Data analysis Project
Dell's daily stock returns

Séries Chronologiques - LSTAT2170

Louvain-la-Neuve, May 16th, 2020

Members	**NOMA**
Julio Pastor Tronch	10471901

Contents

List of Figures

List of Tables

1 Identification and preliminary analysis

1.1 Preliminary visual analysis on daily stock returns

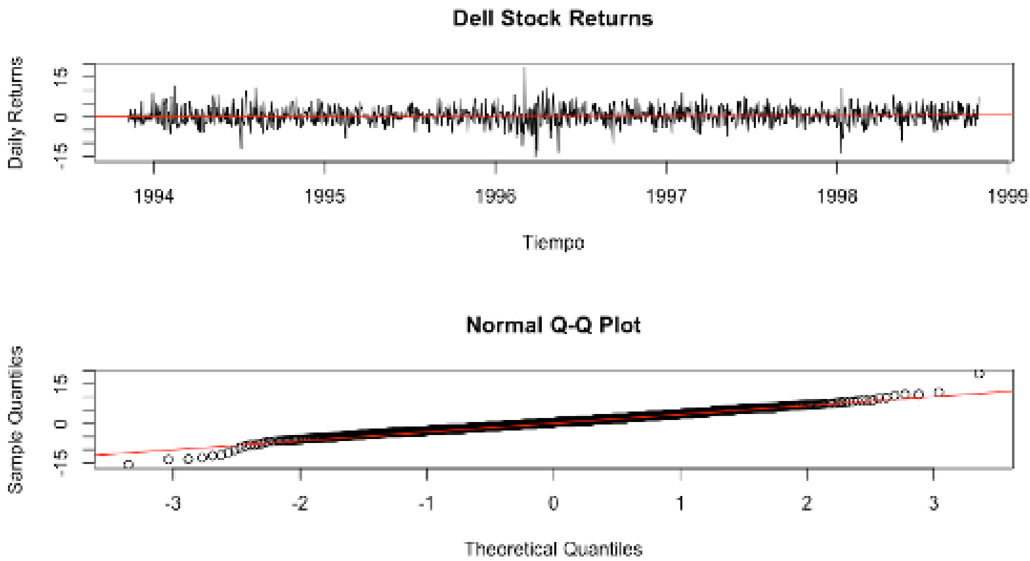


Figure 1: Time plot and normal Q-Q plot comparison of the daily stock returns

In order to start our data analysis we should take a look to the time plot where we can observe how the returns vary along zero with the largest daily stock return around 1996. The time series shows some signs of non-constant conditional volatility as there are some spikes in 1994, 1996 and 1998.

As we can observe by a preliminary visual analysis there's no interesting mean trend as the slope of the line fitted by linear regression (red line in ***Figure 1*** first plot) to the series is close to zero. We can also observe in the normal Q-Q plot from ***Figure 1*** that our distribution fits quite well in a Normal distribution $N(0, \sigma)$.

1.2 Analysis of the possible trends, seasonality, breaks and volatility

It's also easy to conclude that there's not an appreciable seasonality in the data by looking at ***Figure 2*** where there's not a fixed pattern of the distribution during the trading year (which is of 253 days) neither for the daily stock returns nor for the squared series. But we can appreciate as we have stated in the first visual analysis how there are several high peaks(clusters) that might be associated with conditional volatility in 1994,1996 and 1998.

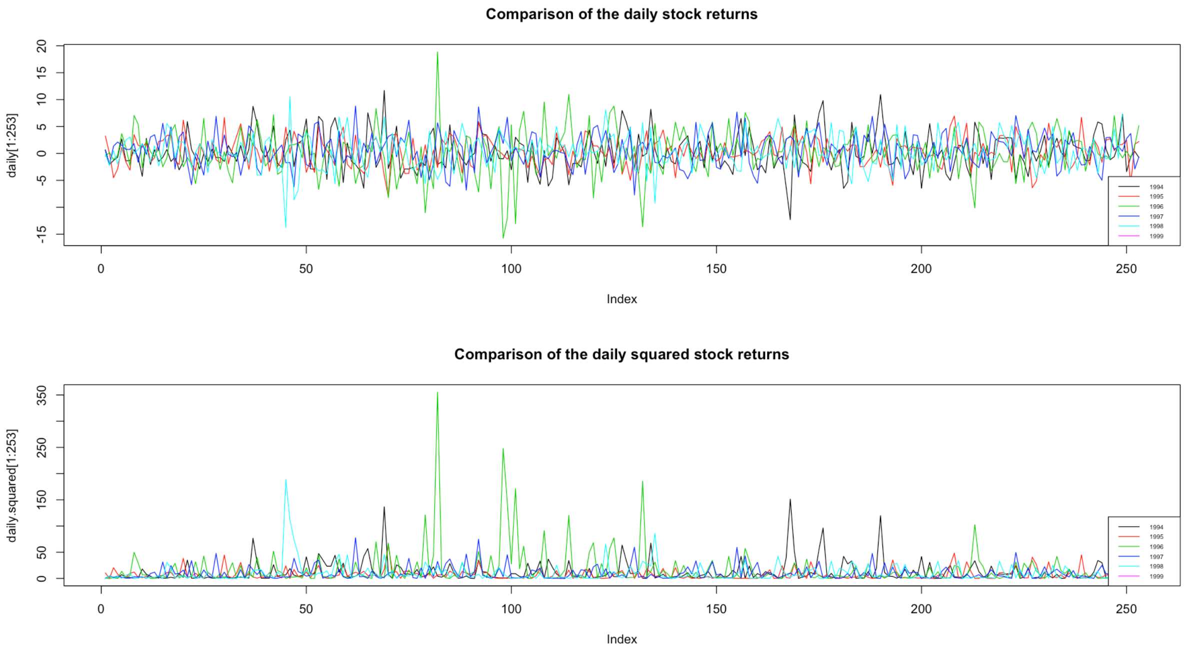


Figure 2: Yearly comparison of the daily returns and squared daily returns

1.3 Preliminary conclusions

To sum up it won't be necessary any kind of transformation on the series as the mean remains constant and there's no appreciable seasonality in the data.

In fact, the use of the log returns in this case would only lead to a difference in the scale, therefore we won't apply any kind of previous transformation. It is interesting to remark that as we are actually treating with the returns we are already working at a difference of order one in the Stock prices.

It is also worth noting that this preliminary analysis suggests a white noise model but due to the financial nature of the data it would be convenient to check if there's correlation in the squared series, as we can observe some volatility clusters and thus a conditional variance, for example in the year 1996.

Stock returns summary	
Length	1261
Mean	0.366
Median	0.365
Variance	11.328
Stdev	3.366
Skewness	-0.053
Kurtosis	1.715

Table 1: Statistical summary of the time series

To end up with this preliminary analysis, and before we start to estimate our model, by looking at the ***Table 1*** can see how our mean and median are nearly around zero, maybe a little bit higher than zero but with no relevant effect in our distribution, and also that the *Kurtosis* is slightly smaller than the one expected for a normal distribution what could also be a sign of an underneath GARCH model.

2 Model estimation

2.1 Estimation of the mean model

In order to choose the model that fits better to our series we will actually follow two procedures:

- Visual Analysis of the Autocorrelation (ACF) and Partial Autocorrelation (PACF) function plots.
- Optimization of the AIC and BIC criteria

We will start by the visual analysis of the ACF and PACF plots in order to get a first idea of the appropriate order of our model.

