LSE Summer School Report

Andrea Ranzato | id. 4709494

andrea.ranzato01@icatt.it

Summary

The general purpose of this project is to use techniques learned during the course in *Statistical methods in risk management* useful to analyze stock price data. In particular, we consider prices concerning ten different stocks recorded during the 2007-2009 financial crisis. The data source is the *Wharton Research Data Services.*

The analysis is structured as follows. In the first section, I show how two extreme events, the rescue of Bear Sterns on March 14 2008 and Lehman Brothers bankruptcy on September 15 2008, affected the price and log-return dynamics. Next, I test both the marginal and joint normality of the ten stock returns by means of the *Jarque-Bera* and *Mardia test* respectively. First, we observe that the normal distribution fails to capture returns happening in the tails. Second, joint normality is not detected. Subsequently, I evaluate some pair-wise stock returns, either from companies of the same sector and different ones, in order to quantify their correlation.

In the second section, I focus on financials stocks with the objective of assessing their dependence using copulas of different types. We see, that the copula which fits the observations best is the t-Copula.

In the third section, I consider a linear portfolio which invest \$1000 in each of the stocks at the beginning of the period, assuming that its composition does not change during the entire period. Afterwards, I evaluate the Value-at-Risk of the portfolio at 95% level using two different strategies. Firstly, I compute the VaR using the empirical profit and loss distribution, secondly using a normal model. The comparison between the two shows that the parametric approach would have led to more conservative capital management decisions compared to empirical one, since it signals higher potential losses. In the end, I add to the analysis 2010 portfolio returns, in order to record the number of violations occurred under the two different frameworks. We see that the normal model incurs in fewer violations with respect to the the historical one.

The last section is devoted to the description of the topics covered during the course.

To conclude, I attach in the appendix the R code used to carry out the analysis.

1 Prices and log returns dynamics

The purpose of this section consists in showing some empirical properties of financial time series using a sample of US assets. First of all, I conduct an exploratory analysis of the price dynamics of equities in Table 1 between 3 January 2007 and 31 December 2009. As we can see from Figure 1, a strong negative trend plays out in correspondence with Bear Stearns rescue on March 14 2008 and terminates around January 20 2009. Table 2 quantifies the percentage price change recorded in that period summarizing the worst downturn of the US financial system since the Great Depression.

Table 1: Composition of the equity portfolio

Company	**Sector**	**Ticker**
Amazon.com, Inc.	Consumer	AMZN
American International Group	Financials	AIG
Bank of America Corp.	Financials	BA
Coca-Cola Company	Consumer	KO
ExxonMobil Corp.	Energy	XOM
Goldman Sachs Group, Inc.	Financials	GS
Microsoft Corp.	IT	MSFT
Morgan Stanley	Financials	MS
Northern Trust Corp.	Financials	NTRS
Wells Fargo & Co.	Financials	WFC

Then, I consider the log returns illustrated in Figure 2 where we can appreciate the presence of the following *stylized facts*, as reported in **embrechts**. (1) conditional expected returns are close to zero; (2) volatility appears to vary over; (3) extreme returns appear in clusters. In particular, within our context, higher volatility is registered afterwards the bankruptcy of Lehman Brothers highlighted in red.

We model log-returns instead of plain prices, in order to accommodate the assumptions underlying the lognormal model, namely that, following the formulation of **ruppert**, $r_1 = log(1 + R_1)$, $r_2 = log(1 + R_2), \ldots, r_n = log(1 + R_n)$ are i.i.d. $N(\mu, \sigma^2)$, where $1+R_t = \frac{P_t}{P_{t-1}}$, for $t = 1, \ldots, n$. Using this model allows easier computations of multiperiod

Table 2: Percentage price change between 14 Mar. 2008 and 20 Jan. 2009

Company	Prices ($)		(%) Change
	14 Mar. 2008	20 Jan. 2009	
AMZN	38.70	48.44	25.17
AIG	72.15	1.37	-98.10
BA	89.17	40.36	-54.74
KO	48.58	42.88	-11.73
XOM	74.11	76.29	2.94
GS	200.72	59.20	-70.51
MSFT	29.86	18.48	-38.11
MS	81.62	13.10	-83.95
NTRS	61.31	43.93	-28.35
WFC	35.74	14.23	-60.18

Figure 1: Stock Price dynamics

returns since

$$
\begin{aligned}
1 + R_t(k) &= (1 + R_t) \cdots (1 + R_{t-k+1}) \\
&= exp(r_t) \cdots exp(r_{t-k+1}) \\
&= exp(r_t + \ldots + r_{t-k+1})
\end{aligned}
$$

where $1 + R_t(k)$ indicates the gross return over the most recent k periods.

As a result,

$$
log(1 + R_t(k)) = r_t + \cdots + r_{t-k+1}
$$

Hence, since each r_t is normal, their sum is normal too.

Moreover, log returns represent a close approximation of the net returns $\frac{P_t - P_{t-1}}{P_t}$, unless $1 + R_t$ is small.

The exploratory analysis follows by testing both marginal and *joint* normality of the log returns. First, the nonnnormality of the marginal distribution is assessed by looking at the *QQ Plots* in Figure 4, which compares theoretical quantiles one would observe in case of normality, with those of the actual sample observations. In all cases, we observe that the normal model provide a faithful description of the returns only with respect to observations close to mean. Instead, observations far away from the mean, the so-called *tails*, are under-represented since the normal model does not set enough probability mass on observations away 2σ from the mean. The results from *Jarque-Bera* test in Table 3 aids the naked eye to confirm the initial beliefs by rejecting the null hypothesis of normality. In addition, I provide the Skewness and Kurtosis Measurements to reinforce the idea that the marginal empirical distributions are not normal (see Table 4). In general, distributions with higher kurtosis and *fat-tails* might provide a better representation of the returns.

From a practical point of view, these results questions the reliability of the Normal Value-at-Risk model to identify extreme negative returns.

Since most of the times the interest is devoted in understanding the behaviour of returns considered simultaneously rather than individually, we test also the normality of the ten-dimensional random vector $X = (X_{AMZN}, X_{AIG}, \ldots, X_{WFC})$. The test chosen

