Università
Ca'Foscari
Venezia

Master's Degree Programme
in Economics and Finance
Finance

Final Thesis

Particle swarm optimization applied to different risk measures in portfolio selection problems

Supervisor
Ch. Prof. Marco Corazza

Graduand
Martino Dal Mas
861299

Academic Year
2022/ 2023

Introduction

It is an ancient aim to transport the wealth from the past to the future. During the centuries a lot of systems was developed in the attempt of reach this goal.

One of the most classic examples is the use of gold: it is well-know that gold was used both as a currency and as a safe haven[1]; this means that, theoretically, an ounce of gold today will be worth the same next year so, its price next year will be the same as today plus the (eventual) inflation. The invention of money could have solved the problem of the wealth's transposition but, money is subject to inflation which reduces its value through the time.

Nowadays, there are better methods to solve this problem than keep the money in form of cash or buy ounces of gold, some examples are the investment in art, commodities, buildings and of course securities. since the purpose of this thesis is to compare different measures of financial risk, securities will be the instrument that will be examined; this is because the amount of data available is certainly greater than that of other tools (such as art or real estate) and the data are more easily comparable with each other. The principal way in which a person can invest his/her money into the securities are through the various capital markets around the world.

Thanks to the evolution of technology, people can purchase stocks and bonds directly from their computers of smartphones assuming very low transaction costs.

Even if this simplification would seem to solve the issue it is not so: there are a lot of choices into the exchanges and the decision among the different kind of

[1] A **safe haven** is an investment that is expected to retain its value, or even increase in value, during times of market turbulence.

securities may be a hard task and, different people have different ideas about what

The modern quantitative finance is the subject that studies this kind of problems. It has its cornerstone in the Harry Markowitz's Modern Portfolio Theory.

Starting from some assumptions on the rational investor and on the efficient Market, Markowitz developed a theory that permit to choose the "best" stocks among a predetermined sample of them given a certain desired return.

As "best" stocks we mean the combination of that which minimize the risk given a certain return. According to MPT, it is sufficient to solve a minimization problem which grants that the portfolio thus constructed is the most efficient in terms of risk-return.

In the following paragraphs, we will synthetically see where Markowitz started his theory i.e. Consumption theory and the Utility function as a method to understand the different preferences of the investors.

Then we will discuss some of its issues and their possible solutions: they are, basically, divided into two main categories:

1. MPT is based on rather simplistic assumptions such as the efficient market hypothesis, the rational investor behaviours, the possibility of short sell and the existence of frictionless markets.

   In his first formulation, Markowitz imposed only two constrains in the minimization problem: "budget constrain" and "returns constraints"[2]. This is, obviously, not enough to take into consideration any limitations an investor might encounter. Solutions of this problem have been proposed in the literature, for example, by adding a larger number of constrains it is possible to relax the majority of the assumptions. This, clearly, implies a greater mathematical

[2] Budget constrain and Return Constrain are mathematical tools to express respectively the possible combinations of purchasable stocks and the minimum desired return.

effort. Indeed, a minimization problem with a large number of constraints cannot be easily solved with "basic" mathematic tools, moreover, it is not granted that an exact solution exists. Instrument such as the meta-heuristics for optimization have been implemented in the attempt of solving this kind of issues[3].

2. The risk measure: Markowitz used variance as proxy of risk. Unfortunately, variance cannot be a proper risk measure for a series of reason.

   Indeed, Variance considers the downside risk as well as the upside potential, this means that it considers risk what it is not. Moreover, it is a dispersion measure based on the Normal distribution but, nowadays, it is well-Know that the returns distributions are, generally, asymmetric and they often present a kurtosis larger than three. Finally, it cannot be considered coherent with respect to what happen in the real markets. As we will see in the next chapters, this is demonstrated by the fact that, variance doesn't respect all the four axioms of coherence (translational invariance, positive homogeneity, monotonicity and subadditivity) which are fundamental to grant that a risk measure makes sense in the real world.

In the end, the aim of this paper will be to demonstrate that starting from the same possible combinations of stocks and through the use of "better" risk measures it is possible to build a better portfolio both in term of returns and in term of risk. In order to reach our goal, we will propose four risk measure that will substitute the Variance in the minimization problem and other four risk measure to, correctly, evaluate their results.

[3] A meta-heuristic is an algorithm that, given a specific optimization problem, try to find the "Best solution" inside the more "promising zone".
.

Contents

CHAPTER 1

MODERN PORTFOLIO THEORY AND UTILITY FUNCTION

1.1 Utility Function

How can we measure the satisfaction into pleasure of gain a positive return or the dissatisfaction of bear a risk?

A possible way is through a utility function: in the consumer theory a utility function is a mathematical tool which connects two factors, the satisfaction and the monetary value. It, somehow, connect the “goodness” of increasing satisfaction with the “badness” of increasing cost.

As in the consumption theory happened with the consumer, also in the portfolio theory the rational investor tries to maximize his/her utility function. But, differently from the consumption theory, in the portfolio theory the factors that modify the utility function are represented by the future returns (satisfaction) and the risks (“cost”).

The shape of the utility function depends on the type of investor, specifically there are three types of investors:

1. Risk-neutral investors: They are indifferent with respect to the risk (their utility is based only on the returns so is linear).

2. Risk-lover investors: They appreciate the risk (their utility function is convex).
3. Risk-averse investors: they prefer to avoid the risk; in order to be meaningful for a risk-averse investor, the utility function must be concave.

The reason behind its form is the following: the value of the function shall increase when the returns increase but, at the same time, it must be offset by the increasing risk.

What is the most common kind of investor?

Without loss of generality, we can assume that ordinary people tend to be "rational investors".

Financial theory states that an investor is rational when:

- The investor respects the so-called non-satiation principle: he or she prefers "to gain more than less".
- He or she is risk-averse: *Ceteris paribus*, the investor prefers to bear less risk.

By taking this assumption on the utility function we can go ahead and see the core of the MPT.

1.2 Modern Portfolio Theory

He developed the first method to measure the risk of a portfolio via the multivariate distribution of returns of all asset- Szego G., Measure of Risk, European Journal of Operational Research, 2005

Modern portfolio theory was developed by Harry Markowitz in the 1952, he created what today is considered the cornerstone of the portfolio theories. Clearly, the world is really complex and the choices that a subject is requested to made are many.

So, in the attempt to create a simplified but significative theory, Markowitz had to make many assumptions:

- The consumer has saved part of his wealth and wants to transpose it into the future; in the attempt to reach his goal, he decides to invest his wealth in $n \geq 2$ stocks.
- The investor is rational, for the reason we saw earlier this means that his/her utility function is increasing and concave.
- The investment decision is done in a single period horizon, there is an evaluation in the time T in order to take a decision about the portfolio composition at the same time T. The portfolio is meant to transpose the wealth from time T to time T+1.
- The market is frictionless, this means that there is no cost of transaction nor cost to obtain information about the stocks.
- All the investment are infinitely divisible.
- The investor is price-taker, this means that he/she cannot affect the prices of the stocks.

Starting from these assumptions he decided to use the most important and well-Known statistical measures of the Gaussian distribution as a proxy of Risk and Reward: **Expected value and Variance**.

1.2.1 Expected value

It is intuitive to understand the choice about expected value ,or better, of its estimator: it is nothing but the weighted sum of the past returns[4]. This means, simply, the higher the expected returns the better for the investor.

In order to compute the expected returns of the portfolio, it is necessary to first compute the expected return for each stock:

$$E(Ri) = \frac{1}{t}\sum_{i=1}^{t} x_i$$

[4] Since it's not possible to know what will happen in the future the past is used as a proxy.

where:

- R_i represents the return of the stock I, it's a random variable.
- t is the number observation during the selected period of time.
- X_t is the value of the return at the time t.

In order to compute the expected return of the whole portfolio we shall do a weighted average of all the expected returns.

$$E(Rp) = \frac{1}{n}\sum_{i=1}^{n} w_i r_i$$

where:

- R_p is the return of the portfolio, it's a random variable.
- N represent the number of stocks in the portfolio.
- w_i is the weight[5] of the i-th stock inside the portfolio.
- r_i is the expected return of the stock i.

1.2.2 Variance

Less intuitive can be understand the Markowitz's decision about the Variance[6]. He chose it as a proxy of Risk. Indeed, Variance can, somehow, represent a risk measure because the higher its value, the more average quadratic distance from the realisations to their mean.

To be more precise, Markowitz decided to use the Variance as a proxy of risk. Moreover, he considered also the Covariance that each stock has had with the others in the same length of time.

To clarify this point, we have to see which are the formulas behind his reasoning:

[5] As weight of a stock inside the portfolio we consider the percentage of starting capital invested in the specific.Clearly the weight of a stock inside a portfolio is a number which is comprised between 0 and 1.

[6] The Variance of a random variable is calculated as the weighted sum of squared distance between the realisations of the variable and its mean.

1. compute the Variance of a single stock during the chosen time period

$$Var(Ri) = \frac{1}{t}\sum_{i=1}^{t}(r_i - \mu)^2$$

Where:

- t is the number of observations during the selected period of time.
- Var(Ri) is the Variance of the stock i.
- R_i is the return of the i-th stock at time t.
- μ is the Expected return of the i-th stock.

2. Compute the Variance of the whole portfolio

The Variance of the portfolio is the weighted sum of all the single variances times their squared weight plus two times the Covariances between each pair of stocks time their weights.[7]

$$Var(Rp) = \sum_{i=1}^{n} wi^2 * Var(Ri) + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} wiwjCov(Ri, Rj)$$

Where:

- Var(Rp) is the variance of the portfolio.
- wi and wj are the weight of the i-th and j-th stock inside the portfolio.
- Cov(Ri,Rj) is the covariance between the i-th and j-th return.

It is worth to know that the use of the Variance as a proxy of risk for the time in which Markowitz wrote his theory was a disruptive innovation.

Variance isn't the best risk measure, but it allows to implement the concept of diversification[8]: indeed, using the variance's formulas it's clear that if we increase the number of the stocks inside a portfolio its risk generally decrease.

[7] This is a basic property of the sum of variances: for example, the Sum of the variance of two random variable

$$Var(\alpha A + \beta B) = \alpha^2 Var(A) + \beta^2 Var(B) + 2\alpha\beta Cov(A, B)$$

[8] The proverb: "don't put all your eggs in one basket" is older than Markowitz as well.

Moreover, another great innovation in the portfolio theories introduced by Modern portfolio theory was the use of covariances: The investors should no longer evaluate if purchase or not a stock only basing their assumption on the single stock's features, but they should consider the relation that a stock has with the others.

1.3 Mean-variance Criteria and Modern Portfolio Theory

As we have seen so far, the two criteria that Harry Markowitz decided to use as proxy of reward and risk were expected return and variance.

How can this be useful for an investor?

A general starting point is the so-called mean-variance criteria. Following the Modern portfolio theory, we know that a positive (and possibly high) expected return represents a point in favour for our investment choice, while it's the opposite for the Variance.

We can formalize this discussion in the following way: Given two different stocks or portfolio (X and Y) we want to know if one of the two is better than the other under the Mean-Variance Criteria.

We can state that, for example, X dominates Y in the M-V criteria if X has a higher expected value than Y and a variance equal or lower than Y or if X has a variance lower than Y and an expected value equal or higher than y.

Mathematically we can express this idea in the following way:

$$\mu_X \geq \mu_Y ;$$
$$\sigma_X^2 \leq \sigma_Y^2 ;$$

At least one of the two its true in the narrow sense.

Clearly, when we want to construct a portfolio, starting from a bunch of stocks, it there is a huge number of possible expected return-variance combinations, each one represented by different participation percentage in each stock or , in other words, different portfolio's combinations. This arose an ulterior issue: Given that we want to obtain a certain return, it is likely that more than one portfolio's combination can satisfy our request, so which one have we to choose? Markowitz solved this problem introducing the so-called efficient frontier. Given a certain desired expected return, on the efficient frontier can lay only the portfolio with the lower Variance.

Starting from the multivariate distribution of returns of all assets, the efficient frontier represents all the best solutions (in terms of mean-Variance) that can exist.

Portfolios that lay under the efficient frontier are called "inferior portfolio" and they are inefficient in terms of mean-Variance.

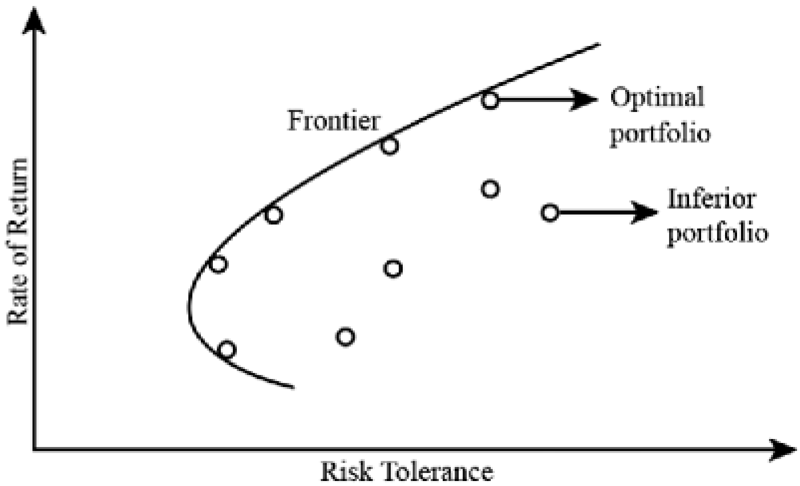


Figure 1 Efficient frontier graph

In order to compute the efficient frontier, the Markowitz's theory purpose to solve a constrained minimization problem. Hence, it is sufficient to minimize the Variance given all the possible expected returns.

$$\min_{x} x'Vx$$

$$\text{s.t.} \begin{cases} x'\bar{r} = \pi \\ x'e = 1 \end{cases}$$

where:

- The objective function is nothing but the minimization of the portfolio variance $\sigma_p = X'VX$

Intended as the matrix multiplication of the vectors of weight times the Variance-Covariance matrix. If the variance-covariance matrix V is positive definite and non-singular – hence invertible – and if there is at least one pair of different mean returns, then the optimization problem admits a unique solution.

- The system represents the 2 constrains: the returns constrain $\mu_p = X'r$

  that indicates that the multiplication between the weights and the expected value of the returns[9]must be equal to a certain value; the second constrain, instead, is the so-called budget constrain. It imposes that the sum of all the stocks weight will be equal to one. In other terms, this means that all the starting capital is invested in the portfolio.

1.4 Problems of the Modern Portfolio Theory

It is incontrovertible that Markowitz's Modern Portfolio theory gave a great contribution to the world of portfolio construction and to the financial world in general. The concept of risk diversification is probably the more significative example of this contribution: through the use of variance as a proxy of risk, he mathematically showed that a diversified portfolio is less risky than a non-diversified one contributing to share the idea (well accepted nowadays) that is dangerous to keep all our saving in" few and concentred" activities[10]. Notwithstanding such a great innovation, the MPT carries on too much assumption and simplifications which strongly affect its usefulness.

[9] The expected value of the returns, computed as the mean of the past return, is used as a proxy of the future returns.

[10] Diversification is not only a matter of numbers, diversified activities means that our investment should be divided into different industries. It is not sufficient to invest in different stocks but they shall be representative of different economic sectors otherwise the correlation among their would eliminate the effect of diversification.

Even if some of these assumptions are plausible (such as the price-taker assumption and the risk-averse assumption[11]) other are, simply, unrealistic. First, the assumption on the investor's rationality: Rational investors are something really closed to the so-called *"homo economicus"*[12], they just try to maximize their expected utility functions given that they, exactly, know what happen in the past perfectly and they rationally take every decision. This is, clearly, a strong simplification of the reality; furthermore, it has been well demonstrated by the Nobel prices Kahneman e Tversky[13] that the individual behaviors are profoundly biased.

Although this argument goes behind the scope of this paper, we want, for the sake of completeness, illustrate what we consider two of the principal behaviors that represent the investor's not full rationality.

- Framing effect: how the decision problem is proposed influences the decision process itself. This means that different investors can choose different solutions in front of the same problem even if only one solution can be "rational" in terms of utility theory.
- Loss aversion: The satisfaction (or, as it is measured in Markowitz's theory, the utility) is generally affected more by the loss than by the gain. This statement seems to be coherent with the Modern Portfolio Theory, but prospect theory[14] showed that people could change their risk-aversion when shifting from the domain of gains to the domain of losses. This means that

---

[11] See the previous section.

[12] A theoretical human being who rationally calculates the costs and benefits of every action before making a decision, used as the basis for a number of economic theories and models- Collins dictionary

[13] Kahneman e Tversky were two important behavioural finance scholars; with their "prospect theory" they demonstrated that the "rational investor" doesn't exist and that the use of use of utility theories to solve problem about uncertain situations may lead to bad solutions.

[14] Prospect theory is a theory proposed by two psychologists (Daniel Kahneman and Amos Tversky) in 1979. This theory tries to explain how individuals take decision in uncertainty conditions.

treating every investor with the same concave-risk-averse utility function may lead to suboptimal portfolio[15].

Secondly, the assumption on the market and on the transactions: Here there is not so much to discuss about, a frictionless market implies that each transaction can be done without pay the cost of the transaction itself. Furthermore, MPT assumes that it is possible to buy/sell each stock in the proportion we prefer but in the real-world stocks are not infinitely divisible and it is not uncommon that they are sold in minimum lots.
The theory assumes also that short selling[16] is always possible, but it may happen that the market's authorities ban this possibility for a certain period[17], moreover, this practice could be allowed only to some kind of investors.

And lastly, the decision on the risk measure.

We want to introduce the problem of the variance as a proxy of risk starting from the "easiest" one: the variance as it is constructed is not a risk measure, but it is, a statistical measure of dispersion. It, indeed, represent the quadratic distance between the realisations and their mean; this means that it considers not only the realisation below the but also the realisation above the mean. Financially this is a contradiction since a rational investor should be happy when the realisations are above the mean.
It is clear that a distinction between positive and negative deviation should be considered. To be honest, the first person who noticed this contradiction was Markowitz himself and he suggested to solve this problem through the use of the Semi-Var. The

[15] Here suboptimal portfolio isn't intended as a portfolio that does not satisfy the requirement of minimum variance given a certain value of the expected return. It is intended as non-optimal portfolio for a certain investor.
[16] Short selling is a common practice in the stock market: it consists in the sale of a stock that we do not own in the moment of the transaction.
[17] As it happened during the first period of the Covid-19 pandemic: market's authorities ban the practice of short sell in order to avoid further deterioration of the stock market.

Semi-Var is a statistical measure of dispersion which is constructed as the variance but considers only the realisation below the mean, the so-called *downside risk*[18]*:*

$$\text{Semi} - \text{var}\ (R_P) = \frac{1}{N} \sum_{i=1;\ R_i < \mu}^{N} (R_i - \mu)^2.$$

Unfortunately, as we will see, the semi-variance too isn't a coherent risk measure.

Another important risk measure was introduced by the investment Bank J.P. Morgan. It is the so-called Value at Risk (VaR).

Definition. *Given a confidence level of α [0;1] and fixed a specific holding period, Value at risk indicates the maximum potential loss associated to a portfolio in α% cases during the holding period.*

It had a large diffusion in the academic and financial fields since it is a simple and understandable tool: The VaR of a certain portfolio is nothing but a "threshold" : the probability that the returns will be lower or equal to the VaR are equal to (1-α). Moreover, it is represented in the same unit of the investment (euro, dollar, yen, etc.) which makes it even easier to interpret.

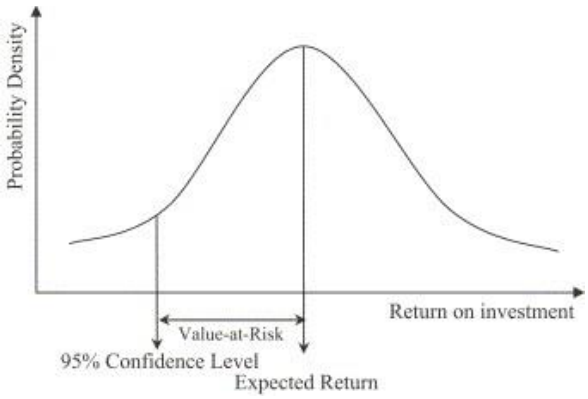


Figure 2 Probability density function of a normal distribution

In mathematical therms, this mean that P(R>VaR)=1- α[19]. Unfortunately, as in the case of Variance, the VaR of a portfolio is a meaningful measure if and only if the

[18] The *downside risk* is the risk of realisations below the mean, its opposite the *upside potential* represents, instead, the realization above the mean.