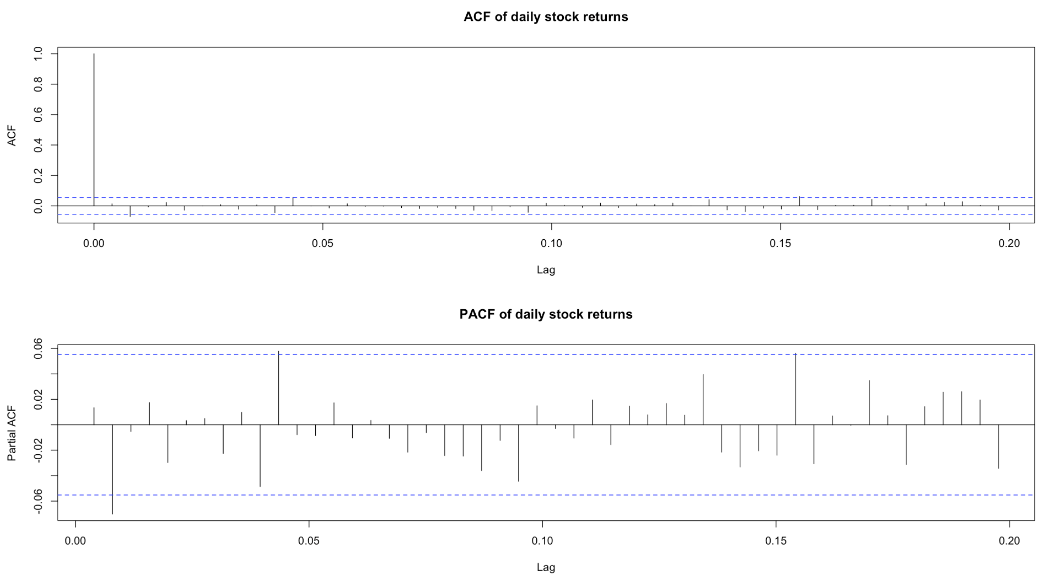


Figure 3: ACF and PACF of the daily stock returns with 95% CI

By taking a look to ***Figure 3*** we can observe how the only plot informative enough is the PACF, for a lag of second order so we could be dealing then with an auto-regressive model of order 2 (***AR(2)***) but as the confidence intervals(CI) we are actually using are quite tight (due to the length of the series) the significance showed by the PACF would not be really relevant and we might be dealing with an ***ARMA(0,0)*** process, which is actually a white noise distribution modeling the mean of the financial time series. In addition there's no further correlation confirming our preliminary condition of no-seasonality or at least not a strong one.

If we actually use a higher confidence probability for the CI we will see how the plots (ACF, PACF) adjust to an ***ARMA (0,0)***(see appendix ***Figure 14***).

Finally, using an iterative comparison of different models by the AIC and BIC criteria, I've come to two useful models such as an ***ARMA(0,0)***(by minimizing the BIC criteria) and an ***AR(2)***(by minimizing the AIC) as the best models, which agrees with our first visual analysis of the ACF and PACF plots.

2.2 Validation of the mean model

In order to check the validity of the models we've used the ***tsdiag()*** and ***coef.p()*** function. The first one gives us a graphic summary of the ACF and the Ljung-Box p-values of the residuals from the fitted model, and the second one performs a statistical Box test in order to check the significance of the coefficients for the model.

First of all we fitted the ***ARMA(0,0)*** and checked the ACF of the residuals(which are actually the series itself) and the p-values of the Ljung-Box test, as we can observe in the ***Figure 4*** the ACF plot shows no significant correlation in our residuals but the p values for lags up to 4 are quite close to 0.05, which makes harder the decision of rejecting the null hypothesis and therefore concluding we're dealing with an uncorrelated time series, there's also again the possibility that this p-values are influenced by the length of the time series.

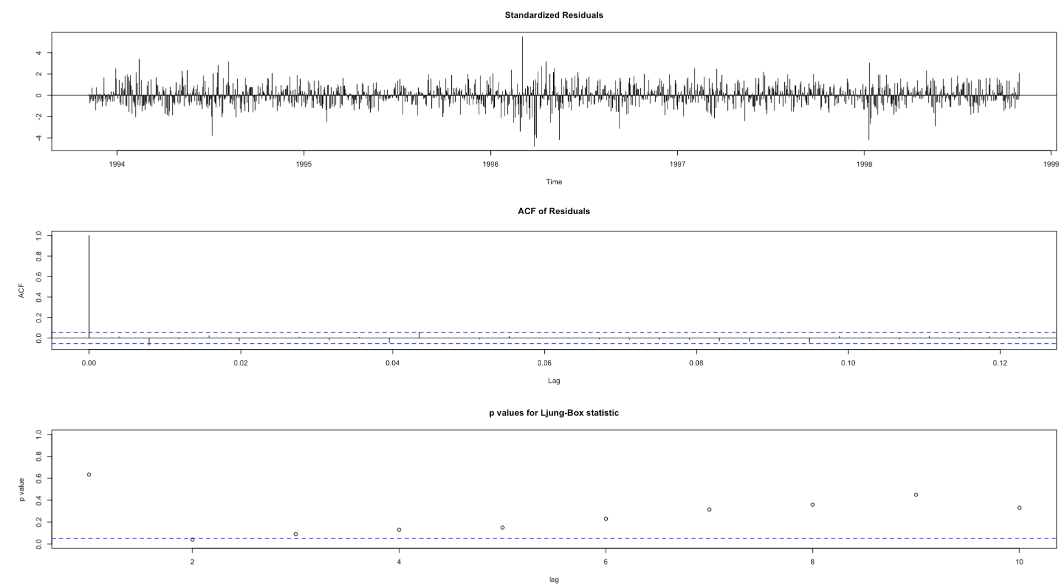


Figure 4: Residuals analysis of the ***ARMA(0,0)*** model

After fitting an ***AR (2)*** model it's easier to observe in the ***Figure 5*** as both the ACF and the p-value of the Ljung-test allow us to state with a 95% probability that there's no correlation between the residuals of the model, what means that our model is able to explain all the correlation of the mean.

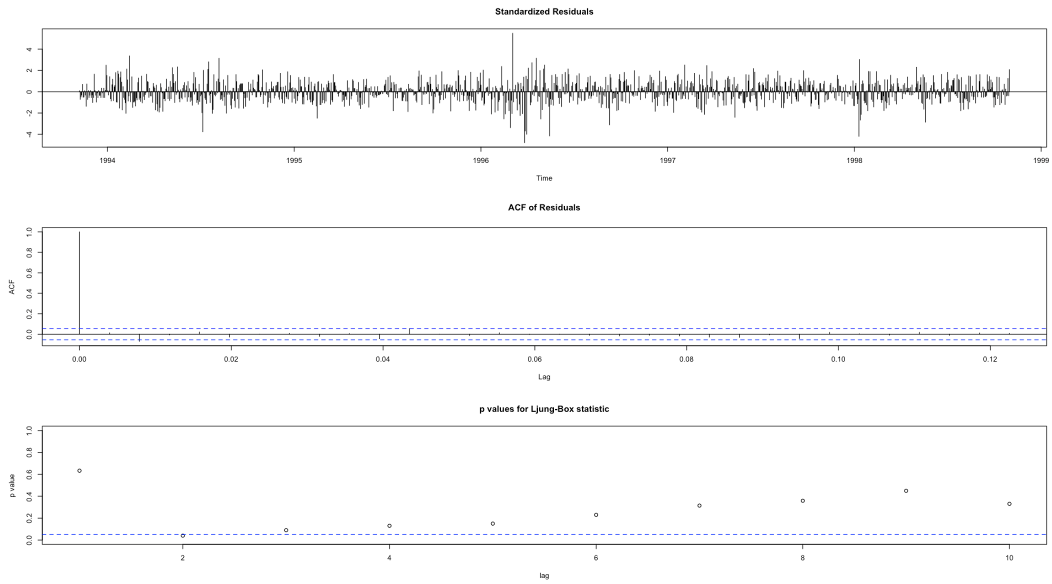


Figure 5: Residuals analysis of the ***AR(2)*** model

By looking at the coefficients importance(***Table 2 and 3***)we can conclude that in the ***AR(2)*** both the ar2 and the intercept terms are significant as their p-values are below the 0.05 whereas the ar1's p-value is over 0.05, indicating us that we should set it to zero when we develop our model. It would be interesting then to check then for a model of smaller order as ***AR(1)*** model and look if it fits the requirements but with just checking the p-values we can conclude that it wouldn't be a good model as its p-value is over the 0.05 therefore we cannot conclude it's a relevant term. So our two best choices for modeling the mean are ***AR(2)*** and ***ARMA(0,0)***

Coefficient	p-value
ar1	$6.079e^{-01}$
ar2	$1.235e^{-02}$
intercept	$4.310e^{-05}$

Table 2: Coefficient importance by the Ljung-Box test for the AR(2) model

Coefficient	p-value
intercept	0.0001134707

Table 3: Coefficient importance by the Ljung-Box test for the ARMA(0,0) model

2.3 Estimation of the variance model

Once we've determined the proposed models for the mean, we'll have to check if there's correlation between the squared residuals which is the main characteristic of the ***GARCH*** model.

