BERGUIGA Oussama

Master Statistic and Big Data

Université Paris Dauphine

TIME SERIES PROJECT N°1

1) This data contains M1 Money Stock data from 01-06-1976 to 23-03-2015

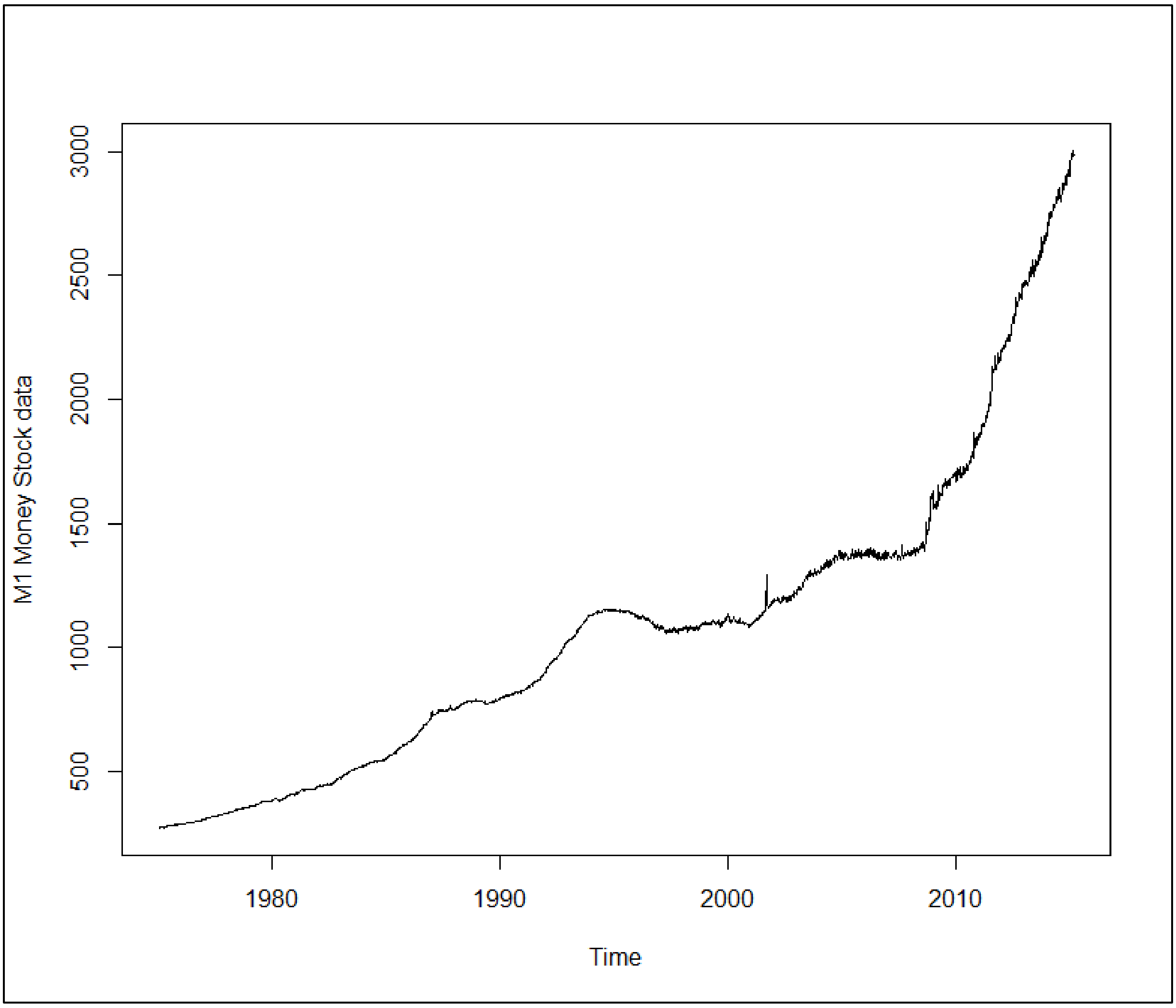
2) What is stationarity? There are two types of stationarity:

- strong stationarity: Z1,Z2,...Zt and Z1+k,Z2+k,..Zt+k have same law

- weak stationarity:

- E[Zt] is constant against t
- Var(Zt) is constant against t AND finite (Var= sigma²< + oo)
- Cov(Zt,Zt+k) is a function that only depends on k

Fig. 1 : plot of the stock M1t



We can see a tendency, so it cannot be stationary. Since the expectation is increasing exponentially.