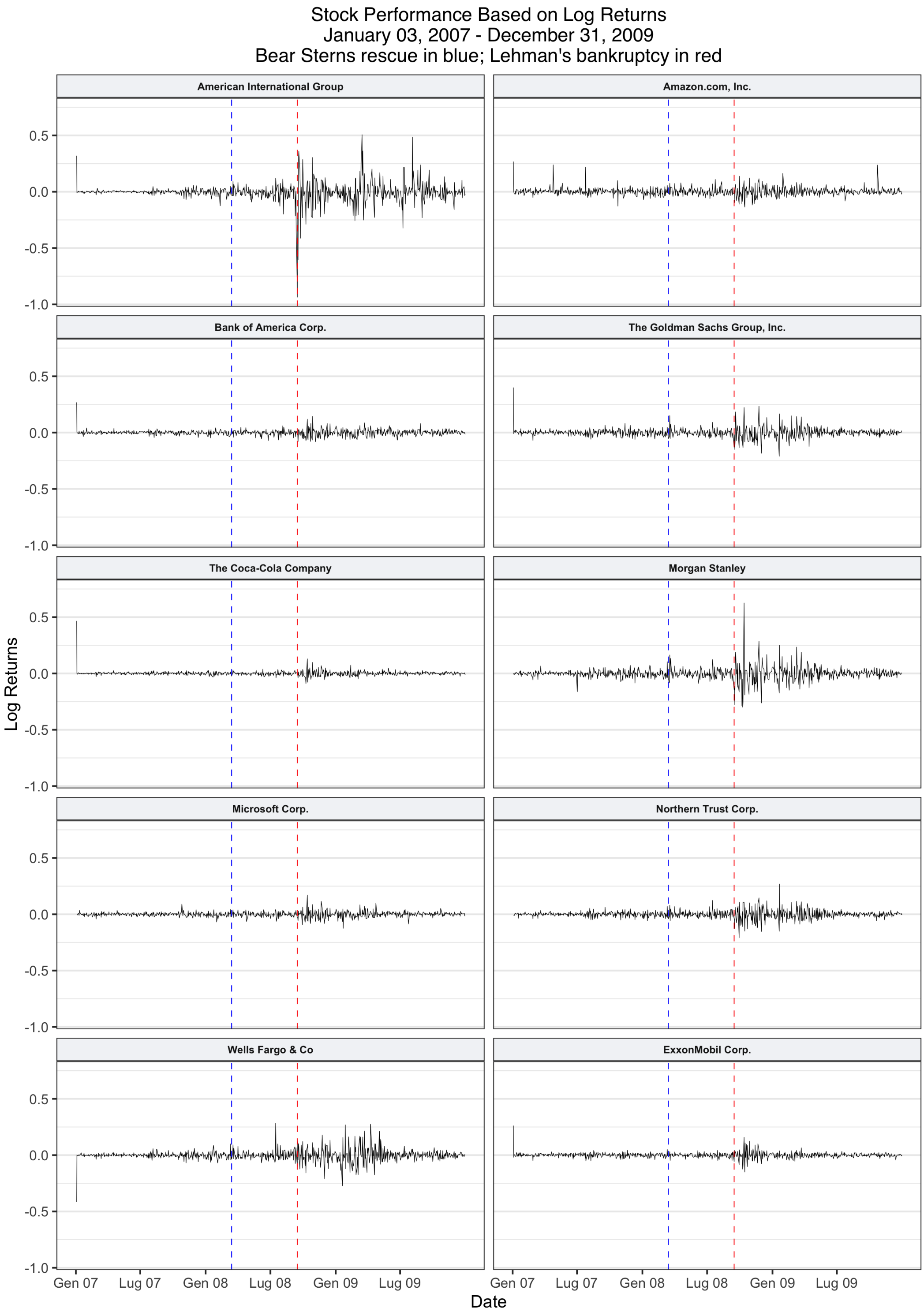


Figure 2: Log returns dynamics

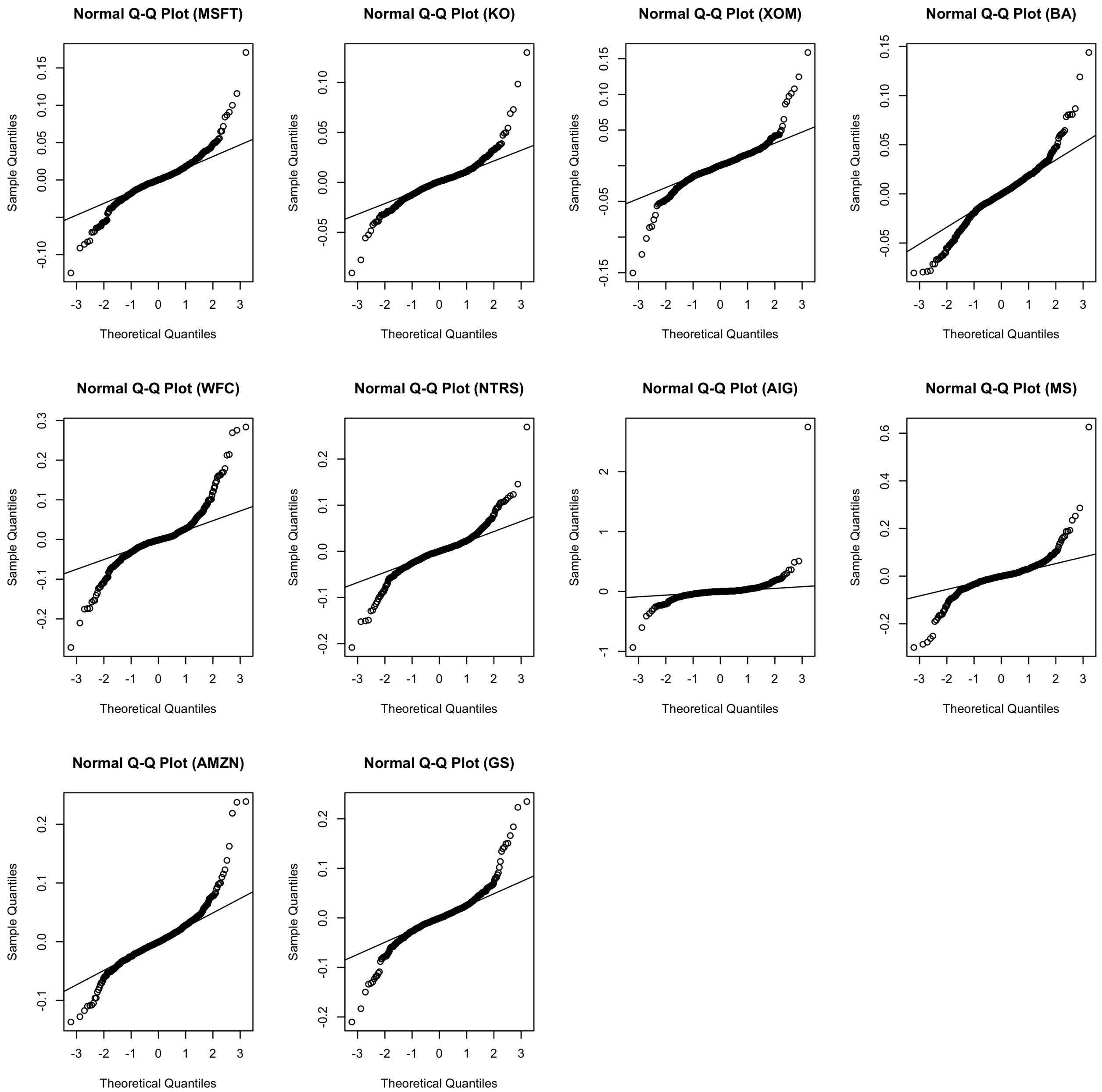


Figure 3: Normal QQ Plot

Table 3: Jarque Bera Test, n. obs. = 756

Stock	**Statistic**	**P-value**	**Normal**
AMZN	2531.4	< 2.2e-16	No
AIG	2102833	< 2.2e-16	No
BA	441.26	< 2.2e-16	No
KO	3459.3	< 2.2e-16	No
XOM	3296	< 2.2e-16	No
GS	1927	< 2.2e-16	No
MSFT	1457.5	< 2.2e-16	No
MS	20344	< 2.2e-16	No
NTRS	2104.5	< 2.2e-16	No
WFC	1971.4	< 2.2e-16	No

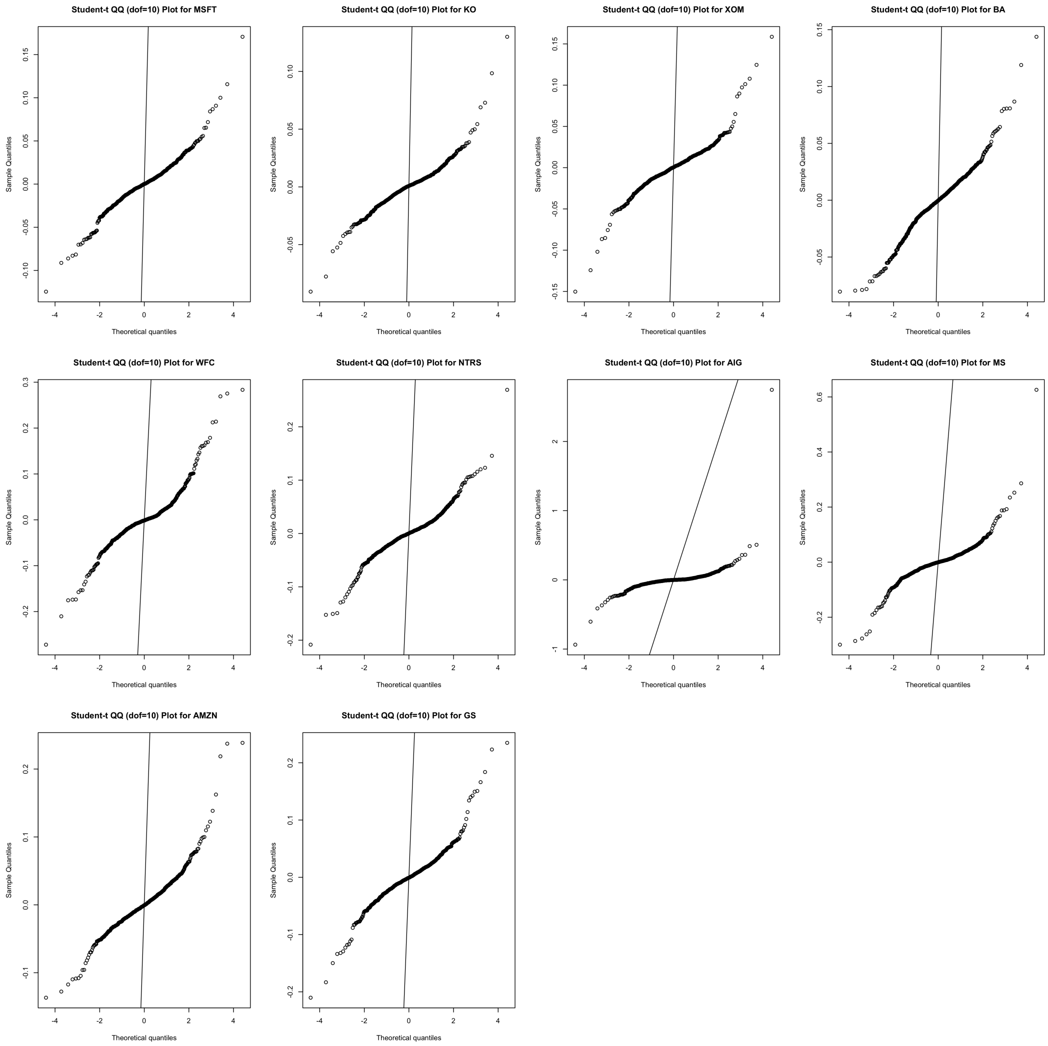


Figure 4: Student-t QQ Plot (dof=10)

Table 4: Log Returns Skewness and Kurtosis, n. obs. = 756

Stock	**Skewness**	**Kurtosis**
AMZN	1.21	8.61
AIG	11.63	256.80
BA	0.25	3.69
KO	0.75	10.34
XOM	0.14	10.20
GS	0.43	7.75
MSFT	0.35	6.74
MS	1.40	25.20
NTRS	0.18	8.14
WFC	0.67	7.77

to assess multivariate joint normality is the *Mardia test* which compares multivariate measures of skewness and kurtosis on the basis of the *Mahalanobis distance*. The results in Table 5 and Figure 5 suggest strong evidence against the non-normality hypothesis.