[19] Practically, the VaR is the α-quantile of the returns distribution.

underlying probability distribution is Normal: if the distribution is not Gaussian this risk measure can possibly underestimate the potential losses. The VaR is just a threshold and without correct assumptions on the tail's behaviours it is meaningless. It is not capable to distinguish among two different investments that have the same α-percentile but different property in the left tails of their distributions. In plan words, the VaR cannot measure "how bad is bad".

Differently from the Variance, the VaR doesn't represent the diversification-effect correctly. This means that, following the VaR summation rules, it may happen that the risk of two aggregated portfolios could be higher than the sum of the two taken alone.

Through the use of variance as a risk measure (or more specifically the Var-covariance matrix of the N-stocks), Harry Markowitz assumed that the returns were normally distributed, hence their behaviour could be well explained by the first and the second moment of the distribution Itself. Such an assumption implies that the returns are symmetrically distributed around their mean and that the probability of the realisation rapidly decrease when we move away to the mean itself.

Mathematically a distribution is symmetric when $F(\mu - x) = F(\mu + x)$ [20]

In order to see if (and how much) a distribution is or is not symmetric we can calculate the empirical skewness:

$$b_1 = \frac{m_3}{s^3} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^3}{\left[\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2\right]^{3/2}}$$

Skewness is the third moment of the distribution, if it is equal to zero the distribution is symmetrical with respect to its mean, while if it is greater or lower than zero the distribution is, respectively, skew to the right or skew to the left.

[20] Where F represent the cumulative density function.

It has been shown by Peirò (1999) that the returns of the stocks tend to have a skewness greater than zero, hence they are skewed to the right. This means that, generally, most of the returns are greater than their mean. This result not only invalidates the use of variance as a proxy of risk but it does affect Markowitz's model because investors could prefer less diversified portfolio (inefficient in mean-Variance terms) rather than optimal-well-diversified portfolio only because the latter could give them higher probability of positive returns. These results suggest that we should consider higher odd moments than the first (mean) in order to "capture" the empirical skewness of the distribution as well.

For what concern the tails of the distribution[21]., we mean that large deviations from the mean are almost impossible under the Gaussian assumption[22]. As in the case of asymmetry also here there is a statistical measure to compute "how fat are the tails" This measure is the so-called empirical Kurtosis

$$k = \frac{m_4}{s_4} = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})}{\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2\right]^2}$$

A higher value of the kurtosis indicates a higher probability of "extreme events". In the case of normal distribution this statistical indicator must be exactly equal to 3. The empirical analysis on the stocks return showed how much the normality assumption was always unsatisfied. The probability of extreme event (either "good" or "bad") in the market is much higher than under the Gaussian assumption. A lot of interesting material about this topic has been written by the trader and statistician Nicola Nassim Taleb[23]; Even if it is not possible to summarize all his contributions to the financial world in just a few sentences, anyway we can say that tails events happen more frequently than

[21] The tails represent, in the probability density function, the events far away from the mean. Fat tails are an indicator that the realisation distant from the mean are likely while for the thin tail is true the viceversa.
[22] In statistical terms we can say that realizations above 4 sigmas are practically impossible.
[23] In his books he speaks about the decisions under uncertainty, probably the most famous one is "Black Swan".

expected, under Normality assumptions, and when they do occur their impact can affect us in an unexpected serious way.

In the end, Markowitz in his modern portfolio theory through the use of the variance covariance matrix as a proxy of risk implied that the returns were elliptically distributed[24]. Hence, he supposed that they can be statistically treated with a normal (or a Student t) distribution, but nowadays we know that this assumption can no longer be considered true. The returns in the stock market behave, generally, as a leptokurtic normal skewed to the right, showing more results above their mean and tails which are larger than the gaussian's one.

Concluding, since we are aware of, general, market behaviours we ca no longer consider Variance as a proxy of risk. We should, instead, search risk measures which take into account the asymmetry and the Kurtosis of the return's distributions.

1.5 Risk measures and coherence

There exist two methods to "improve" the Modern Portfolio theory: one concerns the relaxation of the assumption underlying the model. This can be possible through the use of a larger number of constrains.

As we have seen in the previous chapter, the original Markowitz's theory considered only two constrains for the minimization problem; However, it is possible, to increase the number of constrains in order to reduce the number of assumptions

Since it is not the objective of this paper, we will show just an ideal bunch of constrains that can be used to reach this goal.[25]

[24] An Elittical distribution is any distribution of the broad family of probability distributions that generalize the multivariate normal distribution. The most famous ones are the gaussian, student-t, laplace, Cauchy.. They share some useful characteristics for example, the sum of elliptical distribution is elliptical itself. All the elliptical distributions have a finite variance.

[25] This formula has been taken by "I vincoli a variabili miste intere" Corazza (2002)

$$\min \quad (PLx)'V(PLx)$$

$$\text{s.t.} \quad \begin{cases} (PLx)'r \geq \pi C \\ f_1(x) \leq \alpha C \\ f_2(x) \leq \beta C \\ (PLx)'e \geq (1-\alpha-\beta)C \\ x_i \geq 0 \; \forall i \\ x_j \in \mathbf{N} \; \forall j \in I \end{cases}$$

In this formulation we can see a constrain on:

- Minimum lots that can be purchased: P and L are two matrixes that contain respectively the price and the minimum number of shares that can be bought.
- Maximum cost of transactions and tax's cost: f_1 and f_2 are two functions that indicates that costs. They are expressed as a percentage (alpha and beta) of the total capital: the investor is not disposed to spend more than α*C for the transaction cost and β*C for the Capital gain's taxes.
- Short sell: the second-last constrain indicates that the weight of each stock must be at minimum equal to zero, hence there isn't the possibility to short-sell.

Although by adding constrains the model become more and more realistic, this doesn't happen with any cost. For example, introducing this "mixed integer constrains" increases the mathematical complexity of the problem.

The other methods to "improve" Markowitz's modern Portfolio Theory, instead, acts on the risk measure. As we have seen so far, the minimization problem that use the Variance (or the semi-var) as risk's proxy is no longer acceptable. Thus, we want to use and compare different types of risk measures. In order to reach our goal, we have, firstly, to understand what a risk measure is, and which are its desirable features.

It is a hard task to summarize all the features that a risk measure shall consider in just "one formula" in order to translate it into a comparable number[26]. First of all, we need to give a definition of risk: Even if in the common languages we use the word risk and uncertainty as they were synonymous, there is a substantial difference between the two. Indeed, taking the financial field as example, with the word uncertainty we mean that given a certain prediction about the future return we are **uncertain** about its realisation either if it will be **better** or **worse** than what we predicted. Instead, with the word risk, taking as example the financial field again, we intend that given a certain prediction about future return we **risk** that the result will be **worse** than predicted.

Thus, as a first sight, it seems that risk can be described as a combination of exposure and uncertainty, or in plan words, we risk only when we can bear a loss.

Although risk is for sure a combination of exposure and uncertainty, the concept is broader because it is also a relative (to a given bench-mark), personal, asymmetric and multidimensional concept[27].

So, an ideal risk measure shall also take into account the correlation and the diversification among the sources of risk, the downside risk, the risk's propension of the investor and the computation complexity, all this risk's features (and probably the list is not complete) are necessary in order to create a meaningful risk measure.

In the paper "desirable proprieties of an ideal risk measure in portfolio theory" the work of Artzer et Al[28]. of developing the theories on the risk measures is continued.

The paper affirmed that a risk measure is a function ρ that assigns a non-negative numeric value to a random variable X. It described also which are the features that the risk measures shall share in order to be meaningful:

[26] DESIRABLE PROPERTIES OF AN IDEAL RISK MEASURE IN PORTFOLIO THEORY-SVETLOZAR RACHEV, SERGIO ORTOBELLI, STOYAN STOYANOV, FRANK J. FABOZZI and ALMIRA BIGLOVA;International Journal of Theoretical and Applied Finance.

[27] Asymmetric since loss and gain are perceived in a different way. Multidimensional because the investor can have multiple objectives.

[28] Coherent risk Measures; Philippe Artzner,Freddy Delbaen,Jean-Marc Eber,David Heath

1. Positivity: A risk measure must be positive or at least non-negative, since it is obvious that a "position" can only be risky ρ(X)>0 or not ρ(X)=0[29]
2. Convexity: It is the property that explain the meaning of the diversification. As we have seen with the variance the risk of two combined portfolio must be lower or at least equal than the weighted sum of the two. Given a random variable X.

   $$\rho(\alpha X+(1-\alpha)X)\leq \alpha\rho(X)+ (1-\alpha)\, \rho(X)$$

   Actually, convexity is itself composed by two sub-characteristics:

- Subadditivity: Given two random variables X_1 and X_2

  $$\rho(X_1+X_2)\leq \rho(X_1)+ \rho(X_2)$$

  risk measures shall indicate that the risk of two activities together must be less than the sum of the two activities alone[30].
- Positive Homogeneity: given a random Variable X and an $\partial\geq 0$

  $$\rho(\partial X) = \partial\rho(X)$$

  This means than no-matter the size of the investment, the risk shall be measured as the risk time its size.

3. Translation invariance: A measure of risk should consider that investing a part (beta) of the available capital into the reference instrument[31] reduces the overall risk exactly by the amount beta.

   $$\rho(X+\beta) = \rho(X)- \beta$$

   Where X is a random variable and beta is the part of capital invested in the reference instrument.

   Moreover, this means that adding a risk-free quantity equal to $\rho(X)$ to the risky position, we obtain a risk-free entity.

---

[29] ρ(X)=0 means that X is non-stochastic, hence X is known a priori.
[30] It is the demonstration of the phrase "merges don't create extra risk".
[31] As reference instrument we intend the risk-free instruments such as cash or goverment's bond ecc..)

4. Monotonicity: A risk measure shall underline the preferability of an asset that systematically overperform another.
   ρ(X) ≤ρ (Y) for all r.v. X,Y with X≥Y

In his paper, Ortobelli explained also the differences among the so-called uncertainty measure and the proper risk measure (that will be called coherent risk measure for the reason we will see later). The uncertainty measures only partially respect the characteristics of the ideal risk measures that we have just seen, indeed they respect positive homogeneity, Translation invariance and Positivity.

A particular class of the uncertainty measure is the dispersion measure class. in addition to respect the proprieties listed above, the dispersion measures respect also the subadditivity propriety. In this class we can find statistical measures that we are used to see as a correct proxy of risk such measures are, for example, the standard deviation, the MAD[32] and the semi-standard deviation. They are called dispersion measures because they share one characteristic: they depend on the center of the random variable, in other word, even if in different way, they all measure the distance (dispersion) between the realizations and the mean of the random variable itself.

Unfortunately, the use of the measure of dispersion as a proxy of risk may lead to the wrong conclusions because they don't respect all the coherence axioms. Indeed, the monotonicity one is not satisfied under this class of measures[33] .

Following the path traced by Artzer et al(1999), we can state that it cannot exist a unique risk measure capable of solve all the uncertainty problem, but, instead it is possible to define a bunch of roles that a right measure of risk (for a risk-averse individual) should have. These characteristics are summarized in what we called the

[32] Median absolute deviation. $MAD = \frac{1}{n}\sum_{i=}^{n}|Xi - \overline{Xn}|$

[33] For the mathematical explanation see Appendix

axioms of coherence. If and only if a risk measure respect all this axiom can be a coherent risk measure.

CHAPTER 2
COHERENT RISK MEASURE

2.1 Tail conditional expectation

The firsts coherent risk measures were proposed by Artzer et al in their paper on this subject. They were the Tail Conditional Expectation (TCE) and the Worst-Conditional Expectation (WCE).

It is worth to note that these two measures (together with all the other coherent risk ones) focus their attention on the "extreme-left" side of the return distribution, in this way they try to keep under control what can, truly, damage the investor: extreme negative events.

In this sense, we can note a fundamental difference towards the VaR: Coherent risk measures are not a "threshold", they are, rather, a sort of indicator in the case the "worst situation" happens. Thus, they give an indication about the nature of loss in the case the VaR's limit is overcome.

The TCE_α[34] is expressed as:

$$TCE_\alpha(X) = E[X| \leq VaR_\alpha(X)]$$ [35]

this means that the Tail conditional expectation is the expected value of the random variable X given that it is lower than then Value at risk of X

While the WCE_α :

$$WCE_\alpha(X) = inf\{E[X|A]|P[A] \leq \alpha\}$$

[34] Both TCE and WCE are function of alpha: as in the case of the VaR alpha represents a small probability (usually 0.05 or 0.01); in a certain sense this means that we are considering what may happen in the alpha-percentage of the worst scenarios.

[35] All the formulas that will follow can be find in literatures in different forms, their forms depend on the underlying distributions: they can be either returns or losses distributions. In order to be coherent, we chose to use only the return distribution.

meaning that the Worst-Conditional expectation of the random variable X is the infimum of the expected value of X given A and knowing that the probability that A happens is lower than alpha.

Although both represent a sort of mean of the distribution's left side and look similar, these two formulas behave differently: the TCE is easier to implement for practical purpose, but it may not respect the subadditivity axiom when it is applied to non-continues (discrete)[36] probability distribution; in this sense the WCE is more robust because it doesn't need a continuous distribution function to be coherent. Unfortunately, in order to implement the WCE it is fundamental to know the underlying probability distribution and, as we know, it is not always feasible.

In the attempt to obtain a simple but coherent risk measure Acerbi and Tasche (2002) developed the so-called Expected shortfall (ES). The ES share, with the TCE, the characteristic of simplicity: they both are the mean of the alpha-percentage realisation. This common feature is remarked by the fact that in the continuous case they are represented by the same formula:

$$ES_\alpha(X) = TCE_\alpha(X)$$

this is true if and only if the underlying distribution function is continuous.

If the underlying probability is known, it is possible to implement this tool through formulas. These formulas exist for the majority of the continuous probability density functions such as Student-t, Laplace, exponential and, of course, the Gaussian one. Since this is not the aim of this paper, we will present only the Gaussian formula and its representation in order to better understand this risk measure. The expected shortfall of a gaussian random variable is:

$$\mathrm{ES}_\alpha(X) = -\mu + \sigma \frac{\varphi(\Phi^{-1}(\alpha))}{\alpha},$$

[36] A discrete probability distribution is a probability distribution that doesn't have a finite value in all its points; in the financial field an example can be the return distribution of some kind of exotic options.

where μ and σ represent respectively the mean and the standard deviation while ϕ and φ represent the probability density function and the cumulative density function (small and capital letter respectively). Hence, knowing the two main parameters of the Normal distribution it is possible to compute "the risk in the left tail" or, in other word, the mean of the realisation if the VaR is overcome.

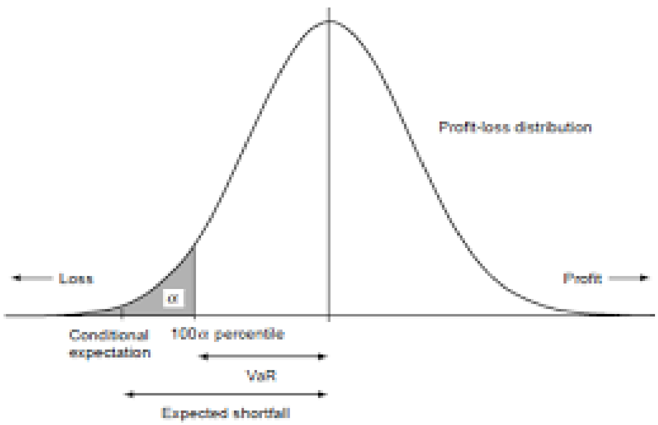


Figure 3 Probability density function of a Gaussian distribution

Nevertheless, it is not always meaningful to fit a probability distribution to the real data, since the true underlying probability density function may be not known with certainty; it is, instead, better start from the true realisations and try to understand which probability density function can better adapt to the reality.

As we have mentioned earlier, another great point in favour of the ES is its simplicity: from an empirical point of view, it can be calculated in the following way: given that we have n realizations of the random variable X, it is sufficient to sort them in ascending order and take the mean of the α-percent ones[37].

It can be mathematically express as:

$$ES_{\alpha}(X) = \frac{\sum_{i=1}^{k} Xi}{k}$$

[37] As smaller ones we mean the realization from the first to the (n*α)-th, considering that the result of n*α must be an integer.

where K represents the integer number of realisations that have a value lower or equal to the Value at Risk of level α.

Expected shortfall was one of the firsts coherent risk measures proposed, its simplicity and its adaptability to the “risk-averse” individual’s behaviours helped its widespread into the financial world. As we have seen, differently from the Var, the ES respects all the coherent axioms including, perhaps, the more important: sub-additivity. This means that, differently from VaR, ES does represent the diversification effect and it ensure a proper protection against the tail risk. Another important feature of this risk measure can be seen in its non-parametric approach which permits to obtain, meaningful, measures even when the underlying probability distribution isn’t know.

2.2 Shortfall deviation risk

This coherent risk measure was introduced by Righi and Ceretta in their paper “Shortfall Deviation Risk: an alternative for risk measurement” Journal of Risk, September 1998,.

This measure owns all the desirable features of the ES, and it takes in consideration also the tail’s dispersion of the data. The consideration behind this risk measure is the following: does it make sense to compute the mean value of the losses that exceed the Value at Risk (at a certain α-level) if they aren't concentrated around the mean itself? Apparently, the answer is no because if the losses which exceed the VaR are not concentrated around their mean (that is nothing but the Expected shortfall), the expected shortfall becomes just a number without any sense, hence a more robust risk measure is needed. In other terms, if two position have the same tail expectation but different tail’s behaviour the Es is no longer useful.

As a first definition, we can state risk measure that represents the expected loss that occurs with certain probability penalized by the dispersion of results that are worse than such an expectation[38]

Mathematically it can be expressed in this way:

$$SDR_{\alpha}(X) = -ES_{\alpha}(X) + (1-\alpha)^{\beta} SD_{\alpha}(X) \quad \text{with } \beta \geq 0$$

It is the combination of two risk measures: the expected shortfall and semi-deviation of the latter.

Indeed, it simultaneously, encompass two risk pillars: possibility of extreme events and their dispersion.

This measure is capable to take in consideration higher moment than the first because, it not only considers the mean of the "worst realization" but also their dispersion hence, it is more robust than the Es taken alone.

As the expected shortfall, it is based on the Value at Risk, hence an α-percentile must be expressed in order to compute the SDR.

The first part of the formula is nothing but the negative of expected shortfall[39] (for more info about see the previous chapter). The second part of the formula, instead, is divided into other two sub-parts: in the first one we can see a (1-α) elevated to the power of β: this part serves as a proxy for the degree of risk aversion of the investor.

Alpha represents the percentile of data that are considered the "worst cases" and, as in the case of VaR, is generally really small (0.01;0.05) while Beta is serves as "penalty-chooser"; The higher the value of Beta the lower the penalty that the semi-deviation applies to the Shortfall deviation-risk. So, for example, an investor with a low degree of risk aversion should have a high Beta.

[38] SHORTFALL DEVIATION RISK: AN ALTERNATIVE FOR RISK MEASUREMENT- Marcelo Brutti Righi-Paulo Sergio Ceretta

[39] In its original formulation there was a positive expected shortfall but for coherency reasons we are reasoning in the domain of return, hence a negative sing must be put in front of the Es.

The second and last part of the formula is represented by the semi-deviation of the "worst realizations". It represents a distance between the realizations of X and the Expected shortfall, if and only if these realisations are smaller than the expected shortfall it-self.

Mathematically it can be expressed as:

$$SD_\alpha(X) = (E[|(X - Es_\alpha(X))^-|^p])^{\frac{1}{p}}$$

where:

$$(X)^- = max(-X; 0)$$

In other words, it is the p-norm of the difference between the expected shortfall of alpha and the realization below the Es itself.

The semi-deviation doesn't respect all the coherency axiom, in particular it doesn't respect the monotonicity axiom, hence it cannot be considered a coherent risk measure.

Semi-deviation belongs to another class of measures, the so-called generalized measure of dispersion (as defined by Rockafellar et al. (2006)). This class of measures shares with the coherent risk measures some important axioms; indeed, they are both positive homogeneous and sub additive. They also own other useful features:

- Relevance: if a position always generates losses, it must be risky.
- Strictness: *ro of x greater or equal than minus expected of x* Generalized risk measures are sufficiently conservative, indeed they exceed the common loss expectation.
- Law invariance: if the underlying random variable probability distributions are equal, so are their risks.