The tendency seems to be exponential. This can be confirmed by a Dickey-Fuller test:

Dickey-Fuller test:

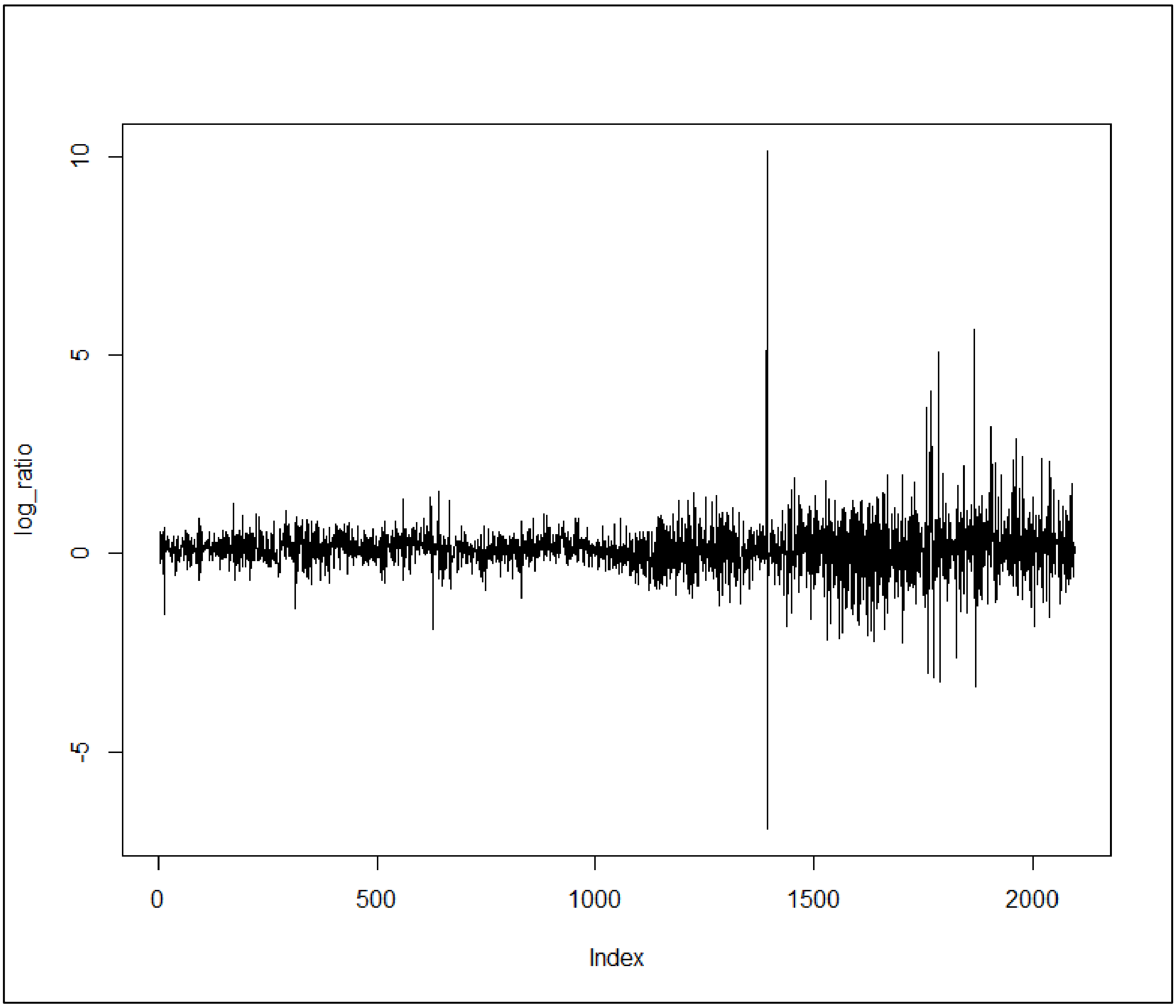
Ho=the time series is not stationary

H1=the time series is stationary

The p-value given by the equals 0.99: it is absolutely not stationary.

3)

Fig. 2 : plot of the log ratio DM1t = 100 ∗ log(M1t/M1t−1)



On that plot we can see:

- that the mean is constant during the time:
- the variance is NOT constant:
- there is heteroscedasticity

- outliers, strong and weak volatility

The log ratio time serie seems to have clusters of volatility (probably because of different financial crisis).

The variance of the 1000 first values is 0.1, the variance of the 500 last values is 0.97..