As we can observe in ***Figure 6*** and ***Figure 7*** the PACF and the ACF of the squared residuals for both models (which actually contain the variance of the time series) show that we may need also a GARCH process to model the Series, as there's a significant correlation for different lags of time showing that the process has a strong non-linear dependence in the variance of the series.

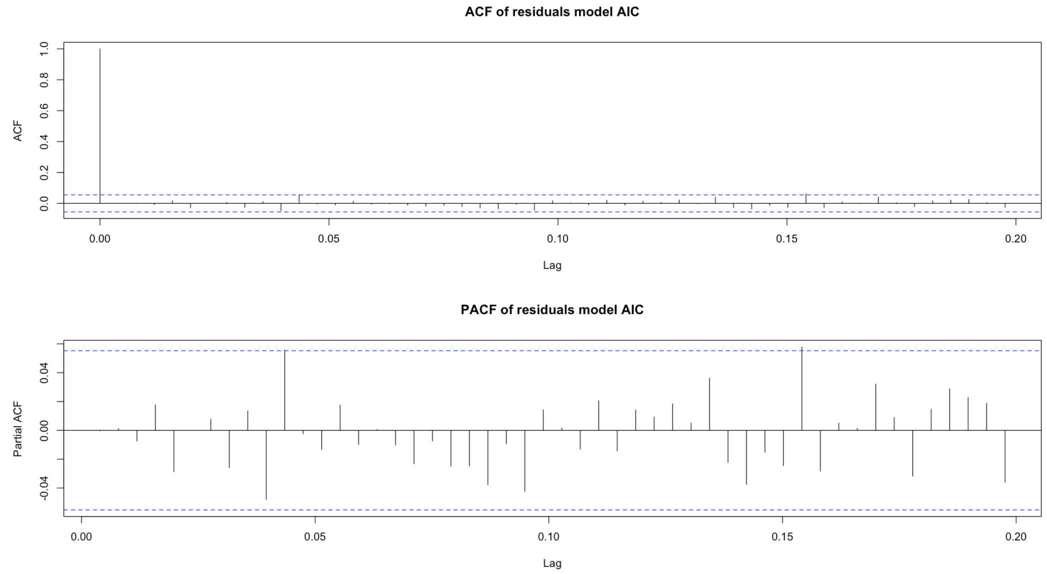


Figure 6: ACF and PACF plots of the squared residuals of the ***ARMA(2,0)*** fitted model

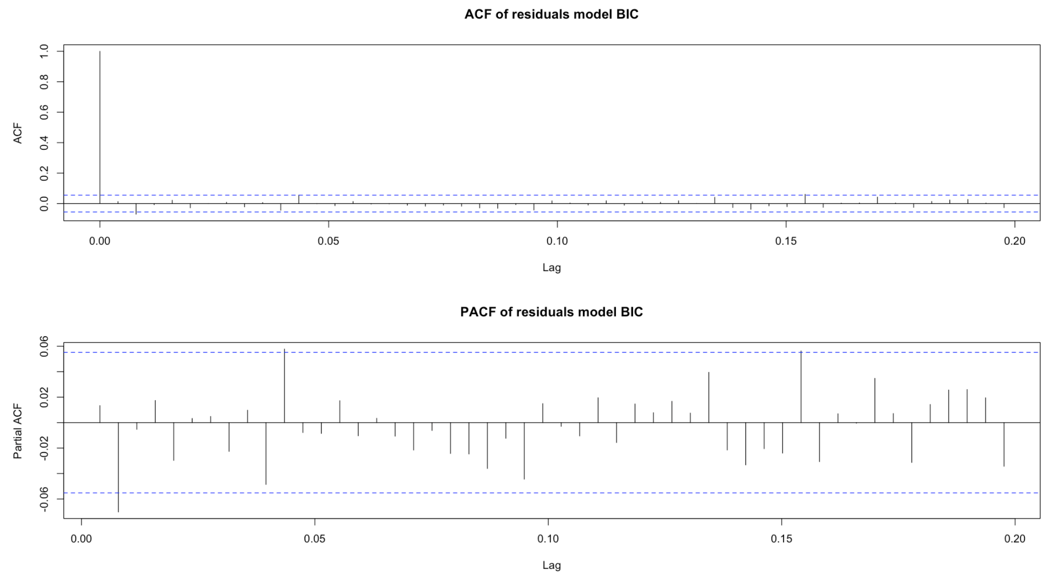


Figure 7: ACF and PACF plots of the squared residuals of the ***ARMA(0,0)*** fitted model

Once we have checked that our squared residuals present a certain correlation, we must find out which is the best order for our ***GARCH*** model. In order to do that we will use the squared residuals to find an optimum ***ARMA*** model by minimizing the AIC and then use this model to infer the correct order of the ***GARCH*** model required as we know that the order of our ***GARCH*** model will have a relation with the ARMA model fitted in the squared residuals of ***ARMA(max(p,q),q)*** $\leftrightarrow$ ***GARCH(p,q)*** applying the function ***s.arima()*** to find the corresponding order for the residuals of the two models we have reached to an ***ARMA(3,3)*** for the squared residuals what is equivalent to a ***GARCH(1-3,3)*** so let's compare different models and see which one fits better.

From a first analysis we have checked the ***ARMA (2,0) x GARCH(1-3,3)*** and the ***ARMA (0,0) x GARCH (1-3,3)*** And our best AIC has been for the ***ARMA (2,0) x GARCH (1,3)***. (***Table 6*** in the appendix)

In order to find out if by applying a different distribution to the conditional part of the model we can improve both the AIC and the log-likelihood we've used a t-student distribution for the conditional variance instead of a normal distribution but as the improvement wasn't as high as expected and we had to include the skewness and the shape making the model more complex we decide to remain with the normal distribution approach to model the conditional variance.

2.4 Validation of the variance model

After a further analysis of the significance of the coefficients we have concluded that the ar2 and ar3 term weren't significant enough and therefore we have removed them reaching some more parsimonious models and an improvement in their AIC values with a***AR(2) x GARCH(1,1)*** and a ***GARCH(1,1) x ARMA(0,0)*** with the best AIC values and log-likelihood values, with the best value for the ***AR(2) x GARCH(1,1)*** in particular.

After fitting our models, we've taken a look to our residuals to check if they follow a normal distribution and if we've accomplished to explain all the correlation in them. If we take a look to the ***Figure 8***, we can appreciate how for both models the residuals do not show any relevant correlation and they fall in the expected normal line in the qq-plot.

By taking a look again to the significance of the coefficients from the ***AR(2) x GARCH(1,1)*** we can conclude that all of them are significant enough (p-values much smaller than 0.05) except the ar1 term that we had concluded before that it wasn't actually relevant so we should take ride of it. On the other hand in the ***ARMA(0,0) x GARCH(1,1)*** model all the coefficients are relevant enough for the model(***Table 4 and 5***).