Unfortunately, they lack monotonicity and translational invariance. For what concern the monotonicity, this means that it may happen that a generalized risk measures indicates as less risky a portfolio which constantly underperforms another one. While the lack of translational invariance implies that adding a risk-free entity to the portfolio does not reduce the overall risk. To be more precise, generalized risk measures respect another axiom in the field of “translation”, indeed they present the so-called translation insensitivity. This implies that adding a risk-free entity “C” to the risky portfolio doesn’t change the value of that this kind of measures give to the risk. Taking as example the semi-deviation, this means that:

$$SD_{\alpha}(X+C) = SD_{\alpha}(X)$$

As we mention, this feature implies the impossibility for the semi-deviation to be a coherent risk measure, but it is fundamental in the construction of the shortfall deviation risk as we will see in a moment.

A question that may arise spontaneously is: is it possible that the sum of a coherent and a non-coherent risk measure would be coherent? Summarizing the characteristics of the expected shortfall and the semi-deviation we will demonstrate that the shortfall-deviation risk is a coherent risk measure.

As we already know, a measure of risk is stated to be coherent if it respects all the four axioms of coherence: translational invariance, sub-additivity, monotonicity, positive homogeneity, hence we are going to show that SDR follows these rules:

1. Translational invariance:

   $$SDR_{\alpha}(X+C) = ES_{\alpha}(X+C) + (1-\alpha)^{\beta} SD_{\alpha}(X+C)$$

   Since the expected shortfall does respect the translational invariance while the semi-deviation respects the translational insensitivity, the formula can be rewritten as:

   $$-Es_{\alpha}(X) - C + (1-\alpha)^{\beta} SD_{\alpha}(X) = SDR_{\alpha}(X) - C$$

2. Subadditivity:

   $$SDR_{\alpha}(X+Y) = -Es_{\alpha}(X+Y) + (1-\alpha)^{\beta} SD_{\alpha}(X+Y)$$

Since both the measures are sub-additive, we can state that:

$$-ES^{\alpha}(X+Y)+(1-\alpha)^{\beta}SD^{\alpha}(X+Y) \leq -ES^{\alpha}(X)+(1-\alpha)^{\beta}SD^{\alpha}(X)-ES^{\alpha}(Y)+(1-\alpha)^{\beta}SD^{\alpha}(Y)$$

Hence, $SDR_{\alpha}(X+Y) \leq SDR_{\alpha}(X)+SDR_{\alpha}(Y)$

3. Monotonicity:

Let X,Y,Z be three random variables which respect the following assumptions:

- X is always lower than Y,
- Z is a random variable,
- X plus Z is equal to Y

$$SDR_{\alpha}(Y) = SDR_{\alpha}(X+Z) \leq SDR_{\alpha}(X)+SDR_{\alpha}(Z) \leq SDR_{\alpha}(X)$$

Due to lower range dominance $(1-\alpha)^{\beta}SD_{\alpha} \leq Es_{\alpha}(X)$ hence, SDR(Z)<0

This propriety grants that a position which always has worst results is indicated as riskier.

4. Positive homogeneity

$$SDR_{\alpha}(\lambda Y) = -Es_{\alpha}(\lambda X)+(1-\alpha)^{\beta}SD_{\alpha}(\lambda X)$$

Since both the measures are positive homogenous, we can state that:

$$-ES_{\alpha}(\lambda X)+(1-\alpha)^{\beta}SD_{\alpha}(\lambda X) = \lambda[Es_{\alpha}(X)+(1-\alpha)^{\beta}SD_{\alpha}(X)]$$

$$= \lambda SDR_{\alpha}(X)$$

As we have seen, the shortfall deviation risk does respect all the axioms of coherence, hence It can be fully considered a coherent risk measure.

In addition to being a consistent risk measure, shortfall deviation risk shares with the two risk measures from which it is formed also the relevance, strictness, and Law invariance axioms. All these features make it a better instrument than the expected shortfall when working with data coming from the real world.

While recognizing the usefulness of the ES as "tail-measure of risk", there is no doubt that the SDR is a superior instrument when dealing with the left tail of the returns; this happen because it is not only a coherent risk measure (as the ES) but it is able to take in consideration the negative deviations from the ES as well. This is done by giving a "weight" to the extremely negative realizations; Even if this may seem a technicism it is not so, these extreme- realizations must be taken into account because if they will happen again in the future they may "hit" so hard the returns to vanish all the efforts done until that moment to build a profitable portfolio.

2.3 Entropic Var

The last measure we want to introduced is the entropic Value at risk. Entropic VaR is a recently introduced risk measure indeed, it was presented in 2011 by Ahmadi-Javid in his paper *"Entropic VaR: a new coherent risk measure*". At a first sight entropic Var can be explained as "a sort of evolution" of the Expected shortfall. Ahmadi demonstrated that as the ES, the EVaR is coherent and it takes in consideration the results "larger than the Value at risk" but, differently from the former, it hasn't got computability issues. Moreover, the Entropic Var is strongly monotone implying that ρ(X)< ρ(Y) if :

- X≥Y all the observations of X are greater than the ones of Y;
- Pr{X>Y}>0 it grants that X is greater than Y;
- Ess sup X> ess sup Y;

Definition*: Entropic Value at risk is the tightest possible upper bound obtained from the Cherenoff inequality*[40] *for the VaR.*

In order to find this new measures Ahmadi utilized two important statistical theorems: *Markow inequality and Chernoff inequality*.

The former states that $P(X \geq a) \leq \frac{E(X)}{a}$

[40] In probability theory, the Chernoff bound or inequality gives exponentially decreasing bounds on tail distributions of sums of independent random variables-Wikipedia

this result is demonstrated by a series of intuitive mathematical passages:

$E(X) = p(X < a) * E(X|X < a) + P(X \geq a) * E(X|X \geq a)$, this means that

$E(X) \geq P(X \geq a) * E(X|X \geq a) \geq a * P(X \geq a)$, and so $P(X \geq a) \leq \frac{E(X)}{a}$

thanks to the probabilities proprieties and by adding the exponential we arrive to the last part of the demonstration. The latter states: $P(X \geq a) = P(e^{tX} \geq e^{ta}) \leq \frac{E^{tx}}{e^{ta}}$

this directly leads to the starting point of the Chernoff inequality for the Value at risk, since we know that E^{tX} is nothing but the moment generating function of X; it is sufficient to compare the two extremes of the formula in order to obtain:

$$P(X \geq a) \leq e^{-ta} * \ M_x(t)$$

In which a is the Entropic VaR; this formula represents the tightest possible upper bound for the Chernoff inequality for the VaR.

From this formula is possible to derive the E_VAR's formula as well: the probability that X is greater than a is, for definition, equal to α hence:

$$\alpha = M_x(t) * e^{-ta}$$

from this point we want to compute the value of a, which is nothing but the Entropic value at risk, by applying both sides to a logarithm we obtain

$$log(\alpha) = log(e^{-ta}) + log(M_x(t))$$

then after some arrangements we obtain:

$$log(e^{-ta}) + log(M_x(t)) - log(\alpha) = 0$$

From which, using the proprieties of the logarithms,

$$-ta + log(\frac{M_x(t)}{a}) = 0,$$

And finally, we got

$$a = t^{-1} \frac{log(M_x(t))}{\alpha}$$

This is the formula to compute the Entropic VaR of a general random variable, given a certain value of $\alpha \epsilon$ [0,1] and t>0

The entropic Var is a more conservative risk measure than VaR or the expected shortfall, ,as from its nature, it is the upper bound of both the previous measures: $CVaR_{1-\alpha}(X) \leq EVaR_{1-\alpha}(X)$ Hence, given a certain confidence level α, the entropic-var is more risk-averse than the other two measures we have seen so far, this implies that, in portfolio optimization, it may lead to portfolios with lower returns.

CHAPTER 3
A GENTLE INTRODUCTION TO THE PSO

The minimization problems described so far are not solvable with "normal" mathematical tools when considering the coherent risk measures presented in the previous chapter. Indeed, all the risk measures we have seen haven't got a close. This is one of the reasons for which Variance as risk measure has been used for a such long time: the other risk measures lack an instrument to compute the joint variability of two or more random Variable. For example, when using the Variance as a proxy of risk, the variance of the sum of two (or more) random variable is nothing but the sum of the Variance plus two times the Covariances while for the other measures it is not possible to compute their "jointly movements". From this it descends that, it's not possible find the solution of the portfolio minimisation problem as developed by Markowitz. Moreover, if in the minimization problem a bunch of restriction is taken into consideration it becomes even harder to solve.

Approximations procedures have been developed throughout time in order to solve this kind of problems. Probably the most famous and effective are heuristics and metaheuristics.

Definition: A metaheuristic is a high-level problem-independent algorithmic framework that provides a set of guidelines or strategies to develop heuristic optimization algorithms (Sörensen and Glover, 2013)

Metaheuristics are therefore developed specifically to find a solution that is "good enough" in a computing time that is "small enough"; in plan words a metaheuristic can be seen as an algorithm that search for optimal solutions that cannot be find through "close formulas". They, generally, don't provide the **true** optimal solution but something which is close to it. In the following part we want to introduce the Particle Swarm Optimization metaheuristic.

3.1 Particle Swarm Optimization

Particle Swarm Optimization (PSO) is a bio-inspired metaheuristic born from an idea of Kennedy and Eberhart in 1995; They tried to create an intelligence method in order to study, and understand, the behaviours of groups of birds (swarm). In their original formulation, they were able to transpose the form and the behaviours of a swarm into a more rigorous “mathematical concept”: Each bird can be seen as a particle so, the swarm is a bunch of particles which moves into a 3-dimensional space in the attempt to find the best possible position[41].

Kennedy and Eberhart[42] noticed that, inside a bird’s swarm, the key to find food (hence survive) was the cooperation among the birds and their ability to “communicate” one with the others. Indeed, a single particle has almost no chance to find the optimal solution, that’s why it arose the need to introduce some kind of communication among the neighbour particles. In the 95’s PSO formulation each particle was meant to iterate with the two closest particle, in order to calculate which one had the higher personal best. The higher personal best[43] was considered the global best (gbest) for those particles.

To be more specific, each particle is represented by 3 vectors **(Xi,Vi,Pi)**, that are, respectively, its position, velocity and **personal** best.

[41] Since the algorithm was born in order to study the behaviours of swarms, their movements are based on food research: the best possible solution is the place that has the highest quantity of food and that is reachable by the swarm.

[42] Kennedy is a social psychologist and Eberhart is an electrical engineer.

[43] As Personal Best we mean the best position that the single particle has visited so far; there can be different parameters in order to compute a function (generally called fictness fuction) which by giving a score to each point visited by the particle permits to compare one point to the other to decide which is the personal best.

Position is a pretty straightforward concept, for what concern velocity, the concept will be clear in the next pages when we will introduce the entire iteration process while the personal best is strictly related to the concept of fitness.

The Fitness function can be seen as the core of the PSO procedure, it clearly depends on the problem we are trying to solve but, in a general fashion, we can state that each particle has a fitness value and, in case of minimization, the lower it is the better is the position, hence, throughout the iterations, in the local best (or in the global best) vector are recorded only the particles that are able to decrease this number.

Before presenting the algorithm in its original formula, it is better to introduce the communication topology which is nothing but the ways in which the particles can communicate each other.

3.2 Communication topology

For what concern the communication topology, it was the authors themselves who presented 2 typologies of interaction among particles: the *global best* and the *local best* topology.

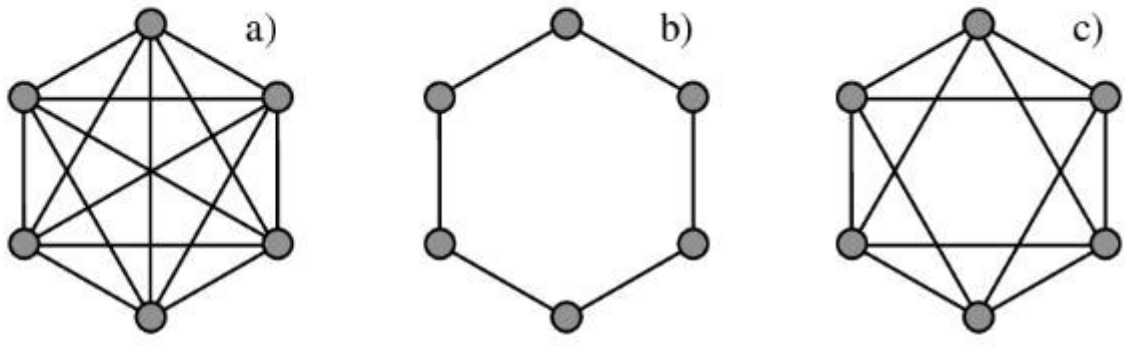
In the local best topology, the trajectory of one particle is influenced only by the closest particles, hence the communication is based only on **proximity** in the search space. This method allows parallel search avoiding to “fail” in local optima. This happen because the interactions between the particles are slow and this consents to stabilize 2 or more bunch of particles around more than one local optima. If one local optima is truly better than the others this information will slowly pass from a particles to another and they will start to converge following the best-particle direction. With respect to the global best topology where, each particle is meant to communicate with all the others, there is more particle dispersion in the search space, this clearly increase the quantity of observed space but, the downside is that this comes with a cost: more time is needed for the particles to converge towards the optimum point[44]. Two classic examples of local best topologies are:

- The ring: Each particle knows only its personal best and the one of the two particles that are closest to itself. The local best is computed comparing these values. Graphically it can be seen as a ring where each particle interacts with the ones that are over and under itself.
- Von Neumann: each particle knows its personal best and the one of four other particles that are in predetermined positions. The local best is computed comparing these values. Graphically it can be seen as a polygon where each vertex is connected with 4 others: each particle interacts with the ones that are over, under, to the right and to the left.

In the global best topology, instead, each particle is meant to communicate with all the others. This means that every single particle known its personal best and it's aware of the pbest of all the others. This model allows the particles to, rapidly, converge towards the optima, unfortunately this comes with a cost: it may happen that some part of the space won't be explored. This kind of topology could be not useful to explain swarms' behaviours, but it is, certainly, meaningful when applied to optimization problem like the portfolio selection one: If the fitness's value of each particle is known there is no reason the other particles shouldn't be aware of it, so, given that each particle knows where the optimal place is, it seems correct to assume that they should converge there.

Graphically, it can be seen as a bunch of point that are fully connected one with others.

[44] The local best topology is, probably, the best representations of swarms' behaviors since the birds when flying are usually able to see and understand only what their neighbors are doing.



A) Fully-connected topology B) Ring topology C) Von Neumann topology

In the Original PSO formulation the algorithm was set as follow:

1. Randomly initialize the vectors of positions and velocities.
2. Start the loop.
3. Compute the fitness value for each particle.
4. Compare each particle's fitness with its own personal best. If the particle 's current fitness value is better than its personal best, then the personal best must be updated with the value of the current position[45].

$$p_i^{k+1} = p_i^k \qquad \textbf{if} \quad f(x_i^{k+1}) \geq f(p_i^k)$$

$$p_i^{k+1} = x_i^{k+1} \qquad \textbf{if} \quad f(x_i^{k+1}) < f(p_i^k)$$

5. Identify the particle that, in its neighbourhood[46], has the better personal best, if this value its better than the local best, the variable local best must be updated with the value of the current local best position.

$$lb_j^{k+1} = lb_j^k \qquad \textbf{if} \quad f(clb_j^{k+1}) \geq f(lb_j^k)$$

[45] In this procedure we are assuming that we want to minimize the fitness function.
[46] The neighbourhood of a particle is formed by the particles with which it can communicate and share information; the form of the neighbourhood depends on the communication topology.

$$lb_j^{k+1} = clb_j^{k+1} \qquad \textbf{if } fclb_j^{k+1} < f(lb_j^k)$$

6. Change velocities and positions according to the following formulas:

$$v_i^{k+1} = v_i^k + c_1 * rand_1 \otimes (p_i - x_i^k) + c_2 * rand_2 \otimes (lb_i - x_i^k)$$

$$x_i^{k+1} = x_i^k + v_i^{k+1}$$

Where c_1 and c_2 are two constants for a motivation that we will see later in this chapter, $rand_1$ and $rand_2$ represent two vectors of randomly generated numbers between zero and one and $\otimes$ represents the tensor product

7. Control if criteria/criterions is/are satisfied. Stopping criteria may depends, for example, on number of iterations or fitness function's best value. If they are satisfied go to step number eight otherwise return to step number two.
8. Exit loop.

This original formulation has been modified by subsequent revisions in order to improve the utility of the PSO. These revisions have taken in consideration all the part of the PSO that influence the particles behavior: **communication topology, velocity and fitness function.**

3.3 Velocity

As we can see in the 5th and 6th step of the PSO's algorithm, velocity and, consequently, position in the step k+1, depends on several factors. For the sake of simplicity but without loss of generality, we will discuss about velocity in the PSO fully connected topology. Hence, each particle will know only its personal best and the global best. This "simplification" is particularly useful because it permits to avoid using different local bests and it focus only on the global best in the velocity computation.

Decomposing the velocity's formula, we note that in the step k+1 it depends on:

- Previous velocity: In the first step it is a randomly generated number while in the following it just the results of its formula.

- C_1 and C_2: they are two parameters that influences, respectively, the importance we give to the local best and the global best. The higher the value the more one or the other best is represented in the velocity computation.
- $Rand_1$ and $Rand_2$: they represent two vectors of randomly generated numbers between zero and one
- P_i-X_i: this component serves as an indicator for the distance between the actual particle's position and its best. If the particle is moving away from its best this difference will be different from zero, hence it will try to "accelerate or decelerate" the particle in the attempt to get it closer from its local best position.
- G_i-X_i: This component represents the distance between the actual particle's position and the global best. As in the previous component, it will be zero if the actual particle is the global best otherwise it will try to attract the particle towards the global best.

It is worth to notice that in the first interaction each particle cannot compare its fitness value with a previous one (because it doesn't exist), for this it descends that every particle lay on its on personal best. For the way in which the algorithm is constructed, without an initial randomly generated velocity the particles would stop after the first interaction. Unfortunately, a velocity function constructed in this way may create a rapidly degenerating velocity, this, in turns, implies that the particles will move "too fast" failing in correctly observe the search space.

Practically, an uncontrolled velocity function may lead to miss the local and the global optima. The authors themselves noticed this issue and they proposed to impose an exogenous factor (V_{max}) to fix the problem. If one hand the introduction of a maximum velocity may solve the degeneration problem, on the other hand it is restrictive for the exploration activity of the particles.

Only three years after the introduction of the PSO, two researchers (Shi and Eberhart on their paper-"Comparing inertia weights and constriction factors in particle swarm optimization") proposed an ingenious solution for this problem based on the *inertia weights*.

In order to understand their idea, we need to have in mind the basic concept of the PSO: The PSO process is, fundamentally, articulated in two "sub-processes":

1. Exploration: In this phase the individual's results are important, each particle moves rapidly into the search space.
2. Exploitation: In this phase the comparisons between the personal results are important: the global best is searched, and the particles should move slowly.

Clearly, the aim of the researchers is to find a balance between these two sub-processes; to be more precise, a perfectly function PSO should "explore" in its first's iterations and "exploit" in its lasts. In this manner during the first phase a good part of the search space will be known and in the second phase all the available information will be share in order to find the effective optimum position.

As suggested by Shi and Eberhart, it possible to achieve this result by acting on the velocity of the process. Practically they added a factor ω(weight) to the velocity formula and they imposed that this factor would be interactions dependent.

$$v^{k+1} = \omega v^k + c_1 rand_1(p_i - x_i) + c_2 rand_2(Gb - x_i)$$

When this factor is greater than one, it will tend to increase the velocity favouring the exploration hence the research of the various personal best. When this factor is lower than one, it will tend to decrease the velocity favouring the exploitation hence the research of the global best.

The first suggestion to link the value of the inertia-weight to the interactions was the following:

$$\omega^k = \omega_{max} + \frac{\omega_{min} - \omega_{max}}{K} * k$$

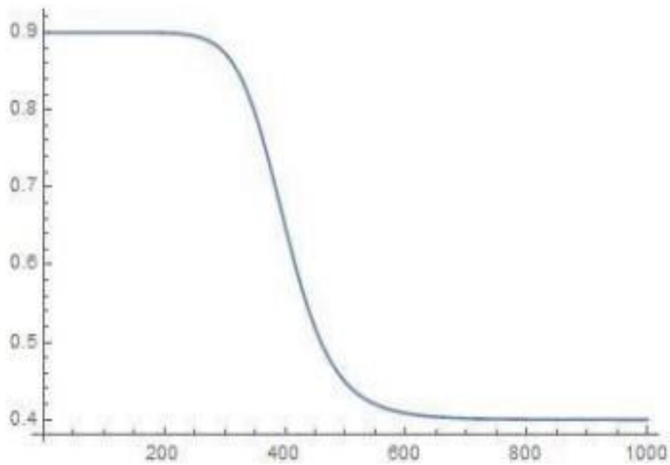
where w_{max} and w_{min} are arbitrarily set[47], while K is the total number of interactions and k is the interaction's number.

