*Dipartimento di Economia e Finanza*

*Cattedra di Applied Statistics and Econometrics*

# Value at Risk and Conditional Value at Risk. An Econometric Analysis

RELATORE

PROF. Giuseppe Ragusa

CANDIDATO:
Pierluigi Vallarino

MATR: 188621

ANNO ACCADEMICO 2016/2017

# Contents

# List of Figures

# List of Tables

## Acknowledgments

I would like to thank Professor Giuseppe Ragusa for having supervised and guided me in approaching the world of research. His advice and presence have made this thesis possible.

I would also like to be thankful to Doctor Giuseppe Brandi for having introduced me to the world of Financial Econometrics and for all the help he has given me in writing of my dissertation.

# 1 Introduction

The first part of this thesis employs some statistical and mathematical tools in order to analyze Value at Risk and Conditional Value at Risk, two of the most employed Market Risk Metrics.
The second part shows how these two Risk Measures can be employed for active and passive portfolio management.

The core of the thesis is structured as follows:

- Chapter 2 introduces the concept of risk as codified by the International Organization for Standardization and presents the different types of risk financial institutions have to face. Theoretical and practical instruments employed to analyze and control risk are introduced along with the mathematical properties of Coherent Risk Measures as described by Artzner et alius (1999). Finally the concept of Quantile Risk Measure is defined.

- Chapter 3 presents the Value at Risk, VaR. The VaR is initially introduced by looking at the historical events that lead financial institutions to develop this measure. After having looked at the history of VaR a formal definition for this market risk measure is provided and three sections of this chapter are devoted to illustrate, both theoretically and empirically, three different ways of computing the Value at Risk: the Delta-Normal Approach, the Historical Approach and the Monte Carlo Method.

  Here it is stressed out out how both the Delta-Normal Approach and the Monte Carlo Method might create a biased estimation of the Value at Risk due to a misspecified model for the behavior of asset prices and returns.

  Finally the chapter presents some drawbacks of the VaR, namely not being always sub-additive and not providing any clue as to what regards losses in exceed of it, and introduces a methodology to backtest the VaR, as imposed by the Basel Committee for Banking Supervision.

- Chapter 4 deals with the Conditional Value at Risk, CVaR, a risk measure presented by Rockafellar and Uryasev (2000) to overcome the second drawback highlighted before, namely the impossibility of using the VaR to forecast stock returns' behavior once the VaR threshold for losses has been breached.

First the CVaR is formally defined and it is shown that, as proved by Acerbi and Tasche (2001), the CVaR follows all the mathematical properties of a coherent risk measure. Then, following what did by Rockafellar and Uryasev (2000), a portfolio is structured by means of minimizing the CVaR of the portfolio subject to the attainment of a precise level of expected returns. An empirical analysis carried out using 10 Italian stocks shows how the portfolio structured in this way outperforms, both in terms of Sharpe Ratio and of returns per unit of CVaR, the one constructed using the classic Mean-Variance Analysis, i.e. minimizing the variance for a given level of expected returns.

The chapter goes on analyzing the problem of forecasting what will happen beyond a certain level of losses by looking at the Extreme Value Theory. More precisely this section presents the method of fitting a Pareto Distribution to the left tail of the empirical distribution of returns through Maximum Likelihood Estimation. The use of the Kolmogorov-Smirnov test allows us to infer that these Pareto distributions actually well fit the data and could be employed to model returns' behavior under "extreme circumstances". This section concludes by illustrating how we could estimate the VaR for a given Pareto Distribution as proposed by McNeil (1999), an approach whose results are backed by Gencay and Selcuk (2004).

- Chapter 5 deals with the issue of employing the CVaR in order to construct an Exchange Traded Fund over an index created by I myself. The approach is very innovative in that it combines the employment of Deep Learning techniques, namely auto-encoding, to decide which stocks to include in the ETF, a process known as "Sampling", and the minimization of the CVaR of the distribution of the Tracking Error of the ETF to weight the selected stocks within this replicating portfolio.

  The sampling methodology, i.e. the application of auto-enconding, follows what did by Heaton et alius (2016). Here the authors shows that Deep Learning might allow for a superior sampling technique in that, differently from other techniques used for dimensionality reduction, Deep Learning algorithms are able to capture hidden non-linear relations between entities.

  The minimization of the CVaR solves a simple Linear Programming problem. The scientific software MatLab has been employed for both steps of this algorithm.

  The last section of this chapter explains why the CVaR has been chosen as the object of the minimization process and references are made to Bamberg and Wagner (2000) and Hansen (2012).

In the *Conclusions* some possible extensions of the work here carried out are proposed.

# 2 Introducing Risk

## 2.1 What is Risk

The 2009 ISO3100, a family of standards relating to risk management, codified by the International Organization for Standardization, defines Risk as the effect of uncertainty on objectives.
Risk Management is the process of identification, analysis and either acceptance or mitigation of uncertainty in investment decision-making. Its tools are widely employed to assure uncertainty does not deflect the endeavor from the business goals.

## 2.2 Classifying Risk

Financial Institutions face different types of risk:

- Credit Risk, the risk of default on a debt that may arise from a borrower failing to meet his payment obligations as they fall due.
- Market Risk, which the Basel Committee on Banking Supervision defines as the risk of losses in financial positions arising from price movements of any asset belonging to one of the following four classes: Equity, Interests Rates related Securities, Commodities and Forex Financial Instruments.
- Operational Risk, the risk of incurring in a loss greater than the expected one due to inadequate or failed internal processes, people and systems or because of external factors.
- Liquidity Risk, the risk that arises from not being able to liquidate some assets without causing substantial price changes of it, hence selling them at a price substantially lower than their intrinsic value.
- Model Risk, the risk of potential losses arising from adopting incorrect models.
- Business Risk, this name refers to those risks that are an integral part of the core business of a firm and therefore simply should be taken on.

Effectively managing all these potential risks is essential to ensure the viability, profitability and good reputation of a financial institution.

## 2.3 Market Risk and its Metrics

This thesis will mostly deal with Market Risk. This form of uncertainty is not an internally generated one, thus firms involved in financial activities cannot directly reduce the sources of market risk but can only take actions aimed at reducing the uncertainty related to market movements. The only way to mitigate market risk is to adopt certain investments strategies, such as Diversification or Hedging, aimed at reducing the potential losses resulting from price movements.

To properly assess the riskiness of any investment Risk Managers employ a batch of Risk Measures aimed at providing them with Risk Metrics, the latter being abstract concepts quantified by Risk Measures. This thesis will be focused over Market Risk Metrics, thus a good starting point is a proper, formal, definition of them.

**Definition 2.1.** *A Market Risk Metric is an abstract measure of the uncertainty in the future value of an investment, i.e. a measure of uncertainty in the investment return or Profit and Loss* ($P\&L$).

Risk Measures work in order to practically quantify these abstract measures, hence a formal definition of them shall be provided too.

**Definition 2.2.** *A Risk Measure is a mapping from a set of random variables to the one of real numbers, where the random variables are the returns of an investment.*

Labelling with X the random variable associated to the returns of an investment and with $\rho(X)$ the risk measure we can write

$$\rho : \Lambda \to \Re \cup \{+\infty\}$$

Where $\Lambda$ denotes a set of random variables $X_i \;\; i \in [1, n]$

Any Risk Measure must satisfy the following properties

- Normalization, $\rho(0) = 0$, i.e. the risk measure associated to a portfolio without any investment must be zero.
- Monotonicity, given $X_1, X_2 \in \Lambda$ and $X_1 \geq X_2$ then $\rho(X_1) \leq \rho(X_2)$
- Translation Invariance, given $a \in \Re$ and $X_1 \in \Lambda$ then $\rho(X_1 + a) = \rho(X_1) - a$

A further, useful, notion is the one of Coherent Risk Measure.
According to Artzner and alius (1999) a Risk Measure is classified as a Coherent Risk Measure when it satisfies both the three properties of all Risk Measures and the following two:

- Subadditivity, given $X_1, X_2 \in \Lambda$ then $\rho(X_1 + X_2) \leq \rho(X_1) + \rho(X_2)$
- Positive Homogeneity, given $a \in \Re$ and $X_1 \in \Re$ the $\rho(aX_1) = a\rho(X_1)$

Let's now explain better these axioms.
Monotonicity implies that if the payoff of investment $X_1$ is always at least the one of $X_2$ then the risk associated to the former will never exceed the one of the latter.
The property of Translation Invariance ensures that adding cash, a value-unchanging security, to a portfolio decreases the portfolio risk of the same amount.
Sub-additivity reflects the fact that the risk associated to the payoff of a portfolio, hence a diversified or hedged investment, will always be lower than the convex linear combination of the risk associated to the payoffs of the individual securities.
Positive Homogeneity grants the rescaling of risk whenever the investment positions are rescaled.
Having introduced what risk management is and how does it work we can start studying one of the most relevant market risk measures, the Value at Risk, VaR by convention. First of all we need two definitions to introduce the concept of Quantile Risk Measure:

**Definition 2.3.** *For any $\alpha \in [0;1] \subset \Re$ the $\alpha$ quantile of a probability distribution function, being it either a continuous or a discrete one, is that $x_\alpha$ such that*:

$$P\{X \leq x_\alpha\} = \alpha$$

**Definition 2.4.** *A Quantile Risk Measure is any risk measure that can be computed as the $\alpha$ quantile of a probability distribution function.*

Assume, for instance, that we have $Z \sim N(0,1)$, then its 5% quantile will be that $z_\alpha$ such that:

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_\alpha} e^{-\frac{u^2}{2}} du = 0.05$$

where $f(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$ is the probability density function of $Z$.

Figure 1 shows the 5% quantile of a Standard Normal Distribution.
More generally, given a random variable $X$, with invertible cumulative distribution function $F(X)$ its $\alpha$ quantile can be computed as:

$$x_\alpha = F^{-1}(\alpha)$$

Should the cumulative distribution function not have a uniquely defined inverse we can still define, for $\alpha \in [0, 1]$, its Generalized Inverse as:

$$F^{-1}(\alpha) = inf\{x \in \Re : F(x) \geq \alpha\}$$

In the framework of market risk analysis $X$ is a random variable that represents the $P\&L$ of a portfolio over a pre-determined time horizon and $\alpha$ is chosen as a sufficiently low percentage, so that, being $\alpha$ the probability of incurring in a return smaller than the $\alpha$ quantile, it is sure that such loss will not be exceeded.