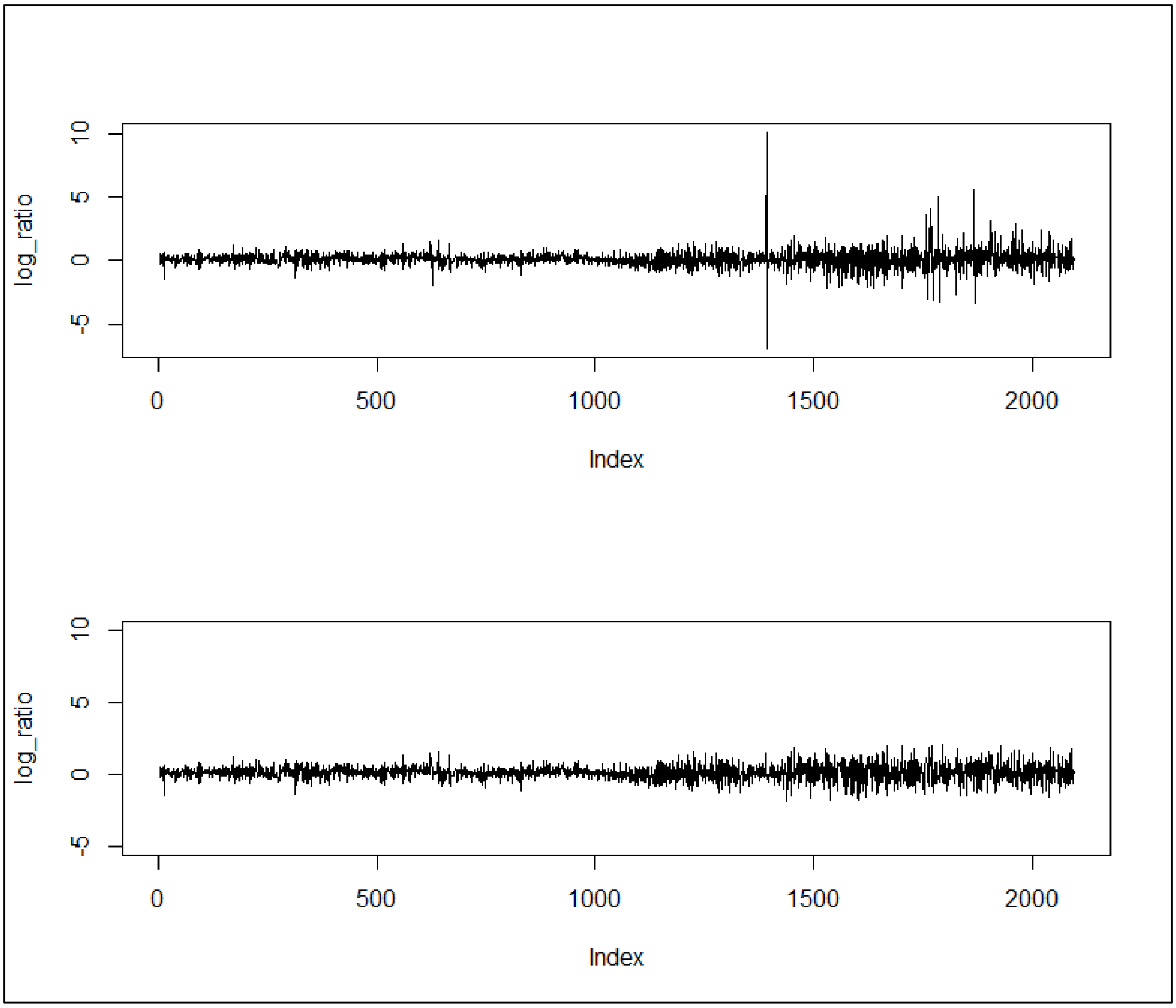
The variance seems to increase a lot.

The most visible outlier seems to be caused by 9-11 terrorist attacks

At first sight the time serie seems to be strongly stationnary with outliers (this means that it does not imply that it is weakly stationnary).

Let's delete the outliers in order than we can consider the time serie to be weakly stationary (we use "tsclean" function in the package "Forecast").

Fig. 3 : plot of the log ratio DM1t = 100 ∗ log(M1t/M1t−1) before and after the cleaning of the Time Serie



- About the Stationarity :

In view of the plot the cleaned Time Serie seems to be weakly stationary.

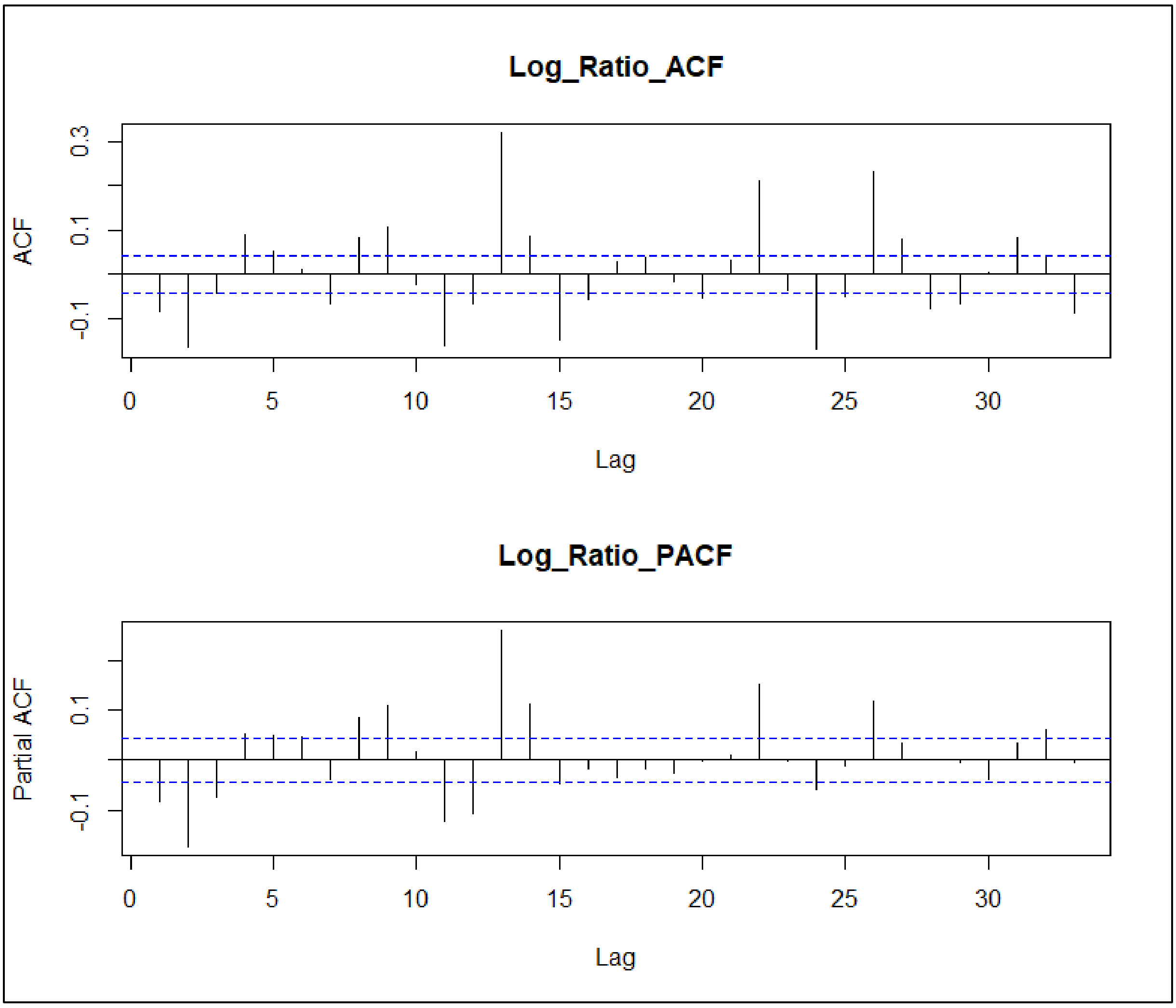
Let's try an adf test to check with R weather it is stationary :

