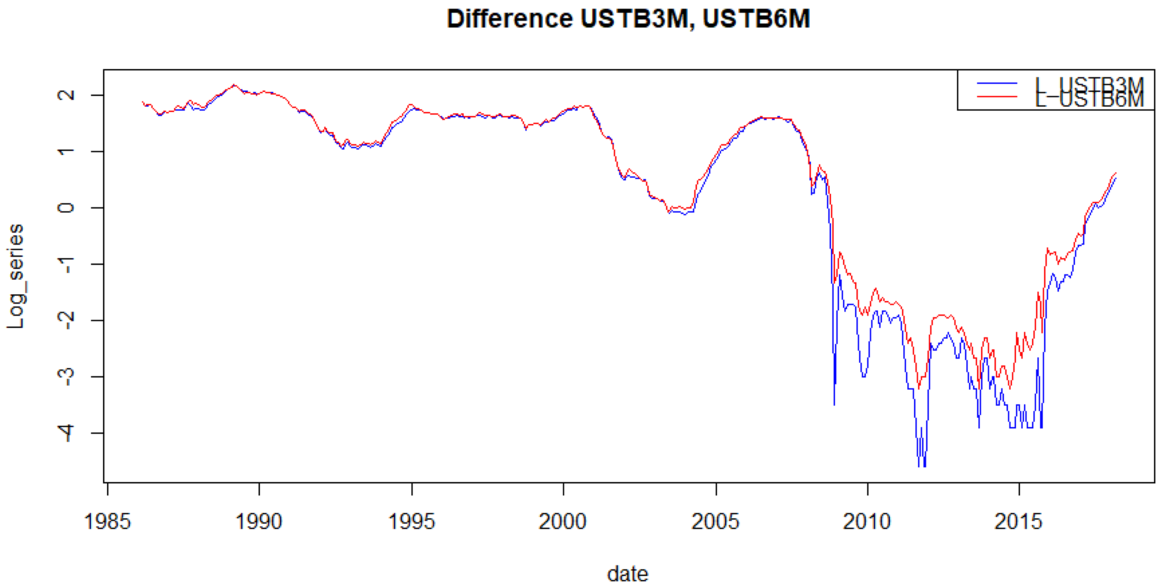


<u>Conditional correlation of the 2 series</u>

We can see that in general there is always a correlation between our 2 series except for some date on which the correlation is at almost 0. Having a time-varying correlation implies that if for example at a certain date there is a huge crisis in our stock exchange or for a firm then the other time series will highly be impacted too. It’s more significant here as we know that Microsoft is part of the SP500 index so any variation of Microsoft stocks at a certain period will impact the SP500.

Task 3

3.1



<u>Plot of the USTB3M and USTB6M series</u>

The two prices of USTB3M and USTB6M are shown on the graph. We can deduce that these series are perfectly cointegrated in terms of price, as the difference between the prices of both has remained the same for thirty years. In fact, we can also observe a difference between 2008 and 2015 which is the short-term deviation. While cointegrating variables may deviate from their relationship in the short-term, they would return to their equilibrium in the long-term like we can see at 2018-19. On the long term we have a clear cointegration with a similar trend.
Though this is a hypothetical example, it perfectly explains the cointegration of two non-stationary time series.