Another example of inertia-weight strategy is the Butterworth inertia weight strategy proposed by Xianming Zhua, Hongbo Wang in their paper "*A new inertia weight control strategy for Particle Swarm"*

Optimization.

$$\omega^k = \omega_{max} + \frac{1}{1 + (\frac{t}{p_1})^{p_2}} + \omega_{min}$$

where p_1 and p_2 are 2 numbers that are needed to increase or decrease the decrement speed of the inertia weight[48], t is the number of interactions and W_{max} and W_{min} are, respectively, the maximum and the minimum desired velocity. As in the other inertia Weight approach, the weight rapidly decreases after the first interactions.



Y-axis=weight

X-axis=interaction's number

Figure 4 Weight function under the Butterworth approach, W_{max}=0.5, W_{min}=0.4, T=1000, p_1=330, p_2=10

[47] Empirical results suggest to use as maximum velocity 0.9 and minimum velocity 0.4.

[48] Empirical results suggest to set p1 equal to a third of the total number of iteration and p2 equal to 10.

Using the inertia-weight approach is possible to slowly reduce the influences of the previous velocity, and interaction by interaction, achieve the desired features of an optimal PSO.

In the end, the inertia-weight approach has the advantages to avoid the degeneration of velocity without using a statical maximum velocity V_{max} and to facilitate the two phasis of the PSO (exploration and exploitation). Although the inertia-weight approaches are useful in order to prevent velocity degeneration, they need, in their computation, the assumption of a maximum weight parameter. As we have seen, this may reduce the maximum possible velocity and lead the particles to avoid some parts of the search space.

Clerk and Kennedy (2002) presented a model able to eliminate the Weight parameters without losing is precious features of convergence and velocity containment: the *constriction coefficient* **χ.**

The idea behind the model is to manipulate the velocity's addenda rather than impose a weight on the first element of the sum (previous velocity).

Practically, the coefficient *chi* is multiplied by the whole velocity formula obtaining:

$$v^{k+1} = \chi(v^k + rand_1(0,\varphi_1) * (p_i - x_i) + rand_2(0,\varphi_2) * (Gb_i - x_i))$$

where $\varphi_1 + \varphi_2 = \varphi > 4$ And $\chi = \frac{2}{\varphi - 2 + \sqrt{\varphi^2 - 4\varphi}}$

The authors themselves presented the optimal value for φ imposing it equal to 4.1, hence the result for χ would be 0.7296.

In this way is possible to create a well-functioning PSO without the implementation of any restrictive parameters. Furthermore, this method can be traced back to the easiest formulation of the inertia weight approach because: if φ=4.1 and φ_1= φ_2=2.05 the maximum value that χ*rand(0, φ_1) = χ*rand(0, φ_2)=1.4961*rand(0,1)

Which is equivalent to write 1.49618*rand(0,1), so the model can be rewritten as:

$$v^{k+1} = \omega v^k + c_1 rand_1(p_i - x_i) + c_2 rand_2(Gb - x_i)$$

Assuming $\omega=\chi$ and $c_1=c_2=1.49618$

This final solution for the calculation of the velocity is the one majorly adopted nowadays for particle swarm optimization algorithm. In addition to the above presented advantages, constriction parameter is static: it doesn't need to be change through the iterations simplifying the implementation of the whole model.

3.4 Fitness function, constrains and penalty

The fitness function can be seen as the core of the PSO, it gives a "valuation" to each particle in the optimization. In this manner is possible to compare the particles with each other and find out which one owns the best characteristics. Since the fitness function is, generally, represented by the argument of the portfolio minimization problem, to find and compare the different fitness functions of the particles a constrained nonlinear optimization problem (CNOP) is addressed. Mathematically the CNOP can be explained as follow:

Find $\vec{x}$ that optimize $f(\vec{x})$ bounded to:

$gi(\vec{x}) \leq 0$, $i=1,\ldots,m$

$hj(\vec{x}) \leq 0$, $j=1,\ldots,p$

where:

- $\vec{x} = [x1, x2, \ldots, xN]'$ represents the vector of solutions;
- m is the number of bounds expressed as inequality
- p is the number of bounds expressed as equality

When constrains are added to a problem the region of possible Solution (S) is divided into two sub-regions: feasible and unfeasible. When the results fall into the unfeasible region it means that one or more constrains have been violated thus the solution must be discarded.

Unfortunately, PSO was not designed to deal with constrains because, in its first formulation, is a technic to solve unconstrained optimization problems, hence it is not able to understand when the results fall in the admissible region or in the other. In order to use the PSO to solve constrained optimization problems it is necessary to add, into the procedure, a “a constraint management mechanism”. Many of this mechanism have been presented so far, they are for example:

- admissibility function and reinitialization of the particles when they are not in the feasible region: in this case the PSO cycle is re-initialized as many times as is necessary for all particles to be in the admissible region.
- random or bounded displacement of particles within the feasible region: in this case if the solution of the PSO optimization lays outside the admissible region is moved inside it through the intervention on the velocity of the particle. In the bounded mode the particle which lays in the unfeasible region is moved only until it reach the feasible one while in the random mode the intervention on the particle’s velocity is, as the name suggest, random.

Both methods have their drawback, for what concern the former the main problem is the computational one: re-initialize the particle as many times as it is necessary for all the swarm to be in the admissible region can be a great computational effort, especially if the swarm is composed by a lot of particles. For the latter, instead, if the intervention on the PSO results becomes too frequent, there is a risk of invalidating the very mechanism on which the algorithm is based, i.e. the self-organization of the swarm of particles.

In this work we will present the mechanism to add the constrains that has been used for the computation of our optimal portfolio: the **exact** penalty function[49].

[49] The exact penalty function is a method to avoid that the PSO results fall into the unfeasible region, its functioning is pretty simple: each time a solution falls outside the admissible region (and so a constrain is violated) a worsening factor (penalty) is added to the fitness function.

This choice has been undertaken for two main reasons: First, the exact penalty function is simple to implement and understand because it acts directly on the PSO’s fitness function. Second, some other methods need to implement another metaheuristic in order to compute the penalty[50] to apply to the PSO’s fitness, this, clearly, generate computational difficulty and circumscription in the solvable problems.

In penalty function each constraint’s violation has a “cost”. This cost directly increase the value of the fitness function: the more a particle violated one or more constrains the more its fitness function will be high. This directly implies that particles which violate constrains become “less attractive” then the ones that remain in the feasible region.

Mathematically the implementation of the exact penalty function to the fitness function can be seen as:

Min $f(X)+\epsilon P(X)$ where:

- ϵ is a positive constant that multiplies the constraint’s violations. The higher the value of epsilon the more the violations become significative in the fitness computation.
- $P(X)$ is the function that represents the sum of the violations of the constrains, it must be continuous.
- $P(X)\geq 0$ for each X

It seems that increasing the value of ϵ may be a good idea to avoid that the PSO’s results fall into the unfeasible region since if ϵ is large also small constrain violations impact on the fitness function; unfortunately, it may happen that as epsilon increase the solution of the unconstrained problem will be different to the solution of the constrained one.

Fortunately, this problem can be avoided trought the use of the **Exact** penalty function as demonstrated by Luenberger and Ye(2008).

[50] The difference between exact penalty and penalty is that the former the solutions of the original constrained problem are equal to the solution of the unconstrained one (the one that use the penalty); while in the latter there is approximation and results may differ.

Theorem. Exact Penalty: Let $x*$ be a point that satisfies all the sufficient conditions of the second order to be a local minimum point of a constrained problem and let λ and μ be two vectors containing the associated Lagrange multipliers with respect to m constraints in $h(x)=0$ and to p constraints in $g(x)\leq 0$. Then, for $\epsilon>\max\{\lambda_i|\mu_j: i=1,\ldots,m\ j=1,\ldots,p\}$, $x*$ is also a local minimum for the unconstrained penalty objective function.[51]

Hence, when exact penalty function is implemented, there is a theorem proving coincidence between unconstrained problem's solutions and original constrained problem's solutions.

3.5 Fitness function of a real portfolio

One of the most interesting things of the PSO's algorithm is that it's sufficient to modify the fitness function in order to adapt it to different problems, hence it is applicable to a wide set of optimization problems.

When dealing with the specific problem of the portfolio optimization, as conceived by Markowitz, the fitness function shall consider the risk of the portfolio and a series of other constrains. For the discussion we have done in the previous chapter the constrains have to be transformed into penalty hence the fitness function, which, practically, correspond to the function we are trying to minimize becomes:

Min (risk measure) + ϵ (sum of constrains' violations)

The scope of the first addendum of the formula is clear: reduce the value of the considered risk measure. Since the risk measures we have presented in this work are part of the coherent risk measures class they respect the positivity axiom hence, the smaller they are the safer is the portfolio.

While for the second addendum another discussion shall be done: if this part is neglected the algorithm will search for the portfolio that minimize the risk without take

[51] "Linear and NonlinearProgramming"- Luenberger (1984)

in consideration any other factors. By adding more and more constrains into the portfolio optimization problem it become more realistic and closer to the rational investor's interest. Obviously, every added restriction implements the mathematically difficulty of the problem. Since the scope of this work is valuing the results (in an out of the sample period) of different portfolios which were constructed utilizing different risk measures, we will focus on few constrains: the ones that the authors consider the "minimum" in order to build a rational portfolio.

In the following part we will show some constrains and their relative transformation for the PSO's algorithm.

1.Budget constrain

The **budget constrain** refers to the bunch of securities (stocks in our case) that the investor owns in his/her portfolio. It ensures that the sum of all the participation will be equal to one, hence the total available wealth is invested. To be more precise, the percentage of participation is expressed as a vector of numbers between 0 and 1 computed as: capital invested in the i-th stock/ total available wealth.

Mathematically it can be expressed as:

$$\sum_{i=1}^{n} x_i = 1$$

Where x_i is the quote of investment gave to the i-th stocks, the equality can be expressed in a more compact way in its matrix form as **x'**e=1

Its conversion into a violation to use as an exact penalty function is straightforward:

$$|\sum_{i=1}^{n} x_i - 1| = 0$$

There is a characteristic that is worth to note: the difference is inside an absolute value, this is due to the fact that when the difference is different from zero it must negatively impact on the fitness function no matter if its negative or positive[52]

2.Return constrain

The other constrain that must be considered in the portfolio optimization problem is the **return constrain**, indeed, as we have seen in the previous chapter, if the return constrain miss the PSO's algorithm (as well as the other portfolio optimization methods) will suggest buying the less risky[53] portfolio without consider the reward.

The return constrain overcomes this problem consenting to choose, on the efficient frontier, the portfolio that has a given return. It can be expressed as[54]:

$$\sum_{i=1}^{n} x_i * \overline{r_i} = \pi$$

where:

- X_i is the percentage of capital invested vector.
- r_i is the average returns, computed as sum of the returns of the i-th portfolio divided by N.[55]
- Π is the desired return.

Practically, this constrain implies that the return of the portfolio (in the sample period) must be equal to a certain value π. This limitation can be relaxed in order to enlarge the feasible region and the constrained transformed into:

$$\sum_{i=1}^{n} x_i * \overline{r_i} \geq \pi$$

This enlargement on the feasible region is due to the fact that each rational investor looking for an investment with a certain return π would accept another investment that,

[52] The equation can be expressed in its matrix form as **x**'e -1=0
[53] Here less risky is intended as the portfolio with the lower value of the relative risk measure.
[54] It can also be expressed in matrix form as **x'**r=π
[55] N is nothing but the number of assets in the portfolio..

ceteris paribus, would give him/her a higher return. The conversion into an exact penalty function can be expressed as:

$$max(0, \sum_{i=1}^{n} x_i * \overline{r_i} - \pi) = 0$$

Utilizing this function, the objective of obtain a return equal or higher than π is achieved. Indeed, the function is different from zero, hence negatively accounted into the fitness function, if and only if the portfolio's return is lower than the desired return pi.

It is notable that, conceptually and mathematically, minimize the risk given a certain return or maximize the return given a certain risk is the same thing. However, it is a common practice to follow the former method, hence minimize the risk given a certain return, reflecting and highlighting the fact that the rational investor is risk averse (hence it prefers to avoid risk). The practice is so rooted into the financial sectors that, in all the developed countries, financial institutions are required by law to interview their clients to try to understand their degree of risk aversion.[56]

The two constrains we have illustrated so far, together with any risk measure, are sufficient to create a meaningful portfolio solving a portfolio optimization problem even if they don't take into account a part of the issues of the real stock markets.

The last two constraints that I decided to include in the PSO optimization concern the lower and the upper bounds for the percentage of participation. This choice had to be done to avoid too big or too small percentage in the single stocks. Indeed, without any restriction on the participation bounds the algorithm will suggest acquiring the portfolio that minimises the risk given a certain return without any consideration about its composition. This, in turn, may lead to 2 different situations:

[56] The European example is the MIFID interview: financial institutions are not only required to understand the risk aversion of their clients but if their clients are experienced on the financial instruments they are buying as well.

1. too much concentration: a portfolio with few securities they have a high level of participation, this situation must be avoided since it doesn't take the advantage from benefit of diversification;
2. too small positions: a portfolio where some securities have small percentage, this situation must be avoided because too small positions are simply meaningless: the transaction costs would not be justified for tiny positions.

Constraints on minimum and maximum participation are:

$x_i \leq |u_b|$ and $x_i \geq |l_b|$ for all i with $|l_b| \leq |u_b|$

where:

- L_b represents the lower bound.
- U_b represents the upper bound.

Both l_b and u_b are represented by a number comprised between zero and one, moreover it is considered the absolute values of u and l since the aim of restrictions is not avoid the short selling but it is rather avoid large or small positions either positive or negative. Furthermore, without the absolute value a constraint on the maximum number of stocks in the portfolio should be included. this happen because the maximum number of stocks is limited by the lower bound $max.n = \frac{1}{l_b}$ and by the upper bound $min.n = \frac{1}{u_b}$.

However, this constrain will not be presented since it is not used in this paper.

The lower an upper bound constraints can be transformed into a violation applicable to exact penalties in the following way:

$$max(0, (|x_i| - u_b)) \qquad max(0, l_b - |x_i|))$$

Where the results negatively impact on the fitness function if and only if X is greater than the upper bound or smaller than the lower bound.

At the end of this discussion about the set of constrains that we have applied to the PSO algorithm we can summarize in a formula what will be the fitness function of our model:

$$f(x) = min\left\{risk\ measure \right.$$
$$+ \varepsilon[|\sum_{i=1}^{n} x_i - 1|$$
$$+ max(0, \sum_{i=1}^{n} x_i * \bar{r} - \pi)$$
$$\left. + \sum_{i=1}^{n} max(0, |x_i| - u_b) + \sum_{i=1}^{n} max(0, l_b - |x_i|)\right\}$$

remembering that its aim is reduce the risk while obtaining a return which is equal or greater than a certain value without take position larger than u_b and smaller than l_b.

CHAPTER 4
CASE STUDY AND METHODOLOGY

The aim of this paper is to compare the results of different portfolios obtained by the PSO of different risk measures in order to understand if one of these can be considered better than the others.

In the attempt to reach this goal, the selected portfolios shall be evaluated in the same periods and starting from the same assumptions, Moreover, a set of considerations need to be done:

1. multiperiod observations are mandatory: indeed, it may happen that in a relatively quiet or growing period a risk measure Performs better than the others while in a recession period another (being more conservative) obtains better results.
2. the results cannot be evaluated only on a performance\return basis, but other indicators are needed to understand the true risk reward trade off. Clearly, it doesn't make any sense compare the different risk measures we used to compute the optimal portfolios Since they start from different assumptions, hence other indexes will be introduced.
3. The PSO's parameters, that we decided to use in the computation of the optimal portfolios, influence the final results.
4. the results are only indicative: even if large samples and statistical methods are implemented, what happened in the past may not be indicative for what will happen in the future.

4.1. Methodology on the samples

The method we will use for the risk measures evaluations is the classical "event studies" for economic evaluations. Practically, it consists into divide the data in two samples: "in the sample data" and "out of the sample data" divided by the event. In the first sample the “Normal Returns” are computed while in the second one this Normal Returns are confronted with the actual returns to get if there is stationarity or, in other words, if there is presence of abnormal returns.

When studying the performance of optimal portfolios, the approach is a little bit different: Data inside the sample are used to compute the optimal portfolio following the rules of the different risk measures’ fitness function while, data “out of the sample” are used to evaluate the performance of the optimal portfolios. The reasonings behind this method are based on the theoretical repetitiveness in the behaviours of the stock market: if the periods "in the sample" and "out of the sample "are not too long and they are closed one with the other, the optimal portfolios should keep their "optimal features" in both periods. Clearly, it is only a theoretical point of view, since it doesn't take into consideration that events such as crisis, external information, institutional breaks, government interventions, etcetera... may affect the market. Anyway, "out of the sample" period can be seen as virtual future where we can test if the optimal portfolios behave in the same way they did in the sample period. The data on which we rely are the closing prices of the 30 stocks that compose the Italian index FTSE MIB[57]. So, we rely on data on a daily basis, the reasons for this decision are simple: having to deal with long periods data with high frequencies are useless because a rational investor, differently from a trader, is interested into hold his\her portfolio for at least the medium

[57] It is an Italian weighted index: Each stock inside the index has a specific weight which depends on its capitalization. It is composed by the 30 biggest (more capitalized) firms of the Italian stock market. Before 2009 its name was MIB30.

term[58]. Moreover, stocks prices on a daily basis are easy to find. Data were gathered in the period 01/01/2006-01/01/2021, and they concern the closing price of the 30's stocks of the FTSE MIB. Since the FTSE MIB composition depends on the capitalizations of the Italian stock market' firms, which are clearly variable, it has changed during time. Hence, the data we gathered pertains to 2 different FTSE MIB as shown in the table behind:

FTSE MIB 2006-2009	FTSE MIB 2014-2021
A2A S.p.A.	Amplifon S.p.A.
Amplifon S.p.A.	Assicurazioni Generali
Assicurazioni Generali	Atlantia SpA
Atlantia SpA	Azimut Holding S.p.A.
Autogrill S.p.A.	Banca Mediolanum S.p.A.
Azimut Holding S.p.A.	Banco BPM S.p.A.
Banca Mediolanum S.p.A.	BPER Banca SpA
Banca Monte dei Paschi di Siena S.p	CNH Industrial N.V.
Banco BPM S.p.A.	Davide Campari-Milano N.V.
BPER Banca SpA	DiaSorin S.p.A.
Buzzi Unicem Spa	Eni S.p.A.
Campari - Milano S.p.A.	Ferrari N.V.
Enel S.p.A.	FinecoBank Banca Fineco S.p.A.
Eni S.p.A.	Interpump Group SpA
Exos N.V.	Italgas S.p.A.
Geox S.p.A.	Leonardo S.p.a.
Interpump Group SpA	Mediobanca Banca S.p.A.
Intesa Sanpaolo S.p.A.	Moncler S.p.A.
Leonardo S.p.a.	Pirelli & C. S.p.A.
Mediobanca Banca S.p.A.	Poste Italiane S.p.A.
Recordati Industria Chimica S.p.A.	Prysmian S.p.A.
Saipem S.P.A.	Recordati Industria Chimica S.p.A.
Snam S.p.A.	Saipem S.P.A.
Stellantis N.V.	Snam S.p.A.
STMicroelectronics N.V.	STMicroelectronics N.V.
Telecom Italia SpA	Telecom Italia SpA
Tenaris S.A.	Tenaris S.A.
Terna S.P.A.	Terna S.P.A.
UniCredit S.p.A.	UniCredit S.p.A.
Unipol Gruppo S.p.A.	Unipol Gruppo S.p.A.

Table 1 Composition of the FTSE MIB during different years

[58] Medium term is a generic definition, usually it can be interpreted as one year or more.

The firms highlighted in yellow were comprised in the 2006' FTSE MIB but not in the 2014's one; while for the firms highlighted in green is true the vice versa.