Table 5: Mardia test for Multivariate Normality

Test	**Statistic**	**P-value**
Skewness	28530.78	0.00
Kurtosis	462.25	0.00

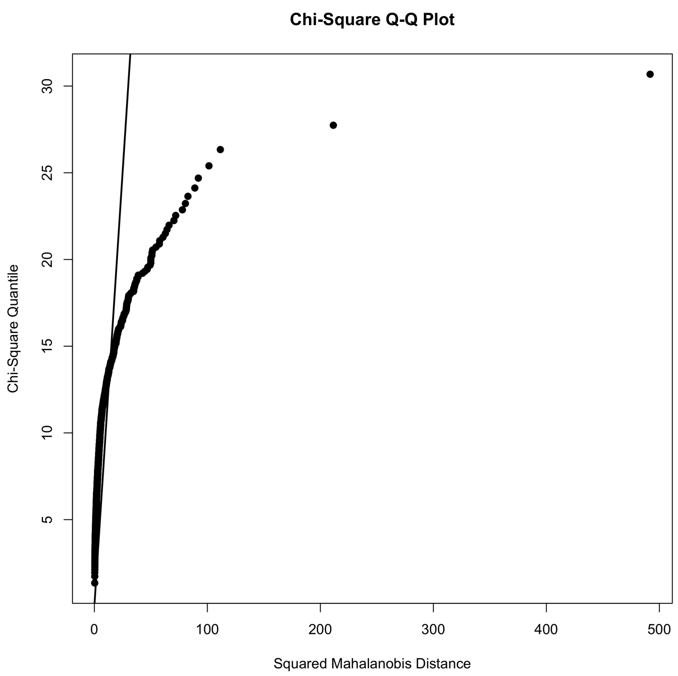


Figure 5: Chi-Square QQ Plot for stock returns

In this section, we have become aware that risk models based on the normal distribution are not satisfactory since they might underestimate the risk of suffering large losses. **embrechts** summarizes the following pitfalls:

1. The tails of its univariate marginal distributions are too thin; they do not assign enough weight to *extreme* events.
2. The joint tails of the distribution do not assign enough weight to *joint extreme* outcomes.
3. The distribution has a strong form of symmetry, known as elliptical asymmetry.

2 Copulas

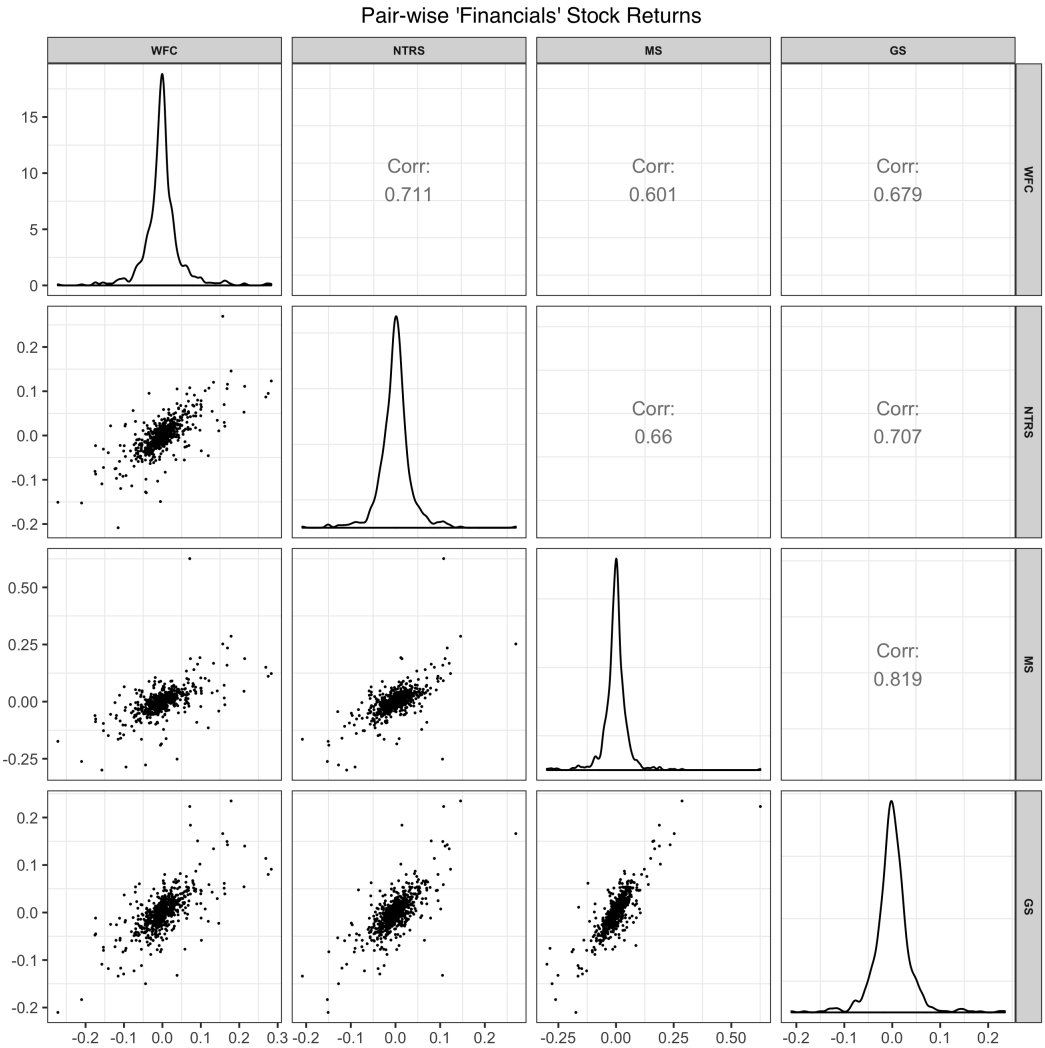


Figure 6: Pair-wise Financials stock log returns

In this section, I begin with considering pair-wise returns of stocks belonging to the financial sector, in order to investigate their dependence structure. Figure 6 represent the empirical marginal distribution, as well as pairs of scatterplots between returns. In addition, the upper off-diagonal quadrants provides a measure of linear correlation between stocks. We can see that there exists a decent amount of linear correlation between pairs stocks. However, we know that this measure is reliable only when the underlying data generating mechanism is ascribable to an elliptical distribution such as the multivariate Normal. For this reason, we assess visually if the data presents the features of elliptical distributions. Instead, we spot within our data: (1) the presence of deviations from ellipsoids of constant density; (2) some *long tails*, especially between `GS` and `MS`; (3) potential presence of non linear behaviours; (4) high kurtosis of the marginals. These feelings are confirmed by the Mardia test on all the combinations of financials stocks which confirms that bivariate distributions between stocks are not Normal (see, Figure 7 which contains a selection of combinations). As a consequence, the Pearson correlation coefficient is a good

```
$WFC_NTRS
             Test          Statistic               p value Result
1 Mardia Skewness 95.4436027366937 9.17082352023949e-20     NO
2 Mardia Kurtosis 76.1676075471859                    0     NO
3             MVN             <NA>                 <NA>     NO

$WFC_MS
             Test          Statistic                p value Result
1 Mardia Skewness 566.853418608482 2.30841826476769e-121     NO
2 Mardia Kurtosis 180.160826706452                     0     NO
3             MVN             <NA>                  <NA>     NO

$WFC_GS
             Test          Statistic               p value Result
1 Mardia Skewness 133.460773521553 7.08181879454237e-28     NO
2 Mardia Kurtosis  68.198135849846                    0     NO
3             MVN             <NA>                 <NA>     NO

$NTRS_MS
             Test          Statistic               p value Result
1 Mardia Skewness 396.375308296942 1.68833194900181e-84     NO
2 Mardia Kurtosis 181.751157414448                    0     NO
3             MVN             <NA>                 <NA>     NO

$NTRS_GS
             Test          Statistic               p value Result
1 Mardia Skewness 59.3714556040696 3.93177496103147e-12     NO
2 Mardia Kurtosis 80.4564559164172                    0     NO
3             MVN             <NA>                 <NA>     NO

$MS_GS
             Test          Statistic               p value Result
1 Mardia Skewness 380.493956079174 4.55388249361014e-81     NO
2 Mardia Kurtosis 159.791721710244                    0     NO
3             MVN             <NA>                 <NA>     NO
```

Figure 7: Mardia test for combinations of financial stocks

measure of association, only when the returns have an elliptical distribution, because it does not takes into account non linearities which might be well present in empirical data. If this is not the case, we should rely on concordance measure such as the *Spearman's rank correlation* and *Kendall's Tau*. For example, given two bivariate distributions, one with stronger lower tail dependence than the other, it might happen that the correlation coefficient is the same, even if one of the two should be perceived as more riskier than the other. Hence, once again, assuming i.i.d. returns from elliptical distributions in the construction of risk models, and relying on their measure of linear correlation, might not represent a reliable representation of the phenomenon. For this reason, we need risk measures able to provide a description of the returns which go beyond the description "when the returns of stock A fall, returns of stock B fall too". Having statistical tools able to account for these behaviours is relevant for risk management purposes.

Copulas are joint distributions defined on the hyper-unit cube with the following characteristics:

1. Provide way of isolating the description of the dependence structure.
2. Express dependence on a *quantile scale*.
3. Facilitate a bottom-up approach to multivariate model building.