# 3 Value at Risk

## 3.1 Introducing the VaR and a bit of history

Loosely speaking the Value at Risk of an investment is a single number that provides investors with information over the highest possible loss at a certain confidence level, $\alpha$ in what follows, over a pre-specified time horizon, $t$. Before approaching a rigorous study of the VaR I think that it is worth spending few words to track down its history.
On Monday, October 19, 1987 stock markets around the world crashed, with the Dow Jones Industrial Average falling as much as 22.61% in one trading day. Today that infamous day is known as *Black Monday*.
After this financial crash the entire basis of quantitative finance started to be questioned, as a matter of fact the statistical possibilities of such a crash were so low that many quant-finance practitioners started reconsidering the whole trading history and came to the conclusion that such unlikely crisis actually had happened far more times than what predicted by those models over which Trading, Investing and Pricing were based.
Risk Managers noted that having included these "unlikely" events in the employed models would have hampered their day-to-day validity since their magnitude was so large that it would have overwhelmed the one of more regular events. On the other hand it was argued that if such events were completely left aside the profits made between two consecutive extremely rare events might have been too low to cover the losses incurred when one happened, hence financial institutions would have failed. A solution to this trade-off was to be found.
The Value at Risk was developed as a systematic way to segregate extreme events, which are qualitatively studied over long-term history and broad market events, from everyday price movements, which are quantitatively studied using short-term data in specific markets. It was hoped that "Black Swans", a term widely employed to define very unlikely events, would be preceded either by an increase in the estimated VaR or by an increased frequency of VaR breaks. This new Risk Management approach, called VaR Risk Management, was widely adopted by quantitative trading groups at many financial institutions prior to 1990.
The financial events of the early 1990s found many firms in trouble because the same underlying bet had been made at many places in the firm. Since many trading desks already employed VaR Risk Management, and it was the only common risk measure that could be both defined for all businesses and aggregated without strong assumptions, it was the natural choice for reporting firm-wide risk. In 1995 J.P. Morgan, now JPMorgan Chase & Co., published the article RiskMetrics Technical Documents, where it made publicly the now

standard methodologies employed to compute the VaR and gave free access to estimates of the necessary underlying parameters.
In 1997 the U.S. Securities and Exchange Commission, the SEC, enacted a new derivatives regulation, according to which public corporations must disclose quantitative information about their derivatives activity. Major banks and dealers chose to implement the rule by including VaR information in the notes to their financial statements.
In 1999 the Basel Committee, with its Second Basel Agreement, elected the VaR, under certain requirements, as the preferred way of computing the Market Risk for Financial Institutions.

## 3.2 A Formal Definition of the VaR

This brief introduction allowed us to understand the crucial role played by the VaR in the modern financial world. We can now proceed with discussing a bit more formally about it.

**Definition 3.1.** *Let $X$ be a random variable representing the returns, over a specified period of time $(t)$, of an investment. Let $\alpha$, a quantile of the probability distribution function of $X$, be the Significance Level. The $\alpha 100\%$ VaR at the time horizon $t$ is the $\alpha$ quantile of the probability distribution function of $X$ and its absolute value represents the highest possible loss at the Significance Level $\alpha$.*

$$VaR_t^\alpha = -x_\alpha$$

where

$$x_a = x \in \Re : P\{X \leq x_a\} = \alpha$$

To have a practical example the Second Basel Agreement states that banks shall measure, using internally developed methods, their VaR at a Significance Level of 1% over a risk horizon of 10 days.
Figure 2 shows the 5% VaR of an investment whose returns are assumed to be normally distributed.
The VaR is commonly computed employing one of the following methods:

- Normal Method, the returns of the investment whose VaR has to be estimated are assumed to be a normal random variable and the VaR is computed as the $\alpha$ quantile of its distribution function. The mean and variance of this normal random variable are estimated by the correspondent sample value of the analyzed returns.

- Historical Simulation, no assumption is made about the distribution followed by the returns of an investment and the VaR is calculated as the $\alpha$ quantile of their empirical distribution.

- Monte Carlo Method, this approach is very similar to the Historical One, but here data are not historical ones but are randomly generated by a Monte Carlo Simulator.

## 3.3 Normal Estimation

A good starting point is the normal estimation of the VaR, even known as Delta-Normal Approach.
Let's assume that the daily returns, i.e. $t$=1, over an investment are an i.i.d normally distributed random variable labelled by $X$ with mean $\mu$ and variance $\sigma^2$

$$X \sim N(\mu, \sigma^2)$$

Then X will have a probability density function, $f()$ given by

$$f_{\mu,\sigma^2}(t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t-\mu)^2}{2\sigma^2}}$$

Hence the $\alpha$ level of significance VaR of this investment will be that value $VaR^\alpha$ such that

$$\alpha = \int_{-\infty}^{VaR^\alpha} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$

Letting $F(t)$ denote the correspondent cumulative distribution function we can re-write the above equation as:

$$\alpha = F(VaR^\alpha) := P\{X \leq VaR^\alpha\} = P\left\{\frac{X-\mu}{\sigma} \leq \frac{VaR^\alpha - \mu}{\sigma}\right\}$$

Where $\mu$ and $\sigma$ are, respectively, the mean and the standard deviation of the distribution of $X$ and they have been used to standardize $X$ so that such probability can be easily computed looking at the normal table.

$$P\left\{\frac{X-\mu}{\sigma} \leq \frac{VaR^\alpha - \mu}{\sigma}\right\} = P\left\{Z \leq \frac{VaR^\alpha - \mu}{\sigma}\right\} = \alpha \quad Z \sim N(0,1)$$

Let $\phi$ denote the Standard Normal Cumulative Distribution Function and $\phi^{-1}(\alpha)$ its $\alpha$ quantile , by definition of $\phi$

$$P\left\{Z \leq \frac{VaR^{\alpha} - \mu}{\sigma}\right\} = P\left\{Z \leq \phi^{-1}(\alpha)\right\} = \alpha$$

It follows that

$$\frac{VaR^{\alpha} - \mu}{\sigma} = \phi^{-1}(\alpha)$$

With a very easy algebraic manipulation

$$VaR^{\alpha} = \left[\phi^{-1}(\alpha)\right]\sigma + \mu$$

Where this last formula is a very easy way to compute the VaR at the $\alpha$ level of Significance for an investment whose returns are i.i.d and Normally distributed.
Note that to convert this value into the "Expected VaR" the only further step that has to be made is multiplying the obtained quantile by the portfolio value at the moment of estimation ($P_t$).
An alternative, yet identical in its nature, formula would be:

$$VaR^{\alpha} = \mu + z_{\alpha}\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}\omega_i\omega_j\sigma_i\sigma_j\rho_{i,j}}$$

where the portfolio standard deviation has been re-written in terms of individual standard deviations of the stocks, $\sigma_i$ and $\sigma_j$, of their weights $\omega_i$ and $\omega_j$ within the portfolio and of the linear correlation, $\rho_{i,j}$ between pairs of securities.

### 3.3.1 An Empirical VaR computation with the Normal Approach

We will now proceed with an empirical computation of the Value at Risk trough the normal approach, in this exercise:

- The VaR will be computed, using a Normal Approach, at the 1% Level of Significane.
- The considered time horizon for P&L will be one trading day.
- The data will be downloaded from Yahoo Finance and will cover the time period between 3/2/2010 and 31/12/2015.

- Two financial companies, JP Morgan Chase and Co. and Monte dei Paschi di Siena, will be considered along with two non-financial ones, Apple and Atlantia.

- RStudio will be employed in order to analyze data and compute the VaR.

The outcomes of this empirical analysis are summarized in Table 1. An interesting point to notice, in my opinion, is that the absolute value fo the maximum negative return experienced by each of these four firms lies well above the VaR, roughly three times for each of them.
This suggests that the VaR on its own is not a comprehensive measure of risk, since it does not allow us to make any prediction about what will happen once the VaR threshold has been breached.

#### 3.3.2 The Normality Bias

In the previous analysis the VaR has been computed on the assumption that stock returns are i.i.d normally distributed random variables. This assumption, however, does not fit the data thus creating a "Normality Bias" in the VaR estimators.
As it is common practice when analyzing data we will start by carrying out a visual analysis of the distribution of these data.s Figures 3, 4, 5, 6 all plot a histogram of the actual returns of each of the stocks against a Normal distribution having as parameters the mean and the variance of the respective stock.
Note how they all present two similar features:

- The empirical distribution is more peaked around the mean than the normal one.

- The tails of the empirical distributions are "fatter", i.e. more observations lie in their tales than what suggested by the normal approximation.

A further step of this visual analysis is to look at the Quantile-Quantile (Q-Q) plots of the daily returns of each distribution. In a Q-Q plot an empirical distribution is compared with a normal one by plotting their quantiles against each other.
On the horizontal axis we have the quantiles of a standard normal distribution and on the vertical one the quantiles of our empirical distributions. If the latter were normal their Q-Q plot would be a straight line, however none of Figures 7,8,9,10 exhibits such behavior, hence, from a visual point of view, we can reject the hypothesis that the daily returns of these four stocks follows a normal distribution.

To analytically test whether the daily returns of these four stocks are normally distributed we can compute the Skewness and Kurtosis of the empirical distributions of our four stocks and compare them with the one of a normal distirbution, who has a skewness of 0 and a kurtosis of 3.
The Skewness, or Third Standardized Moment of a distribution, encloses its asymmetry and is computed through the formula:

$$S = \frac{\sum_{i=1}^{n} (x_i - \mu)^3}{(\sum_{i=1}^{n} (x_i - \mu)^2)^{\frac{3}{2}}}$$

distributions with a negative skewness are said to be negatively-skewed or left-tailed distributions, whereas distributions showing a positive third standardized moment are called positively-skewed or right-tailed distributions.
The Fourth standardized Moment, or Kurtosis, of a distribution measures the "fatness" of its tails, i.e. how many observations lie within them. Distributions whose Kurtosis exceeds 3, the value of the Normal, go by the name of name of leptokurtotic.
The formula for computing such moment is:

$$K = \frac{\sum_{i=1}^{n} (x_i - \mu)^4}{(\sum_{i=1}^{n} (x_i - \mu)^2)^2}$$

As it can be read in Table 2 JPM, AAPL and MPS all show both a Skewness different from 0 and a Kurtosis greater than 3, hence their distributions do not follow a Normal one. ATL has a non-zero Skewness too but its Kurtosis is smaller than 3, thus its distribution is a mesokurtotic rather than a leptokurtotic one.
Finally we can employ a test of hypothesis to check whether the daily returns of our four stocks follow a normal distribution. The Jarque-Bera test is the most commonly employed test of hypothesis when it comes to check the normality of a dataset., Its null hypothesis is that the distribution from which data are being drawn is a normal one and the alternative states that such distribution is not Normal. The Jarque-Bera test uses the skewness and the kurtosis as inputs and generates a test statistic according to the following formula:

$$JB = \frac{n}{6}\left(S^2 + \frac{(K-3)^2}{4}\right)$$

Where $n$ is the number of observations, $S$ the skewness and $K$ the kurtosis of the distribution. This test statistic follows a $\chi^2$ distribution with two degrees of freedom. Table 3 sums up the results of the Jarque-Bera test performed over the returns of these four stocks. None of them presents a normal distribution since their JB-statistic always exceeds any critical value of the $\chi^2$ table and the associated p-values are always extremely low.

### 3.3.3 Conclusions over the Normal VaR

The normal approximation provides risk-practitioners with a very easy and general way of computing the Value at Risk, however it can can create serious biases to their computations. This is why new methodologies must be developed in order to properly compute the Value at Risk.