Coefficient	p-value
mu	$4.05e^{-06}$
ar1	0.52886
ar2	0.00154
omega	0.04444
alpha1	0.00327
beta1	$2e^{-16}$

Table 4: Coefficient importance by the Ljung-Box test GARCH(1,1)

Coefficient	p-value
mu	$2.63e^{-05}$
omega	0.04444
alpha1	0.00225
beta1	$2e^{-16}$

Table 5: Coefficient importance by the Ljung-Box test AR(2)xGARCH(1,1)

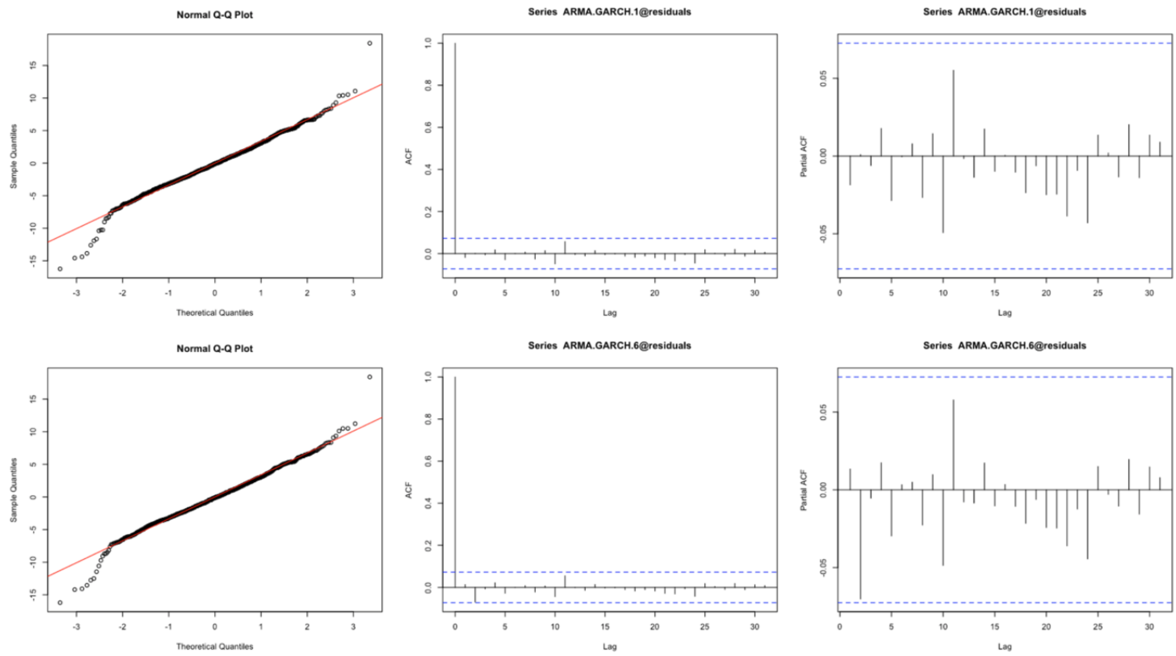


Figure 8: analysis of the residuals of the AR(2)xGARCH(1,1) model in the first line and ARMA(0,0)xGARCH(1,1) in the second one with 99% CI

To confirm these results I've applied for both models a few statistical tests where we can observe (***Table 7 and 8*** in the appendix) that since the Jarque-Bera and the Shapiro-Wilk test have a p-value smaller than 0.05 we can reject the null hypothesis and conclude the residuals don't follow a normal distribution, even our first visual analysis seems to agree with a normal distribution. In terms of correlation of the residuals by the Ljung-Box Test we can conclude that as the p-values are above the 0.05 we cannot reject the null hypothesis that they are uncorrelated with a 95% of probability, the same conclusion follow for the squared residuals by applying the same test. And last after applying the LM Arch test we cannot reject the null hypothesis, and therefore our residuals have no remaining ARCH effect to include in the model.

Finally in order to check if we have accomplished to model the conditional volatility we can also take a look at the ACF of the standardized squared residuals dividing the estimated squared residuals by the estimated conditional variance. As we can observe in the ***Figure 9*** where there's no correlation left to explain.

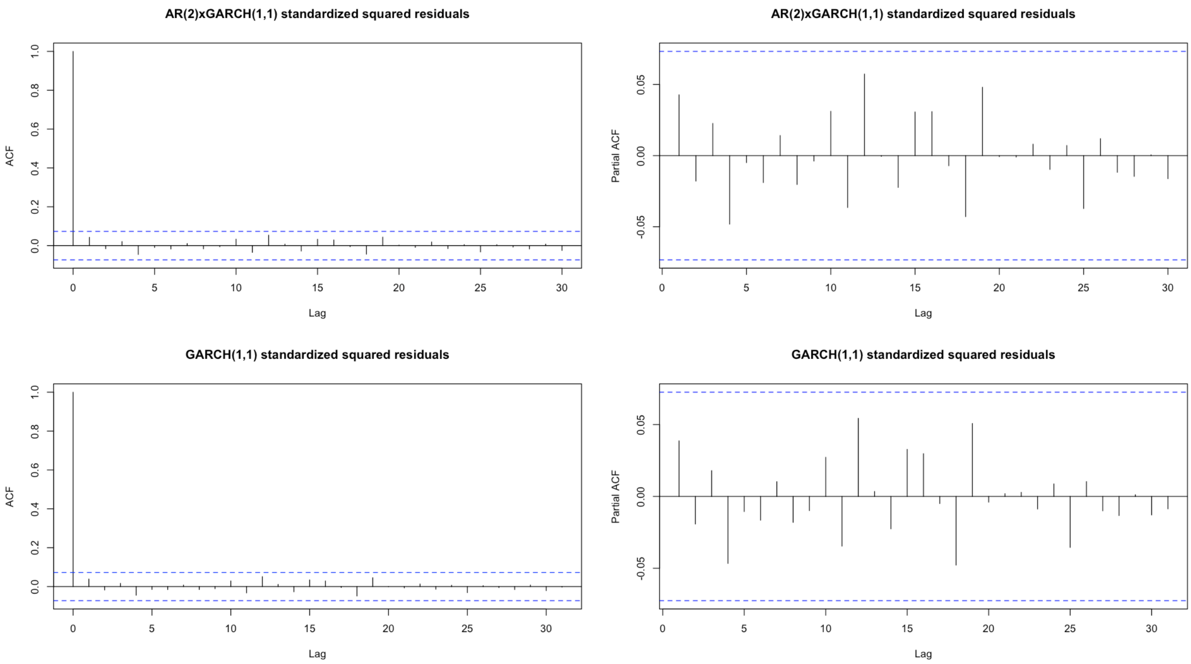


Figure 9: ACF and PACF of the standardized squared residuals with 99% CI

2.5 Final models

(a) AR(2) x GARCH(1,1):

$$A(B)X_t = \varepsilon_t$$

$$A(B)X_t = X_t - 0.476565 + 0.070838X_{t-2}$$

$$\varepsilon_t = GARCH(1,1)$$

$$\varepsilon_t^2 = \sigma_t^2 = 0.496903 + 0.0671\varepsilon_{t-1}^2 + 0.888849\sigma_{t-1}^2$$

(b) GARCH(1,1):

$$X_t - 0.459459 = \varepsilon_t$$

$$\varepsilon_t = GARCH(1,1)$$

$$\varepsilon_t^2 = \sigma_t^2 = 0.4927255 + 0.066943\varepsilon_{t-1}^2 + 0.889133\sigma_{t-1}^2$$

It is interesting to take a look to the coefficients of the GARCH part of the model and see how the sum of both of them is smaller than 1, otherwise the model would have had a non-stable volatility, making it unreliable for prediction. And we can also observe that the coefficients for the GARCH part are quite similar in both models.