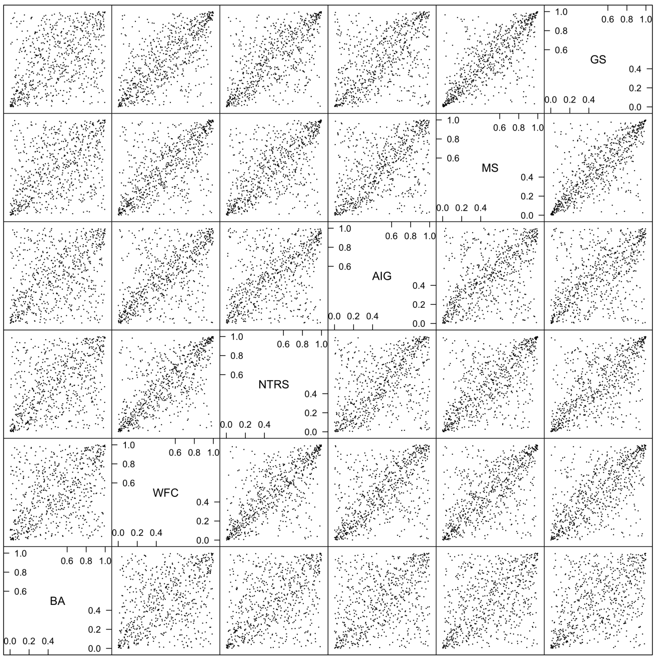
As any other joint distributions copulas have density function and cumulative distribution function. Generally speaking, copulas are fitted to the data as follows: firstly, the so called margins (i.e. the stand alone marginal distributions) are specified, namely they are known distributions. If this is the case, we proceed by estimating the parameters of interest from empirical data using MLE. Then, we apply the corresponding probability integral transform to each margin to obtain Uniforms margins. This is possible since the parameters and the inverse function are known. Finally, we estimate the copula density with MLE.

However, in practice, the margins are unknown. As a consequence, the margins are estimated non-parametrically, by means of the empirical distribution function. This is done in order to recover a sample of *pseudo-observations* with uniform distributed margins where the copulas densities are fitted. Table 6 reports the output of the maximum likelihood regarding four different types of multivariate copulas fitted on the financials stocks. Student-t coupla fits the data better compared to the others

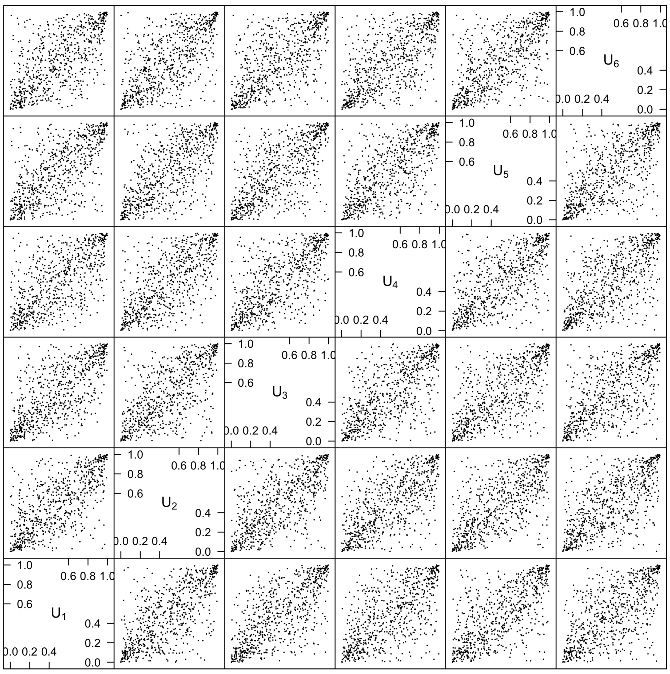
Table 6: Results from MLE of four different multivariate copula (d = 6) for stocks of the financial sector

Normal Copula	**t-Copula**	**Gumbel**	**Clayton**
1254	1596	911.2	884.6

Then, we consider bivariate pseudo-observations arising from both actual data and simulated ones (see, Fig. 8). The Appendix shows provide the fitting procedure of both bivariate combinations of stocks as well as of multivariate ones.

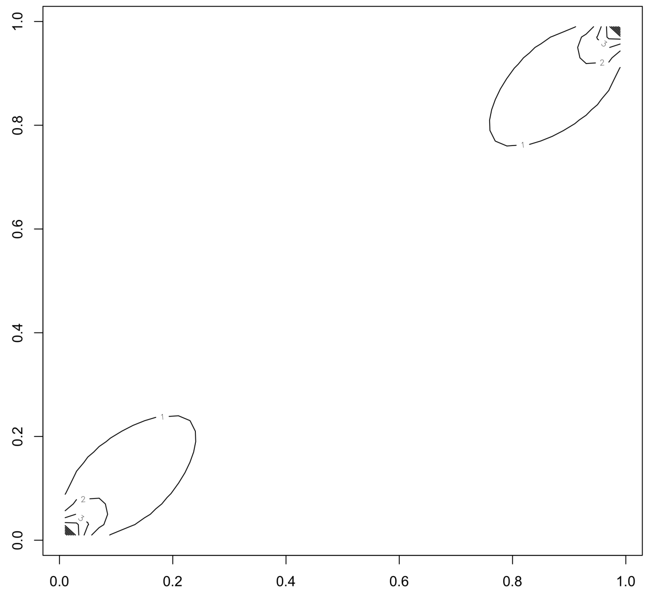


(a) Pseudo observations with uniform margins obtained from actual data, with non parametric estimation of the cdf

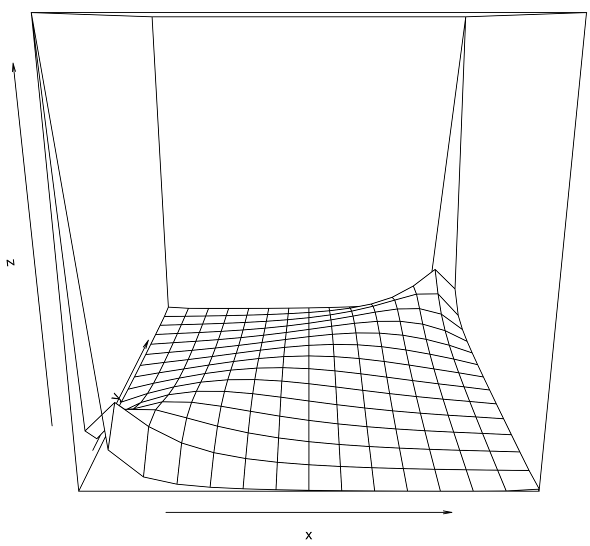


(b) 709 *simulated* pseudo observations obtained from a multivariate Student-t distribution with 4 dof and equicorrelation dispersion matrix with rho = 0.7

Figure 8: Comparison between actual data and simulated ones



(a) Contour plots of a t-Copula cdf with rho = 0.7



(b) Bivariate cdf of a t-Copula with equicorrelation dispersion matrix with rho = 0.7

Figure 9: t-Copula

3 Value-at-Risk

In this section we consider a linear portfolio which allocates \$1000 on each of the ten assets at the beginning of the period. We denote $(n_1, n_2, \ldots, n_k)$ the vector of portfolio holdings at the beginning of the period where n stands for units of asset i bought at $t = 0$. We assume that the holdings (units) remain constant over the entire period. Equation 1 represent the value of the portfolio at any time time t which is the sum of each asset.

$$P_t = \sum_{i=1}^{k} n_i p_{it} \tag{1}$$

The portfolio weights on asset i are

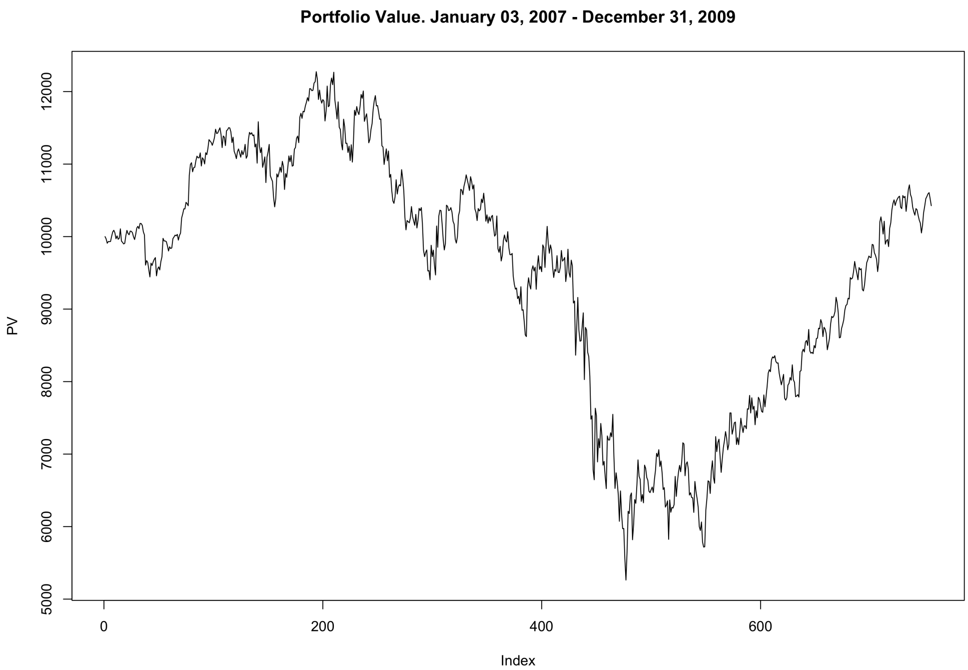


Figure 10: Portfolio value during 3-years period

$$w_{it} = \frac{n_i p_{it}}{P_t} \tag{2}$$

When each n_k remain constant over time, then the portfolio weights changes over time (see Figure 11).