## 3.4 Historical Estimation

Among all the possible Non-Parametric approaches to computing the VaR this thesis will present only the Historical Method, which, according to a *McKinsey & Co.* report of 2012, is the most widely adopted method, being it employed by roughly 85% of financial institutions. The historical method considers the empirical distribution of the returns, over a specified time horizon, of the investment whose VaR has to be measured and computes its $\alpha$ quantile. This way of calculating the VaR produces results that are usually higher, in absolute value, than the ones obtained using the normal approach. Such difference is due to the fact that empirical distributions, as already seen, generally have much fatter tails than normal ones. Let $t$ denote today and call $\{R_{t+1-\tau}\}_{\tau=1}^{m}$ a sequence of $m$ past daily portfolio returns, computed using historical prices but with today's weights.
Then the histogram that represents this sequence of past returns may provide us with an idea of what will be the behavior of future returns.
The $\alpha$ % VaR of this distribution will be:

$$VaR^{\alpha} = -\{x_{\alpha} \in \{R_{t+1-\tau}\}_{\tau=1}^{m} \ : \ P\{X < x_{\alpha}\} = \alpha \ \ \forall x \in \{R_{t+1-\tau}\}_{\tau=1}^{m}\}$$

The large diffusion of this computational methodology for the VaR come from two features of this approach:

- It is very easy to estimate the VaR in this way. Because no parameter has to be estimated no numerical optimization problem has to be solved, thus there is less room for possible computational mistakes.
- The historical approach is a model-free technique, i.e. it does not rely on any model to work, thus its performances cannot be hampered by the poorness of the underlying model. This advantage implies that this methodology does not create model risk.

Despite these advantages this methodology is not free of drawbacks. If we think about $m$, the sample length, for instance we can see how the problem of determining the right value of $m$ arises. Indeed choosing a too large $m$ might hamper the validity of the estimation because most recent data, which are presumed to be the most relevant ones when it comes to forecasting tomorrow's behaviors, will be heavily underweighted. On the other hand a too small $m$ might produce a dataset which is not so representative, hence the computed $VaR^{\alpha}$ would not be so useful.

The Weighted Historical Simulation Method allows risk managers to select a sufficiently large $m$ without hampering the validity of the estimated $VaR^{\alpha}$ in that it assigns relatively more weight to more recent data and relatively less to less recent ones. Note that, since no estimations are required, the method can still be implemented very easily.
Such methodology is put in practice in the following way:

1. Probability Weights that decline over time are assigned to the sequence of past return $\{R_{t+1-\tau}\}_{\tau=1}^{m}$ as follows:

$$\lambda_{\tau} = \left\{ \lambda^{\tau-1} \frac{1-\lambda}{(1-\lambda^{m})} \right\}_{\tau=1}^{m} \qquad \lambda \in (0,1)$$

2. The observations, along with their assigned weights, are sorted in ascending order, with respect to their values.

3. The desired quantile is computed, thus the $VaR^{\alpha}$ is derived.

In an another variation to the standard historical approach Hull and White (1998) proposes to compute the VaR by implementing the historical approach over a volatility weighted sequence of returns. Such proposal relies on the idea that data taken from previous periods of relatively low volatility may underestimate future risk in periods of relatively high volatility. The other way round might happen, i.e. previous periods of relatively high volatility might overestimate risks in periods of relatively low uncertainty if a basic historical approach is applied.
Thus the $m$ returns will have to be substituted by $m$ $R^*$ rescaled returns, defined as

$$R^* = \frac{\sigma_N}{\sigma_j} R$$

where $\sigma_N$ stands for the most recent forecast of returns volatility and $\sigma_j$ is the historically estimated returns volatility.

### 3.4.1 An Empirical VaR computation with the historical approach

We will now proceed with a practical computation of the VaR employing the historical approach. Then we will compare the results with the ones obtained with the normal approach.

Table 4 shows the VaR computed using the historical method. As suggested by the theory the figures obtained using this approach exceed, in all four cases, the ones we calculated under the assumption of normality. Moreover these figures are closer to the ones of the highest loss ever experienced by each of the four considered stocks, this is a clear example of how the normality assumption would have created problems in properly managing risk.

## 3.5 Monte Carlo Simulation

A Monte Carlo Simulation corresponds to a computational algorithm that generates numerical results out of repeated random sampling. These numbers are then used to solve a formula that does not have a complete analytical form, i.e. it cannot be computed with a finite number of passages.
Let's now analyze how this computational methodology can be employed to compute the VaR of an investment.
Estimating the VaR with a Monte Carlo simulation is very similar to applying the historical approach. The main difference lies in the first step of the algorithm, as a matter of fact, rather than employing historical data regarding the returns of an investment, a Monte Carlo simulation is employed in order to obtain a batch of numerical outcomes that will represent the data starting from which the VaR will be estimated.
Once such data have been generated the computational process follows the very same path of the historical simulation approach, hence the data are sorted in ascending order with respect to their value, then the desired quantile of their distribution is computed.

### 3.5.1 An Empirical VaR computation using Monte Carlo Simulation

We will now estimate the Value at Risk of our four stocks using a Monte Carlo Simulation. To do so we will assume that stock prices move over time following a Geometric Brownian Motion.
Let $S_t$ denote the stock price at time $t$, $\mu$ the average return for the time period considered $dt$ and $\sigma$ the standard deviation over such time horizon.
Using this notation we will have that stock prices change over time according to the following stochastic differential equation:

$$dS_t = \mu dt S_t + \sigma S_t dW_t$$

where $W_t$ is a Brownian Motion, or Wiener Process, such that:

$$dW_t = \epsilon\sqrt{dt}$$

Where $\epsilon$ is a random variable distributed as a standard normal, i.e. $\epsilon \sim N(0,1)$. Thus we have that:

$$dS_t = \mu dt S_t + \sigma S_t \epsilon\sqrt{dt}$$

Approximating this stochastic differential equation with a finite difference one we get:

$$\Delta S_t = \mu\Delta t S_t + \sigma S_t \epsilon \Delta t$$

As before we are interested in estimating the daily 1% Value at Risk, hence $\Delta t = 1$.
Figure 11 shows the simulated patterns of stock prices over 100 trading days.
Table 5 reports the VaR of the four stocks compute using the method of Monte Carlo Simulation.

### 3.5.2 Some Theoretical Remarks

Readers might be interested in noticing how the maximum loss experienced by each of the four stocks in the simulated scenario is much lower, in absolute value, than the one actually recorded by historical data. Possibly this is due to the fact that we have carried out our simulation under the assumption of constant volatility, which is the case for a plain Geometric Brownian Motion. Possibly the results would have been closer to reality if we had studied the behavior of volatility over time employing either a local volatility model, such as a GARCH, for instance, or a model of stochastic volatility, i.e. modeling volatility as a separate stochastic process.
Furthermore note that the Monte Carlo simulation relied on the assumption that stock prices follow a Geometric Brownian Motion, a continuous stochastic process that does not allow for any discontinuity, hence cannot model the jumps we usually observe when dealing with stock prices. A possible way of overcoming this issue could be employing a Levy's Process, a stochastic process which can be decomposed as the sum of a Wiener process, a linear drift and a pure jump process that captures all the discontinuities of the process.
Finally our simulation assumed that stock prices are Log-Normally distributed. Indeed, the stochastic differential equation:

$$dS_t = \mu S_t dt + \sigma W_t dt$$

has the following analytic solution:

$$S_t = S_0 e^{\left[\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t\right]}$$

Note, however, that stock prices might actually behave differently, hence our Monte Carlo simulation could provide biased results due to a mistake in the assumed model.

## 3.6 Further Drawbacks of the VaR

The normality bias is not the only drawback of the Value at Risk, as a matter of fact the VaR has two more pitfalls:

- The VaR is not, in general, sub-additive, hence it is not a coherent risk measure as defined by Artzner and alius (1999).
- The VaR does not tell us anything about what happens in the left tail of the returns distribution, thus it does not allow risk-practitioners to understand what will happen once the VaR loss is exceeded.

### 3.6.1 VaR non sub-additivity

As already mentioned in Section 1 for a risk-measure $\rho$ to be sub-additive, the following condition must hold:

$$\rho : \Lambda \rightarrow \Re \cup \infty$$

$$X_1, X_2 \in \Lambda \text{ then } \rho(X_1 + X_2) \leq \rho(X_1) + \rho(X_2)$$

The following example will show that the VaR is not always a sub-additive measure, hence it cannot be generally considered as a coherent risk measure.
Assume that we are a financial institution issuing the following two binary options:

- Option A, which pays out \$ 10000 if the S&P500 returns exceed 20% at the option expiration and costs \$1000.
- Option B, which pays out \$ 10000 if wheat returns exceed 20% at the option expiration and costs \$1000.

the returns are assumed to be stochastically independent and each security will return to his buyer at least 20% at expiration with Probability $p = 0.02532$.
First of all let's write down the discrete random variable that represents our payoffs from these two derivatives:

$$\Pi_i = \begin{cases} -9000 & P = 0.02532 \\ 1000 & P = 0.97468 \end{cases} \qquad i \in \{A, B\}$$

Since this random variable is a purely binary discrete one, and not a binary approximation of a continuos one, its 5% quantile is \$1000, hence its 5%$VaR$ is -\$1000.
Assume now that we issue both options and we detain a portfolio of them, the discrete random variable that models the payoff of such portfolio will be:

$$\hat{\Pi} = \begin{cases} -18000 & P = 0.00064 \\ -9000 & P = 0.049358 \\ 2000 & P = 0.95 \end{cases}$$

where

$$P\{(R_{S\&P500} > 20\%) \cap (R_{wheat} > 20\%)\} = 0.00064$$

$$P\{(R_{S\&P500} > 20\%) \cup (R_{wheat} > 20\%)\} = 0.049358$$

$$P\{(R_{S\&P500} \leq 20\%) \cap (R_{wheat} \leq 20\%)\} = 0.95$$

the 5% quantile of this distribution is $-\$9000$, hence its 5% VaR is \$9000.
As we can see this figure is much greater than the one of the investment strategy where we were not detaining a portfolio comprehensive of both the issued options, hence the axiom of sub-additivity has been violated .
But why does this matter so much? In order to understand it we shall think about the investment technique known as Diversification.
Diversification is a Long-Only Investment Technique that reduces an investor's exposure by investing in many , possibly lowly correlated, securities, so as to minimize the idiosyncratic component of the uncertainty over the $P\&L$ of a certain portfolio .
Note, however, that such technique does not provide any element to deal with Market Risk

which works as a "Risk Floor" beneath which the riskiness of our portfolio cannot be brought without other investing techniques, the latter involving sort-selling and the use of derivative products.

Since the VaR does not comply with the many benefits provided to investors by diversification it cannot be properly used in the framework of portfolio optimization, hence other risk measures have to be created.

Moreover, when the VaR is used to compute the business risk of a firm the fact of not being sub-additive could lead to conclude that the business as a whole is riskier than the sum of individual business lines.

Furthermore, from a regulation point of view, the lack of sub-additivity might lead financial institutions to hold less capital than the actually required one.

A last remark regarding the VaR non sub-additivity. It happens that the VaR is always a sub-additive, thus coherent, risk measure when its computation is restricted to a Gaussian Family of random variables.