3 Forecasting

3.1 In sample forecast

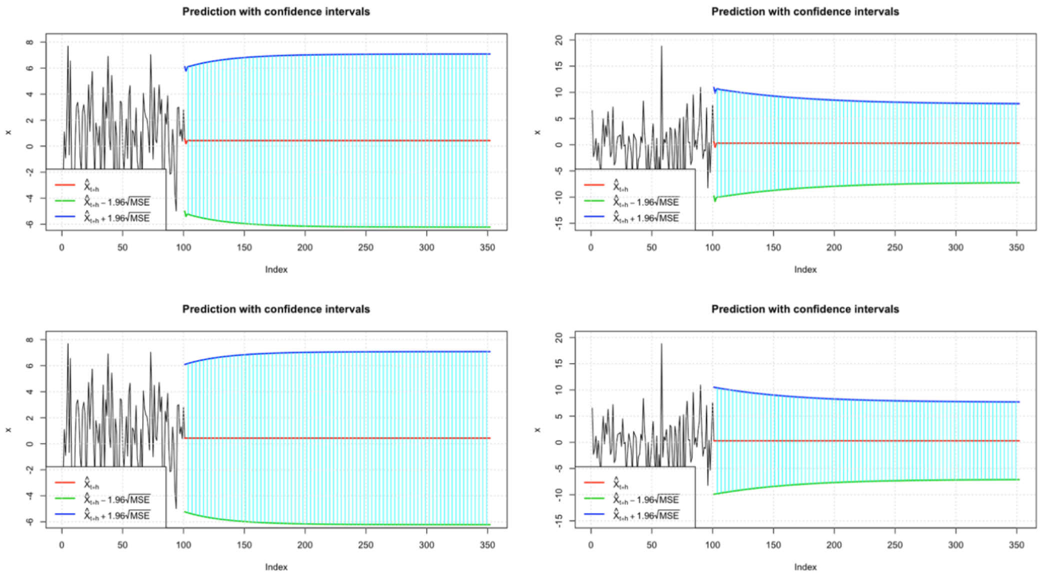


Figure 10: In sample prediction of the series with the AR(2) x GARCH model in the first line and the GARCH model in the second line with 95% CI

In GARCH prediction the mean predicted by the model has no real interest as it is supposed to be around 0, it contains a little bit more of information when we combine it with an AR(2) model but even though it doesn't provide us a reliable prediction. Instead of using the mean it would be much more interesting to take a look to the trajectory of the prediction intervals to make us an idea of the possible change in the volatility in nearly steps.

Thus, analyzing those intervals in the ***Figure 10*** we can appreciate how they are influenced in the proximities of the series by the conditional variance in their surroundings but as soon as they are further away, they start to tend to the unconditional variance being constant over time and providing no information. It is worth noting how depending on the conditional variance at the starting point the volatility has one tendence or another, if it's a small value the volatility tends to increase as we go further in time whereas if we are dealing with a higher conditional variance the volatility is supposed to tend to decrease. We can establish then a relationship between the trajectory of the confidence intervals and the trajectory of the volatility.

We can also appreciate how the AR (2) term of the mean model changes a little bit the behavior of the prediction by adding a small decrease and maybe a more reliable prediction for the initial steps of the prediction. Both cases are quite similar so we cannot establish which model could give us the better performance in terms of prediction. But the addition of an AR (2) to the mean would help to give a better short-term prediction.

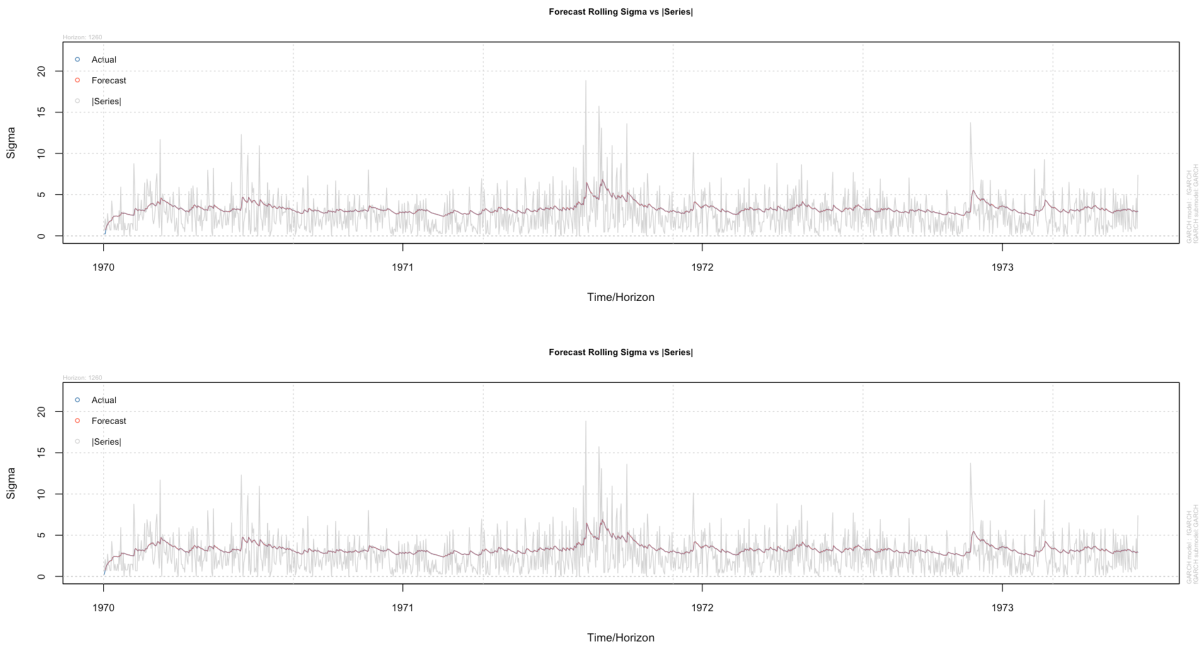


Figure 11: One-step-ahead in-sample prediction of the volatility of the time series,on the first line the AR(2)xGARCH(1,1) and in the second line the pure GARCH(1,1) model

By applying an One-step-ahead in sample prediction and comparing the forecasted volatility with the real one we can see how both models predict almost the same behavior(which is logical as they have almost the same GARCH model to explain the conditional variance) and even they don't allow us to predict the volatility value itself they can give us an idea about the time length of a volatility spike.

3.2 Future forecast

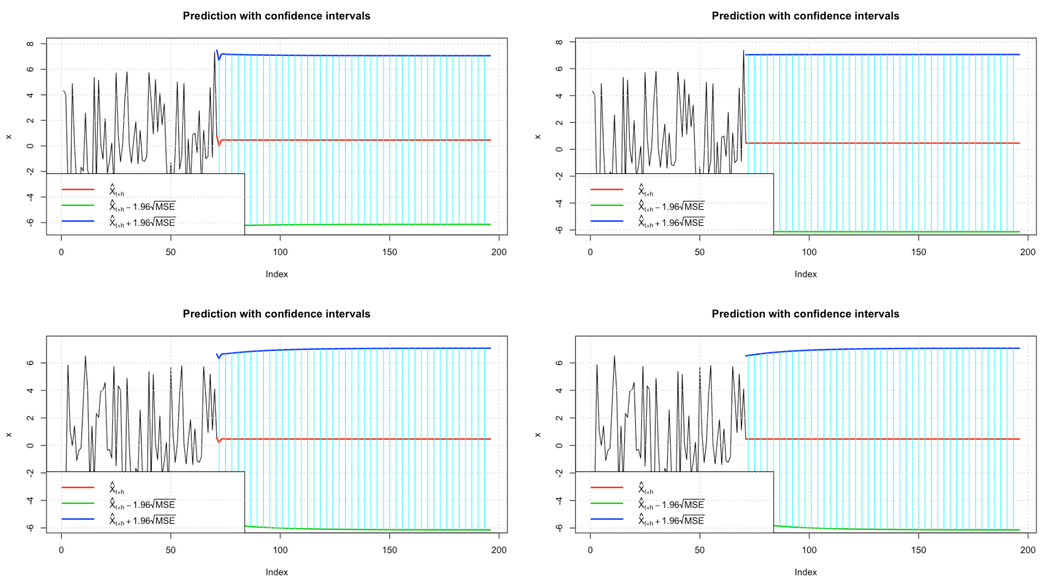
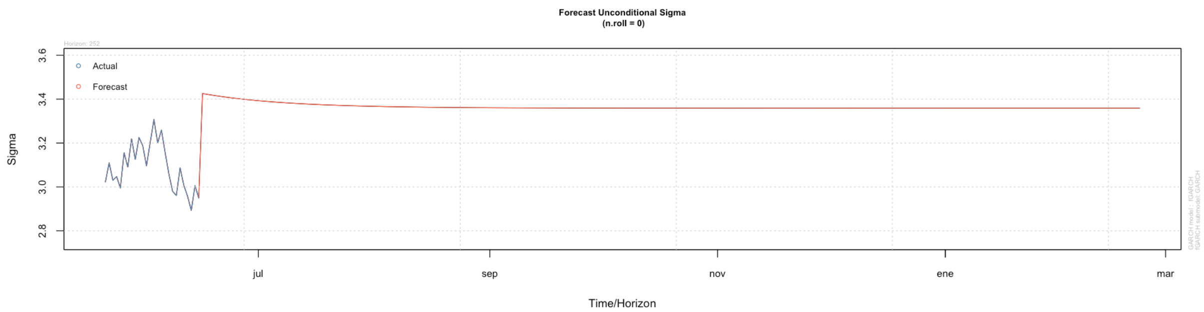