To be coherent with what we said at the beginning of this chapter about the necessary evaluation in times that are substantially different, 3 diverse periods will be evaluated. the first period starts the 1st of January 2006 and it ends the 17st of December 2008, the second starts 1st of January 2014 and it ends the 1st of January 2017 while the last one start the 1st of January 2018 and it ends the first of January 2021[59].

The period's choice has been undertaken considering that the former and the latter were trouble periods since in the first case there where the Lehman Brothers default and the sequent financial crisis while in the second one, due to the pandemic outbreak, the world economic suffered a huge shock. The period in the middle instead can, somehow, represents a relatively quiet period. The sample periods' length was set equal for the three cases, and it will last the two third of the total observation days (two years), consequently the evaluation period (Virtual future) was set Equal to 1/3(one year).

Period n.	SAMPLE PERIODS (Start-end)	OUT OF THE SAMPLE PERIODS (Start-end)
1	1st January 2006-20th December 2008	20thDecember2008-19th December 2008
2	1st January 2014-1st January 2016	1st January 2016–1st January 2017
3	1st January 2018-1st January 2020	1st January 2020-1st January 2021

This decision, made in the light of the fact that the two subsamples shall be sufficiently long and consecutive, consents to satisfy the intentions of the general rational investors.

[59] Clearly the composition of the FTSE MIB is changed during the years.

Clearly, the closing prices need to be transformed into daily returns for simple comparative reasons.[60] In the attempt to be as consistent as possible we choose to use the logarithmic returns rather than then normal return since, the observations over the prices are continuous. This means that the principles which rules over their computation are the ones of the continuous compounding[61].

Mathematically, the logarithmic returns can be expressed as: $R_{i,t} = ln(\frac{P_{i,t}}{P_{i,t-1}})$

Where $P_{i,t}$ and $P_{i,t-1}$ represent, respectively ,the price of the i-th stock at the time T and T minus 1

.

4.2 Index as performance measures

Since various risk measures have been implemented in the particle swarm optimization algorithm in order to compute the optimal portfolios, the results they gave in the out of the sample periods are not completely and directly compatible. This depends on how the risk measures are constructed: for example, given a certain alpha the expected shortfall will be always bigger (in absolute value) than the value at risk; the same would happen if we took in consideration the shortfall deviation risk or the entropic var.

Clearly, we could decide that one of the risk measures we used to create the optimal portfolios is our risk index; unfortunately, following this way the optimal portfolios constructed with the decided risk measure would certainly perform (in terms of risk) better than the others. Hence, we cannot compare the risk measures one with the others nor we can pick just one of them. That's the reason why we decided to implement a set of new measures as indexes for the risk of the optimal portfolios in the out of the sample

[60] returns are expressed in percentage while prices are not comparable.

[61] in the continuous compounding $P_t=P_{t-1}*e^{rt}$

periods. For the reasons will see in the next chapter these measures are Sharpe ratio, Sortino ratio, Farinelli Tibletti ratio and Traynor ratio.

4.3 Sharpe ratio and Sortino ratio

The Sharpe ratio is one of the most famous and widespread performance index for portfolio evaluations. It was named after the 1990 Nobel Price William Sharpe.

The idea that Sharpe developed in 1966 is based one the "reward-to-volatility" concept: It can be summarized in the question "how many units of risk should I take in order to obtain a unit of reward?".

Since we are going to use this index for portfolio evaluation we will present only the *ex-post* formulation, remembering that all the following indexes can be used for portfolios constructions as well[62].

$$Sharpe = \frac{r_p - r_f}{\sigma_p}$$

where:

- R_p is the portfolio return in the observed period, computed as the mean of the daily returns.
- R_f is the risk-free return, it is nothing but the return of an asset that has no risk.[63]
- Sigma is the standard deviation of the portfolio.

The Sharpe ratio strength is its simplicity, it is computable without any assumption on the source of profitability only knowing the returns' time series of the portfolio. Moreover, it is easy to understand: *Ceteris paribus*, a portfolio with a higher Sharpe ratio is preferable to another with a lower one. Unfortunately, such a simplicity comes with some important drawbacks:

---

[62] It is sufficient to substitute the returns with their expectation in order to compute the *ex-ante* sharpe ratio.

[63] It is generally assumed to be a bond issued by Governments, such as the American treasury bond or the Italian BTP.

1. If the Sharpe ratio is negative when the variance increases its value decrease conducting to the wrong conclusion that a higher variability can be safer.
2. The results of the Sharpe ratio are clearly scalars, they are not easily comparable. For example, how much better is a Sharpe ratio of 0.5 then one of 0.1?
3. It is hard to find the true rich risk-free rate although it is generally assumed to be interest rate of the treasury bonds.
4. It can be easily manipulated: for illiquid assets their prices, and consequently their returns, can rapidly increase without the true possibility of cash out. This would lead Sharpe ratio to increase as well indicating a safer position even if it is not the case. the most famous example is the Ponzi scheme.[64]
5. The denominator of the formula is the standard deviation and, as we know, it considers both downside risk and upside potential, hence the eventual presence of positive skewness will decrease the Sharpe ratio.

For what concern the drawbacks 3 and 5 a possible solution has been developed by Frank Sortino: the Sortino ratio. At the first sight it can be seen as a direct evolution and improvement of the Sharpe ratio.

Mathematically, it can be expressed as:

$$Sortino = \frac{R_p - T_r}{TDD}$$

where:

- R_p is the portfolio return.

[64] A Ponzi scheme is a form of fraud that lures investors and pays profits to earlier investors with funds from more recent investors

- T_r is the target return: it substitutes the hardly definable risk-free rate. it is nothing but the required rate of return for the investment strategy under consideration[65].
- TDD is the target deviation return: it solves the problem of considering the positive skewness as negative for a portfolio. Practically, it computes the standard deviation only for returns that are smaller than the target return.

$$TDD = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Min(0, x_i - T))^2}$$

for what concern drawback number 4 it can be ignored since the portfolio we are looking for would be composed by the 30th biggest firms of the Italian stock market hence illiquidity is unlikely. Unfortunately, drawbacks number one and #2 aren't easily surmountable.

4.4 Jensen's alpha and Treynor ratio

Jensen's alpha, Treynor ratio and Sharpe ratio derive from the well know capital asset pricing model (CAPM). CAPM is a theory developed to price assets, it states that expected value of an asset or a portfolio should depend on the risk-free rate and, partially, on the benchmark[66]. So, if the market was efficient the expected returns of stock or a portfolio could be expressed as:

$$E[Ri] = Rf + \beta[Rm - Rf]$$

[65] It can be set by the investor depending on his willingness, only in the particular case in which $T_r=r_f$ the numerator of the equation becomes equal to the numerator of the Sharpe Ratio. in its earlier formulation the target return's name was minimum acceptable return.

[66] risk free rate and benchmark are generally assumed to be the rate of treasury bonds and the return of a large stock index respectively. this is due to the fact that a large stock index such as the standard and poor 500 can be seen representative of the stock market.

where Rm represents the return of the market over the period, Rf is the return of the risk-free asset over the period, E[Ri] Is the expected return of the considered asset or portfolio and beta represents the correlation between Ri and the market. This model, by assuming the existence of efficient markets, tells us that there is no room for arbitrages since every stock's returns is based only on the risk free rate and on the correlation between the market and the asset. Making some simple modifications, it is possible from this model derive the so-called equilibrium relation: $r_i = \alpha_i + \beta r_m + \varepsilon_i$

where:

- α represents the influence that the stock has on its own returns, as we have already seen it should not be statistically different from zero since the market are assumed to be efficient. Actually, it is always different from zero.
- βr_m This part of the formula explains the dependence the returns of the stock have with respect to the benchmark. Mathematically, it is the covariance between the stock's returns and the benchmark's returns everything divided by the variance of the benchmark's returns.
- ϵ Represent the eventual error terms.

It is precisely from this formula that two of the more used performance indexes have been derived: the former, Jensen's alpha, is nothing but the capital asset pricing model's alpha; practically, it represents the percentage excess return earned in addition to the required average return over the period. The latter is the Traynor's ratio which represents the percentage of abnormal return earned the unit of systematic risk.

$$TR_i = \frac{\alpha_i}{\beta_i}$$

It is worth to note that the CAPM and the Treynor's ratio (not the Jensen's alpha) take into consideration both TDs idiosyncratic risk and systematic risk[67]. This is a huge

[67] Risks of price changes due to the unique circumstances of a specific security, as opposed to the overall market risk (systematic risk), are called idiosyncratic risks. This specific risk, also called unsystematic, can be nullified out of a portfolio through diversification

difference with respect to the Sharpe and Sortino ratio which are based on a “more simple” total risk.

The four indexes presented so far share a common characteristic, they are based on gaussian assumptions. Sharpe Ratio has, in its denominator, the standard deviation which is the square root of the variance: a measure that is meaningful only if the underlying distribution is elliptical.

For the Sortino ratio the discourse is similar, its denominator’s component (target deviation risk) is pretty similar to the standard deviation.

And, for what concern Jensen’s alpha and Treynor ratio we can state that they are based on Normality assumptions since the theory from which they derive is based on the Gaussian assumptions as well. Hence, these indexes may be in disagreement with what we said in the previous chapters about the non-normality of the returns’ distributions: if these probability distributions are different from the elliptical all the performance indexes presented so far are useless. The last performance index that will be presented is Farinelli and Tibletti Ratio and it is untied from probability’s assumptions, however for the sake of completeness and correctness some simple normality tests have been performed in order to understand if all the other indexes can be properly used to compare the results of the optimal portfolios. The results are presented in the appendix.

4.5 Farinelli Tibletti Ratio

As demonstrated in the appendix, normality assumptions should be left behind: behaviours such as Skewness and Kurtosis are always present in our sets of data.

Hence, performance indexes such as Sharpe or Sortino Ratio become meaningless because the underling probability functions don’t behave like a normal (nor as an elliptical). The reason for their uselessness depends on the fact that they are based on a two-side risk measure (standard deviation) which captures, both, upside risk and downside potential at the same moment. As is widely demonstrated in the appendix the 2 major concerns when dealing whit return's distributions that don't have a gaussian behaviour are:

1. Asymmetry with respect to the mean returns (Skewness)
2. Presence of large deviations (extreme events) in the tails, especially in the left one (Kurtosis).

The impact of these 2 concerns can be evaluated throughout the use of the so-called one-side risk moments: they are nothing but ***coherent*** risk measures that are representative of the random variable's excess returns[68].

The left-side moment can be expressed as: $\rho^{p}_{-,b}(X) = E^{\frac{1}{q}}\{[(X-b)^{-}]^{q}\}$

The right-side moment can be expressed as: $\rho^{p}_{+,b}(X) = E^{\frac{1}{p}}\{[(X-b)^{+}]^{p}\}$

Starting from these assumptions Farinelli and Tibletti were able to build a new coherent risk measure for portfolio evaluations, this risk measure is able to give different weights to "good" and "bad" volatility without any assumptions on the return's distribution.

This risk measure can be mathematically written as:

$$\phi_b^{p,q}(X) = \frac{E^{\frac{1}{p}}\{[(X-b)^{+}]^{p}\}}{E^{\frac{1}{q}}\{[(X-b)^{-}]^{q}\}}$$

It is nothing but the weighted ratio between positive and negative deviation from the benchmark[69]; p and q serve as weights for the investor's preferences:

The more the investor is risk averse the more q should be elevated since extreme events should affect him/her more. The more the investor is interested in a positive skewness ("higher probability of positive returns") the more p should be high.

It is worth to remember a particular case of the Farinelli-Tibletti ratio, the omega ratio.

---

[68] The returns above or below a certain benchmark.

[69] Without any loss of generality, we can assume the benchmark equal to the mean of the distribution.

In the Omega ratio p=q=1 and it represents nothing but the ratio between expected gains and expected loss.

$$\Omega_b(X) = \frac{E[(X-b)^+]}{E[(X-b)^-]}$$

4.6 PSO parameters

As mentioned earlier the decisions about the parameters influence the results and they need to be taken carefully.

In order to make a fair comparison between the different risk measures the common parameters must be equal; in particular I'm speaking about the PSO's parameters:

- Number of interactions:250
- Number of particles: 50
- Epsilon: 0.0001
- Inertia Weight ω: 0.7298
- Individual acceleration coefficient c_1:1.49618
- Social acceleration coefficient c_2:1.49618

Moreover, the constrains parameters shall be equal as well. They are:

- Desired return π:0.0002 per days, which means an annual return of 5% (circa).[70]
- Minimum and maximum participation in a single share: 0.02 and 0.2 respectively.
- α: the level of significance is common to all the three risk measures, and it has been set to 0.95.

[70] The computation has been done using compound interest and assuming 252 opening days a year.

Finally, the problem specific parameters have been decided as well. In particular, for what concern the Shortfall deviation risk, β and p have been set equal to 0.2 and 2 respectively.[71]While for the entropic Var, the t-Var is equal to 1.[72]

[71] The decision derives to the fact that I have assumed that the general investor is risk averse, hence its β is small.

[72] As advised by the author Ahmadi-Javid (2017) in its paper on portfolio optimization.

CHAPTER 5
RESULTS

As mentioned in the previous chapter, three different periods have been chosen for the comparison. In each period and for each risk measure the PSO algorithm is run 5 times; this means that, for each risk measures, five optimal portfolios will be constructed. Then, for each risk measures, only one portfolio has been selected using as criteria the portfolio's fitness value.

In this way for each period only three portfolios have been chosen each one representative of a different risk measure (entropic VaR, Expected Shortfall, Shortfall Deviation Risk). Finally, using the indexes we presented in the previous chapter, the optimal portfolios have been compared.

5.1 First period

The first period starting from the first of January 2006 and lasting until the 20th of December 2008 is divided into a 12-month long sample period and a 6-month long out-of-the-sample period.

PSO has been applied in order to solve portfolio selection problems as presented by Harry Markowitz; Mean-Expected shortfall, Mean-Shortfall deviation Risk and Mean-Entropic VaR have been used instead of Mean-Variance as criteria for the selections.

In the following tables are shown the results of the PSO and the comparison between the optimal portfolios, In the tables are reported the value of the particles (portfolios) which reach the best fitness value and the violations of the constrains as well:

Table 2. PSO results on Mean-Expected shortfall optimization problem, 2006-2009

OUTPUTS	Run 1	Run 2	Run 3	Run 4	Run 5
Best fitness value	110,49896	4,58636	0,01283	4,73367	17,164155
Return constrains	0	0	0	0	0
Budget constrains	5,98E-09	6,37E-12	0	2,5E-13	2E-14
Min. investment constrains	0,01105	0,00046	0	0,00047	0,00171
Max. investment constrains	0	0	0	0	0

Since we are in front of a minimization problem, the lower the value of the fitness function the better is the optimal portfolio in terms of constrain violations[73]: only if the constrain is not violated they are equal to zero and the fitness function can reach its minimum only if all the constrains are not violated.[74] It is straightforward to note that the first run has a high value of the fitness function with respect to the others, hence, it cannot be the best solution, this high value is due to the fact that it strongly violates the minimum investment constrain. Run 2, 4 and 5 are good approximations of what the algorithm was searching since the fitness function is small and the constrains are equal to zero or really close to zero. However, run 4 is, for sure, the best solution of the optimization problem since it hasn't violation on any of the constrains; this is highlighted by the fact that the fitness function is equal to the Expected Shortfall of the optimal portfolio.

Table 3 . PSO results on Mean-Shortfall deviation risk optimization problem, 2006-2009

OUTPUTS	Run 1	Run 2	Run 3	Run 4	Run 5
Best fitness value	110,60244	0,01754	76,06154	13,74675	76,50440
Return constrains	0	0	0	0	0
Budget constrains	4,88E-15	2,22E-16	1,44E-15	9,61E-12	2,65E-12
Min. investment constrains	0,01106	0	0,00760	0,00137	0,00765
Max. investment constrains	0	0	0	0	0

In this case (with the shortfall deviation risk as risk measure) is extremely evident which run has been selected as the best solution: all the constrains of the second run are, practically, equal to zero.

[74] The minimum value that the fitness function can assume vary every run and it is equal to the value of the optimal-portfolio risk measure.

Table 4. PSO results on Mean-Entropic VaR optimization problem, 2006-2009

OUTPUTS	Run 1	Run 2	Run 3	Run 4	Run 5
Best fitness value	123,45703	42,09697	86,15766	57,38787	8,30977
Return constrains	0	0	0	0	0
Budget constrains	2,51E-08	2,22E-08	3,52E-05	8,59E-06	2,91E-08
Min. investment constrains	0,01230	0,00416	0,00853	0,00568	0,00078
Max. investment constrains	0	0	0	0	0

For what concern the last risk measure, the five run of the algorithm have not found a solution able to satisfy all the constrains. However, the fifth run can be taken as best solution's good approximation since the constrains are not violated so much.

In the end the three optimal portfolios are constructed as follow:

Table 5. Optimal portfolios 2006-2009

Securities	Expected Shorfall	Shortfall Deviation Risk	Entropic VaR
Atlantia S.p.a.	0,035	0,053	0,026
Recordati industria chimica	0,054	0,036	0,027
Interpump Group S.p.a.	0,024	0,038	0,02
Buzzi Unicem S.p.a.	0,028	0,023	0,024
Telecom Italia	0,028	0,022	0,025
Leonardo S.p.a.	0,029	0,06	0,066
A2A	0,034	0,033	0,023
Autogrill S.p.a.	0,026	0,02	0,045
Medio Banca	0,031	0,034	0,02
Banco monte dei paschi di Siena	0,039	0,02	0,04
Azimut holding S.p.a.	0,02	0,02	0,024
Assicurazioni generali	0,048	0,047	0,019
Stellantis S.p.a.	0,021	0,025	0,033
Campari	0,055	0,048	0,026
STMicroelectronics N.V.	0,033	0,031	0,033
Saipem S.p.a.	0,038	0,024	0,021
ENEL	0,046	0,023	0,039
Geox	0,048	0,027	0,047
Terna S.p.a.	0,034	0,03	0,031
Banca mediolanum S.p.a.	0,027	0,02	0,037
Ampifon S.p.a.	0,02	0,021	0,047
Unipol gruppo SPA	0,033	0,028	0,02
Intesa san Paolo	0,03	0,025	0,027
ENI	0,05	0,048	0,041
Unicredit	0,02	0,052	0,026
BPER Banca	0,02	0,039	0,029
Banco BPM	0,024	0,024	0,021
Tenaris S.p.a.	0,023	0,024	0,052
Snam S.p.a.	0,061	0,029	0,069
Exor N.V.	0,022	0,076	0,039

The 3 portfolios so constructed were projected into the virtual future (out of the sample period) in order to study which one would have had the better behaviours. As discuss in the first part of this chapter, not only the returns were compared but also some performance indexes.

Table 6. Comparison between the 3 optimal portfolios, 2006-2009

	Expected Shortfall	Shortfall Deviation Risk	Entropic VaR
Daily returns	-0,002264626	-0,002470505	-0,002454248
Total return	-0,568421193	-0,620096871	-0,61601627
Standard deviation	0,008389573	0,009401976	0,008709013
Sharpe ratio	-0,274658685	-0,266980921	-0,286357448
Sortino ratio	-0,0422719211	-0,0444169021	-0,0473074062
Farinelli Tibletti ratio	0,437828906	0,411554512	0,425278691
Treynor ratio	-0,07583852	-0,080798214	-0,085861735

Finally, a graph showing the returns of each optimal portfolio if 10000 euro were invested. In order to have a better sight of the portfolios results in term of returns, the out-of the sample period was divided into 2 sub-parts, hence there are 2 graphs.