First of a useful definition:

**Definition 3.2.** *A family of random variables, $X_1 ..... X_n$, is said to be gaussian if, and only if, any linear combination of them is a gaussian distribution.*

Let $\Lambda$ be a gaussian family of random variables and let $X_1$ and $X_2$ be two random variables belonging to $\Lambda$, then it has to be proved that:

$$VaR\left[X_1 + X_2\right] \leq VaR\left[X_1\right] + VaR\left[X_2\right]$$

Since $X_1, X_2 \in \Lambda$ we have that $Y$ defined as $Y := X_1 + X_2$ is a Gaussian distribution with mean

$$\mu_Y = \mu_{X_1} + \mu_{X_2}$$

and standard deviation

$$\sigma_Y = \sqrt{\sigma_{X_1}^2 + 2\rho\sigma_{X_1}\sigma_{X_2} + \sigma_{X_1}^2} \qquad \rho \in (-1, 1)$$

then

$$VaR_\alpha\left[X_1 + X_2\right] = -\left(\mu_Y + \sigma_Y z_\alpha\right)$$

where $z_\alpha$ denotes the $\alpha$ quantile of the standardized normal distribution.
Note that:

$$\sigma_Y < \sigma_{X_1} + \sigma_{X_2}$$

as a matter of fact

$$\sigma_Y^2 = \sigma_{X_1}^2 + 2\rho\sigma_{X_1}\sigma_{X_2} + \sigma_{X_2}^2$$

and

$$(\sigma_{X_1} + \sigma_{X_2})^2 = \sigma_{X_1}^2 + \sigma_{Y_2}^2 + 2\sigma_{X_1}\sigma_{X_2}$$

hence the right-hand side is greater than the left-hand one for any pair of securities that are not perfectly linearly correlated, i.e. $\rho = 1$.
Since $z_\alpha < 0$ for any $\alpha$ smaller than 0.5 we have that:

$$-\mu_{X_1} - \mu_{X_2} + \sigma_Y(-z_\alpha) < -\mu_{X_1} - \mu_{X_2} + \sigma_{X_1}(-z_\alpha) + \sigma_{X_2}(-z_\alpha)$$

where

$$VaR_\alpha\left[X_1 + X_2\right] = -\mu_{X_1} - \mu_{X_2} + \sigma_Y\left(-z_\alpha\right)$$

and

$$VaR_\alpha\left[X_1\right] + VaR_\alpha\left[X_2\right] = -\mu_{X_1} - \mu_{X_2} + \sigma_{X_1}\left(-z_\alpha\right) + \sigma_{X_2}\left(-z_\alpha\right)$$

Thus the VaR is a sub-additive risk measure when it is computed for random variables that belong to a gaussian family of random variables.
Danielsson et alius (2010) argues that sub-additivity is violated only under two, not so relevant, conditions:

- The probability density function of the returns over a certain investment has particularly fat tails, classified as *super-fat tails*.
- When the confidence level is set so as to lie within the interior of the probability distribution function.

The VaR is otherwise expected to be sub-additive, hence a coherent risk-measure.

In practice the second case is of no relevancy since the VaR is computed at very low $\alpha$s. Moreover it is true that assets' return commonly exhibit a probability density function with tails fatter than the ones of a normal, i.e. a Kurtosis greater than 3, however to have a non sub-additive VaR tails must be super-fat ones, where with super-fat Danielsson et alius (2010) meant that the tails do not present a smooth behaviour but the observations lying within them might be inter-dispersed by very rare events that result in extreme losses.
Since assets' returns do not usually present such distributions the VaR might still be employed as a market risk measure without any lack of sub-additivity. However this is not a general rule, since certain securities actually present this non-smooth behaviour in the tails of their p.d.f.

## 3.7 Backtesting the VaR

As always in statistics the "true" VaR, i.e. the population parameter, cannot be directly observed and must be estimated using a certain methodology. The estimates must undergo a Backtesting procedure in order to be considered as valid ones.
The Basel Committee for Banking Supervision defines Backtesting as *"an ex-post comparison of the risk measure generated by the risk model against actual daily changes in portfolio values over longer periods of times, as well as hypothetical changes based on static positions.*
A typical Backtesting procedure involves setting a time period over which the number of VaR violations will be computed, then their number will be divided by the number of total observations and such ratio gives the violation rate for the considered time horizon.
The Kupiec test, which also goes by the name of POF-test (Proportion of Failures), is the most widely employed statistical tests in this contest and examines whether the observed violation rate is statistically equal to the expected one.
The null hypothesis of the test is:

$$H_0 : p = \hat{p} = \frac{x}{T}$$

where $x$ stands for the number of observed violations and $T$ for the total number of observations.
The test statistic for the Kupiec test, known as the Likelihood-Ratio, will be:

$$LR = 2ln\left(\left(1-\frac{x}{T}\right)^{T-x}\left(\frac{x}{T}\right)^{x}\right) - 2ln\left((1-q)^{T-x}\,q^{x}\right)$$

exploiting the properties of logarithm functions

$$LR = 2ln\left(\frac{\left(1-\frac{x}{T}\right)^{T-x}\left(\frac{x}{T}\right)^{x}}{(1-q)^{T-x}\,q^{x}}\right)$$

and such statistic is asymptotically $\chi^2$ distributed with one degree of freedom. Table 6 presents the LR statistics for the 1% Non Parametrically computed Value at Risk of the four stocks analyzed so far.

As we can see none of the four LR statistics actually exceeds the critical value of 6.635, thus, for the four analyzed stocks, we cannot reject the Null Hypothesis that the estimated violation rate is equal to the true one at the 1% level of significance.

# 4 Beyond the VaR: the CVaR and Tail Distributions

## 4.1 The Conditional Value at Risk

The Value at Risk presents decision makers with a loss which is unlikely to be exceeded, however they cannot figure out anything about what will happen once this loss has been exceed if they are not provided with further tools.
The Conditional Value at Risk (CVaR), presented in Rockafellar and Uryasev (2000), is another Risk Measure that tells us which is the expected loss once the VaR threshold has been breached.
More formally:

**Definition 4.1.** *Let $X$ denote the returns, over a defined time horizon, of an investment. For any $\alpha \in (0,1)$ the Conditional Value at Risk, or Expected Shortfall or Expected Tail Loss, of $X$ is defined as* :

$$CVaR^{\alpha}\left[X\right] := -E\left[X|X \leq VaR^{\alpha}\right]$$

Given a sample $X$ of $n$ returns of a certain investment $x_i,\ i \in [1,n]$, let $x_\alpha$ be its $\alpha$ VaR then the population $CVaR^{\alpha}$ can be estimated by the following sample statistic:

$$\widehat{CVaR}^{\alpha}\left[X\right] = -\frac{\sum_{i=1}^{\alpha} x_i}{\alpha}$$

Acerbi and Tasche (2001) has proven that this estimator asymptotically converges to the population CVaR.
Differently from the $VaR^{\alpha}$, the $CVaR^{\alpha}$ is a coherent risk measures, since it satisfies all the axioms of coherence, hence it can be used without all the drawbacks arising from the lack of sub-additivity.
Acerbi and Tasche (2001) provides an easy proof of this axiom adopting the sample CVaR, which is here reported:

$$\widehat{CVaR}^{\alpha}\left[X_1 + X_2\right] = -\frac{\sum_{i=1}^{\alpha} \left(X_1 + X_2\right)_i}{\alpha}$$

$$\widehat{CVaR}^{\alpha}\left[X_2\right] + \widehat{CVaR}^{\alpha}\left[X_2\right] = -\frac{\sum_{i=1}^{\alpha} \left(X_{1,i} + X_{2,i}\right)}{\alpha}$$

since $\sum_{i=1}^{n} \left(a_i + b_i\right) = \sum_{i=1}^{n} a_i + \sum_{i=1}^{n} b_i$ , we have that:

$$-\frac{\sum_{i=1}^{\alpha}\left(X_{1,i}+X_{2,i}\right)}{\alpha}=-\frac{\sum_{i=1}^{\alpha}X_{1,i}}{\alpha}-\frac{\sum_{i=1}^{\alpha}X_{2,i}}{\alpha}$$

hence

$$-\frac{\sum_{i=1}^{\alpha}\left(X_{1}+X_{2}\right)_{i}}{\alpha}\leq-\frac{\sum_{i=1}^{\alpha}\left(X_{1,i}+X_{2,i}\right)}{\alpha}$$

thus

$$\widehat{CVaR}^{\alpha}\left[X_1+X_2\right]\leq\widehat{CVaR}^{\alpha}\left[X_1\right]+\widehat{CVaR}^{\alpha}\left[X_2\right]$$

The central column of table 7 shows the 1%CVaRs computed under the normality assumption, whereas the right column contains the estimates obtained using the historical 1% VaR. Figure 12 provides an illustration of the CVaR in a loss distribution function.

### 4.1.1 Properties of the CVaR

- By construction the $CVaR^{\alpha}$ will always be lower, or at most equal to, the $VaR^{\alpha}$.
- For $\alpha=1$ the $CVaR^{\alpha}$ has the same value of the Portfolio Expected Return.
- As $\alpha$ increases the $CVaR^{\alpha}$ increases too.
- The $CVaR^{\alpha}$ is a Coherent Risk Measure.

### 4.1.2 CVaR based Portfolio optimization

In 1952 the Nobel Prize Economist H. Markowitz introduced the Modern Portfolio Theory, or Mean-Variance Analysis, a mathematical framework for constructing an asset portfolio either by minimizing its variance, $\sigma_P^2$, for a given Expected Return, $E\left[r_P\right]$ or by maximizing the latter for a given level of risk.
Let $\boldsymbol{\omega}$ be a normalized vector, i.e. $\sum_{i=1}^{n}\omega_i=1$, whose $i$-th component is the portfolio weight of the $i$-th security.

$$\boldsymbol{\omega}=\begin{bmatrix}\omega_1\\ \omega_2\\ \vdots\\ \omega_n\end{bmatrix}$$

$E[r_i]$ is the expected return over the $i$-th security and the $i$-th component of the vector $\boldsymbol{r} \in \Re^n$:

$$\boldsymbol{r} = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix}$$

$\sigma_i^2$ is the variance of the returns of the $i$-th security and measures the risk associated to it. As to what regards the portfolio $P$, composed of the $n$ assets, we have that:

$$E[r_P] = \boldsymbol{\omega}^t \mathbf{r}$$

$$\sigma_P^2 = \boldsymbol{\omega}^t \boldsymbol{\Sigma} \boldsymbol{\omega}$$

Where $\boldsymbol{\Sigma} \in Mat\{n, \Re\}$ is a Symmetric square Matrix known as Variance-Covariance Matrix such that its entry $\sigma_{i,i}$ is the variance of the $i$-th security and $\sigma_{i,j}$ is the covariance between the $i$-th and the $j$-th investment.

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{1,2} & \dots & \sigma_{1,n} \\ \sigma_{2,1} & \sigma_2^2 & \dots & \sigma_{2,n} \\ \vdots & \vdots & \dots & \vdots \\ \sigma_{n,1} & \sigma_{n,2} & \dots & \sigma_n^2 \end{bmatrix}$$

s

Then, in a mean variance analysis, the optimization problem that will be solved is either:

$$\begin{cases} \max\limits_{\omega} & \boldsymbol{\omega}^t \mathbf{r} \\ \text{sub to} & \boldsymbol{\omega}^t \boldsymbol{\Sigma} \boldsymbol{\omega} = \overline{\sigma}^2 \end{cases}$$

Where $\overline{\sigma}^2$ denotes a specified level of $\sigma^2$.