Figure 12: Future prediction of the series ,on the first column the AR(2)xGARCH(1,1) and in the second column the pure GARCH(1,1) model



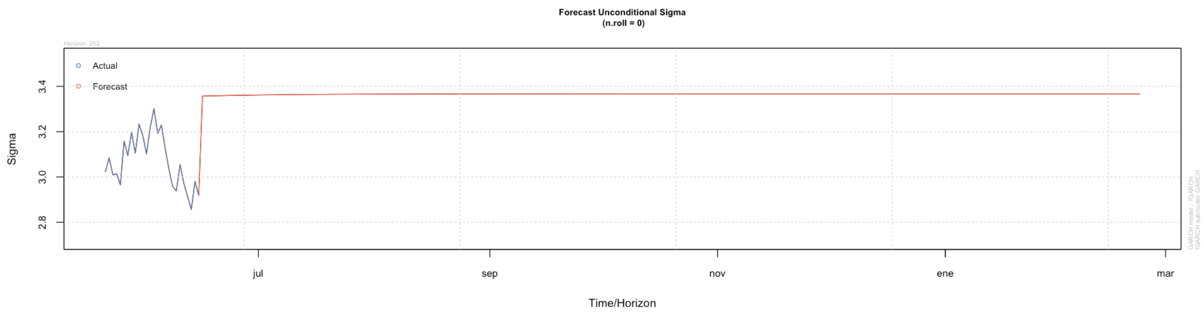


Figure 13: Volatility prediction for a 20 % of the length of the time series,on the first line the AR(2)xGARCH(1,1) and in the second line the pure GARCH(1,1) model

We can observe how our prediction depends on the last point we've selected to use as our time series end like we've mentioned before, giving us an idea of the complexity of the future prediction in financial time series. In the first line of the ***Figure 12*** the prediction intervals remain constant as they start with a value relatively close to the unconditional variance whereas in the second line where we've tried to start a few samples before the end of the series, we can observe how the prediction intervals tend to increase before reaching a plateau what would indicate that the series are probably about to experience a period of increased volatility. We can observe that in the ***Figure 13*** where we've predicted the volatility of the time series, and we can see how after a period of small volatility it is supposed to follow a period with a higher conditional variance.

4 Conclusions

Once we've performed all of the previous steps we must choose the best model for our purposes, due to the nature of both processes, in order to get a more reliable prediction it would be better to include the ***AR(2)*** part in the model, but as we've seen the ar2 term is very small, it has almost no effect in the estimation, and as the improvements of the AIC values are of a really small order (0.003 difference) we should consider the use of the more parsimonious one and rely in the ***GARCH(1,1)*** model as our final choice. This model indeed gives us a good in sample estimation of the volatility behaviour of the series.

5 Appendix

5.1 Other tables and figures

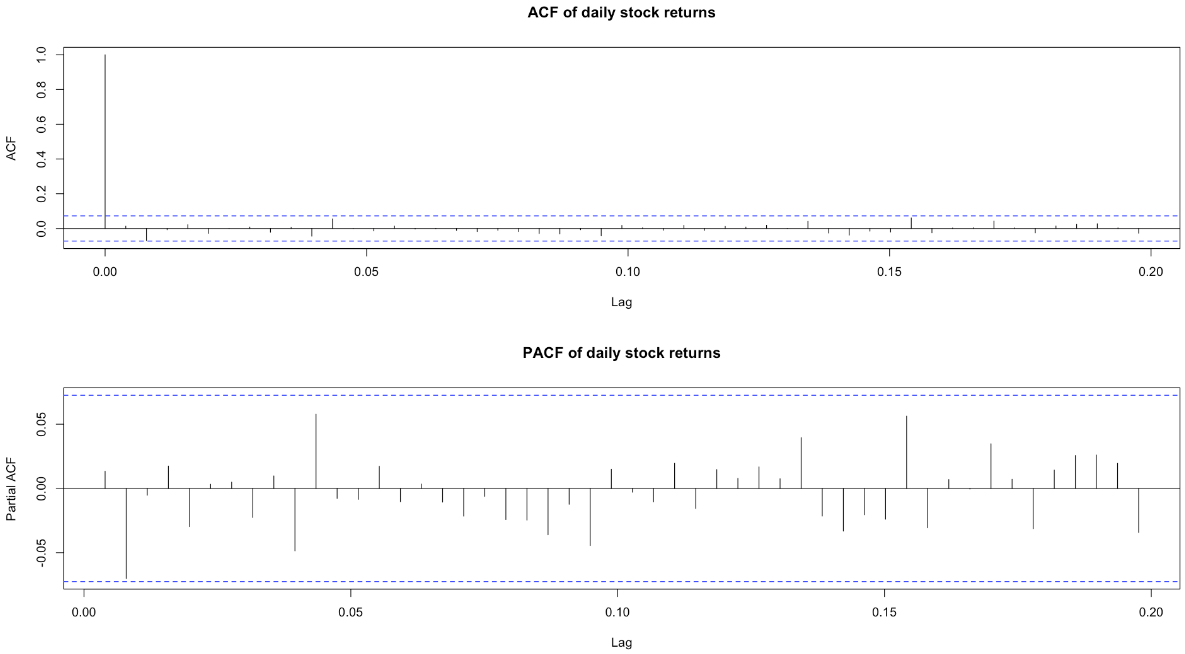


Figure 14: ACF and PACF of the time series with 99% CI

MODEL	AIC	BIC	Log-likelihood	Conditional distribution
arma(2, 0) + garch(1, 1)	5.202606	5.227061	-3274.243	Normal
arma(2, 0) + garch(1, 2)	5.202903	5.231434	-3273.43	Normal
arma(2, 0) + garch(1, 3)	5.204737	5.237344	-3273.587	Normal
arma(0, 0) + garch(1, 1)	5.205572	5.221876	-3278.113	Normal
arma(0, 0) + garch(1, 2)	5.205488	5.225867	-3277.06	Normal
arma(0, 0) + garch(1, 3)	5.207187	5.231642	-3277.131	Normal

Table 6: Comparison of different ARMA x GARCH models

Standardised Residuals Tests	Statistic	p-Value
Jarque-Bera Test R Chi^2	27.58032	$1.025673e^{-06}$
Shapiro-Wilk Test R W	0.994871	0.0002674259
Ljung-Box Test R Q(10)	9.192824	0.513907
Ljung-Box Test R Q(15)	10.6706	0.7755765
Ljung-Box Test R Q(20)	11.79253	0.9230317
Ljung-Box Test R^2 Q(10)	8.463892	0.5836209
Ljung-Box Test R^2 Q(15)	15.78322	0.3966036
Ljung-Box Test R^2 Q(20)	22.88692	0.2943813
LM Arch Test R TR^2	14.26065	0.2843744

Table 7: Standardised residuals tests for ARMA(0,0) x GARCH(1,1)

Standardised Residuals Tests	Statistic	p-Value
Jarque-Bera Test R Chi^2	27.87829	8.83703e^{-07}
Shapiro-Wilk Test R W	0.9951465	0.0004426611
Ljung-Box Test R Q(10)	3.915358	0.9510846
Ljung-Box Test R Q(15)	5.782448	0.9831749
Ljung-Box Test R Q(20)	7.181419	0.9960463
Ljung-Box Test R^2 Q(10)	8.159263	0.6132836
Ljung-Box Test R^2 Q(15)	14.35577	0.4987359
Ljung-Box Test R^2 Q(20)	22.27675	0.3256841
LM Arch Test R TR^2	12.63997	0.395743