Once, we have the value of the portfolio at each time period the *profit and loss* gives

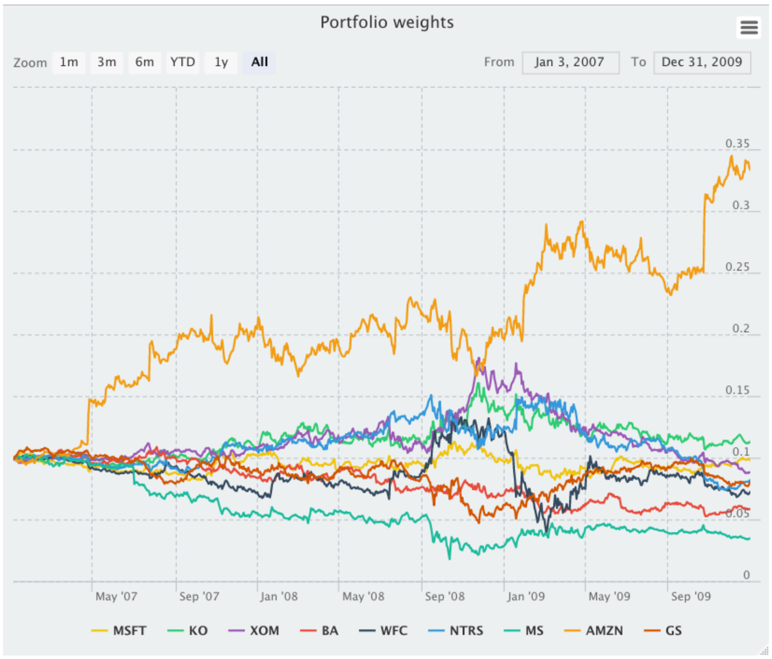


Figure 11: Portfolio's weights dynamics

the change in value of the portfolio between two consecutive times

$$\Delta P_t = P_t - P_{t-1}$$

The box plot in figure 12 displays the historical distribution of the losses obtained until

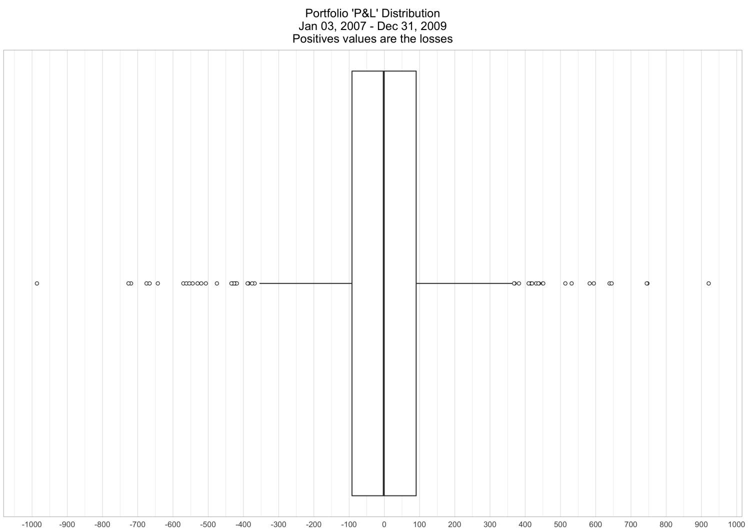


Figure 12: Boxplot of the P&L distribution

31 December 2009. Overall, assuming expected returns equal to zero, seems to be a reasonable assumption.

The next step consists in computing the 95% Value-at-Risk for the portfolio, over a risk horizon of one day.

Two strategies are adopted. Firstly, we assume that the percentage returns of the portfolio are i.i.d. normally distributed. The resulting VaR is expressed as a percentage of the portfolio's current value, since it is estimated from the portfolio return distribution. Secondly, I adopt a non-parametric approach based on the 95-th quantile of the empirical cdf. In this case, the advantage of adopting such approach relies on the fact that no assumption are made on the functional form of the model which generates the data. On the other hand, we are implicitly assuming that empirical cdf estimated from past returns will be representative for the returns one day ahead from now. Needless to say, this assumption might not be correct. The difference between the two approaches is illustrated in Figure 13.

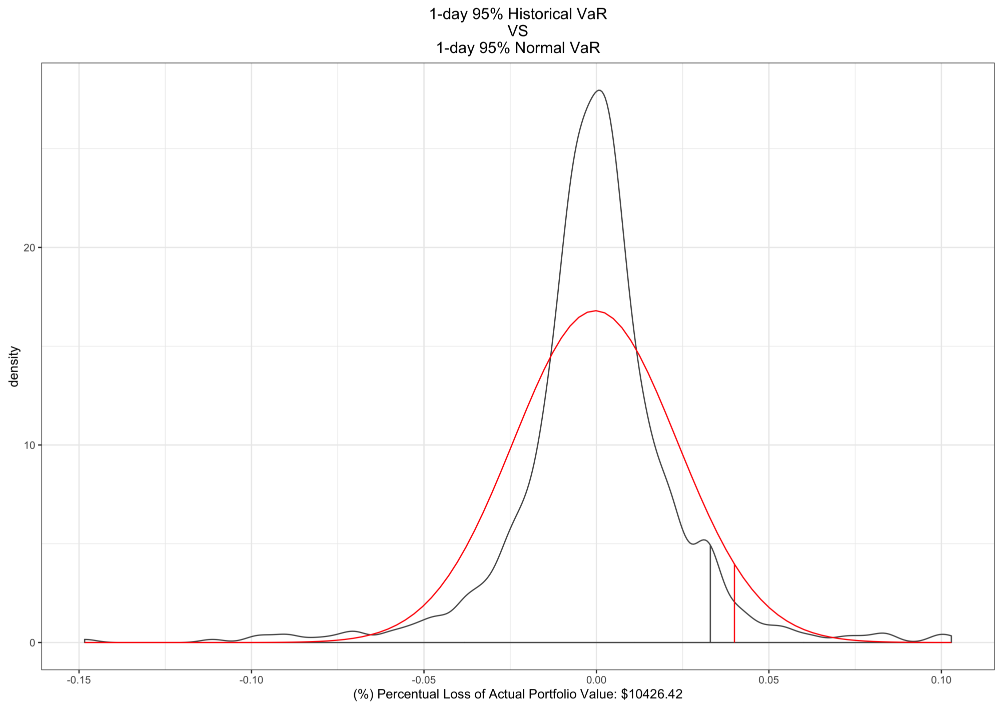


Figure 13: Historical and Normal parametric VaR

Table 7: 95% Value-at-Risk calculated on 31 December 2009 for a one-day horizon under parametric and non-parametric approach

Gaussian VaR	**Historical VaR**
0.040	0.033

The VaR results should be interpreted as follows: (1) under the parametric normal VaR, we expect to incur in a loss of more than \$417.05 with probability 5%, over a one-day horizon; (2) under the historical VaR, we expect to incur in a loss of more than \$344.05

Table 8: Actual 95% Value-at-Risk in $

Gaussian VaR	**Historical VaR**
`$417.05`	`$344.10`

with probability 5%, over a one-day horizon. Hence, the parametric approach leads to more cautious decisions in terms of capital management, because it signals higher potential losses compared to the historical one.