Or:

$$\begin{cases} \min\limits_{\omega} & \boldsymbol{\omega}^t \boldsymbol{\Sigma} \boldsymbol{\omega} \\ \text{sub to} & \boldsymbol{\omega}^t \mathbf{r} = \overline{E}[r_P] \end{cases}$$

Where $\overline{E}[r_P]$ denotes a specified level of $E[r_P]$. If the equality signs in the constraints of both problems are substituted with an "at most", i.e. $\leq$, in the first problem and with an "at least", i.e. $\geq$, in the second one the two optimization problems become two Soft Constrained Problems.

The problem with such theory is that it assumes stock returns to be normally distributed.

Furthermore it uses the variance as a proper measure of risk. This parameter, however, tells us how dispersed the returns are, thus assigning equal importance to extreme profits and extreme losses, not taking into account that risk managers should care more about abnormal losses rather than abnormal profits.

Starting from Rockafellar and Uryasev (2000) a new approach for optimizing a portfolio so as to reduce the risk of high losses has been introduced. Instead of using the portfolio variance or Value at Risk the two authors decided to build a portfolio aimed at minimizing the CVaR for a given level of returns.

Minimizing the CVaR has been proven to be computationally more efficient than minimizing the VaR, moreover a portfolio obtained through CVaR minimization will be a VaR minimizing one too, due to the strict relation between the two measures given by the CVaR definition.

However it has been shown that, under the normality assumption, optimizing a portfolio using VaR minimization, CVaR minimization or variance minimization yields the same results.

Krokhmal et al (2001) extended the CVaR minimization approach to a return maximization problem with one, or more, CVaR constraints.

We will now proceed with the optimization of a portfolio in the following three ways:

- Mean-Variance Analysis.
- 1% VaR Minimization, with the Historical VaR.
- 1% CVaR Minimization, with the Historical CVaR.

The 10 chosen stocks all belong to the FTSE.MIB Index, which comprises the 40 largest Italian firms by Market Cap. For the sake of simplicity the Risk-free rate has been set to 0. Daily returns have been considered and RStudio has been employed, more specifically the package "ghyp".

Table 8 shows the weights of each of the stocks in the three different portfolios, note that a negative weight denotes the fact that the stock is being shorted. As we can see from Table 9 the portfolio obtained by the CVaR-minimization approach is the one that performs better

both in terms of Sharpe Ratio, i.e. the return per unit of standard deviation, and in terms of returns per unit of CVaR.
Moreover the portfolio obtained in such a way has the same VaR of the one obtained by minimizing the VaR. The latter is clearly a poor approach, as a matter of fact not only it has the highest CVaR among the three, but its VaR is even higher than the one of the portfolio obtained through the Mean-Variance Analysis.
Figure 13 shows three dots, each of them corresponding to one of the three optimal portfolios we have just found. More precisely the blue one corresponds to the CVaR minimizing portfolio, the red dot represents the Mean-Variance one and the green one is the VaR minimizing portfolio. As we can see the first portfolio is the one that, for the lowest CVaR, offers the highest possible expected return, hence it is the best choice in terms of hedging against possible abnormal losses.

### 4.1.3 Backtesting CVaR

Despite the strict link between the CVaR and the VaR the backtesting procedure we applied in Chapter 3 for the Value at Risk cannot be applied to theCVaR. Yamai and Yoshiba (2004) consider backtesting the Conditional Value at Risk as a much more difficult procedure to implement than the one for the VaR.
Tasche et al. (2013) proposed to backtest the CVaR by backtesting several VaR at different level of confidence.
Given a loss function $L$:

$$CVaR^{\alpha}\left[L\right] \sim \frac{1}{4}\left[VaR^{\alpha}\left[L\right] + VaR^{0.75\alpha+0.25}\left[L\right] + VaR^{0.5\alpha+0.5}\left[L\right] + VaR^{0.25\alpha+0.75}\left[L\right]\right])$$

If the four different VaR are successfully backtested then the estimate of the $CVaR^{\alpha}$ can be considered as a reliable one too. This backtesting procedure is particularly interesting because it shows that the CVaR cannot be directly backtested, thus plenty of data are needed to reliably backtest it.

## 4.2 Tails Distribution

Extreme Value Theory, EVT, is the branch of Statistics that deals with extreme deviations from the central tendency of a distribution and so so by employ several probability distribution functions.

Recently risk-practitioners have started employing EVT to create models that well fit the tails of stock returns distributions. Since many securities present empirical distributions with particularly heavy left-tails a widely used approach to the problem of finding a proper distribution model for the returns is fitting the tails separately from the interior of the distribution .

The first step of this separate fitting process is to divide the returns distribution into three segments:

- The Left Tail
- The Interior
- The Right Tail

Then a specific distribution function is fitted to each of the three segments.

The Pareto Distribution is a power law probability distribution largely employed to model the tails of other distributions and is widely employed to model extreme situations in the world of finance.

If $X$ has a Pareto Distribution with parameters $x_m$ and $\alpha$ its cumulative distribution function will be:

$$F(X) = \begin{cases} 1 - \left(\frac{x_m}{x}\right)^{\alpha} & x \geq x_m \\ 0 & x < x_m \end{cases}$$

It follows, by differentiation, that its probability density function is:

$$f(x) = \begin{cases} \frac{\alpha x_m^{\alpha}}{x^{\alpha+1}} & x \geq x_m \\ 0 & x < x_m \end{cases}$$

Where $\alpha$ is the Shape Parameter and $x_m$is the Scale Parameter, or Lower Bound Parameter, of the distribution.

Maximum Likelihood Estimators can be used to estimate such parameters. Given $n$ i.i.d. observations $\{x_1, x_2, ....., x_n\}$ coming from an unknown population whose probability density function is $f_\theta$, defined by a parameter $\theta$.

Thanks to the fact that observations are i.i.d. we have that the corresponding joint density function can be written as:

$$f(x_1, x_2, ...., x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

We can look at it from a different perspective, keeping the $x_i$ fixed and letting $\theta$ be the independent variable of our function. In this case the joint density function becomes a so-called likelihood function:

$$L(\theta|x_1, x_2, ...., x_n) = \prod_{i=1}^{n} f(x_i|\theta)$$

For the sake of algebraic simplicity we can take the natural logarithm of these function and work with it:

$$ln[L(\theta|x_1, x_2, ..., x_n)] = \sum_{i=1}^{n} ln[f(x_i|\theta)]$$

And this is the log-likelihood function.
The maximum likelihood estimator $\hat{\theta}$ is that $\theta$ such that the log-likelihood function, $l(\theta|x_1, x_2, ...., x_n)$, is maximized with respect to $\theta$:

$$\hat{\theta} \in \underset{\theta}{\operatorname{argmax}}\, ln[L(\theta|x_1, x_2, ...., x_n)]$$

Note that the maximum likelihood estimator of $\theta$ does not change if we maximize the likelihood function rather than the log-likelihood one in that the latter is monotonic transformation of the former.
Back to the Pareto distribution, the shape and scale, $\alpha$ and $x_m$ parameters can be estimated by:

$$\begin{cases} \hat{\alpha} = \frac{n}{\sum_{i=1}^{n} log\left(\frac{x_i}{\hat{x}_m}\right)} \\ \hat{x}_m = \min(x_i) \qquad i \in [1, n] \end{cases}$$

Column two and column three of Table 10 report the maximum likelihood estimators for the Pareto Distributions used to model the left tails of the four stocks in the example of section 2.

The third one reports the KS statistic for each of the empirical distributions with respect to the Pareto Distribution. The Kolmogorov-Smirnov is a test of the equality of continuous, one-dimensional, probability distributions that can be used to compare a sample with a reference probability distribution.
Its null hypothesis is that the variable at stake is distributed as the distribution against which the empirical results are being compared, here the Pareto one.
The KS statistic is given by:

$$KS_n = \sup_n |F_n(X) - F(X)|$$

Where $n$ is the number of observations in the sample, $F_n(X)$ represents the empirical distribution of the sample and F(X) is the continuos distribution against which we are testing the empirical one.
As the outcomes of the test shows we can never reject the null hypothesis that the returns beyond the VaR can be fitted by a Pareto Distribution, hence the latter can be used to model their behavior under "extreme circumstances", i.e. once the VaR threshold has been broken.
Having found a model that beautifully fits the tails distribution we now have to find a way to model what happens in the interior of the distribution. However the issue at stake here is not so relevant and a simple non-parametric, historical, distribution can be employed to model such segment.
Following McNeil (1999) the $VaR^q$ estimate of a given Pareto Distribution with scale parameter $x_m$ and shape parameter $\alpha$ is given by:

$$VaR^q = u + \hat{x}_m \left[ \left( \frac{n}{N_u} (q)^{\frac{1}{\hat{\alpha}}} - 1 \right) \right]$$

.

where $u$ is a loss threshold, $n$ is the total number of observations, $N_u$ is the number of observations exceeding $u$ and $\hat{x}_m$, $\hat{\alpha}$ are the MLE estimators of $x_m$ and $\alpha$, respectively.
As Gencay and Selcuk (2004) shows and EVT based approach to the VaR computation performs much better than other, more conventional one, when dealing with extreme levels of significance, i.e. very low $\alpha$s ($\leq 1\%$).

# 5 A further Application of the CVaR: Index Tracking.

## 5.1 Introduction to the Problem

Recently the word of Asset Management has experienced a drastic increase in the demand for index-investing. Where with the term ”index-investing” we mean the purchase of securities that closely track the performances of an index, thus granting investors consistently positive returns while providing them with a largely diversified investment obtained for very low transaction costs.
Investors can choose between two types of index-tracking securities to invest in :

- Mutual Funds, which are actively managed funds that pool money from many investors in order to invest in securities. Because professionals, namely Portfolio Managers, actively manage these funds investors have to pay fees in order to keep their share of the fund,

- Exchange-Traded Funds (ETFs). Investors purchase or sell a share of these passively managed funds directly on an exchange. Because ETFs are traded on organized markets they are much more liquid instruments than actively managed funds, which, on the other hand, are exchanged on OTC markets.

Both these instruments closely follow an index, and they can do it in two ways: either Physical or Synthetic Replication.
Physical Replication comprises all those techniques that directly invest in the assets, all or only part of them, belonging to the index. Synthetic Replication matches the performances of an index by using derivative products structured over assets belonging to the index.

## 5.2 Creating an Index

In this work we will deal with the creation of an Not-Leveraged Exchange Traded Fund that aims at tracking a stock-index, whose assets are all stocks belonging to the European aviation industry (airlines, airport managers and aircrafts manufacturers). The ETF will partially be created trough a CVaR-minimization approach, applying what proposed by Rockafeller and Uryasev (2002).
The first step to create an ETF is constructing an index to track. Indexes can be created in many different ways, some examples are:

- Price Weighting, where stocks' weights are assigned proportionally to their relative price in the index.
- Market Capitalization Weighting, here each stock is given a weight equal o the ratio between its Market Capitalization, i.e. Price of the stock times number of them, and the sum of all the Market Capitalizations of the assets belonging to the index.
- Equal Weighting, all assets are given the same weight within the index.
- Fundamental Weighting, in this approach each asset is given a weight proportionally to one, or more, fundamental such as revenues, book value and so on.