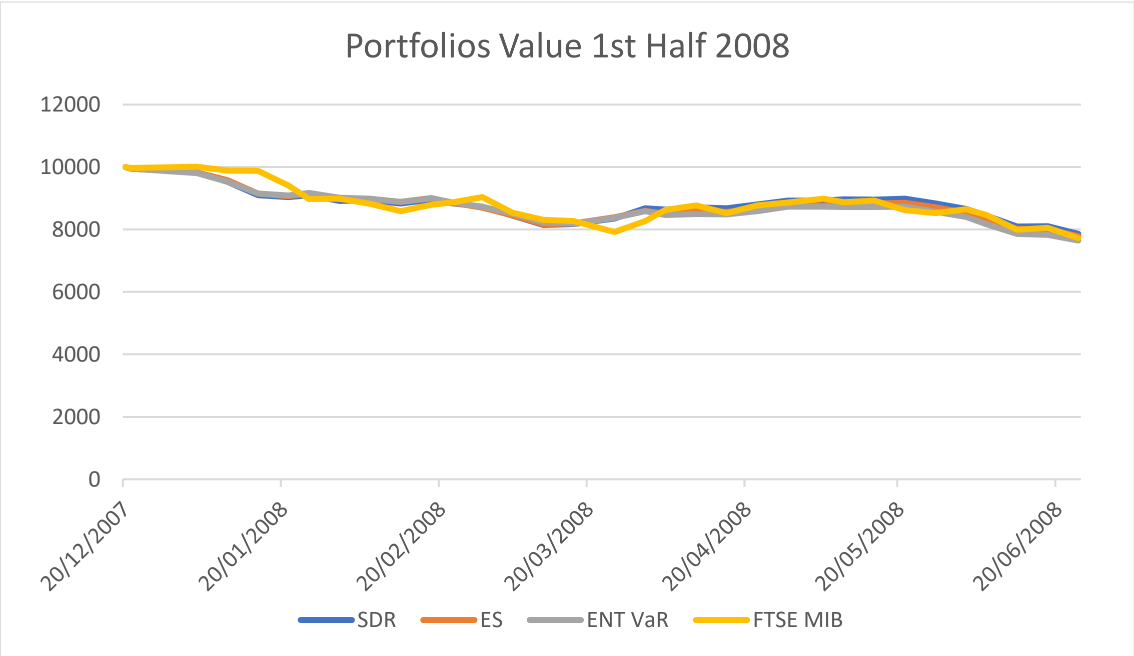


Figure 1 Portfolios Value from an initial investment of 10000 euro; first half of 2008

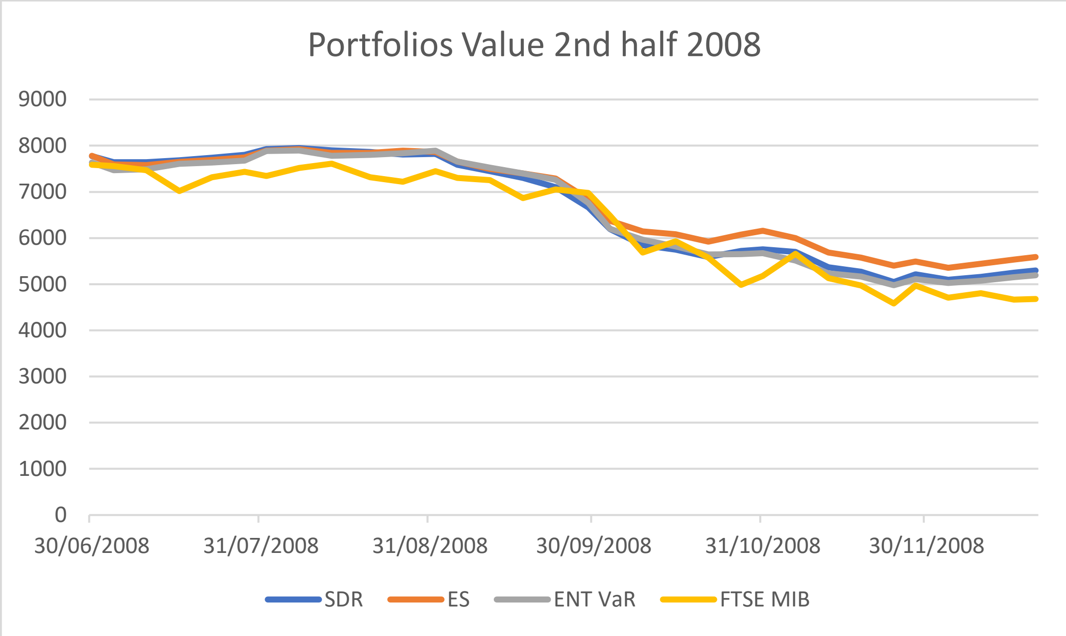


Figure 2 Portfolios Value from an initial investment of 10000 euro; second half of 2008

There is no evidence that one portfolio has outperformed the others in a serious way; all the portfolios had a strongly negative return in the period; anyway, for the second half of the year all the portfolios have outperformed the benchmark. Another noticeable think is the superior behaviour, in the end of the period, of the portfolio constructed using the expected shortfall as risk measure since it has not only a greater return but also better indexes results. For what concern the indexes the 3 portfolios behaved in a similar fashion.

5.2 Second period

The second period starting from the first of January 2014 and lasting until the 1st of January 2017 is divided into a 12-month long sample period and a 6-month long out-of-the-sample period.

In the following tables are shown the results of the PSO and the comparison between the optimal portfolios, In the tables are reported the value of the particles (portfolios) which reach the best fitness value and the violations of the constrains as well:

Table 6. PSO results on Mean-Expected shortfall optimization problem, 2014-2017

OUTPUTS	Run 1	Run 2	Run 3	Run 4	Run 5
Best fitness value	0,02797	0,02833	0,02900	0,02903	0,02908
Return constrains	0	0		0	0
Budget constrains	3E-15	9E-15	1,7E-14	1,3E-13	8E-14
Min. investment constrains	0	0	0	0	0
Max. investment constrains	0	0	0	0	0

In this period the PSO found optimal portfolios (under Mean-Expected Shortfall criteria) that, practically, did not violate any constrain. As in the previous cases the optimal portfolio has been selected using the fitness value criteria hence, the first portfolio has been chosen.

Table 7 . PSO results on Mean-Shortfall deviation risk optimization problem, 2014-2017

OUTPUTS	Run 1	Run 2	Run 3	Run 4	Run 5
Best fitness value	0,03767	0,03763	0,03703	0,03788	0,03771
Return constrains	0	0	0	0	0
Budget constrains	2,98E-14	1,11E-16	6,00E-15	8,88E-16	3,06E-14
Min. investment constrains	0	0	0	0	0
Max. investment constrains	0	0	0	0	0

As for the Mean-Expected Shortfall minimization problem, the PSO easily found five good and similar solutions for the optimal portfolios; the third has been selected as the best.

Table 8. PSO results on Mean-Entropic VaR optimization problem, 2014-2017

OUTPUTS	Run 1	Run 2	Run 3	Run 4	Run 5
Best fitness value	0,40384	0,40925	0,39826	0,40042	0,39598
Return constrains	0	0	0	0	0
Budget constrains	2,22E-16	2,22E-16	1,11E-16	0,00E+00	2,22E-16
Min. investment constrains	0	0	0	0	0
Max. investment constrains	0	0	0	0	0

The algorithm has found five good solutions, the third has been chosen as the best since its fitness value is the lowest.

In the end the three optimal portfolios are constructed as follow:

Table 9 Optimal portfolios 2014-2017

Securities	Expected Shorfall	Shortfall Deviation Risk	Entropic VaR
Atlantia S.p.a.	0,028	0,074	0,036
Recordati industria chimica	0,025	0,02	0,041
Interpump Group S.p.a.	0,036	0,042	0,032
Campari	0,055	0,125	0,029
Telecom Italia	0,032	0,028	0,024
Leonardo S.p.a.	0,034	0,02	0,058
Moncler S.p.a.	0,072	0,04	0,033
Medio Banca	0,022	0,04	0,033
Azimut holding S.p.a.	0,037	0,025	0,069
Generali	0,044	0,02	0,037
Snam S.p.a.	0,035	0,042	0,025
CNH industrial S.p.a.	0,029	0,029	0,06
STMicroelectronics N.V.	0,023	0,04	0,025
Saipem S.p.a.	0,045	0,057	0,039
Terna S.p.a.	0,085	0,031	0,043
Banca mediolanum S.p.a.	0,02	0,028	0,021
Diasorin S.p.a.	0,113	0,092	0,07
Amplifon S.p.a.	0,032	0,022	0,027
Prysmian S.p.a.	0,032	0,025	0,039
Unipol gruppo SPA	0,027	0,046	0,042
Eni	0,032	0,02	0,031
Unicredit	0,02	0,029	0,058
BPER Banca	0,02	0,04	0,055
Banco BPM	0,038	0,02	0,032
Tenaris S.p.a.	0,063	0,047	0,04

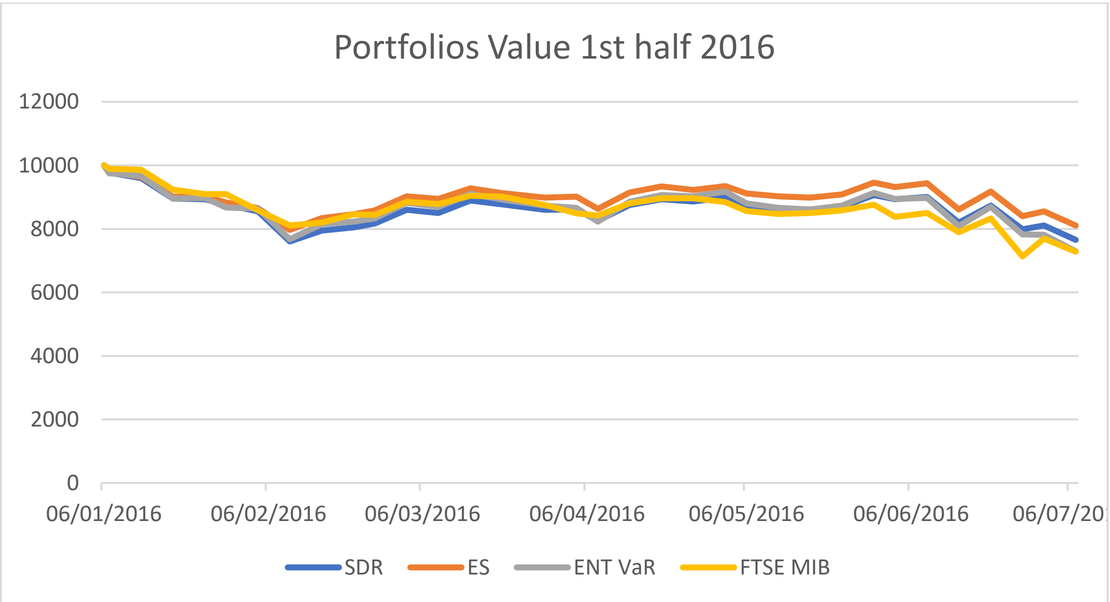
The 3 portfolios so constructed[75] were projected into the virtual future (out of the sample period) in order to study which one would have had the better behaviours.

	Expected Shortfall	Shortfall Deviation Risk	Entropic VaR
Daily returns	-2,74481E-05	-1,40E-05	-0,000203142
Total return	-0,00697183	-0,003557193	-0,051598162
Standard deviation	0,016738064	0,017323665	0,019615035
Sharpe ratio	-0,00398031	-0,003069745	-0,012353631
Sortino ratio	-3,75102E-05	-2,84E-05	-0,000117085
Farinelli Tibletti ratio	0,372173347	0,345813531	0,356821686
Treynor ratio	0,03570241	0,268742955	0,144372975

Table 10. Comparison between the 3 optimal portfolios, 2014-2017

As for the first period, two graphs showing the returns of each optimal portfolio if 10000 euro were invested. In order to have a better sight of the portfolios results in term of returns, the out-of the sample period was divided into 2 sub-parts, hence there are 2 graphs.

Figure 4 Portfolios Value from an initial investment of 10000 euro; first half of 2016



[75] The portfolios constructed in this period have less securities than the ones constructed in the others period, this is due to the fact that 5 securities listed in the FTSE Mib during the 2021 did not exist yet during the 2014.

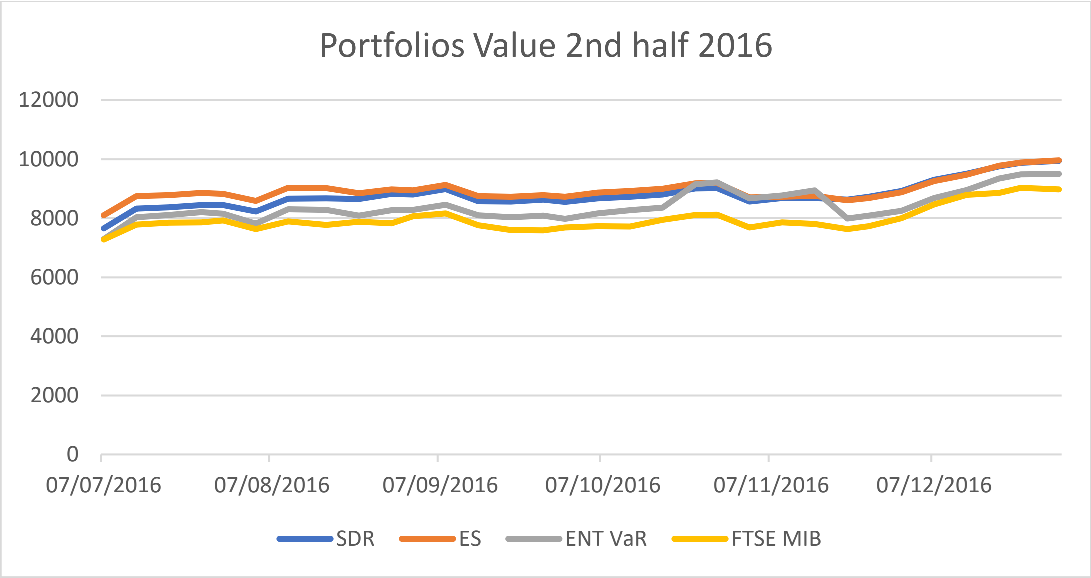


Figure 5. Portfolios Value from an initial investment of 10000 euro; second half of 2016

As in the first period, all the portfolios obtained a negative return, although, in this case, smaller than the previous one. In terms of return all the portfolios constructed with coherent risk measures outperformed the benchmark while, watching to the other indexes, no one of the portfolios seems to behave strongly different from the others.

For sure the worst portfolio is represented by the Entropic VaR since it has a return much lower than the other two and a higher standard deviation as well. It is not possible to state which one between the Expected Shortfall portfolio or the Shortfall deviation risk portfolio behaved better since the former obtained a slightly higher return while the latter had better indexes.

5.3 Third period

The second period starting from the first of January 2014 and lasting until the 1st of January 2017 is divided into a 12-month long sample period and a 6-month long out-of-the-sample period.

In the following tables are shown the results of the PSO and the comparison between the optimal portfolios, In the tables are reported the value of the particles (portfolios) which reach the best fitness value and the violations of the constrains as well:

Table 11 .PSO results on Mean-Expected Shortfall optimization problem, 2018-2021

OUTPUTS	Run 1	Run 2	Run 3	Run 4	Run 5
Best fitness value	93,53366	226,66125	202,20399	231,86508	43,45683
Return constrains	0	0	0	0	0
Budget constrains	2,4E-09	7,5E-06	2,3E-10	0,000116	3,6E-06
Min. investment constrains	0,009351	0,022656	0,020218	0,023069	0,00434
Max. investment constrains	0	0	0	0	0

In this period the PSO did not find a good solution for the optimization problem, especially for what concern the run 2,3 and for 4 the budget constrain and the constrain on the minimum investment were violated (in some case strongly with a 2% or more of violation). The decision about the optimal portfolio to choose followed, as usual, the fitness value criteria, indeed the last portfolio (the one selected) has smaller constrains violation.

OUTPUTS	Run 1	Run 2	Run 3	Run 4	Run 5
Best fitness value	57,63699	121,47178	219,95557	113,76152	466,38351
Return constrains	0	0	0	0	0
Budget constrains	3,89E-15	4,20E-09	3,39E-04	9,93E-09	2,47E-12
Min. investment constrains	0,0057606 4	0,0121440 8	0,0216533 4	0,0113730 2	0,0466349 9
Max. investment constrains	0	0	0	0	0

Table 12 .PSO results on Mean-Shortfall Deviation Risk optimization problem, 2018-2021

For what concern the optimal portfolio constructed following the Mean-Shortfall deviation Risk minimization problem the situation is analogous to the previous one: PSO didn't find a good solution hence the best in terms of fitness value was selected.

Table 13 PSO results on Mean-Entropic VaR optimization problem, 2018-2021

OUTPUTS	Run 1	Run 2	Run 3	Run 4	Run 5
Best fitness value	0,36553	55,76002	41,03038	0,42140	50,97818
Return constrains	0	0	0	0	0
Budget constrains	2,22E-16	1,63E-08	1,85E-04	6,89E-07	1,72E-05
Min. investment constrains	0	0,00553	0,00387	0	0,00503
Max. investment constrains	0	0	0	0	0

Finally, the PSO applied to Mean-Entropic VaR optimization problem obtained excellent results indeed, it found two portfolios (run 1 and run 4) with only a tiny violation on the budget constrain. The first run has been chosen as the best since its violation and fitness are smaller. In the end the three optimal portfolios are constructed as follow:

Securities	Expected Shorfall	Shortfall Deviation Risk	Entropic VaR
Atlantia SpA	0,03	0,02	0,046
Recordati Industria Chimica S.p.A.	0,043	0,059	0,068
Interpump Group SpA	0,033	0,051	0,071
Davide Campari-Milano N.V.	0,043	0,019	0,028
Telecom Italia SpA	0,05	0,029	0,041
Leonardo S.p.a.	0,061	0,02	0,033
Moncler S.p.A.	0,033	0,024	0,026
Ferrari N.V.	0,039	0,058	0,028
Mediobanca Banca S.p.A.	0,045	0,039	0,033
Italgas S.p.A.	0,029	0,024	0,043
Azimut Holding S.p.A.	0,033	0,04	0,026
Assicurazioni Generali	0,028	0,034	0,048
Snam S.p.A.	0,046	0,023	0,024
CNH Industrial N.V.	0,049	0,021	-0,059
STMicroelectronics N.V.	0,051	0,057	0,096
Saipem SpA	0,024	0,028	0,021
Poste Italiane S.p.A.	0,02	0,02	0,063
FinecoBank Banca Fineco S.p.A.	0,024	0,036	0,067
Terna	0,031	0,036	0,034
Banca Mediolanum S.p.A.	0,047	0,082	0,068
DiaSorin S.p.A.	0,027	0,035	0,039
Amplifon S.p.A.	0,026	0,021	0,03
Prysmian S.p.A.	0,02	0,022	0,044
Unipol Gruppo S.p.A.	0,022	0,015	0,027
Pirelli & C. S.p.A.	0,029	0,047	0,028
Eni S.p.A.	0,022	0,028	-0,11
UniCredit S.p.A.	0,022	0,022	0,046
BPER Banca SpA	0,031	0,036	0,034
Banco BPM S.p.A.	0,026	0,034	0,028
Tenaris S.A.	0,026	0,02	0,025

Table 14 Optimal portfolios 2018-2021

The 3 portfolios so constructed[76] were projected into the virtual future (out of the sample period) in order to study which one would have had the better behaviours.

	Expected Shortfall	Shortfall Deviation Risk	Entropic VaR
Daily returns	-0,000336171	-0,000216455	-1,84799E-06
Total return	-0,085051167	-0,05476324	-0,000467543
Standard deviation	0,02245038	0,021874966	0,02102919
Sharpe ratio	-0,016725774	-0,01169304	-0,001958105
Sortino ratio	-0,000601266	-0,000390284	-0,00067846533825
Farinelli Tibletti ratio	0,640647322	0,611215007	0,761945264
Treynor ratio	0,029511144	0,021926138	-0,00347343

Table 15 Comparison between the 3 optimal portfolios, 2018-2021

[76] The portfolios constructed in this period have less securities than the ones constructed in the others period, this is due to the fact that 5 securities listed in the FTSE Mib during the 2021 did not exist yet during the 2014.

Finally, as for the other periods, two graphs showing the returns of each optimal portfolio if 10000 euro were invested. In order to have a better sight of the portfolios results in term of returns, the out-of the sample period was divided into 2 sub-parts, hence there are 2 graphs.