3.2

Correlation value: 0,9922891

Slope (Beta): 1,179854

Intercept (Alpha) : -0,2943569

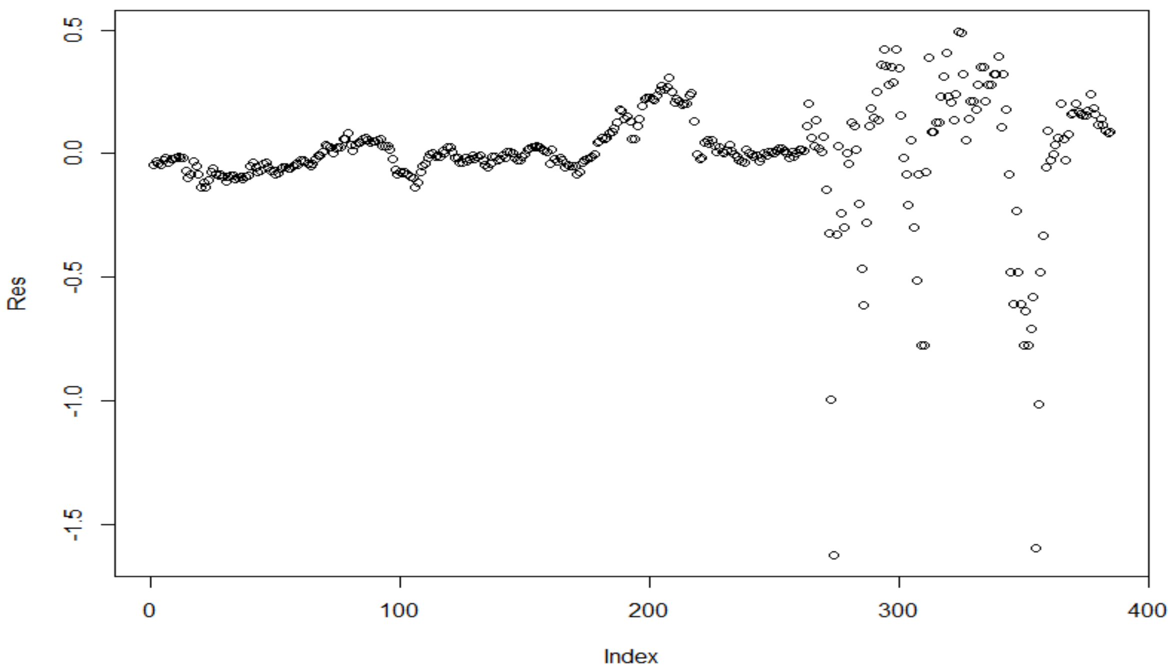
- The correlation coefficient (ρ) is a measure that determines the degree to which the movement of two different variables is associated. The most common correlation coefficient, generated by the Pearson product-moment correlation, is used to measure the linear relationship between two variables. However, in a non-linear relationship, this correlation coefficient may not always be a suitable measure of dependence.

→ The correlation coefficient is equal to 0.992 and the coefficient of the linear regression is equal to 1.1799 which confirm that the time series vary in the same direction. Furthermore, when the yield of US T-Bills 6M goes down, the 3M goes down too in the long term.

- The possible range of values for the correlation coefficient is -1.0 to 1.0. In other words, the values cannot exceed 1.0 or be less than -1.0. A correlation of -1.0 indicates a perfect negative correlation, and a correlation of 1.0 indicates a perfect positive correlation like in our result. If the correlation coefficient is greater than zero, it is a positive relationship like in our case. Conversely, if the value is less than zero, it is a negative relationship. A value of zero indicates that there is no relationship between the two variables.

→ to test non-stationarity, we look at the p-value if it is greater than 0,05 we cannot reject the null hypothesis which is non stationarity. It is possible to say that the six-month yields on the three-month yields are cointegrated meaning that they move together.

Regarding to the OLS --> The sum of residuals is not equal to 0, one of the 5 assumptions of OLS is violated. Furthermore, it is not possible to use only an OLS estimator in this case. We must use the Error correction model.



Residual plot

OLS TEST

```
Call:
lm(formula = L_USTB3M ~ L_USTB6M)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.62285 -0.04574  0.00247  0.08887  0.49337 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.29436    0.01228  -23.98   <2e-16 ***
L_USTB6M     1.17985    0.00753  156.68   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2292 on 383 degrees of freedom
Multiple R-squared:  0.9846,	Adjusted R-squared:  0.9846 
F-statistic: 2.455e+04 on 1 and 383 DF,  p-value: < 2.2e-16
```

```
> summary(OLS, confint = TRUE, ci.width = .95)

Call:
lm(formula = L_USTB3M ~ L_USTB6M)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.62285 -0.04574  0.00247  0.08887  0.49337 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.29436    0.01228  -23.98   <2e-16 ***
L_USTB6M     1.17985    0.00753  156.68   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2292 on 383 degrees of freedom
Multiple R-squared:  0.9846,	Adjusted R-squared:  0.9846 
F-statistic: 2.455e+04 on 1 and 383 DF,  p-value: < 2.2e-16
```