The index we will create is a market capitalization weighted one, thus the $i$-th stock will be given a weight, $\omega_i$ according to the following rule:

$$\omega_i = \frac{P_i Q_i}{\sum_{j=1}^{n} P_j Q_j}$$

where $P_i$ stands for the price of the $i$-th stock and $Q_i$ for the number of its outstanding shares.

## 5.3 Creating the ETF

The ETF will be a Physically Replicating one, and it will adopt the Sampling Replication Methodology, i.e. only some of the assets belonging to the index will be bought by the fund. The sampling and weighting procedure will take place in two steps:

1. First of all a restricted number of stocks will be chosen to be part of the ETF. Stocks will be sampled using deep-learning techniques.
2. Once the stocks have been picked their weights within the ETF will be the result of the linear programming problem associated to the minimization of the 1% CVaR of the fund tracking error

### 5.3.1 Sampling with deep learning

As proposed by Heaton et al. (2016) we will proceed with a deep learning approach to the problem of selecting which stocks to buy. First of all we shall understand how deep learning works.

As a form of machine learning it trains a model over a set of data in order to make predictions, but, differently from classic machine learning techniques, here data features are passed through subsequent layers of abstraction so that factors extracted from the features of the $j$-th layer become the features of the $(j+1)$-th layer.
This is why deep learning is often referred to as a hierarchical algorithm.
Given a dataset $X \in \{Mat\,(m, n, \Re)\}$, where $n$ is the number of observations and $m$ the number of variables observed for each of them, and a $P$-layers deep learning algorithm, $F(X)$, we can define $P$ univariate, element-wise, activation functions, $f_i$ $i \in [1, P]$, so that $F(X)$ can be seen as the iterated composition of the P functions from layer 0, denoted by $Z^0$, to the desired output, which can be seen as the *P+1* Layer $Z^{P+1}$ or, as always in literature, Y. Having said this we can write our predictor Y as:

$$Y = F\,(X) = (f_P \circ .... \circ f_1)\,(X)$$

Which, labelling the $P$-th layer with $Z^P$, can be extensively written as:

$$\boldsymbol{Z^1} = f_1\left(\boldsymbol{W^0 X} + \boldsymbol{b^0}\right)$$

$$\boldsymbol{Z^2} = f_2\left(\boldsymbol{W^1 Z^1} + \boldsymbol{b^1}\right)$$

$$\vdots$$

$$Y = \boldsymbol{Z^{P+1}} = \boldsymbol{W^P Z^P} + \boldsymbol{b^P}$$

Where $\boldsymbol{W^p}$, $p \in [1, P]$, are $(d * m)$ weight matrices. $\boldsymbol{b^p}$ are thresholds vectors of $\Re^m$. These activation vectors play a key role in deep learning structures, as a matter of fact consider:

$$\boldsymbol{C^p} = \boldsymbol{W^p X} \qquad \boldsymbol{C^p} \in \{Mat\,(d, n)\} \quad \forall p \in [1, P]$$

and

$$\boldsymbol{b^p} = \begin{bmatrix} b_1^p \\ b_2^p \\ \vdots \\ b_d^p \end{bmatrix} \qquad \forall p \in [1, P]$$

Then the activation function of the $p$-layer will be applied if, and only if, the elements of the $i$-th row of $\boldsymbol{C^p}$ are all greater than the $i$-th entry of $\boldsymbol{b^p}$. Namely, fixed a row $i$,:

$$c_{i,j} > b_i \quad \forall j \in [1, n]$$

$Z^p$ are hidden features (or factors) extracted by the algorithm and used as inputs for the next transformation or Layer.
P, the number of Layers, is the Depth of the algorithm.
This characteristic hierarchical structure enables deep learning predictors to be much less biased than common machine learning ones. As a matter of fact the latter suffer more from the over-fitting problem, i.e. capturing noise rather than actual connections only, since their one-layer structure does not allow successive filtering of information.
Moreover deep learning is not based on linear-only models, thus it is able to capture more complex and non-linear relations between the observed data.
Given a dataset $X$ an auto-enconder is a deep learning routine that works to approximate $X$ by itself, i.e. $X = Y$, via a bottleneck structure. A model, $F_W(X)$, is selected so as to concentrate in a small number of nodes, within one or more hidden layer, the information required to recreate $X$. Then, starting from the nodes of the lasy hidden layer, all the initial observations are recreated in their auto-encoded versions.
Fig 14 might give the reader a better idea of how a deep learning algorithm is structured.
Since we always want transformations to take place we will set $\boldsymbol{b^p} = \mathbf{0}\ \forall p \in [1, P]$, where $\mathbf{0}$ is the zero vector of $\Re^d$.
Such a model is specified in order to solve the following minimization problem:

$$\min_{\mathbf{W}} \quad ||\mathbf{X} - F_{\mathbf{W}}(\mathbf{X})||^2$$

where $W$ is, as always, a weight matrix for the activation functions applied to the data of $X$.
Once the stocks have been auto-encoded we select the 10 whose Auto-Encoding Euclidean Norm, i.e. the norm of the vector of the difference between the original data and the new ones, is lower. As a matter of fact the auto-encoder reduces the total information to a compressed information set from which all stocks can be re-created. Thus, for indexing purposes, the stocks whose auto-encoded reproduction is closer to the original one can be thought as a basis for the complete set of information contained in the original dataset.
Note that the algorithm has been applied to a matrix of daily returns for the 20 stocks of

our index, spanning over the time horizon 12/03/2012-31/12/2015.
Table 11 shows the 20 stocks of our index sorted in ascending order with respect to the norm of the vector of the difference between the original data and the new ones. The first 10 ones are those that would be bought by our ETF at it inception.

### 5.3.2 Weighting the Stocks

As already said we will assign weights to stocks according to a minimization procedure of the ETF Tracking Error. This way of weighting stocks has beeb chosen
First of all the notion of tracking error shall be introduced, the tracking error, *TE*, is a measure of the difference between the performances of the Index and the ones of the ETF. Let $R_{I,t}$ be the index returns at time $t$, over a pre-specified period of time, and $R_{P,t}$ the ones of the ETF over the same time period, we can denote the tracking error at time $t$, $TE_t$, as:

$$TE_t = R_{I,t} - R_{P,t} = R_{I,t} - \sum_{j=1}^{10} w_j r_{j,t}$$

Where $\omega_j$ is the weight of the $j$-th stock within the replicating portfolio and $R_{j,t}$ denotes its return at time $t$. Since we want to minimize the absolute value of the tracking error its sign does not matter. Hence we will deal with $|R_{i,t} - R_{P,t}|$.
We will now proceed with minimizing the $CVaR_\alpha$ of the distribution of $|R_{i,t} - R_{P,t}|$.
To do so we shall consider the $CVaR_\alpha$ as proposed by Rockafeller and Uryasev (2000) for optimization purposes:

$$CVaR_\alpha = \epsilon + \frac{1}{T\left(1-\alpha\right)}\left[\sum_{t=1}^{T} max\left(0, |TE_t| - \epsilon\right)\right]$$

$$max\left(0, |TE_t| - \epsilon\right) = \begin{cases} |TE_t| - \epsilon & if \quad TE_t \geq \epsilon \vee TE_t \leq \epsilon \\ 0 & if \ -\epsilon < TE_t < \epsilon \end{cases}$$

Where $\epsilon$ represents the $\alpha$ % Value at Risk, $T$ is the number of time periods over which the Tracking Error has been recorded and $\alpha$ specifies the probability that the Tracking Error is greater than $\epsilon$.

We can now formulate the Linear Programming Problem whose solutions will provide us with the optimal weights of our ten stocks in this ETF:

$$\begin{cases} \min_{\boldsymbol{W},\epsilon} & CVaR_\alpha = \epsilon + \frac{1}{T(1-\alpha)}\left[\sum_{t=1}^{T} max\left(0, |TE_t| - \epsilon\right)\right] \\ sub\,to & \sum_{j=1}^{10} w_j = 1 \\ & w_j \geq 0 \quad \forall j \in [1, J] \\ & t \in [1, T] \end{cases}$$

Where the first two conditions are the standard ones for a long-only portfolio, i.e. the weights must sum-up to one and cannot be negative, i.e. short-selling is forbidden.
The vector of weights that solves the above constrained minimization problem is reported in table 12.

## 5.4 Why This Method

One might ask why we have chosen to minimize the Tracking Error CVaR to create our ETF. As a matter of fact there are plenty of more common techniques, all based on the minimization of a quality measure of the TE with respect to the vector of weights $\boldsymbol{\omega}$.
Namely:

$$\min_{\boldsymbol{\omega}} \; measure\,(TE)$$

To choose which measure to minimize we can consider the regression of the returns of the benchmark index against the ones of the tracking portfolio.
Using the same notation as before, and considering the Tracking Error as the error term, we have:

$$R_{I,t} = R_{P,t} + b + \epsilon_t = \sum_{i=1}^{n} \omega_i r_i + b + TE_t^*$$

Where $b$ is a constant difference between the two returns, usually set so as to be 0. Re-writing The Tracking Error in its explicit form, $R_{I,t} - R_{P,t}$, we can get a measure of the Tracking Error adjusted for this bias, $b$.

$$TE_t^* = R_{I,t} - R_{P,t} - b = TE_t - b$$

It is possible to show that the best tracking quality measure to minimize depends on the Probability Distribution of the error term.
Despite being common practice, estimating weights through this regression approach has two main drawbacks:

- First of all, as shown in Bamberg and Wagner (2000), the assumptions of the Linear Regression Model are being violated, as a matter of fact:
  - $E[TE^*] \neq 0$
  - Tracking Errors are not independent.
  - The Variance of Tracking Errors is not constant.
  - The explained and the explanatory variables are correlated.
  - The regression coefficients, here the weights, are not time invariant.
- As explained in Hansen (2012) the constraints used in the optimization problem will add bias to the estimators.

To get rid of these issues the literature proposes to minimize robust statistics, such ad the CVaR.

# 6 Conclusions

The thesis showed how relevant is for financial institutions to properly deal with market risk. In this regard it presented some quantitative tools practitioners commonly employ to measure and control market risk. Furthermore the thesis showed how quantitative risk measures can be efficiently used for both active and passive investing.

As to what concerns active portfolio management the Conditional Value at Risk turned out to be an efficient substitute for the variance in the framework of Modern Portfolio Theory. For what regards passive investing, namely an Exchange Traded Fund that tracks an index, minimizing the Conditional Value at Risk of the distribution of the Tracking Error has proved to be an efficient way of weighting stocks within the ETF.

I can see at least three ways of further expanding the ETF problem:

- When an ETF is structured to replicate the performance of an index composed of stocks with low liquidity its managers might find troublesome to adjust its composition to keep on closely tracking the benchmark. Indeed lowly liquid assets are characterized by higher trading costs, and it would be interesting, in my opinion, to quantify how these increasing transaction costs would hamper investors' returns.