Table 8: Standardised residuals tests for ARMA(2,0) x GARCH(1,1)

5.2 Implemented code for the project

```
setwd("/Users/juliopastor/OneDrive - UPV/UCL/Time Series/Group 14-20200415");
#Load the required packages:
require(forecast)
library(tibbletime)
library(dplyr)
require(xts)
require(TSstudio)
require(fGarch)
library(rugarch)
library(tseries)
library(fBasics)
require(aTSA)
source('/Users/juliopastor/OneDrive - UPV/UCL/Time Series/FonctionsSeriesChrono.r')
## needs to be in the working directory
#We create the time series
Stock <- ts(read.table('dell.txt', header = F)$V1); ## or "import Dataset" in Rstudio
basicStats(Stock)# We obtain an aproximation of the mean and the variance

# Stock
# nobs         1261.000000
# NAs             0.000000
# Minimum       -15.741931
# Maximum        18.849372
# 1. Quartile   -1.823277
# 3. Quartile    2.482740
# Mean           0.365698
# Median         0.365047
# Sum          461.144737
# SE Mean        0.094782
# LCL Mean       0.179749
# UCL Mean       0.551646
# Variance      11.328433
# Stdev          3.365774
# Skewness      -0.052559
# Kurtosis       1.715090  we can see that the distribution would be
```

```
#more centered than an standard normal one


shapiro.test(Stock)
# Shapiro-Wilk normality test
#
# data:  Stock
# W = 0.98584, p-value = 9.801e-10 as the p value of the test is <<<
#than p=0.05 thus we night reject  the null hypothesis that the series is based on a normal distribution

daily <- (ts(Stock, start = c(1993,304*253/356), frequency = 253))
# we create a time series starting from 1993 and with trading year frequency
# We use 253 as they are the trading days per year
daily.squared=daily^2;
# As we are dealing with highly volatile finantial data it is important to take a look to the squared ts
#Plot both:
par(mfrow=c(1,1))
plot(daily,main="Daily stock returns ") ## plot time series which is actually a previous log
plot(daily.squared,main='Square of the daily stock returns')
#Now we must check if any kind of ARMA model fits for the daily stock returns

#Check if any of them has a mean tendence:
par(mfrow=c(1,1))
plot(daily,main="Daily stock returns ") ## plot time series
t.daily <- seq(1993 + 304/365, by = 1/253, length = 1261)
abline(reg=lm(daily~t.daily),col=2)

#Check for stationary conditions
par(mfrow=c(2,1))
plot(daily[1:253], type="l", ylim=c(min(daily), max(daily)),main="Comparison of the daily stock returns")
# we plot every year one compared to each other so that we can establish their trends
for(i in 2:5){
  lines(daily[253*(i-1)+(1:253)],col=i)
}
legend("bottomright", legend = 1994:1999, lty = 1, col = 1:12, cex = 0.5)

plot(daily.squared[1:253], type="l", ylim=c(min(daily.squared), max(daily.squared)),main="Comparison of the dai
# we plot every year one compared to each other so that we can establish their trends
for(i in 2:5){
  lines(daily.squared[253*(i-1)+(1:253)],col=i)
}
legend("bottomright", legend = 1994:1999, lty = 1, col = 1:12, cex = 0.5)

#Check correlations for a white noise:
?acf
par(mfrow=c(2,1))
acf=acf(daily, lag = 50, main = 'ACF of daily stock returns')
par(new=TRUE)
plot(acf,ci=0.99)

pacf=pacf(daily, lag = 50, main = 'PACF of daily stock returns')
par(new=TRUE)
plot(pacf,ci=0.99)
```

```
# Looks like there's some partial autocorrelation for the lag 2
# Fit an ARIMA model:
Model_Select=expand.grid(order_p=0:4,order_q =0:4,Aic=0,Bic=0)
n=1
while(n<length(Model_Select$order_p)+1){
  suppressWarnings( model <- arima(daily,order=c(Model_Select$order_p[n],0,Model_Select$order_q[n])))
  Model_Select$Aic[n] <- Aic.arima(daily, model ) ## AIC order i
  Model_Select$Bic[n] <- Bic.arima(daily, model) ## BIC order i
  n=n+1
}
suppressWarnings(Comp.Sarima(daily, d=0, saison=0, D=0, p.max=4, q.max=4, P.max=0, Q.max=0))
Best_Aic=which.min(Model_Select$Aic)
Best_Bic=which.min(Model_Select$Bic)
modelAIC <- arima(daily,order=c(Model_Select$order_p[Best_Aic] ,0,Model_Select$order_q[Best_Aic] ))
modelBIC <- arima(daily,order=c(Model_Select$order_p[Best_Bic] ,0,Model_Select$order_q[Best_Bic] ))
modelAr1<- arima(daily,order=c(1,0,0))
#Check the importance of the coefficients
coef.BIC<-coef.p(modelBIC$coef, diag(modelBIC$var.coef))
coef.AIC<-coef.p(modelAIC$coef, diag(modelAIC$var.coef))
coef.ar1<-coef.p(modelAr1$coef,diag(modelAr1$var.coef))
coef.BIC
coef.AIC
coef.ar1
# Check the residuals as white noise
tsdiag(modelBIC,main='BIC')
tsdiag(modelAIC,main='AIC')
par(mfrow=c(2,1))
acf(modelAIC$res, lag = 50, main = 'ACF of residuals model AIC')
pacf(modelAIC$res, lag = 50, main = 'PACF of residuals model AIC')
par(mfrow=c(2,1))
acf(modelBIC$res, lag = 50, main = 'ACF of residuals model BIC')
pacf(modelBIC$res, lag = 50, main = 'PACF of residuals model BIC')

qqnorm(modelAIC$res) ## check for normality of residuals with a QQplot
abline(0, 23/7, col = 2)
qqnorm(modelBIC$res) ## check for normality of residuals with a QQplot
abline(0, 23/7, col = 2)

?tsdiag()
#Check the squared residuals as white noise
par(mfrow=c(2,1))
acf((modelAIC$res)^2, lag = 50, main = 'ACF of squared residuals model AIC')
pacf((modelAIC$res)^2, lag = 50, main = 'PACF of  squared residuals model AIC')
par(mfrow=c(2,1))
acf((modelBIC$res)^2, lag = 50, main = 'ACF of squared residuals model BIC')
pacf((modelBIC$res)^2, lag = 50, main = 'PACF of  squared residuals model BIC')
par(mfrow=c(1,2))


#Fit a (G)ARCH model to our two previous models:
suppressWarnings(Comp.Sarima(modelAIC$residuals^2, d=0, saison=0, D=0, p.max=4, q.max=4, P.max=0, Q.max=0))
```

```
# mod?le (p,d,q)x(P,D,Q)_saison :  3 0 3 x 0 0 0 _ 0 :  nb param:  6     AIC: 0
# mod?le (p,d,q)x(P,D,Q)_saison :  3 0 4 x 0 0 0 _ 0 :  nb param:  7     AIC: 10.59227
# mod?le (p,d,q)x(P,D,Q)_saison :  4 0 4 x 0 0 0 _ 0 :  nb param:  8     AIC: 3.783044
suppressWarnings(Comp.Sarima(modelBIC$residuals^2, d=0, saison=0, D=0, p.max=4, q.max=4, P.max=0, Q.max=0))
# mod?le (p,d,q)x(P,D,Q)_saison :  2 0 4 x 0 0 0 _ 0 :  nb param:  6     AIC: 12.67532
# mod?le (p,d,q)x(P,D,Q)_saison :  3 0 3 x 0 0 0 _ 0 :  nb param:  6     AIC: 0
# mod?le (p,d,q)x(P,D,Q)_saison :  4 0 4 x 0 0 0 _ 0 :  nb param:  8     AIC: 5.767767
#Thus we will use an ARMA(3,3) to model the residuals which implies and GARCH(1-3,3)