p-value= 1%, it is < than 5% so the log_ratio serie is probably stationary

We can suppose that the log_ratio is weakly stationary since the variance is finite with no more outliers, BUT we can see that the variance is increasing in the second part of the dataset) it seems that there are still volatility clusters ...

4)

Fig. 4 : plot of the Log-ratio's ACF and PACF



If acf is equal to zero after q lags, we could fit an MA(q) model. If pacf is equal to zero after p lags, we could fit an AR(p) model. Here, both pacf and acf seem to decrease to zero after some lags so perhaps we can fit an AR, or an MA model.

5) The last important lag in the pacf is for p=13, so we can fit an AR model (with “AR” function in R).

We find AR (32) !

6) The last important lag in the acf is for q=26 so we can fit an MA (26) model.

7)

Fig. 5 : plot of the Log-ratio’s AR and MA residuals

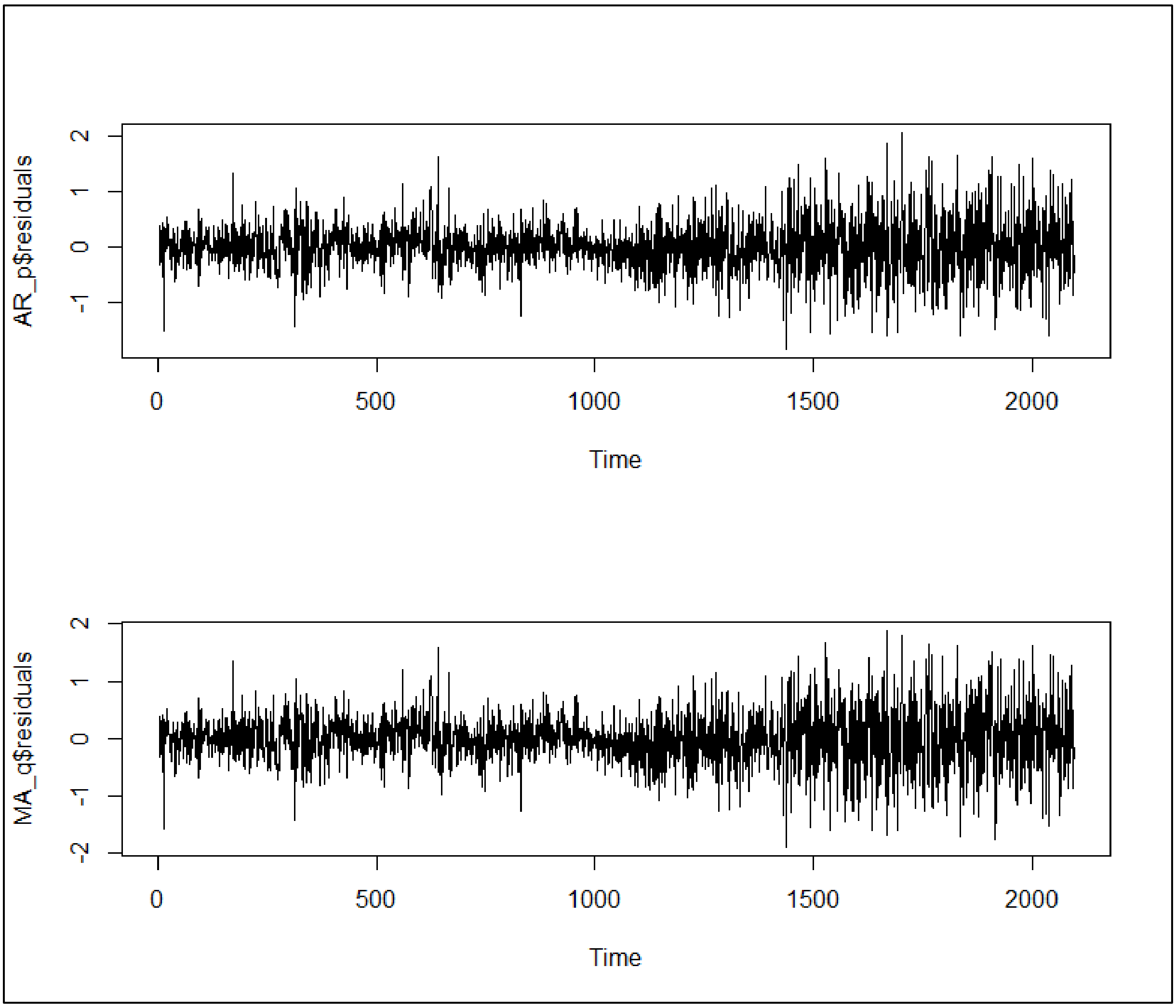
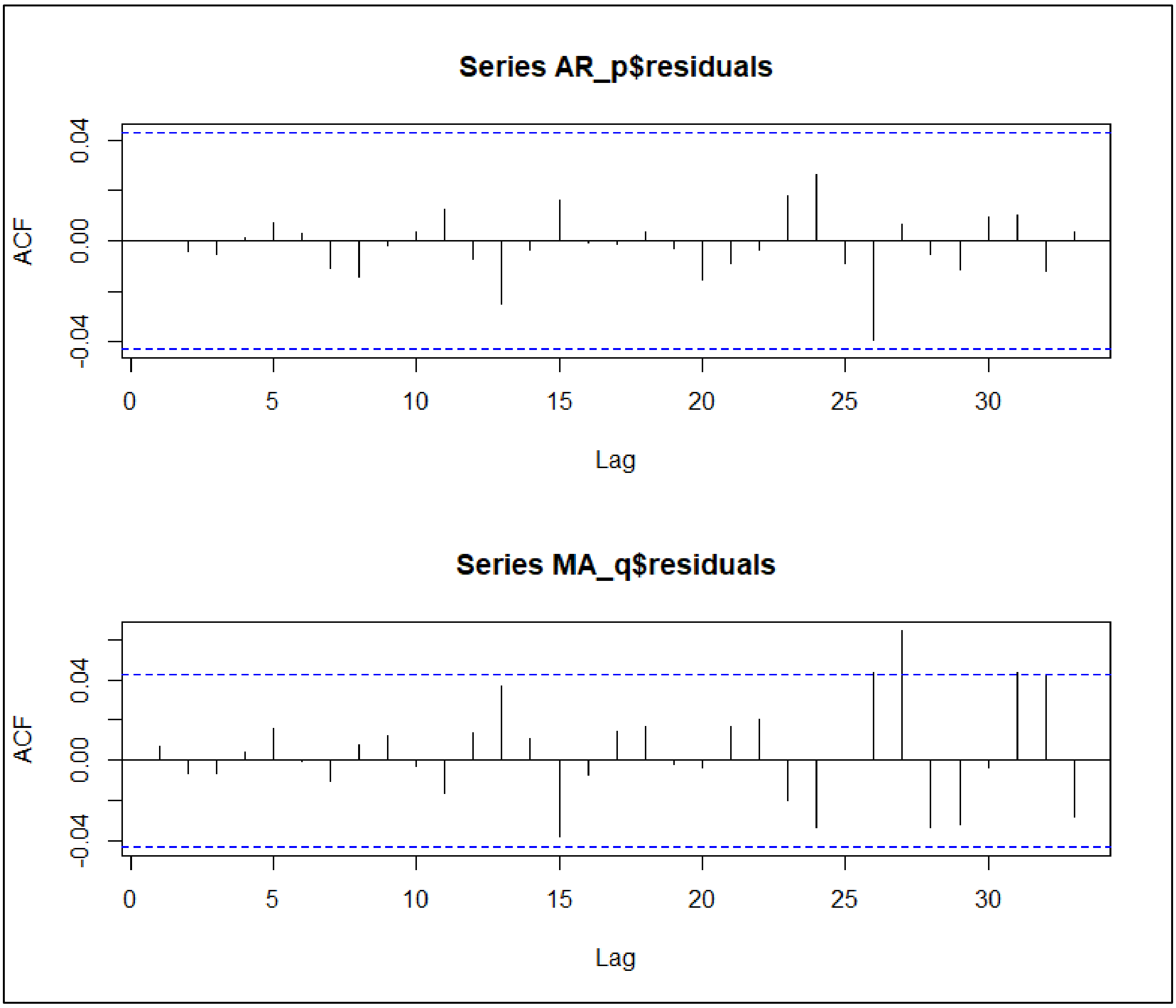


Fig. 6: plot of those residuals ACF



In view of the plot, the residuals look like uncorrelated, BUT the variance of the residuals does not seem to be constant (volatility)...