- In my analysis I created an ETF which exactly tracks the benchmark, thus providing investors with the same return. What if the ETF was created in a way so as to incorporate both historical data regarding the stocks and public and private information concerning the underlying assets? Possibly the replicating fund will be able to outperform the benchmark while still exhibiting an appreciable co-movement with the underlying's returns

- One might try to sample stocks using different statistical approaches, such as Principal Component Analysis or Bayesian Shrinkage for large samples,Koop et al. (2016), and see whether the ETFs built in this way are able to outperform the one structured with the two-step algorithm here developed.

For what regards the empirical computation of the Value at Risk using Monte Carlo simulations I see at least two ways of extending what presented in the thesis:

- It could be possible to carry out the Monte Carlo simulation implementing a model, either stochastic or deterministic, to include dynamic volatility, i.e. volatility that changes over time.

- One might even change the underlying model for stock prices. Indeed more recent results in quantitative finance rely on the application of Levy Processes, an already introduced class of stochastic processes, to model the returns on financial securities . See Wu (2008), which surveys the use of Levy Processes in finance, to gain better knowledge of the issue.

# 7 References


1. *Non-normality of Market Returns. A Framework for Asset Allocation Decision-Making.* H. Qiao, A.Z. Sheikh, 2009
2. *Empirical Study of Value-at-Risk and Expected Shortfall Models with Heavy Tails.* F. C. Harmantzis, L. Miao Y. Chien,2005
3. *Value at Risk Models.* C.Alexander, 2008
4. *Econometric Modeling of Value-at-Risk.* T. Angelidis, S. A. Degiannakis.
5. *Some Remarks on the Value at Risk and the Conditional Value at Risk.* G.C.Pflug, 2000
6. *Optimization of Conditional Value at Risk.* R.T. Rockafeller, S.Uryasev, 2000
7. *Portfolio Optimization with Conditional Value at Risk Objective and Constraints.* P. Krokhmal, J.Palmquist, S.Uryasev, 2001
8. *On the Coherence of Expected Shortfall.* C.Acerbi, Tasche, 2001
9. *Exploring Heavy Tails, Pareto and Generalized Pareto Distributions.* K.Hornik
10. *Elements of Financial Risk Management.* P. F. Christoffersen, 2011
11. *Incorporating Volatility Updating into the Historical Simulation Method for Value at Risk*, J.Hull, A. White, 2001
12. *Extreme Value Theory for Risk Managers*, A.J.McNeil, 1999
13. *Extreme value theory and Value-at-Risk: Relative performance in emerging markets*, R. Gencay, F. Selcuk, 2004
14. *Subadditivity Re-Examined, the Case for Value at Risk*, J.Danielsson et alius, 2010
15. *Techniques for Verifying the Accuracy of Risk Management Models*, P.Kupiec, 1995
16. *Comparative Analysis of Expected Shortfall and Value at Risk: Their Validity Under Markte Stress*, Y. Yamai, T.Yoshiba, 2004
17. *What is the best risk measure in practice? A comparison of standard measures*, Tasche et al.,2013

18. *Conditional Value-at-Risk for general loss distributions.* R.T. Rockafeller, S.Uryasev, 2002

19. *Deep Learning in Finance.* J.B. Heaton et alius, 2016

20. *Deep Portfolio Theory* . J.B. Heaton et alius, 2016

21. *Econometrics.* Hansen, 2012

22. *Equity index replication with standard and robust regression estimators.* G. Bamberg, N. Wagner, 2000.

23. *Bayesian Compressed Vector Autoregression.* G.Koop, D. Korobilis, D. Pettenuzzo, 2016.