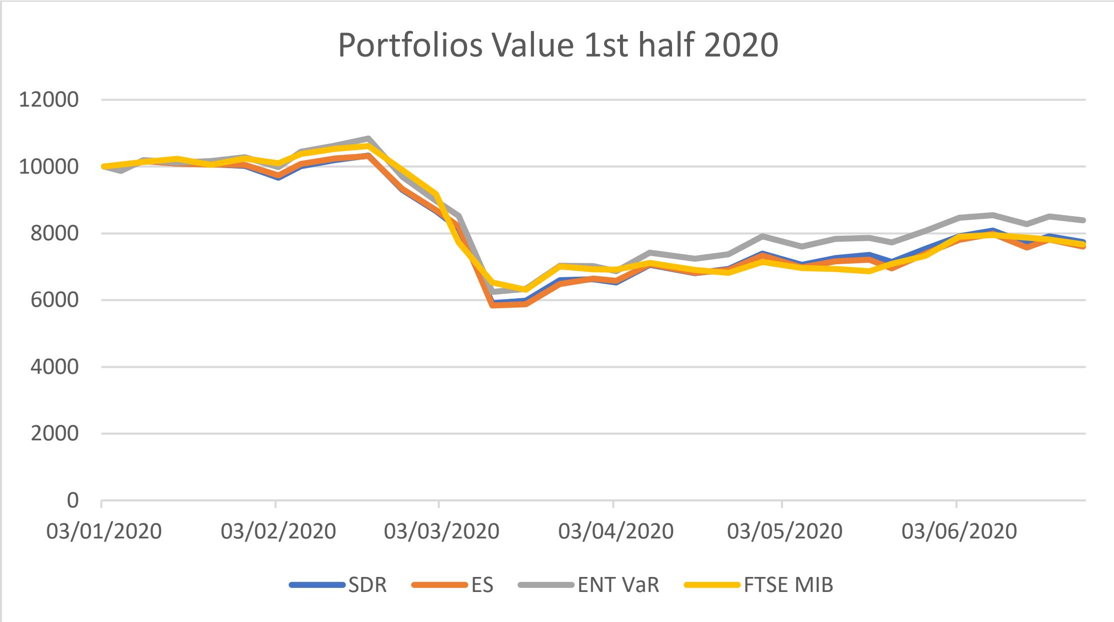


Figure 6 Portfolios Value from an initial investment of 10000 euro; first half 2020

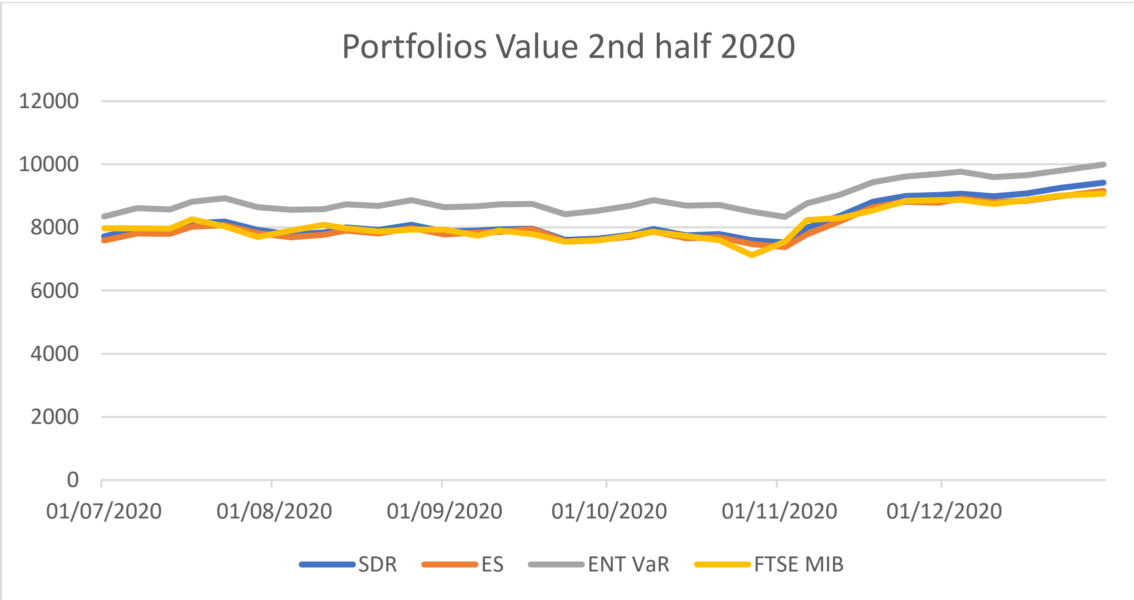


Figure 7 Portfolios Value from an initial investment of 10000 euro; second half 2020

Although for the first quarter of the 2020 all the portfolios performed well, at the end of the year, they all ended with a loss: this is, probably, due to the fact that in the 2020 second quarter there was the Covid-19 outbreak. It is interest to note that after the start of the pandemic, and especially in the 2020 second half, the portfolio constructed using the Mean-Entropic VaR criteria outperformed all the others and the benchmark as well.

5.4 Comparison with Markowitz portfolios

For the sake of completeness, a comparison with Mean-Variance optimal portfolios will be done as well. The approach in computing the Mean-Variance optimal portfolio is slightly different from the same computation for the other risk measures; This descends from the fact that the Mean-Variance optimization problem has close form[77] hence it is not necessary to utilize the Particle Swarm Optimization.

For what concern the comparison, only the worst[78] portfolio of each period will be confronted with the Mean-Variance one; this choice has been undertaken to demonstrated that, even in the worst case, portfolios constructed with Mean-coherent risk measure criteria outperformed the ones built with the Mean-Variance criteria.

The comparison will be done presenting a table similar to the ones presented in the previous chapter and using 3 graphs that will show the returns of each optimal portfolio if 10000 euro were invested at the begging of each out of the sample period.

	Shortfall deviation risk	Variance
Daily returns	-0,002470505	-0,002475018
Total return	-0,620096871	-0,621229465
Std. dev.	0,009401976	0,012763839
Sharpe ratio	-0,266980921	-0,193763792

Table 16 Comparison between the Expected shortfall and Variance Portfolio, 2006-2009

[77] For more information see Chapter 2

[78] In terms of return during the out of the sample period

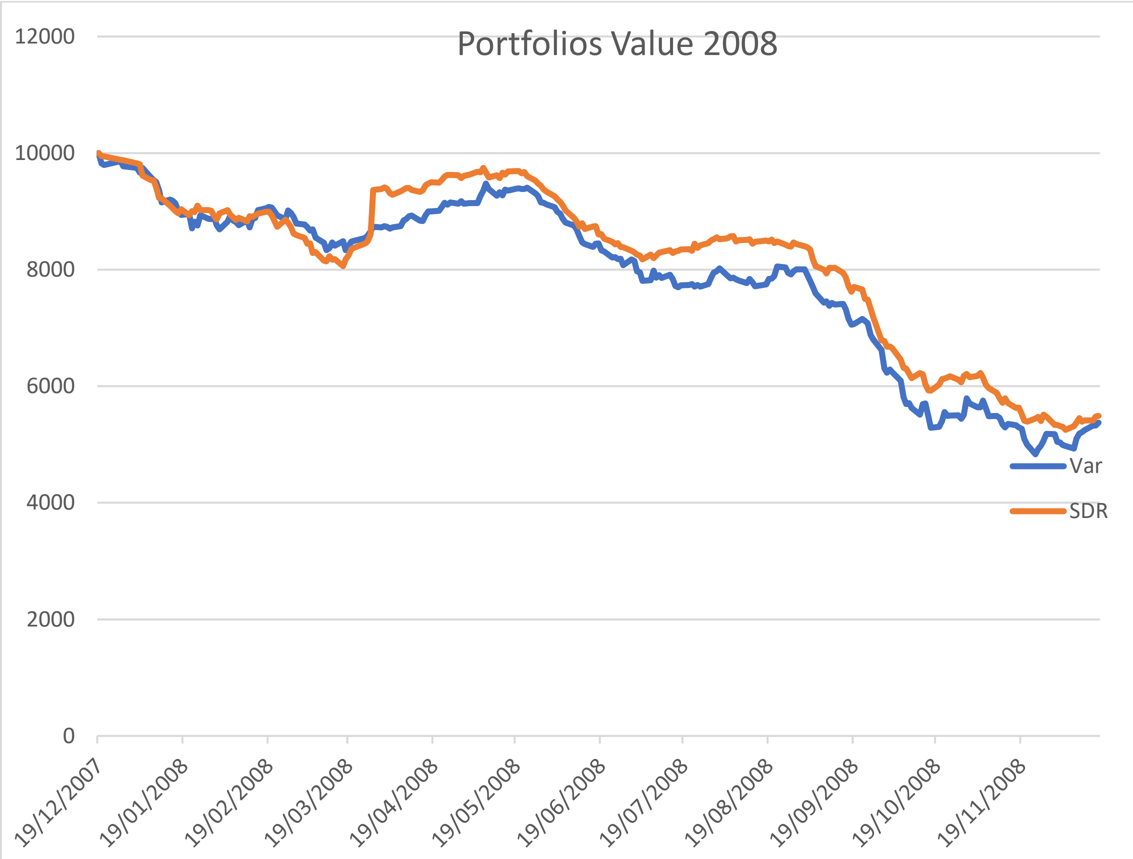


Figure 8 .Portfolios Value from an initial investment of 10000 euro; 2008

In the first out of the sample period (2008), the portfolio computed using the coherent risk measure outperformed the one calculated with the mean-variance criteria even if for a small amount. This behaviour is coherent with what we saw in the previous chapters: Shortfall deviation risk is a more conservative risk measure than the Variance, hence the portfolios constructed using this coherent risk measure as a proxy of risk are safer.

	Entropic VaR	Variance
Daily returns	-0,000203142	3,16992E-05
Total return	-0,051598162	0,00801991
Std. dev.	0,019615035	0,018152916
Sharpe ratio	-0,012353631	0,001848036

Table 17 Comparison between the Entropic VaR and Variance Portfolio, 2014-2017

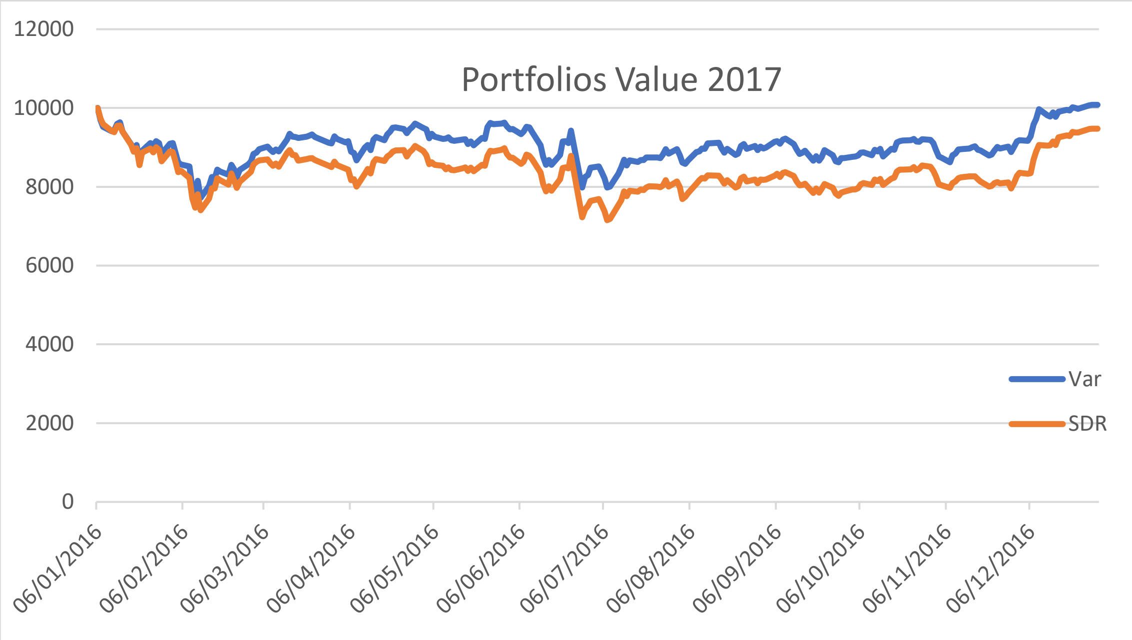


Figure 9 Portfolios Value from an initial investment of 10000 euro; 2017

Even in this period (2014-2017), theory is confirmed by practise: In a relatively calm and growing period a portfolio constructed with a riskier risk measure tends to lead to a higher return than a portfolio constructed with a conservative one.

	Expected Shortfall	Variance
Daily returns	-0,000336171	-0,000897271
Total return	-0,085051167	-0,227009653
Std. dev.	0,02245038	0,022234336
Sharpe ratio	-0,016725774	-0,040272098

Table 18 Comparison between the Expected shortfall and Variance Portfolio, 2018-2021

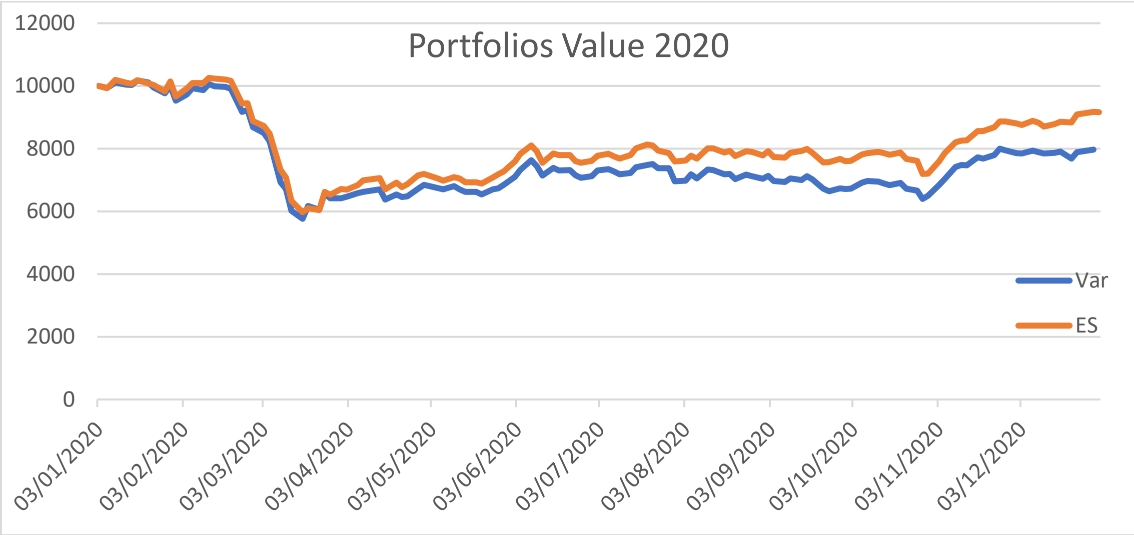


Figure 10 Portfolios Value from an initial investment of 10000 euro; 2020

In the first part of this period (2020) in which the portfolios results were very close, but, after the covid 19 outbreak the Expected shortfall portfolio outperformed the Variance one.

CONCLUSION

The aim of this thesis was to establish whether, among the three chosen risk measures (entropic Var, expected shortfall and shortfall deviation risk), one could be established superior to the others for what concern the construction of financial portfolios following the Modern Portfolio Theory; In order to have results as general as possible, data of three periods were taken into consideration: two periods that includes a major financial crisis (2008's financial crisis, covid-19 outbreak) and a period of relative calm in the financial markets. The periods were in turn divided into a period for the computations, where Particle Swarm Optimization was used to calculate the optimal composition of the portfolios (sample period) and in an evaluation period (out of the sample period), where the optimal portfolios were projected into a virtual future. In order to establish if one of these risk measure can be considered superior to the others in the construction of optimal portfolios, the results obtained by the optimal portfolios in the virtual future were compared both in terms of return and in terms of risk using the indices proposed in the 4th chapter.

As for the use of the PSO in the computation of the minimization problems, it can be said that the use of the algorithm has been effective, indeed, the only constraint that is violated in almost all the particles (portfolios) is the one on minimum quotas but, it is violated to such a small extent that it is never significant, moreover, even the optimal portfolio with a high fitness value do not show significant violations of the constraints.

For what concern the returns int the out-of the sample period, it is possible to state that the portfolio generated by the coherent risk measures has always outperformed the benchmark although they never had positive results. On the contrary, it is not possible to establish if there is a risk measure better than the other nor in terms of risk nor in terms of returns. This happened for two reasons:

1. The risk measure which generates the best portfolio among the optimal portfolios is different for each period.

2. Since the risk measures that generates the optimal portfolio similar one with the others, the portfolios results tend to be similar as well.

What, instead, is confirmed it’s the effectiveness of the coherent risk measures with regard to unexpected shock conditions: both in the period 2008 and in the period 2020 the portfolios generated by the coherent risk measures outperformed the benchmark and the portfolio generated with the Mean-variance model Proving the theory that this kind of risk measure is more conservative.

A last important observation on the results concerns the original Markowitz Model: even if the use of Variance as a proxy of risk is wrong (as demonstrated in the first chapter), the results obtained by the Markowitz model are not so bad. In 2 over the 3 periods, it outperformed the Benchmark and in one of this 2 cases it outperformed all the other risk measures as well. This indicates that the Markowitz model is not that outdated especially because it grants a certain level of diversification which generally, permits, to safeguard the portfolios result in turbulence periods.

Appendix A

A.1 Demonstration that a measure of dispersion may not respect all the coherence axioms

Assume X and Y are returns from two different Portfolios. Let X = c * B with B being a Bernoulli distributed random variable and c > 0 being a constant. In other words, the return of portfolio X is +c or -c. The portfolio might for example consist of some kind of digital option[79].

The Y portfolio shall consist of 2 digital options of the same kind and some fixed income position, so that the return is given by Y = 2 * X + 2 * c. This means, if X is +c than Y is 4 * c and if X is -c than Y is 0.

Clearly X < Y for all possible outcomes. We also know that var(Y) = 4 * var(X) > var(X)

In other words,Y dominates X, but the risk of Y measured by the variance would be greater than the risk of X.

this is a contradiction to the monotonicity property.

[79] A digital option is a form of option that allows traders to manually set a strike price. The digital option provides the traders with a fixed payout if the market price of the underlying asset exceeds the strike price while if the underlying asset market price goes below the strike price the trader loses the initial investment.

A.2 Normality test

As we have seen, some of the performance indexes we decided to use in the portfolios' evaluation are based on the gaussian assumptions[80]. To be more precise the ones based on normality or elliptical assumptions are Sharpe Ratio, Sortino Ratio, Jensen's alpha and Treynor Ratio.

Hence, it is fundamental to know if the underlying distributions behave like a normal or, on the other hand, understand if and how it's possible to use these indexes.

In the attempt to reach our aim some simple tests need to be performed: Normality Test. In these simple cases normality tests are used to "compare" the returns' distributions with the Normal distribution in order to grant, with a certain likelihood, that the returns' distributions would be equal (or very similar) to a normal distribution.

The tests shall be performed on the returns in the evaluation periods since these periods are the ones where the performances are computed so, the ones where the performance indexes have been utilized.

Moreover, it is not sufficient to perform the tests on the optimal portfolios generated by a single risk measure hence, they need to be done on a series of portfolios created by all the three risk measures (Expected Shortfall, Shorfall deviation risk and Entropic VaR).

Test

The first and "easiest" test that can be done in order to assess if a probability distribution behave like a normal is the comparison between the histogram of the interested probability distribution and the gaussian. It is a graphical test where the probability density function of a normal distribution is superimposed on the *p.d.f.* of the interested distribution. Practically, in order to create a lower and an upper bound for the

[80] To be more precise they are based on elliptical distribution assumptions.

bins of the histogram the mean and the variance of the return's distribution are computed, and the bounds are calculated as:

$$Lowerbound = \mu - X * \sigma$$

$$Upperbound = \mu + X * \sigma$$

Where X depends on the authors' interest on the tails of the distribution (in this paper it is considered equal to four).

Then a gaussian distribution with the computed mean and standard deviation is generated and finally, the range between the two bounds is divided in bins (in this paper it is considered equal to eleven) in order to compute the frequencies of the gaussian and of the interested probability distributions.

The histogram's normality tests have been performed for the results of the optimal portfolio selected by the Particle Swarm optimization algorithm in the three different cases of Expected shortfall, shortfall deviation risk and Entropic Value at risk.

For each risk measure two histograms have been created.