<u>Results of our OLS test</u>

3.3

```
Call:
lm(formula = DL_L_USTB6M ~ DL_L_USTB3M + ResidsAdj)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.45265 -0.02546 -0.00092  0.02355  0.47650 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.001493   0.004035  -0.370    0.711    
DL_L_USTB3M  0.500479   0.015884  31.507  < 2e-16 ***
ResidsAdj    0.084189   0.018116   4.647 4.64e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.07905 on 381 degrees of freedom
Multiple R-squared:  0.7241,	Adjusted R-squared:  0.7227 
F-statistic:   500 on 2 and 381 DF,  p-value: < 2.2e-16
```

Error correction model

Based on our previous results we can expression a relationship between the six-month and three-month yield because there is a high correlation between these 2 series. It’s possible to check this with an ECM because our series are co-integrated which mean that there is a long-term equilibrium relationship between all of our variables and that the short-term dynamics are driven by how this equilibrium deviates. Let’s check the value of our coefficients for the interpretations. The p-value of DL_L_USTB3M and ResidsADJ are below 0.05 so they are not significant. However, the intercept coefficient is significant and negative which means that there is both a short and long run equilibrium in our model. It means that at a certain period the 3 months and 6months maturity will have an equilibrium and so the supply and demand will be balanced.

3.4

```
> summary(Coint_Eq)

Call:
lm(formula = L_USTB3M ~ L_USTB6M)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.62285 -0.04574  0.00247  0.08887  0.49337 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.29436    0.01228  -23.98   <2e-16 ***
L_USTB6M     1.17985    0.00753  156.68   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2292 on 383 degrees of freedom
Multiple R-squared:  0.9846,	Adjusted R-squared:  0.9846 
F-statistic: 2.455e+04 on 1 and 383 DF,  p-value: < 2.2e-16
```

```
> summary(ECM)

Call:
lm(formula = DL_L_USTB3M ~ DL_L_USTB6M + ResidsAdj)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.07666 -0.02499  0.00070  0.03203  0.94297 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.001165   0.006854   0.170    0.865    
DL_L_USTB6M  1.443920   0.045828  31.507  < 2e-16 ***
ResidsAdj   -0.194456   0.030021  -6.477 2.89e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1343 on 381 degrees of freedom
Multiple R-squared:  0.7374,	Adjusted R-squared:  0.736 
F-statistic: 534.9 on 2 and 381 DF,  p-value: < 2.2e-16
```

Summary of our model

On the ECM we can see that the Intercept coefficient is still significant but positive which means that there is no short and long run equilibrium in our model. The values of residual standard error and Multiple R-squared are lower than in the previous model. It means that the USTB3M times series is more likely to have an equilibrium than the USTB6M series.

Task 4

4.1

First of all we need to find the appropriate number of lags to compute our VAR model. To do so we are going to check values of AIC, HQ, SC, FPE criterias.

```
> VARselect(VARData, lag.max=5)
$selection
AIC(n)  HQ(n)  SC(n) FPE(n)
     5      2      1      5

$criteria
                   1             2             3             4             5
AIC(n) -1.373550e+01 -1.380783e+01 -1.379161e+01 -1.383178e+01 -1.387412e+01
HQ(n)  -1.368602e+01 -1.372125e+01 -1.366793e+01 -1.367098e+01 -1.367622e+01
SC(n)  -1.361083e+01 -1.358965e+01 -1.347993e+01 -1.342660e+01 -1.337543e+01
FPE(n)  1.083304e-06  1.007726e-06  1.024223e-06  9.839436e-07  9.432196e-07
```

Values of our varselect model

We need to take the lag which maximizes our criterias and so we obtained 1 as the appropriate number of lags.

```
Estimation results for equation Yield_3M:
=========================================
Yield_3M = Yield_3M.l1 + Yield_6M.l1 + Yield_10Y.l1 + Yield_3M.l2 + Yield_6M.l2 + Yield_10Y.l2 + Yield_3M.l3 + Y
ield_6M.l3 + Yield_10Y.l3 + Yield_3M.l4 + Yield_6M.l4 + Yield_10Y.l4 + const

             Estimate Std. Error t value Pr(>|t|)    
Yield_3M.l1   1.16107    0.13094   8.867  < 2e-16 ***
Yield_6M.l1   0.23644    0.14336   1.649  0.09995 .  
Yield_10Y.l1  0.05123    0.04747   1.079  0.28126    
Yield_3M.l2  -0.51314    0.16877  -3.040  0.00253 ** 
Yield_6M.l2   0.04657    0.19110   0.244  0.80761    
Yield_10Y.l2 -0.09101    0.07283  -1.250  0.21221    
Yield_3M.l3   0.41950    0.17059   2.459  0.01439 *  
Yield_6M.l3  -0.18474    0.18826  -0.981  0.32709    
Yield_10Y.l3  0.10504    0.07200   1.459  0.14544    
Yield_3M.l4  -0.35638    0.13066  -2.728  0.00669 ** 
Yield_6M.l4   0.18532    0.14300   1.296  0.19582    
Yield_10Y.l4 -0.06474    0.04659  -1.390  0.16546    
const        -0.01893    0.02647  -0.715  0.47486    
---
```