# 8 Figures

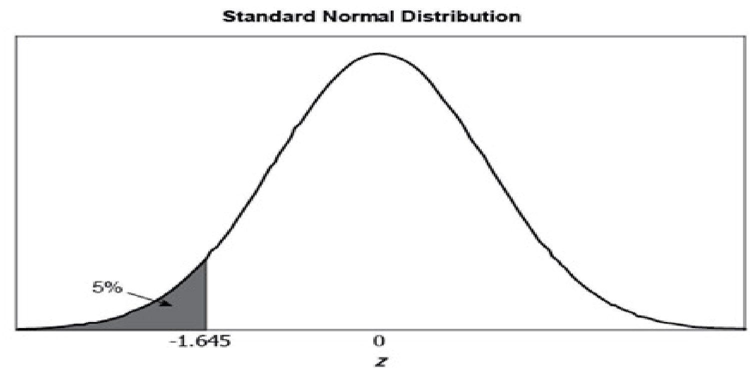


Figure 1: 5% quantile of a SND

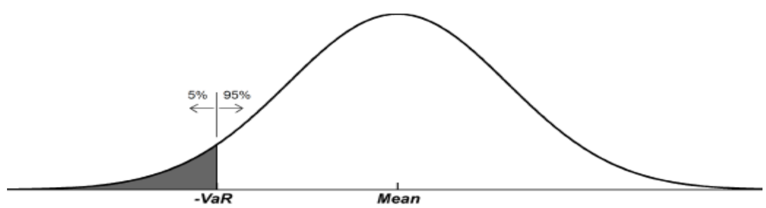


Figure 2: 5% VaR

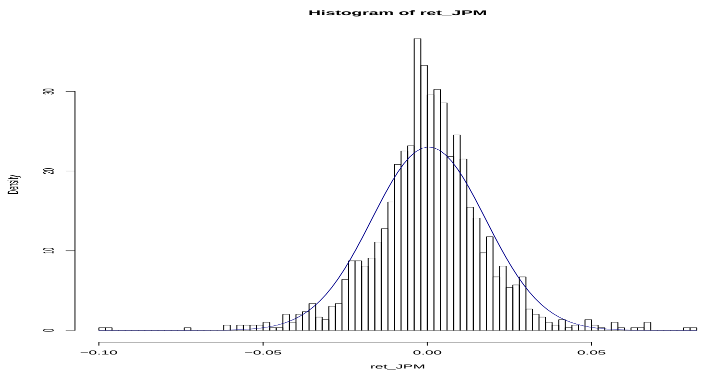


Figure 3: JPM Histogram

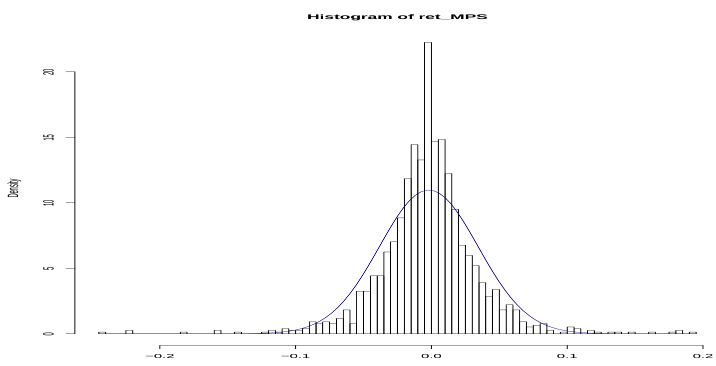


Figure 4: MPS Histogram

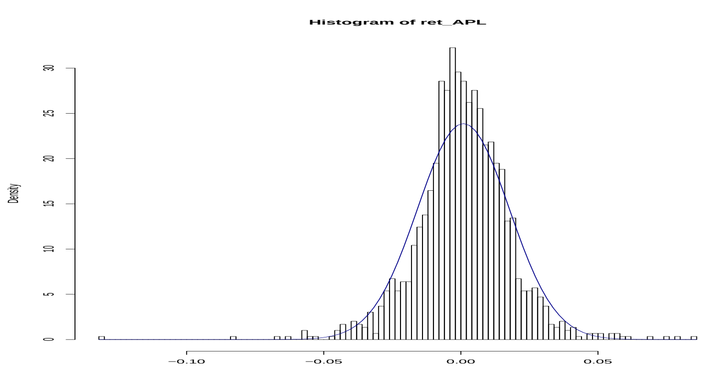


Figure 5: APL Histogram

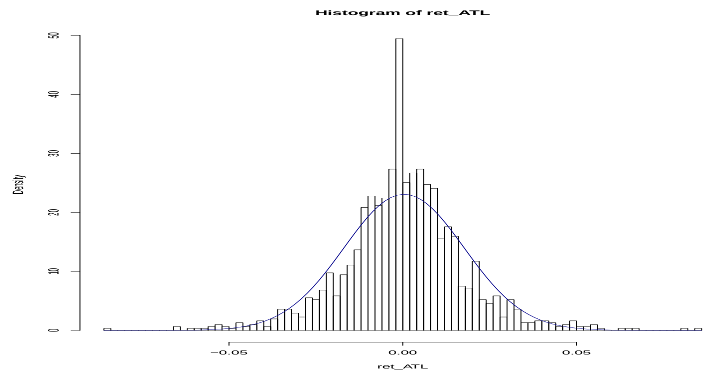


Figure 6: ATL Histogram

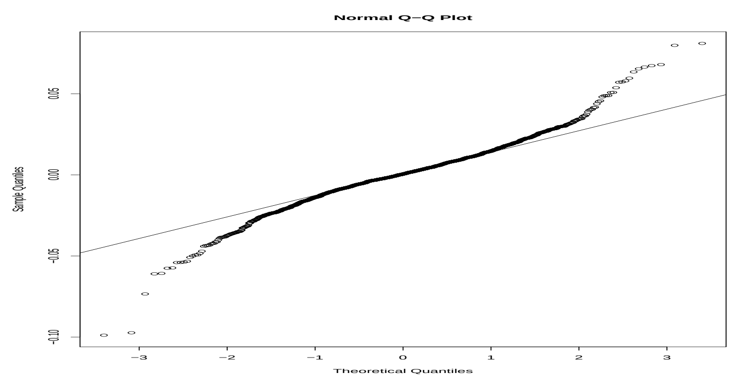


Figure 7: JPM Q-Q Plot

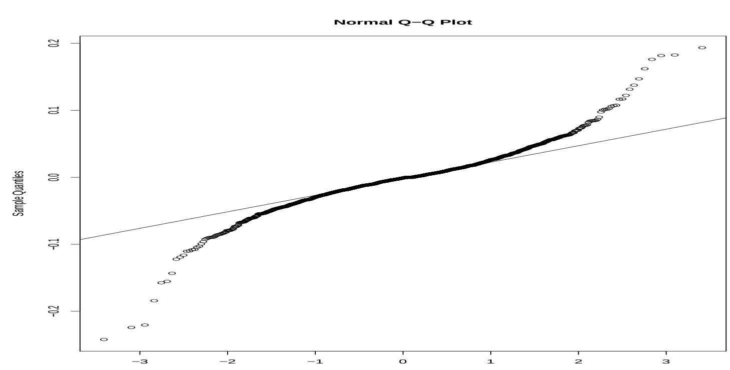


Figure 8: MPS Q-Q Plot

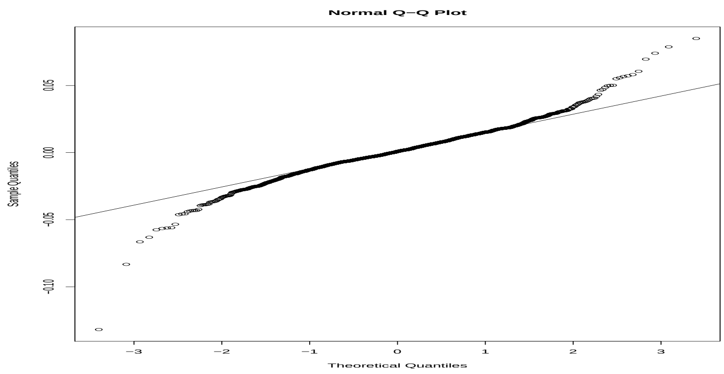


Figure 9: APL Q-Q Plot

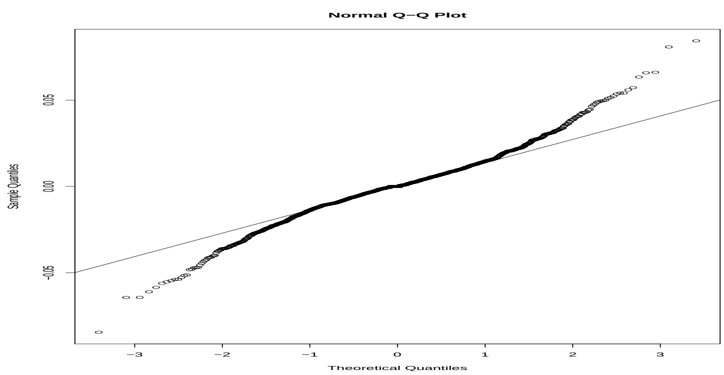


Figure 10: ATL Q-Q Plot

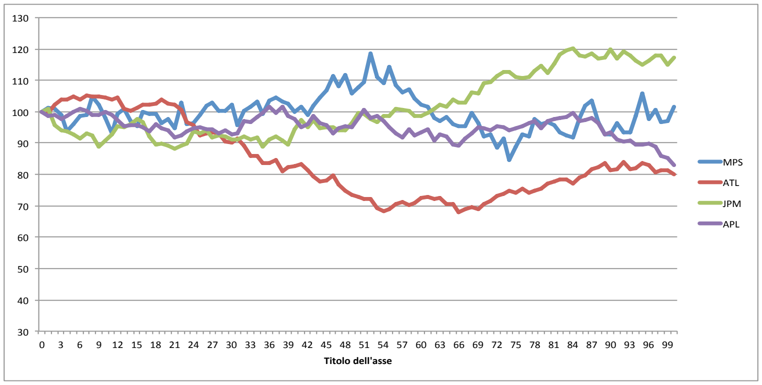


Figure 11: Simulated Stock Prices

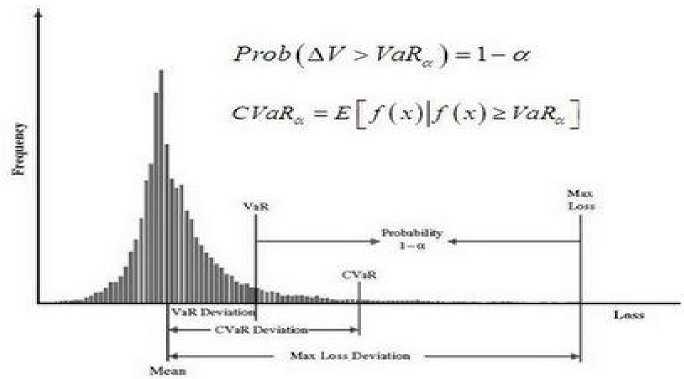


Figure 12: CVaR

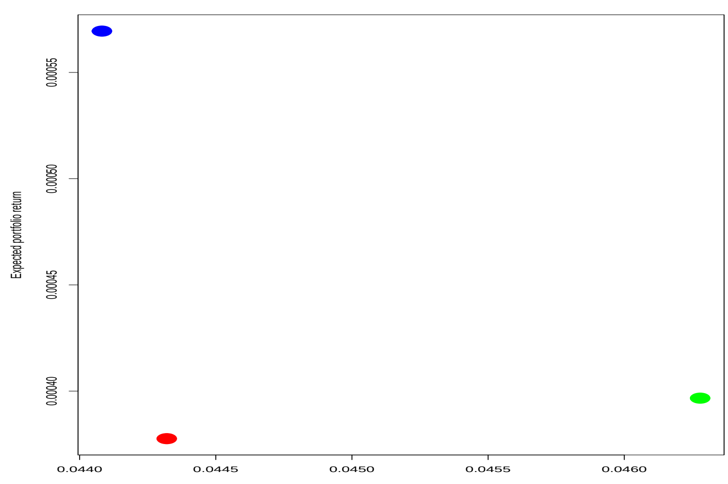


Figure 13: PTF returns VS. CVaRs

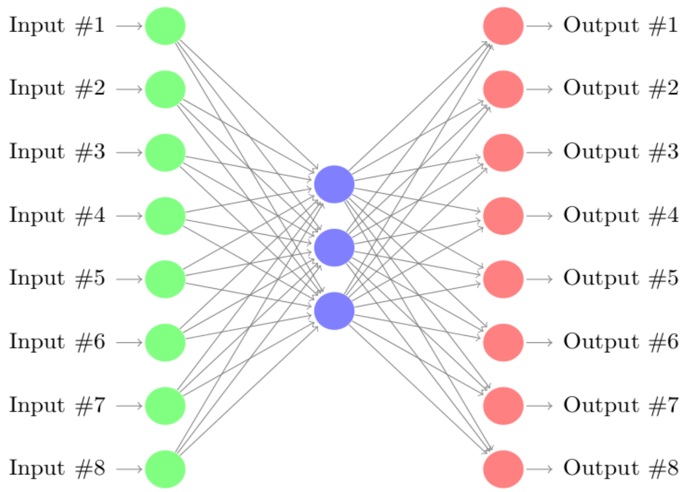


Figure 14: An auto-encoding structure with one hidden layer and three nodes

# 9 Tables

Table 1: Normal 1%VaR and Maximum loss

| *Stock* | *VaR* | *Maximum Loss* |
|---|---|---|
| Atlantia | 3.99 % | 8.44% |
| Apple | 3.79% | 13.19% |
| JPMorgan | 3.99% | 9.99% |
| MPS | 8.65% | 24.21% |

Table 2: Third and Four Moments

| *Stock* | *Skewness* | *Kurtosis* |
|---|---|---|
| Atlantia | 0.05 | 2.21 |
| Apple | -0.27 | 4.78 |
| JPMorgan | -0.15 | 3.23 |
| MPS | -0.19 | 6.14 |

Table 3: JB test

| *Stock* | *JB statistic* | *p-value* |
|---|---|---|
| Atlantia | 315.38 | $< 2.2 * 10^{-16}$ |
| Apple | 1442.7 | $< 2.2 * 10^{-16}$ |
| JPMorgan | 654.5 | $< 2.2 * 10^{-16}$ |
| MPS | 2430.3 | $< 2.2 * 10^{1-6}$ |

Table 4: Historical 1%VaR and Maximum loss

| *Stock* | *VaRt* | *Maximum Loss* |
|---|---|---|
| Atlantia | 4.71% | 8.44% |
| Apple | 4.31% | 13.19% |
| JPMorgan | 4.85% | 9.99% |
| MPS | 10.16% | 24.21% |

Table 5: MC 1%VaR and Maximum loss

| *Stock* | *VaRt* | *Maximum Loss* |
|---|---|---|
| Atlantia | 4.05% | 5.67% |
| Apple | 3.63% | 4.68% |
| JPMorgan | 3.23 % | 4.45% |
| MPS | 9.75 % | 12.09% |

Table 6: 1% Historical VaR LR statistics

| *Stock* | *LR statistic* |
|---|---|
| Atlantia | 0,11 |
| Apple | 0,001 |
| JPMorgan | 0,001 |
| MPS | 5,51 |

Table 7: 1% CVaR

| *Stock* | *1% CVaR Normal* | *1% CVaR Historical* |
|---|---|---|
| Atlantia | 5.06% | 5.65% |
| Apple | 5.23% | 5.95% |
| JPMorgan | 5.34% | 6.14% |
| MPS | 12.46% | 14.55% |

Table 8: Portfolio weights

| | $\omega$ *M-V* | $\omega$ *1%VaR* | $\omega$ *1%CVaR* |
|---|---|---|---|
| Atlantia | 27.78% | 27.42% | 15.08% |
| Telecom | 7.04% | 13.09% | 8.17% |
| Unicredit | -11.97% | -9.25% | -17.65% |
| Intesa San Paolo | -16.08% | -10.74% | -5.87% |
| Generali | 22.37% | 13.81% | 16.07% |
| Monte dei Paschi | 0.17% | -2.46% | -2.75% |
| Eni | 36.85% | 52.54% | 32.48% |
| Enel | 7.58% | -8.31% | 11.14% |
| Mediobanca | -0.97% | -2.56% | 6.41% |
| Luxottica | 28.81% | 24.87% | 36.93% |

Table 9: Portfolio Statistics

| | *M-V* | *1%VaR* | *1%CVaR* |
|---|---|---|---|
| $E[r_P]$ | 0.04 % | 0.04% | 0.06% |
| $\sigma_P$ | 1.29% | 1.34% | 1.33% |
| VaR | 3.22% | 3.39% | 3.39% |
| CVaR | 4,43% | 4.63% | 4.41% |
| Sharpe Ratio | 0.03 | 0.03 | 0.05 |
| Return per unit of CVaR | 0.009 | 0.009 | 0.014 |

Table 10: Pareto Distributions Parameters and KS statistics

| *Stock* | $\alpha$ | $x_m$ | *KS for the Pareto* | *p-value* |
|---|---|---|---|---|
| Atlantia | 6 | 0.048 | 0.19 | 0.6234 |
| Apple | 4.13 | 0.043 | 0.21 | 0.4264 |
| JPMorgan | 5.04 | 0.048 | 0.11 | 0.9912 |
| Monte dei Paschi | 3.31 | 0.102 | 0.14 | 0.9037 |

Table 11: Outcomes of the auto-encoding algorithm

| *Ticker* | *Norm* | *Ticker* | *Norm* |
|---|---|---|---|
| ADP.PA | 0.3976 | FAI.F | 0.6669 |
| FRA.F | 0.4218 | EADS.F | 0.6875 |
| UZZA.F | 0.4358 | LDO.MI | 0.7050 |
| FLW1.F | 0.4840 | TAVHL.IS | 0.7235 |
| LHA.DE | 0.5144 | NAS.OL | 0.7591 |
| RYA.L | 0.5249 | DTG.L | 0.7617 |
| EZJ.L | 0.5268 | AFR.F | 0.7706 |
| IAG.L | 0.5380 | AB1.F | 0.8354 |
| RR.L | 0.5531 | 32A.F | 0.9159 |
| AM.PA | 0.5599 | AFLT.ME | 1.5939 |

Table 12: Weights of the selected stocks

| *Ticker* | *weight* |
|---|---|
| ADP.PA | 0.1907 |
| FRA.F | 0.1297 |
| UZZA.F | 0.1226 |
| FLW1.F | 0.0789 |
| LHA.DE | 0.0578 |
| RYA.L | 0.0802 |
| EZJ.L | 0.0916 |
| IAG.L | 0.1174 |
| RR.L | 0.0281 |
| AM.PA | 0.1029 |