To test the correlation of the residuals: we can use a Box-Ljung test (or Box Pierce or portmanteau test). Box-Lung test:

- Ho=the residuals are not correlated

- H1=the residuals are correlated

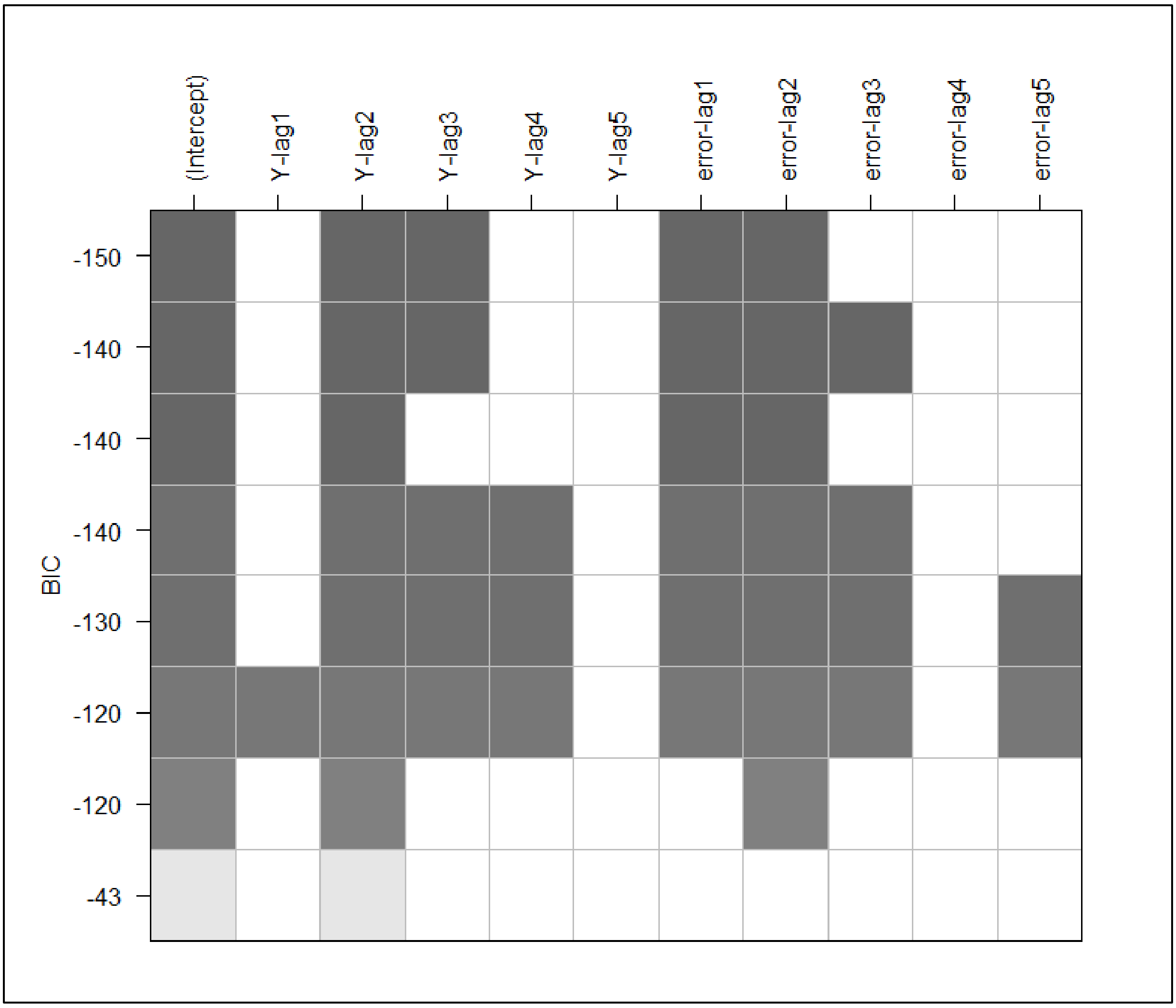
The p-values are nearly equal to 1. Both p-values are >> 5 %: so we can assume that the residuals are uncorrelated. Yet the variance of the residuals does not seem to be constant...The variance is more important at the end of the Time Serie.

8)To fit the ARMA model we can use 2 different function in R:

- auto.arima gives us p=4 and q=2

- armasubset in the package "TSA" :

Fig. 7: plot of Armasubset



Given that plot, the miminum BIC (-150) is got whith p=2 or 3, q=1 or 2. We have less than 5 parameters in the ARMA model whereas we had 32 and 13 parameters in the AR and MA models !

9)Let us compare the different models obtained:

#AR_p : AIC=2854 , log_likelyhood=-1394 , Box Test p value =0.99

#MA_q : AIC=2948, log_likelyhood=-1447 , Box Test p value =0.74

#ARMA_1 : AIC=3170 , log_likelyhood=-1577 , Box Test p value =0.99

#ARMA_2 : AIC=3205 , log_likelyhood=-1597 , Box Test p value =0.88

#ARMA_3 : AIC=3080 , log_likelyhood=-1534 , Box Test p value =3.3e-5

#ARMA_4 : AIC=3199 , log_likelyhood=-1593 , Box Test p value =1

#ARMA_5 : AIC=3054 , log_likelyhood=-1520 , Box Test p value =0.75

Based on thoses statistics indicators, the most preferable model COULD BE AR_p (lowest AIC, highest log likelyhood). BUT HERE p = 32 : it's a trap : there are too much parameters !