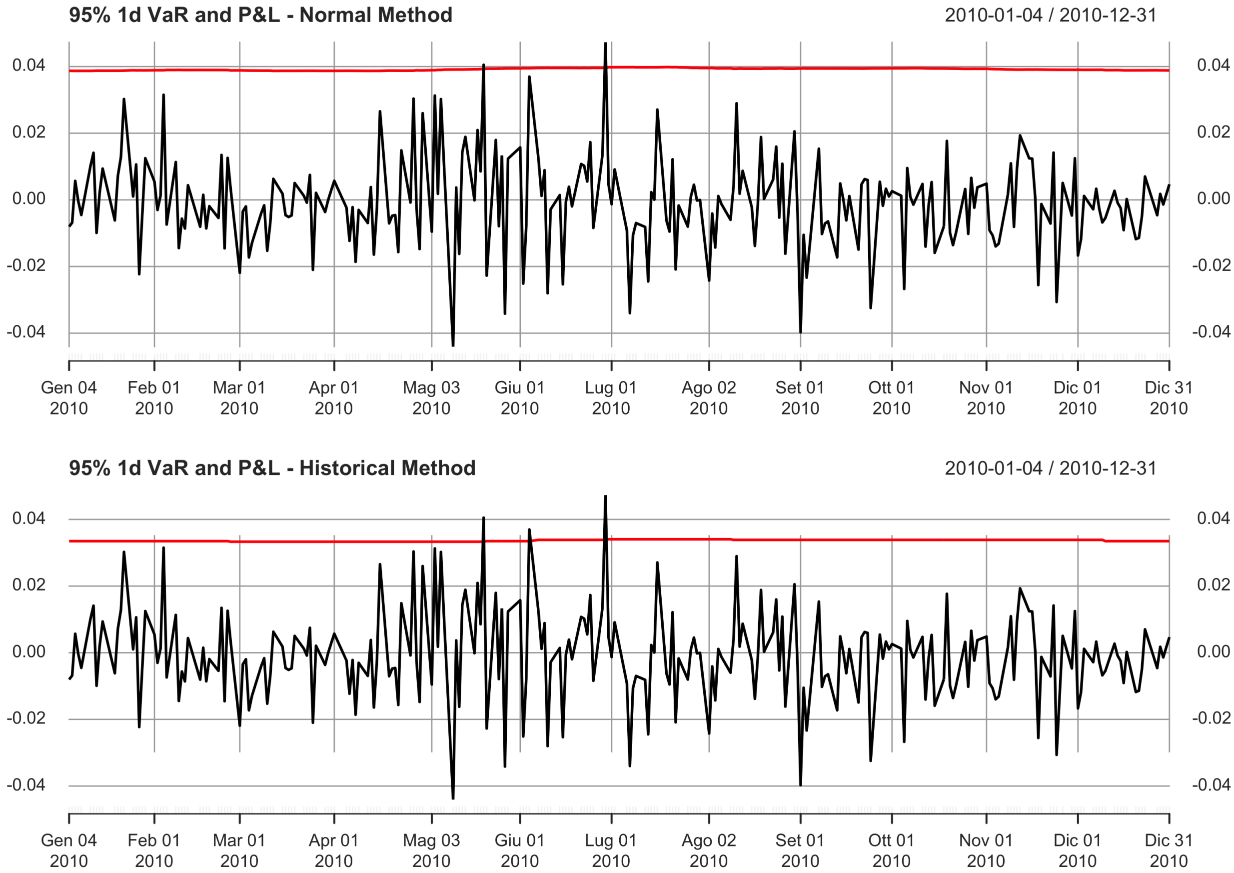


Figure 14: VaR Backtest

The last step of the VaR analysis consists in testing the two models on portfolio returns of 2010 using a rolling window of 755 days. The first step of the test compares the VaR forecast made on 31 December 2009, with the realized loss/return obtained on 2 January 2010. In other words, the realized loss/return on $t+1$ is compared with the VaR predicted for that day in t. If the loss is higher than the threshold signaled by the VaR, then a *violation* is recorded. The procedure proceeds by shifting the estimation window ahead by one day until exhaustion. In the end, we find that both the historical and parametric VaR are almost constant during the period. This might be a pitfall of this metric since it might not be flexible enough to changing condition, especially in presence of higher volatility. Figure 14 shows in red the daily VaR and in black the actual loss/return. When the red line intersects the black one we record a violation where the VaR fails to predict the actual loss. Notice that losses are indicated as positive quantities. In conclusion, Table 9 summarizes the number of violations.

Table 9: Number of violations for 2010 returns under the two different framework

N. violations - Parametric Gaussian	**N. violations - Parametric VaR**
3	2

4 Course contents and organization

From the 29th of July to the 16th of August 2019, I have attended the course in *Statistical methods in risk management*, which took place at the London School of Economics and Political Sciences. The course comprised 36 hours of lectures taught by Dr. Beatrice Acciaio and Dr. Gelly Mitrodima from the Dept. of Statistics, and 18 hours of computer workshops held by Dr. Gianluca Giudice.

Reading material for the course consisted mainly on slides and R scripts posted on the course website. In addition, the suggested book was:
McNeil, A., Frey, R. and Embrechts, P. (2015) *Quantitative Risk Management: Concepts, Techniques, Tools*; Princeton Series in Finance.
As an additional resource, I found the series of volumes by Carol A., *Market Risk Analysis*, useful in providing valuable introduction and further intuitions on the subject.

Assessment consisted of a two-hour written exam accounting for 75% of the final mark. The remaining 25% was evaluated by an individual project in 'R', taking into account the following criteria:

- Task achievement
- Concepts knowledge and Critical analysis

- Clarity of expression and Use of graphs
- Understanding of R and Coding ability
- Timely submission

The goal of the project was to apply statistical techniques learned in class, in order to analyze stock price data during the 2007-08 financial crisis. Time-series data were downloaded from the *Wharton Research Data Services*.

A typical day consisted of 90 minutes of workshop in the morning, followed by lecture in the afteroon starting at 1:30pm until 5pm. Most of the times, evenings were spent on campus studying at the library, which opened at 8am until midnight, Saturdays and Sundays included.

Morning seminars were structured in smaller groups made of ten students. Hence, it was a moment, not only to practice in 'R', but also to revise and exercise theoretical foundations of the course.

The course contents were structured in the following way:

1.1 Introduction and overview (Monday, Week 1)

- Main themes and challenges in risk management
- Course overview

1.2 Loss and its distributions

- Profit and loss of a portfolio
- Three examples (stocks, options and bonds)

2.1 Different approaches of risk management (Tuesday, Week 1)

2.2 Introduction to risk measures

2.3 Distribution function and its quantile

3.1 Value-at-risk (VaR) (Wednesday, Week 1)

- Definition and economic meaning of VaR
- Drawbacks of VaR
- Calculation of VaR

4.1 VaR in practice (Thursday, Week 1)

- Market capital requirement under Basel II
- VaR Backtesting

5.1 Expected shortfall (Friday, Week 1)

- Definition of Expected Shortfall
- Calculation of Expected Shortfall

5.2 Coherent measures of risk

6.1 Multivariate normal models (Tuesday, Week 2)

- Multivariate normal distribution
- Normality test

6.2 Extensions of multivariate normal models

- Normal variance mixtures
- Normal mean-variance mixtures

7.1 Factor models and dimensional reduction (Wednesday, Week 2)

- What is a factor model?
- Advantages of dimensional reduction
- Principal component analysis

8.1 Copula and dependence (Thursday, Week 2)

- Definition of copula
- Sklar's Theorem
- Examples of copulas

8.2 Sampling copulas and meta distributions

- Sampling of copulas
- Sampling of meta distributions
- Fitting of copulas

9.1 Rank correlations (Friday, Week 2)

- Linear vs. rank correlations
- Calculations of rank correlations

10.1 Coefficients of tail dependence (Monday, Week 3)

- Definition of coefficients of tail dependece
- Coefficients of tail dependece and copulas

11.1 Introduction to Extreme Value Theory

- Block maximum
- Threshold exceedance

In the end, I found this course extremely rewarding because it gave me the opportunity of being exposed to a solid overview of a new subject which I have always been curious about. From my point of view, its main strength relies on the fact that it combines theoretical aspects with hands-on training on R every day. As a consequence, I feel that my R fluency has improved in a small amount of time due to this intense pace. Moreover, I had the chance of being surraunded by motivated international students and to take advantage of the amazing LSE facilities of which I attach below some pictures taken during my stay.