Results for equation Yield_3M

By checking the p-value we can see that Yield_3M.l1, Yield_3M.l2, Yield_3M.l3, Yield_3M.l4 are significant. It means that the Yield_3M times series is not impacted over time by Yield_6M and Yield_10Y times series.

```
Estimation results for equation Yield_6M:
=========================================
Yield_6M = Yield_3M.l1 + Yield_6M.l1 + Yield_10Y.l1 + Yield_3M.l2 + Yield_6M.l2 + Yield_10Y.l2 + Yield_3M.l3 + Y
ield_6M.l3 + Yield_10Y.l3 + Yield_3M.l4 + Yield_6M.l4 + Yield_10Y.l4 + const

             Estimate Std. Error t value Pr(>|t|)
Yield_3M.l1   0.36771    0.13302   2.764  0.00599 **
Yield_6M.l1   1.03880    0.14564   7.133 5.25e-12 ***
Yield_10Y.l1  0.12878    0.04822   2.670  0.00791 **
Yield_3M.l2  -0.36763    0.17145  -2.144  0.03266 *
Yield_6M.l2  -0.09054    0.19413  -0.466  0.64121
Yield_10Y.l2 -0.16861    0.07398  -2.279  0.02324 *
Yield_3M.l3   0.28080    0.17330   1.620  0.10601
Yield_6M.l3  -0.10965    0.19125  -0.573  0.56678
Yield_10Y.l3  0.14456    0.07314   1.977  0.04884 *
Yield_3M.l4  -0.40479    0.13273  -3.050  0.00246 **
Yield_6M.l4   0.27215    0.14527   1.873  0.06181 .
Yield_10Y.l4 -0.09628    0.04733  -2.034  0.04263 *
const        -0.01494    0.02689  -0.556  0.57865
```

<u>Results for equation Yield_6M</u>

P-value is significant only for the Yield_6M.l1, Yield_10Y.l1, Yield_3M.l2, Yield_10Y.l2, Yield_10Y.l3,Yield_3M.l4 and Yield_10Y.l4 coefficients which means that the Yield_6M times series is impacted over time by the 2 others one.

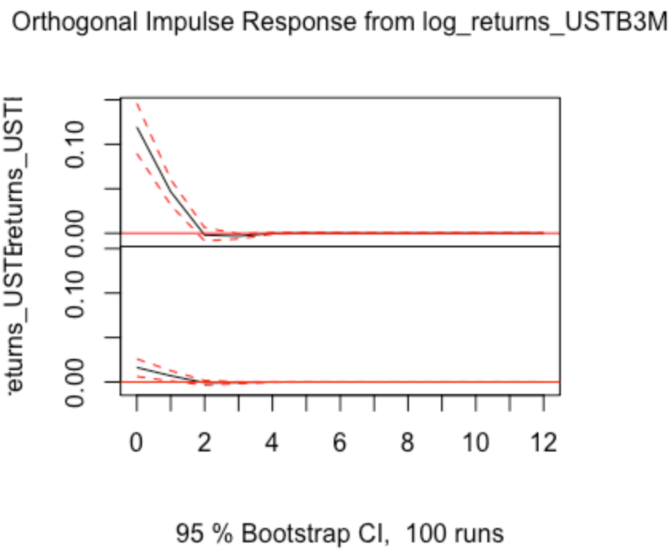
`

```
Estimation results for equation Yield_10Y:
==========================================
Yield_10Y = Yield_3M.l1 + Yield_6M.l1 + Yield_10Y.l1 + Yield_3M.l2 + Yield_6M.l2 + Yield_10Y.l2 + Yield_3M.l3 +
 Yield_6M.l3 + Yield_10Y.l3 + Yield_3M.l4 + Yield_6M.l4 + Yield_10Y.l4 + const

             Estimate Std. Error t value Pr(>|t|)
Yield_3M.l1   0.18350    0.18358   1.000 0.318170
Yield_6M.l1  -0.16858    0.20100  -0.839 0.402158
Yield_10Y.l1  1.31506    0.06656  19.759  < 2e-16 ***
Yield_3M.l2   0.23500    0.23662   0.993 0.321273
Yield_6M.l2  -0.18379    0.26792  -0.686 0.493155
Yield_10Y.l2 -0.49091    0.10211  -4.808 2.23e-06 ***
Yield_3M.l3   0.05399    0.23917   0.226 0.821515
Yield_6M.l3  -0.22874    0.26395  -0.867 0.386720
Yield_10Y.l3  0.38637    0.10094   3.828 0.000152 ***
Yield_3M.l4  -0.36830    0.18318  -2.011 0.045099 *
Yield_6M.l4   0.48422    0.20049   2.415 0.016216 *
Yield_10Y.l4 -0.22523    0.06531  -3.448 0.000629 ***
const         0.05302    0.03711   1.429 0.153866
---
```