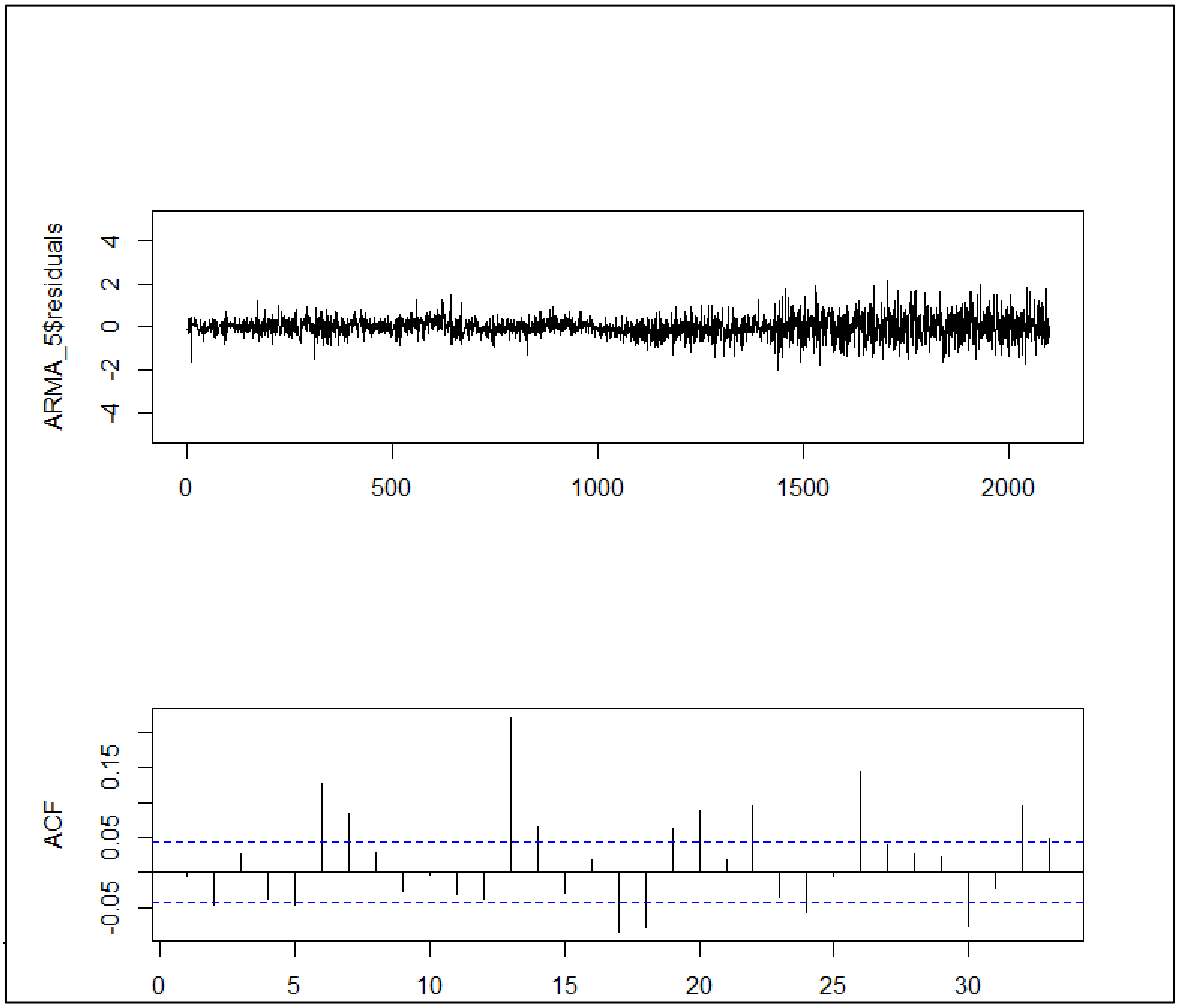
So we should better keep the ARMA_5(ARMA (3,2)) model (p+q=5 << 32 or 13)

(lowest AIC, highest log likelyhood, uncorrelated residuals based on Box Test p value)

Is it a White Noise ? We have to check if :

- the residuals mean equals 0
- their variance IS CONSTANT
- and if they are uncorrelated

Fig. 8 plot and ACF of the ARMA (3,2)



Based on the plot, it seems that the variance is increasing ...

Since the variance is not constant(heteroscedasticity), it cannot be a white noise

test for uncorrelation of the residuals:

We already saw that the residuals where uncorrelated whith the Box-Ljung test

The p-value equals 0.75 >> 5 % so we can consider that the residuals of that ARMA(3,2) model are uncorrelated.

Let us try another test to confirm that information.