(a) Old Building

(b) New Building

(c) Library

Figure 15: LSE Facilities

A R Code

```
rm(list=ls())
library(xts)
library(stringr)
library(tidyverse)
library(highcharter)
library(GGally)

dataraw <- read.csv("/Users/Andrea/__LSE_RISK/Data_project_
   adj_2.csv")
ids <- unique(dataraw[,1])
nAssets <- length(ids)
nObs <- nrow(dataraw)/nAssets
data <- matrix(NA, nrow = nObs, ncol = nAssets + 1)
data[,1] <- dataraw[dataraw[,1] == ids[1], "date"]
for (i in 1:nAssets){
  data[,i+1]<-dataraw[dataraw[,1] == ids[i], "PRC"]
}
ticker <- as.character(unique(dataraw[,"TICKER"]))
dates <- as.Date(as.character(data[,1]), "%Y%m%d")
data <- data[,-1]

colnames(data) <- ticker
ts <- xts(data, dates)
#### prices ####
highchart(type = "stock") %>%
  hc_title(text = "Prices") %>%
  hc_add_series(ts[, ticker[1]], name = ticker[1]) %>%
  hc_add_series(ts[, ticker[2]], name = ticker[2]) %>%
  hc_add_series(ts[, ticker[3]], name = ticker[3]) %>%
```

```
  hc_add_series(ts[, ticker[4]], name = ticker[4]) %>%
  hc_add_series(ts[, ticker[5]], name = ticker[5]) %>%
  hc_add_series(ts[, ticker[6]], name = ticker[6]) %>%
  hc_add_series(ts[, ticker[8]], name = ticker[8]) %>%
  hc_add_series(ts[, ticker[9]], name = ticker[9]) %>%
  hc_add_series(ts[, ticker[10]], name = ticker[10]) %>%
  hc_add_theme(hc_theme_flat()) %>%
  hc_navigator(enabled = FALSE) %>%
  hc_scrollbar(enabled = FALSE) %>%
  hc_exporting(enabled = TRUE) %>%
  hc_legend(enabled = TRUE)

ts["2008-09-15"]

sub_ts <- window(ts, startdate="2008-03-14",end="2009-01-20")
head(sub_ts)
tail(sub_ts)
first <- apply(sub_ts,2,first)
last <- apply(sub_ts,2,last)
((last-first)/first)*100

#### log returns and percentage returns ####
log_rts <- apply(ts, 2, quantmod::Delt, type = 'log')
net_rts <- apply(ts,2,quantmod::Delt,type = 'arithmetic')
head(log_rts)
head(net_rts)

#### Test for multivariate normality (Mardia) ####
normality_test <- MVN::mvn(data = log_rts, mvnTest = "mardia"
   )
```

```
normality_test$multivariateNormality
normality_test <- MVN::mvn(data = log_rts[-1,], mvnTest = "
   mardia", multivariatePlot = "qq")


#### Test for marginal normality (Jarque - Bera) ####
JBTest <- function(dataMatrix){
  ncol <- dim(dataMatrix)[2]
  pvals <- array(NA, ncol)
  for (i in 1:ncol) pvals[i] <- tseries::jarque.bera.test(
     dataMatrix[,i])$p.value
  if (!is.null(names(dataMatrix))) names(pvals) <- names(
     dataMatrix)
  return(pvals)
}
JBTest(log_rts[-1,])
apply(log_rts[-1,], 2, tseries::jarque.bera.test)

#### Test for marginal student t with 1 dgf ####
uniques <- apply(log_rts, 2, unique)
values_t_test <- vector(mode="list",length = dim(log_rts)[2])
for(i in 1:dim(log_rts)[2]){
  values_t_test[[i]] <- ks.test(uniques[[i]], "pt", df = 1)
}
results_t_test <- list()
results_t_test[colnames(log_rts)] <- values_t_test
results_t_test

par(mfrow=c(3,4))
for(i in 1:dim(log_rts)[2]){
```

```
  plot(qt(ppoints(log_rts[-1,i]),10), sort(log_rts[-1,i]),
       main=str_c("Student-t␣QQ␣(dof=10)␣Plot␣for",
                  colnames(log_rts)[i],sep="␣"),
       xlab="Theoretical␣quantiles",ylab="Sample␣Quantiles")
  abline(a=0,b=1)
}

#### FINANCIAL STOCKS ####
fin.stocks <- as.data.frame(log_rts[,c("BA","WFC","NTRS","AIG
   ","MS","GS")])
fin.scatter <- ggpairs(fin.stocks,
                       lower = list(continuous = wrap("points
                         ", size = 0.1))) +
  theme_bw() +
  theme(panel.grid.major = element_blank(),
        strip.text = element_text(face="bold", size=rel(0.60)
          ),
        plot.title=element_text(size = rel(1.1),
                                lineheight = .9,
                                family = "Helvetica",
                                face = "plain",
                                colour = "black",
                                hjust = 0.5)) +
  ggtitle("Pair-wise␣Stock␣Returns␣Financial␣stocks")
head(fin.stocks)



#### Mardia test for combinations of bivariate distributions
   of stocks ####
combs <- t(combn(1:dim(fin.stocks)[2],2))
```

```
values <- vector(mode="list", length = dim(combs)[1])
names <- c()
for (i in 1:dim(combs)[1]) {
  names[i] <- str_c(colnames(fin.stocks[,combs[i,]]),
     collapse="_")
  values[[i]] <- MVN::mvn(fin.stocks[-1,combs[i,]],
                           mvnTest = "mardia")$
                              multivariateNormality
}
results <- list()
results[names] <- values
results


#########################
######## COPULAS ########
#########################
# Semi-parametric estimation -> specify: method="mpl"
############# MULTIVARIATE COPULA OF DIMENSION 6 #########
# Keep only the rows in which returns for each stock are
   different
# from zero.
cond <- apply(fin.stocks, 1, function(x) any(x == 0))
zero_index <- which(cond == TRUE)
fin.stocks <- fin.stocks[-zero_index,]
dim(log_rts)
library(QRM)
library(copula)
#### PSEUDO-OBSERVATIONS ####
#pseudo_obs <- apply(fin.stocks, 2, QRM::edf, adjust = 1)
```

```
pseudo_obs <- copula::pobs(fin.stocks)
splom2(pseudo_obs, cex = 0.2, col.mat = 'black', pch = 16)


par(mfrow = c(3,3))
for(i in 1:6){
  hist(pseudo_obs[,1])
}


#### NORMAL ####
normal_copula <- fitCopula(normalCopula(dim = 6, dispstr = "
   un"),
                           data = pseudo_obs[-1,],
                           method = "mpl") # 1254
summary(normal_copula)
#  We are looking for the linear correlation matrix that max
   the log-lik func


#### STUDENT-T ####
# We search over the correlation matrices and degrees of
   freedom
t_copula <- fitCopula(tCopula(dim = 6, dispstr = "un"), data
   = pseudo_obs[-1,],
                      method = "mpl")
summary(t_copula)
# 1596


#### GUMBEL  ####
summary(fitCopula(gumbelCopula(dim = 6), data = pseudo_obs
   [-1,],
                  method = "mpl"))
```

```
# 911.2

#### CLAYTON ####
summary(fitCopula(claytonCopula(dim = 6), data = pseudo_obs
   [-1,],
                  method = "mpl"))
#884.6



########### BIVARIATE COMBINATIONS ################
combs <- t(combn(1:dim(fin.stocks)[2],2))

#### BIVARIATE STUDENT T COPULA ####
values_biv_t_cop <- vector(mode="list", length = dim(combs)
   [1])
for (i in 1:dim(combs)[1]) {
  values_biv_t_cop[[i]] <- fitCopula(tCopula(dim = 2, dispstr
      = "un"),
                                     data = pseudo_obs[-1,
                                        combs[i,]],
                                     method = "mpl")
}
results_biv_t_cop <- list()
results_biv_t_cop[names] <- values_biv_t_cop
results_biv_t_cop
# Simulation of t Copula #
nObs <- 709
tNVar <- 6
cor.par <- 0.7
# dispersion matrix
```

```
tSigma <- (1-cor.par)*diag(tNVar)+array(cor.par,c(tNVar,tNVar
   ))
t.dof <- 4
# multivariate dim 6 student t distribution
tData <- rmt(nObs,t.dof,mu=array(0,tNVar),tSigma)
head(tData)
tCopulaData <- apply(tData,2,pt,df=t.dof)
splom2(tCopulaData, cex = 0.2, col.mat = 'black', pch = 16)


# For displaying contour charts
distribution.limits <- c(0.01,0.99)
Sigma.contour <- equicorr(2,0.5)
threed.limits <- c(-3,3)
par(mfrow = c(3,3))

for (i in 1:6){
  hist(tCopulaData[,i])
}

par(mfrow = c(1,2))
BiDensPlot(func=dcopula.t,xpts=distribution.limits,ypts=
   distribution.limits,
           df=t.dof,Sigma=Sigma.contour, type = 'contour')
BiDensPlot(func=dcopula.t,xpts=distribution.limits,ypts=
   distribution.limits,
           df=t.dof,Sigma=Sigma.contour, type = 'persp',npts
              =15)
```

```
#### BIVARIATE NORMAL COPULA ####
values_biv_normal_cop <- vector(mode="list", length = dim(
   combs)[1])
for (i in 1:dim(combs)[1]) {
  values_biv_normal_cop[[i]] <- fitCopula(normalCopula(dim =
     2),
                                          data = pseudo_obs
                                             [-1,combs[i,]],
                                          method = "mpl")
}
results_biv_normal_cop <- list()
results_biv_normal_cop[names] <- values_biv_normal_cop
results_biv_normal_cop

#### BIVARIATE GUMBEL COPULA ####
values_biv_gumbel_cop <- vector(mode="list", length = dim(
   combs)[1])
for (i in 1:dim(combs)[1]) {
  values_biv_gumbel_cop[[i]] <- fitCopula(gumbelCopula(dim =
     2),
                                          data = pseudo_obs
                                             [-1,combs[i,]],
                                          method = "mpl")
}
results_biv_gumb_cop <- list()
results_biv_gumb_cop[names] <- values_biv_gumbel_cop
results_biv_gumb_cop

#### BIVARIATE CLAYTON COPULA ####
values_biv_clayton_cop <- vector(mode="list", length = dim(
```

```
   combs)[1])
for (i in 1:dim(combs)[1]) {
  values_biv_normal_cop[[i]] <- fitCopula(claytonCopula(dim =
      2),
                                                    data = pseudo_obs
                                                       [-1,combs[i,]],
                                                       method = "mpl")