<u>Results for equation Yield_10Y</u>

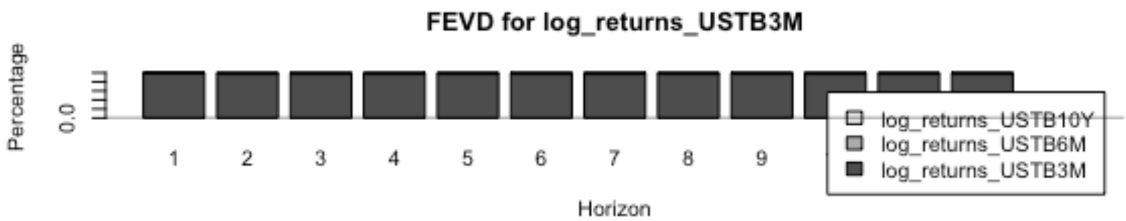
P-value is significant for the Yield_10Y.l1, Yield_10Y.l2 Yield_10Y.l3, Yield_3M.l4,Yield_6M.l4,Yield_10Y.l4 which means that the Yield_10y times series is impacted over time by the 2 others one.

4.2



<u>Impulse response from log returns of USTB3M</u>

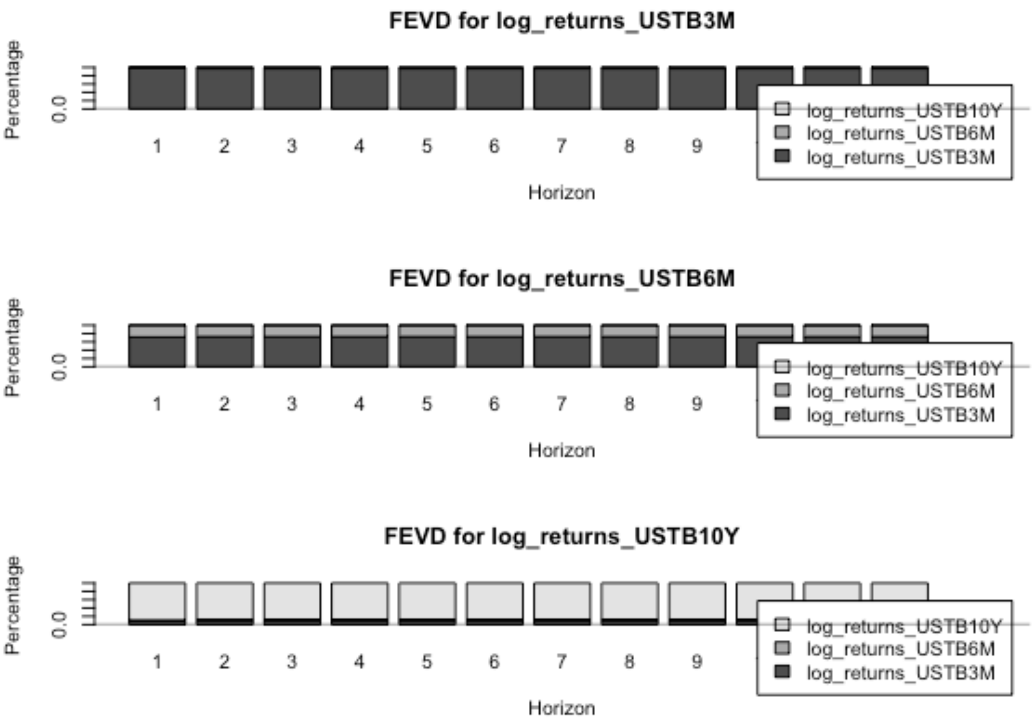
We can see that for a shock observed on today to decay 2, the decreasing trend is similar for USTB3M and USTB6M times series so the impact will be observed for both time series.