n=1
#With a simple normal distribution
ARMA.GARCH.1 <- garchFit(data ~ arma(2,0) + garch(1,1), data = daily, trace = F)# We have reduced the order due
capture.output(summary(ARMA.GARCH.1)) ## print only the interesting part// best one according to AIC and BIC
ARMA.GARCH.2 <- garchFit(data ~ arma(2,0) + garch(1,2), data = daily, trace = F)
 capture.output(summary(ARMA.GARCH.2)) ## print only the interesting part
ARMA.GARCH.3 <- garchFit(data ~ arma(2,0) + garch(1,3), data = daily, trace = F)
 capture.output(summary(ARMA.GARCH.3)) ## print only the interesting part

 ARMA.GARCH.4 <- garchFit(data ~ arma(0,0) + garch(1,1), data = daily, trace = F)
 capture.output(summary(ARMA.GARCH.4)) ## print only the interesting part
 ARMA.GARCH.5 <- garchFit(data ~ arma(0,0) + garch(1,2), data = daily, trace = F)
 capture.output(summary(ARMA.GARCH.5)) ## print only the interesting part
 ARMA.GARCH.6 <- garchFit(data ~ arma(0,0) + garch(1,3), data = daily, trace = F)
 capture.output(summary(ARMA.GARCH.6)) ## print only the interesting part

 #We apply a conditional t-student distribution
  ?garchFit
 ARMA.GARCH.1 <- garchFit(data ~ arma(2,0) + garch(1,1), data = daily, trace = F,include.skew = FALSE,include.s
 capture.output(summary(ARMA.GARCH.1)) ## print only the interesting part// best one according to AIC and BIC
 ARMA.GARCH.2 <- garchFit(data ~ arma(2,0) + garch(1,2), data = daily, trace = F,include.skew = FALSE,include.s
 capture.output(summary(ARMA.GARCH.2)) ## print only the interesting part
 ARMA.GARCH.3 <- garchFit(data ~ arma(2,0) + garch(1,3), data = daily, trace = F,include.skew = FALSE,include.s
 capture.output(summary(ARMA.GARCH.3)) ## print only the interesting part

 ARMA.GARCH.4 <- garchFit(data ~ arma(0,0) + garch(1,1), data = daily, trace = F,include.skew = FALSE,include.s
 capture.output(summary(ARMA.GARCH.4)) ## print only the interesting part
 ARMA.GARCH.5 <- garchFit(data ~ arma(0,0) + garch(1,2), data = daily, trace = F,include.skew = FALSE,include.s
 capture.output(summary(ARMA.GARCH.5)) ## print only the interesting part
 ARMA.GARCH.6 <- garchFit(data ~ arma(0,0) + garch(1,3), data = daily, trace = F,include.skew = FALSE,include.s
 capture.output(summary(ARMA.GARCH.6)) ## print only the interesting part


 plot((abs(daily)))
 1#Interesting till here # Best model AIC-log-likelihood: ARMA.GARCH1; if we want the most parsimonious one GAR
 # Now we must check the residuals
par(mfrow=c(2,1))
acf(ARMA.GARCH.1@residuals,main='AR(2)xGARCH(1,1) residuals',ci=0.99)
pacf(ARMA.GARCH.1@residuals,main='AR(2)xGARCH(1,1) residuals',ci=0.99)

par(mfrow=c(2,1))
acf(ARMA.GARCH.4@residuals, main='GARCH(1,1) residuals',ci=0.99)
pacf(ARMA.GARCH.4@residuals,main='GARCH(1,1) residuals',ci=0.99)

# now we check for the standardized residuals
```

```

par(mfrow=c(3,5))
acf(ARMA.GARCH.1@residuals^2/ARMA.GARCH.1@h.t,main='AR(2)xGARCH(1,1) standardized squared residuals',ci=0.99)
pacf(ARMA.GARCH.1@residuals^2/ARMA.GARCH.1@h.t,main='AR(2)xGARCH(1,1) standardized squared residuals',ci=0.99)

plot(ARMA.GARCH.1,which='all')


acf(ARMA.GARCH.4@residuals^2/ARMA.GARCH.1@h.t, main='GARCH(1,1) standardized squared residuals',ci=0.99)
pacf(ARMA.GARCH.4@residuals^2/ARMA.GARCH.1@h.t,main='GARCH(1,1) standardized squared residuals',ci=0.99)

 #Comparison of their prediction behaviour

par(mfrow=c(2,1))
model<-ugarchspec(variance.model = list(model = "fGARCH",submodel='GARCH', garchOrder = c(1, 1)),
                  mean.model = list(armaOrder = c(2, 0), include.mean = TRUE), distribution.model = "sstd")
modelfit<-ugarchfit(spec=model,data=daily)

spec = getspec(modelfit);
setfixed(spec) <- as.list(coef(modelfit));
length(daily)
forecast = ugarchforecast(spec, n.ahead = 1, n.roll = 1260, data = daily[1:1261], out.sample = 1260);

plot(forecast,which=4)
model<-ugarchspec(variance.model = list(model = "fGARCH",submodel='GARCH', garchOrder = c(1, 1)),
                  mean.model = list(armaOrder = c(0, 0), include.mean = TRUE), distribution.model = "sstd")
modelfit<-ugarchfit(spec=model,data=daily)

spec = getspec(modelfit);
setfixed(spec) <- as.list(coef(modelfit));
length(daily)
forecast = ugarchforecast(spec, n.ahead = 1, n.roll = 1260, data = daily[1:1261,drop=FALSE], out.sample = 1260)

plot(forecast,which=4)

#Future prediction:
ARMA.GARCH.1 <- garchFit(data ~ arma(2,0) + garch(1,1), data = daily[1:round(1*length(daily))], trace = F)# We
par(mfrow=c(2,2))
pred1=fGarch::predict(ARMA.GARCH.1,n.ahead=round(0.1*length(daily)),plot=T,nx=70)
ARMA.GARCH.1 <- garchFit(data ~ arma(0,0) + garch(1,1), data = daily[1:round(1*length(daily))], trace = F)# We
pred2=fGarch::predict(ARMA.GARCH.1,n.ahead=round(0.1*length(daily)),nx=70,plot=T)

ARMA.GARCH.1 <- garchFit(data ~ arma(2,0) + garch(1,1), data = daily[1:round(0.98*length(daily))], trace = F)#

pred1=fGarch::predict(ARMA.GARCH.1,n.ahead=round(0.1*length(daily)),plot=T,nx=70)
ARMA.GARCH.1 <- garchFit(data ~ arma(0,0) + garch(1,1), data = daily[1:round(0.98*length(daily))], trace = F)#
pred2=fGarch::predict(ARMA.GARCH.1,n.ahead=round(0.1*length(daily)),nx=70,plot=T)




#Volatility prediction

par(mfrow=c(2,1))
```

```
model<-ugarchspec(variance.model = list(model = "fGARCH",submodel='GARCH', garchOrder = c(1, 1)),
                  mean.model = list(armaOrder = c(2, 0), include.mean = TRUE), distribution.model = "norm")
modelfit<-ugarchfit(spec=model,data=daily)

spec = getspec(modelfit);
setfixed(spec) <- as.list(coef(modelfit));
length(daily)
forecast = ugarchforecast(spec, n.ahead = round(0.2*length(daily)),data = daily[1:1261]);

plot(forecast,which=3)
model<-ugarchspec(variance.model = list(model = "fGARCH",submodel='GARCH', garchOrder = c(1, 1)),
                  mean.model = list(armaOrder = c(0, 0), include.mean = TRUE), distribution.model = "norm")
modelfit<-ugarchfit(spec=model,data=daily)

spec = getspec(modelfit);
setfixed(spec) <- as.list(coef(modelfit));
length(daily)
forecast = ugarchforecast(spec, n.ahead = round(0.2*length(daily)), data = daily[1:1261,drop=FALSE]);

plot(forecast,which=3)
```