}
results_biv_clayton_cop <- list()
results_biv_clayton_cop[names] <- values_biv_clayton_cop
results_biv_clayton_cop


###### PORTFOLIO constant holdings ######
# The value of the portfolio at the beginning of the period
   will be 10.000$
# I.e. the first observation in our time-series
prices <- data
head(prices)
single_k <- rep(1000, dim(prices)[2])
pf_value_1 <- sum(single_k)

# We calculate how many stocks we get at the first period (
   units or portholdings).
# THE NUMBER OF ASSET IS DETERMINED AT T=1, THEN IT IS FIXED
   #
# ALONG THE SUCCESIVE PERIODS #

n_asset <- as.matrix(single_k/prices[1,])
```

```
# PORTFOLIO VALUE at each time period
PV <- prices%*%n_asset

# We calculate the matrix of weights
n_asset_b <- matrix(rep(t(n_asset), dim(prices)[1]),
                    ncol = dim(prices)[2],
                    nrow = dim(prices)[1],
                    byrow = TRUE)

PV_b <- matrix(rep(PV, dim(prices)[2]),
               ncol = dim(prices)[2],
               nrow = dim(prices)[1],
               byrow = FALSE)

weights <- (prices*n_asset_b)/PV_b
ts_weights <- xts(weights,dates)
highchart(type = "stock") %>%
  hc_title(text = "Portfolio␣weights") %>%
  hc_add_series(ts_weights[, ticker[1]], name = ticker[1])
     %>%
  hc_add_series(ts_weights[, ticker[2]], name = ticker[2])
     %>%
  hc_add_series(ts_weights[, ticker[3]], name = ticker[3])
     %>%
  hc_add_series(ts_weights[, ticker[4]], name = ticker[4])
     %>%
  hc_add_series(ts_weights[, ticker[5]], name = ticker[5])
     %>%
  hc_add_series(ts_weights[, ticker[6]], name = ticker[6])
     %>%
```

```r
  hc_add_series(ts_weights[, ticker[8]], name = ticker[8])
     %>%
  hc_add_series(ts_weights[, ticker[9]], name = ticker[9])
     %>%
  hc_add_series(ts_weights[, ticker[10]], name = ticker[10])
     %>%
  hc_add_theme(hc_theme_flat()) %>%
  hc_navigator(enabled = FALSE) %>%
  hc_scrollbar(enabled = FALSE) %>%
  hc_exporting(enabled = TRUE) %>%
  hc_legend(enabled = TRUE)

# time plot of portfolio value
png("portfolio_value.png",res=320,height=210,width=297,units
   = "mm")
plot(PV, type = 'l',main = "Portfolio␣Value.␣January␣03,␣2007
   ␣-␣December␣31,␣2009")
dev.off()

### P&L ###
losses <- -diff(PV)
losses <- as.data.frame(losses)
ggplot(losses, aes(y=V1)) +
  geom_boxplot(outlier.size=1.5,outlier.shape=21) +
  theme_light() +
  scale_y_continuous(breaks=seq(-1000,1000,100)) +
  scale_x_continuous(breaks=NULL) +
  theme(axis.title.x=element_blank(),
        axis.title.y=element_blank(),
        axis.ticks.x = element_blank(),
```

```
      plot.title = element_text(hjust=0.5)) +
  coord_flip() +
  ggtitle("Portfolio␣'P&L'␣Distribution\nJan␣03,␣2007␣-␣Dec␣
    31,␣2009\nPositives␣values␣are␣the␣losses")
  #ggsave("boxplot_profit_and_losses.png",dpi = 320, width
    =297,height = 210, device = "png",units = "mm")

#### VAR for the return distribution ####
pv_rts <- apply(PV,2,quantmod::Delt, type = 'arithmetic')
mean_pv_rts <- mean(pv_rts[-1,])
sd_log_rts <- sd(pv_rts[-1,])

#Var as a percentage of portfolio's current value
var_gauss <- qnorm(0.95)*sd_log_rts - mean_pv_rts
var_hist <- quantile(pv_rts[-1,],p=0.95)

# Difference btw parametric and non parametric var
pv_rts <- as.data.frame(-pv_rts)
colnames(pv_rts) <- "Losses"
mean_loss <- mean(pv_rts$Losses, na.rm = TRUE)
sd_loss <- sd(pv_rts$Losses, na.rm = TRUE)

dx <- density(pv_rts$Losses, na.rm = TRUE)
xnew <- 0.033
y_hist <- approx(dx$x, dx$y, xout = xnew)

df <- data.frame(x1 = 0.033, y1 = y_hist$y,
                 x2 = 0.04, y2 = dnorm(0.04, mean = mean_loss
                   , sd = sd_loss))
```

```
ggplot(pv_rts, aes(x = Losses)) +
  geom_density(colour = '#525252') +
  theme_bw() +
  xlab('(%) Percentual Loss of Actual Portfolio Value: $
     10426.42') +
  stat_function(fun = dnorm,
                args = list(mean = mean_loss, sd = sd_loss),
                colour = "red") +
  geom_segment(aes(x = x1, xend = x1, y = 0, yend = y1),
     colour = "#525252", data = df) +
  geom_segment(aes(x = x2, xend = x2, y = 0, yend = y2),
     colour = 'red', data = df) +
  #annotate('text', x = 0.05, y = 5, label = 'VaR 95%\nNormal
     ', colour = 'red') +
  #annotate('text', x = 0.025, y = 5, label = 'VaR 95%\
     nHistorical', colour = 'black') +
  ggtitle('1-day 95% Historical VaR\nVS\n1-day 95% Normal VaR
     ') +
  theme(plot.title = element_text(hjust = 0.5)) +
  #ggsave("historical_vs_normal_2.png",dpi = 320, width=297,
     height = 210, device = "png",units = "mm")

##### VAR backtest #####
dataraw_test <- read.csv("/Users/Andrea/__LSE_RISK/VaR_
   Backtesting.csv")
ids <- unique(dataraw_test[,1])
nAssets <- length(ids)
nObs <- nrow(dataraw_test)/nAssets
data_test <- matrix(NA, nrow = nObs, ncol = nAssets + 1)
data_test[,1] <- dataraw_test[dataraw_test[,1] == ids[1], "
```

```
   date"]
for (i in 1:nAssets){
  data_test[,i+1]<-dataraw_test[dataraw_test[,1] == ids[i], "
     PRC"]
}
ticker <- as.character(unique(dataraw_test[,"TICKER"]))
colnames(data_test)[2:(nAssets+1)] <- ticker
colnames(data_test)[1] <- 'Date'
dates_test <- data_test[,1]
dates_test <- as.Date(as.character(dates_test), format = "%Y%
   m%d")
data_test <- data_test[,c(-1,-10)] # remove an empty column
   and dates

###### Merging 'data' and 'data_test' - entire #####
entire <- rbind(data, data_test)

###### PORTFOLIO including TEST series ######
# Portfolio value at each time t #
PV_test <- entire%*%n_asset

# Instrumental matrix to compute the weights #
n_asset_c <- matrix(rep(t(n_asset), dim(entire)[1]),
                    ncol = dim(entire)[2],
                    nrow = dim(entire)[1], byrow = TRUE)

PV_c <- matrix(rep(PV_test, dim(entire)[2]),
               ncol = dim(entire)[2],
               nrow = dim(entire)[1],
               byrow = FALSE)
```

```
# Matrix containg weights #
weights_test <- (entire*n_asset_c)/PV_c

sum(apply(weights_test, 1, sum)) == dim(weights_test)[1]

# Net Portfolio returns and losses #
entire_ret <- apply(PV_test, 2, quantmod::Delt, type = '
   arithmetic')
entire_losses <- -entire_ret

#### TIME SERIES of THE LOSSES ####
dates_entire <- as.Date(c(date, dates_test))
dates_entire <- lubridate::ymd(dates_entire)
entire_losses_ts <- xts::xts(entire_losses, dates_entire)

##### BACKTEST - rolling window #####
VaR_t_Normal <- function(vec, alpha) {
  mu.hat = mean(vec, na.rm = TRUE)
  sd = sd(vec, na.rm = TRUE)
  VaR = mu.hat + sd * qnorm(alpha)
  return(VaR)
}

VaR_t_hs <- function(vec, alpha) {
  VaR = quantile(vec, alpha, na.rm = TRUE)
  return(VaR)
}

# I remove the first from the first row of entire_loss_ts
```

```
   which is NA #
entire_losses_ts <- entire_losses_ts[-1,]

# I set the rolling window to be of size 755
Time_t <- dim(data)[1]-1

#### NORMAL VAR ####
# By setting align 'right', the function rollapply start
   counting
# from the first row in order to define the sample on which
   it will
# estimate the parameters needed to compute the VaR
Daily_norm_Var <- zoo::rollapply(entire_losses_ts, width =
   Time_t,
                                  align = 'right', FUN = VaR_t
                                     _Normal, alpha = 0.95)

# We shift forward by one day the series because we are
   calculating the VaR
# over one-day risk horizon. However, rollapply locates the
   first result
# in correspondence of the time instant from which we account
    for the risk
# horizon. This operation allow to make a comparison between
   the realized
# loss and the one-day VaR.
Daily_norm_Var <- lag(Daily_norm_Var, k = 1)

#### HISTORICAL VAR ####
Daily_hs_Var <- zoo::rollapply(entire_losses_ts, width = Time
```

```
  _t, align = 'right',
                                   FUN = VaR_t_hs, alpha = 0.95)
Daily_hs_Var <- lag(Daily_hs_Var, k = 1)

# We select the range to be the entire 2010 because it is our
    test
loss.2010 <- window(entire_losses_ts, start = as.Date('
  2010-01-04'),
                    end = as.Date('2010-12-31'))
norm.Var_2010 <- window(Daily_norm_Var, start = as.Date('
  2010-01-04'),
                        end = as.Date('2010-12-31'))
hs.Var_2010 <- window(Daily_hs_Var, start = as.Date('
  2010-01-04'),
                      end = as.Date('2010-12-31'))

par(mfrow=c(2,1))
plot(merge(loss.2010, norm.Var_2010), main = "95% 1d VaR and 
  P&L - Normal Method")
plot(merge(loss.2010, hs.Var_2010), main = "95% 1d VaR and P&
  L - Historical Method")

# We count and compare the violations
n_violations_hs <- sum(entire_losses_ts > hs.Var_2010, na.rm
  = TRUE)
n_violations_norm <- sum(entire_losses_ts > norm.Var_2010, na
  .rm = TRUE)

df_violations <- data.frame(hs = n_violations_hs, norm = n_
  violations_norm)
```

```
colnames(df_violations)[1] <- "Numb. of violations - Historical"
colnames(df_violations)[2] <- "Numb. of violations - Normal"
```