FEVD from log returns of USTB3M

We can see that the plots have a Y-range from 0 to 25% and it shows the contribution of every individual shock as a portion of our total area during all the time periods. In the initial time period, most of the variation of the USTB3M times series is explained from shocks of USTB6M and a small part is from itself. Over the time period, the repartitions stay constant. As we examine the same stock, the only difference is the period of maturity. A 3 month maturity is still quite closed from a 6 months maturity as only 3 months separates our both time series maturity. Due to this, the shocks of the USTB3M is more likely to be explained from shock of USTB6M rather than the USTB10Y.

4.3



FEVD from log returns of the 3 time series

For the USTB6M we can see that most of its shock is from USTB3M time series and for the USTB10Y time series most of its shock is from itself and a small part is from USTB3M so the results are sensitive to the variable ordering. From a financial standpoint, we can see that the time period difference between a 3 months, 6 months or 10 years yields can have an impact on the shocks results.

4.4

```
######################
# Johansen-Procedure #
######################

Test type: trace statistic , with linear trend

Eigenvalues (lambda):
[1] 0.100241902 0.024467512 0.004501816

Values of teststatistic and critical values of test:

          test 10pct  5pct  1pct
r <= 2 |  1.72  6.50  8.18 11.65
r <= 1 | 11.19 15.66 17.95 23.52
r = 0  | 51.54 28.71 31.52 37.22

Eigenvectors, normalised to first column:
(These are the cointegration relations)

             Yield_3M.l3 Yield_6M.l3 Yield_10M.l3
Yield_3M.l3   1.00000000    1.000000     1.000000
Yield_6M.l3  -1.02211689   -2.537968    -1.129775
Yield_10M.l3  0.02635925    1.586405     0.884769
```

<u>Johansen-procedure with trace approach</u>

For the Trace approach, we will be focus on the test and the 5pct columns which means to the 5% significant level. R represents the rank of our error correction terms matrix and they will tell us how many cointegrating relationships in our system. For r=0, the test value is 51.54 which is greater than the 5% (31.52) so we reject the null hypothesis which suggests that there is at least one cointegrating relation. For r<= 1 and r<= 2 we don't reject the null hypothesis because the test value is lower than the 5pct value. It means that there are at most 2 cointegrating relationships present in this case.

```
# Johansen-Procedure #
######################

Test type: maximal eigenvalue statistic (lambda max) , with linear trend

Eigenvalues (lambda):
[1] 0.100241902 0.024467512 0.004501816

Values of teststatistic and critical values of test:

          test 10pct  5pct  1pct
r <= 2 |  1.72  6.50  8.18 11.65
r <= 1 |  9.46 12.91 14.90 19.19
r = 0  | 40.35 18.90 21.07 25.75

Eigenvectors, normalised to first column:
(These are the cointegration relations)

             Yield_3M.l3 Yield_6M.l3 Yield_10M.l3
Yield_3M.l3   1.00000000    1.000000     1.000000
Yield_6M.l3  -1.02211689   -2.537968    -1.129775
Yield_10M.l3  0.02635925    1.586405     0.884769
```

<u>Johansen-procedure with eigen approach</u>

For the eigen approach, we have the same results that there are at most 2 cointegrating relationships present in our model by following the same logical. In light of the empirical evidence in Shea (1992) which helped for the interpretation of this test and the number of cointegration relationships. "*<u>First, Johansen (1989) gave two test statistics for rank (II) or the number of independent cointegrating vectors",</u>* Shea (1992)