Turning Point Test :

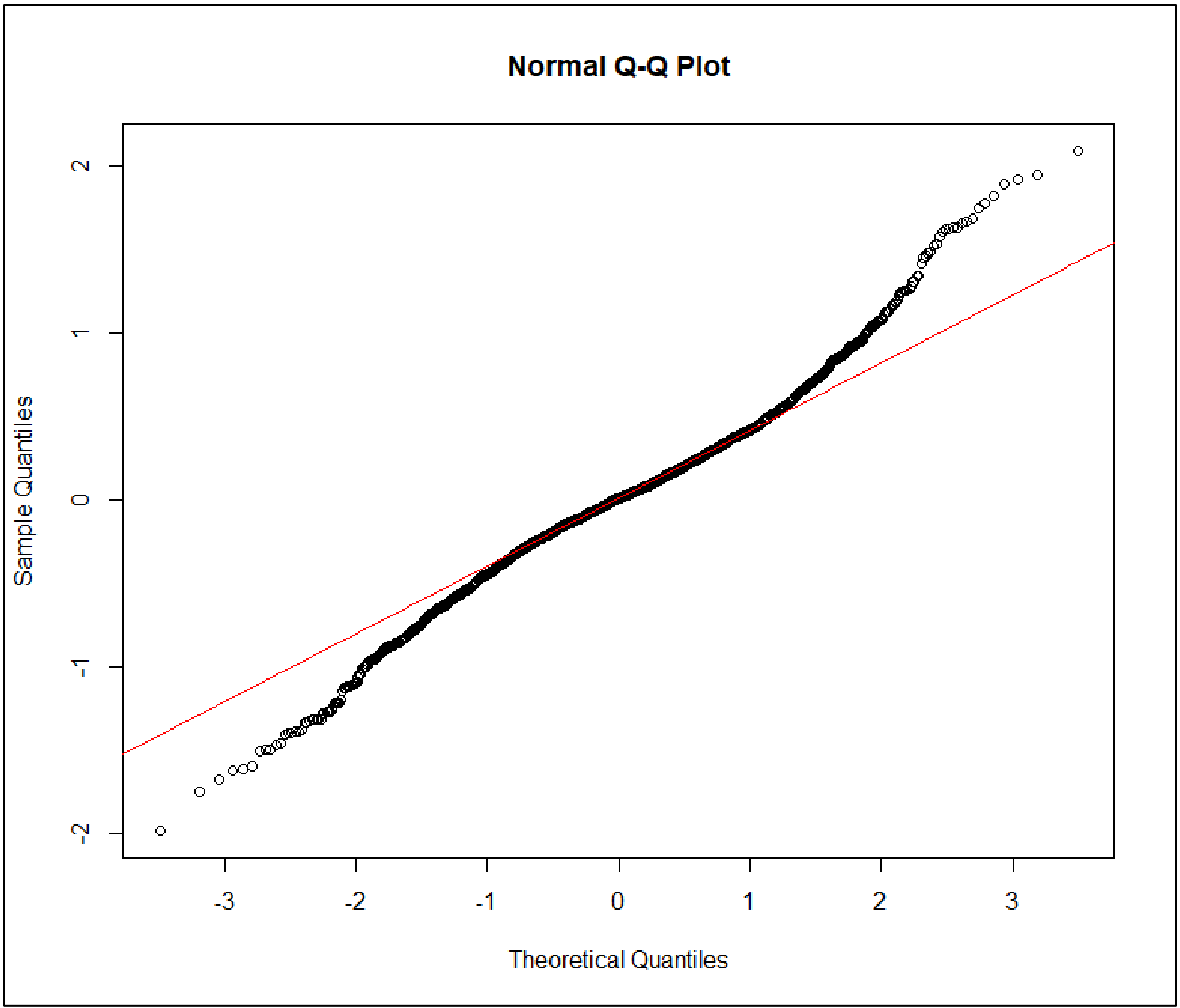
Ho= the residuals are not correlated

H1=the residuals are correlated

p-value is 0.67, > > 5%: we can assume that residuals are not correlated BUT given the residual plot, the variance seems to increase there is no homoscedasticity

- Are the residuals gaussian?

Fig. 9: QQ plot of ARMA's residuals



The qq-plot is not fitted good at its extremity. It fits well only for the theoretical Quantiles between -1 and 1.This may be a default of normality.

Let's try a test for normality of the residuals : we can use a Shapiro-Wulk test.

Shapiro-Wulk test:

Ho= the statistical serie follow a normal distribution

H1=the statistical serie does not follow a normal distribution

p-value = 2.5e-15 << 5%, so shapiro test shows that it is NOT gaussian

Based on the Shapiro test and the qqplot, we can conclude that the residuals of the ARMA (3,2) model are uncorrelated, their mean is zero, but they are not gaussian.

Moreover, since the variance does not seem to be constant, the residuals cannot be a white noise the ARMA model may not be the best thing to do here (no homoscedasticity). We will probably see in the next course that the GARCH models have been discovered by Robert Engle in 1982 (he got the Nobel Prize in Economics in 2003 for that) to treat financial time series whose variance evolves with time (volatility clustering, conditional heteroscedasticity).