The results are showed in the following images:

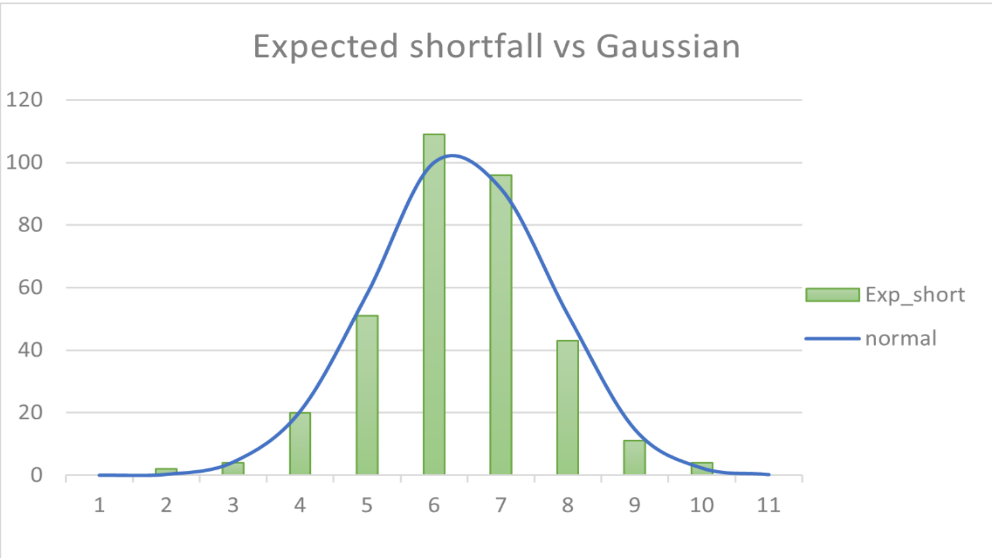


Figure 11 Comparison between the return's histogram of the portfolio constructed with the

expected shortfall and a gaussian obtained starting from the same data, 1st test

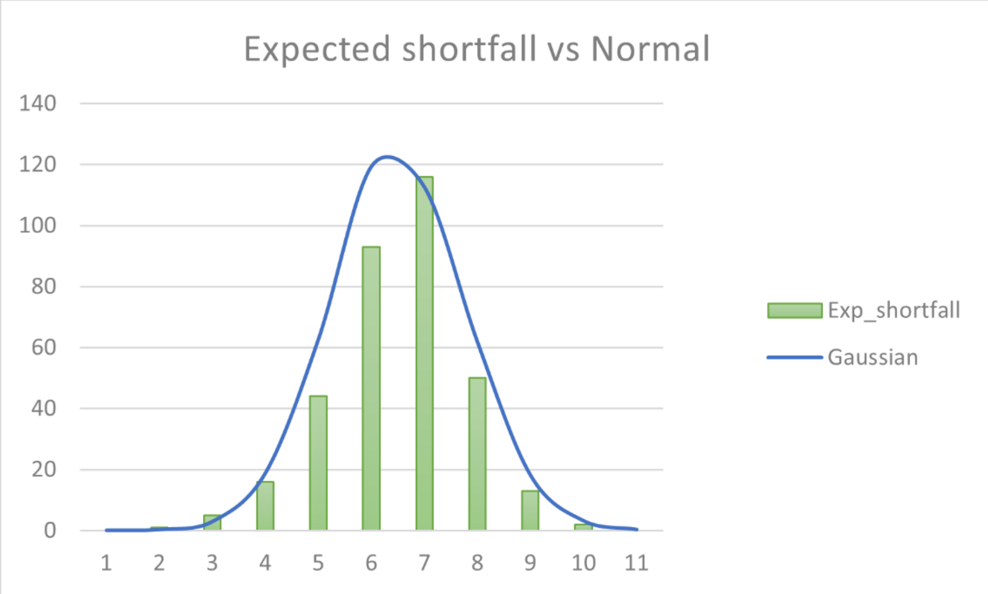


Figure 12 Comparison between the return's histogram of the portfolio constructed with the expected shortfall and a gaussian obtained starting from the same data, 2nd test

At a first sight we could state that the returns distribution of the portfolio selected using the expected shortfall has risk measure would seems pretty close to the normal distribution.

But, in both test data are not normally distributed in the centre of the distribution showing the presence of skewness, in the first case concentrated in the domain of losses (left tail) while in the second concentrated in the domain of gains (right tail)

Moreover, in the first test the data on the tails are not gaussian as well.

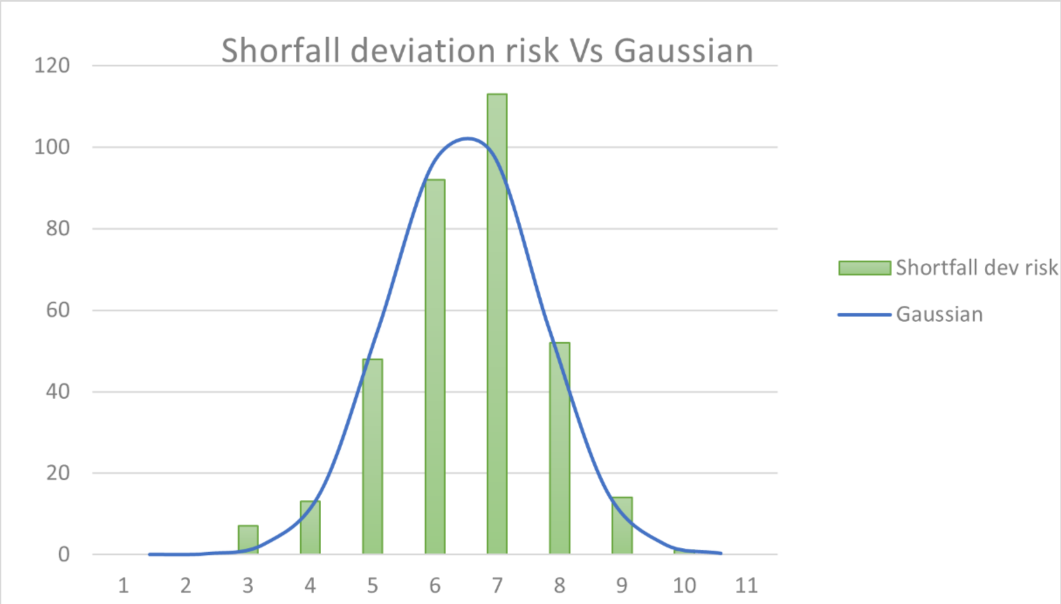


Figure 13 Comparison between the return's histogram of the portfolio constructed with the Shortfall deviation risk and a gaussian obtained starting from the same data, 1st test

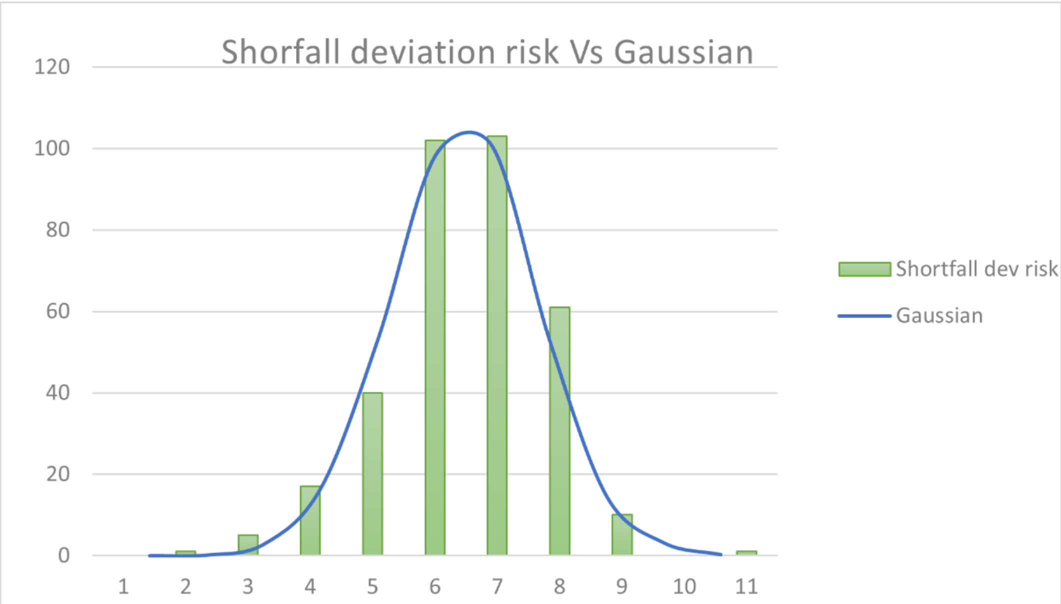


Figure 14 Comparison between the return's histogram of the portfolio constructed with the Shortfall deviation risk and a gaussian obtained starting from the same data, 2nd test

In the shortfall deviation risk tests the presence of kurtosis is evident showing a higher frequency in the tails with respect to the gaussian case.

For what concern the skewness the first test exhibits a clear "right-side" skewness while for the second we can not state the same thing.

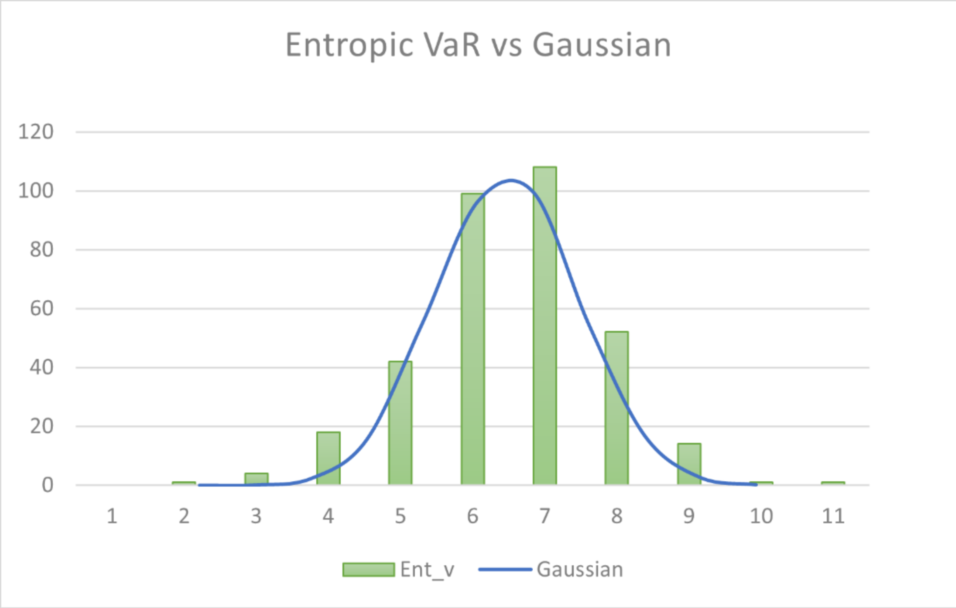


Figure 14 Comparison between the return's histogram of the portfolio constructed with the Entropic VaR and a gaussian obtained starting from the same data, 1st test

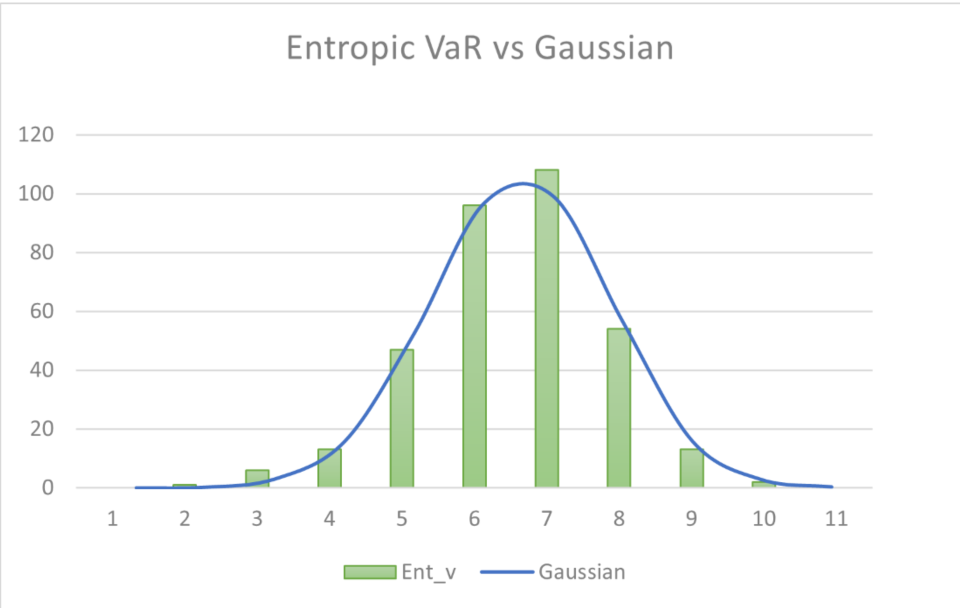


Figure 15 Figure 14 Comparison between the return's histogram of the portfolio constructed with the Entropic VaR and a gaussian obtained starting from the same data, 2nd test

Tests on the entropic Var optimal portfolios manifests the same results: Skewness in the right tails and presence of Kurtosis.

These graphical tests are useful to get an idea of whether the distribution of returns is Gaussian or not but they are not sufficient to state that for sure.

That's why other more rigorous tests have been implemented.

A.2.1 Skewness, kurtosis and Jarque-Bera test

If a distribution behaves in a way which is similar to the Gaussian, a possible easy system to get if it is properly a normal is to check its empirical Skewness and Kurtosis.

Or, in other words, control the features of the distribution in the centre and in the tails.

From what we said in the previous part of this thesis we know that a normal distribution has a skewness of zero and a kurtosis of three, every departure from these "basis" point are a declaration of a possible non-normality.[81]

Clearly, the departure itself is not a sufficient proof to state that the distribution doesn't behave like a gaussian, and more specific test shall be implemented.

We decided to implement the Jarque-Bera Test for normality for reason of simplicity and ease of understanding.

Tests have been implemented 5 times since the optimal portfolios were different one with respect to the other even for the same risk measure.

In order to be meaningful all the test have been implemented "*ceteris paribus*" : the variables used to create the optimal portfolios were kept stable: epsilon=0.001, desired return= 0.0002 (daily), n. particles= 50, n. iterations = 100, alpha =0.05, ω=0.7298, C1=C2=1.46915.

The results are showed in the following tables starting from the Expected shortfall' s tests, then the Shortfall deviation risk and, finally, the entropic VaR.

[81] Also the entity of the departure shall be considered.

Table 19 Expected Shortfall tests results

	Value 1	Value 2	Value 3	Value 4	Value 5
Skew	-0,15173	-0,35689	-0,36419	-0,32737	-0,18815
kurt	0,93474	0,645284	0,947141	0,932663	0,904518
JB	13,68249	13,11644	20,22472	18,39589	13,59657
Pval	0,001069	0,001418	4,06E-05	0,000101	0,001116

Table 20 Shortfall deviation risk tests results

	Value 1	Value 2	Value 3	Value 4	Value 5
Skew	-0,23131	-0,33212	-0,36153	-0,05962	-0,27398
kurt	0,158266	0,948121	1,369268	1,114765	0,364352
JB	3,386805	18,9854	33,96774	17,80635	6,134285
Pval	0,183893	7,54E-05	4,21E-08	0,000136	0,046554

Table 21 Entropic VaR tests results

	Value 1	Value 2	Value 3	Value 4	Value 5
Skew	-0,18161	-0,3193	-0,29384	-0,26997	-0,50354
kurt	0,730685	0,609589	-0,2094	0,332548	0,893481
JB	9,432657	11,04148	5,513708	5,696794	25,67751
Pval	0,008948	0,004003	0,063491	0,057937	2,66E-06

The results are pretty simple to interpret: All the three risk measures in all the 5 tests show skewness and Kurtosis different from zero.

For what concern the skewness, it is always small and negative indicating a relatively low data’s distortion in the left tail.

What is more impactful, in these tests, is the kurtosis: it strongly affects the distributions indicating a non-gaussian behaviour in the tails. Except for the 3rd test on the entropic VaR optimal portfolio, it is always large and positive representing larger probabilities than the gaussian, in the tails of the distributions (fat tails).

Finally, a more statistical test has been implemented, Jarque-Bera test is a statistical tool to assess if the underling p.d.f. behave like a normal; it is based on Skewness and Kurtosis.

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

Clearly, the larger its value the less probable is that the underling probabilities behave like a normal.

As we can see for all the test, Jarque-Bera's values are high and their p-value as well, hence we can conclude that, at this alpha level, we shall reject the normality assumption for the returns' distributions of the optimal portfolio in the evaluation period.

APPENDIX B

In this part I provide the Python codes that I used in order to make the necessary calculations for this thesis.

Code applied to the Expected Shortfall-based portfolio selection model:

```
#import library
import pandas as pd
import numpy as np

#import return's data
ritorni=pd.read_pickle("ritorni_it 2018")

#it divides into in the sample and out of the sample period
ndati=len(ritorni)
n_sample=int(ndati/3)*2
ret_first=ritorni.iloc[0:n_sample,:]
ret_second=ritorni.iloc[n_sample:len(ritorni),:]
ndati_1=len(ret_first)
ndati_2=len(ret_second)
nstocks=len(ret_first.columns)

##it sets alpha
alpha=0.05
n_alpha=int(ndati_1*alpha)

##Particle Swarm Optimization
#number of interactions, particles, epsilon and Pi (for the whole
period)
iterazioni=250
particelle=50
epsilon=0.0001
rit_giorn=0.0002 #desired daily return
rit_voluto=np.exp(rit_giorn*ndati_1)-1 #desired return for the whole
period

#it sets the weight for velocity and constrains
w=0.7298
c1=1.49618
c2=1.49618

#loop's supports
loc_best=np.zeros((particelle,nstocks))
global_best=np.ones((1,nstocks))
Best_loc_ftn=np.ones(particelle)*(1000)
best_glob_ftn=np.ones(1)*(1000)
velocità=np.zeros([iterazioni,particelle,nstocks])
Pos_in=np.zeros([iterazioni+1,particelle,nstocks])
fitness=np.zeros([particelle,iterazioni])
ret_first=np.array(ret_first)
ritorni_st=np.zeros([nstocks,n_sample,particelle,iterazioni]).T
r_port=np.zeros([iterazioni,particelle,n_sample])
ritorno_port=np.zeros([iterazioni,particelle])
ordine=np.zeros([iterazioni,particelle,n_sample])
Es_P=np.zeros([iterazioni,particelle])
```

```
vinc_1=np.zeros([iterazioni,particelle])
vinc_2=np.zeros([iterazioni,particelle])
vinc_2i=np.zeros([iterazioni,particelle])
vinc_3i=np.zeros([iterazioni,particelle,nstocks])
vinc_3c=np.zeros([iterazioni,particelle,nstocks])
vinc_3=np.zeros([iterazioni,particelle])
vinc_4i=np.zeros([iterazioni,particelle,nstocks])
vinc_4c=np.zeros([iterazioni,particelle,nstocks])
vinc_4=np.zeros([iterazioni,particelle])
best_fit=np.zeros(iterazioni)
index_gb=np.zeros(iterazioni)
arrotonda=np.zeros(nstocks)
per_min=np.ones([particelle,nstocks])*0.02
per_max=np.ones([particelle,nstocks])*0.2

##initialization for velocity and weight
for i in range (particelle):
  Pos_in[0,i,:]=np.random.dirichlet(np.ones(nstocks),size=1)#it
generates initial positions for all stocks, all interactions, for the
first particle
  velocità[0,i,:]=np.random.dirichlet(np.ones(nstocks),size=1)#it
generates initial velocities for all stocks, all interactions, for the
first particle

#it computes the daily returns for all stocks, for all particles and
for all interactions
for it in range(iterazioni):
 for p in range (particelle):
  for m in range(nstocks):
   ritorni_st[it,p,:,m]=Pos_in[it,p,m]*ret_first[:,m]

   #it computes minimum participation constrain
   vinc_3i[it,:,:]=per_min-abs(Pos_in[it,:,:])
   vinc_3c[it,p,m]=max(0,(vinc_3i[it,p,m]))

   #it compute maximum participation constrain
   vinc_4i[it,:,:]=abs(Pos_in[it,:,:])-per_max
   vinc_4c[it,p,m]=max(0,(vinc_4i[it,p,m]))

   #it computes the returns of each particle (portfolio) for each
iteration and the return constrain
 for part in range(particelle):
    r_port[it,part,:]=np.sum(ritorni_st[it,part,:,:],axis=1)
    r_porto=np.sum(ritorni_st[it,part,:,:],axis=1)
    ritorno_port[it,part]=np.sum(r_port[it,part,:],axis=0)
    vinc_2i[it,part]=max(0,rit_voluto-ritorno_port[it,part])

##expected shortfall portfolio and constrains
ordine[it]=np.sort(r_port[it])
vinc_1[it,:]=abs((np.sum(Pos_in[it],axis=1))-
np.ones(particelle))#budget constrian
vinc_2[it]=vinc_2i[it] #return constrain
vinc_3[it]=np.sum(vinc_3c[it],axis=1)#minimum partecipation constrain
vinc_4[it]=np.sum(vinc_4c[it],axis=1)#maximum partecipation constrain

Es_P[it]=np.mean(ordine[it,:,0:n_alpha],axis=1)#expected_shortfall of
all particles
```

```
##fitness function
fitness[:,it]=-
Es_P[it]+((1/epsilon)*vinc_1[it])+((1/epsilon)*vinc_2[it])+((1/epsilon
)*vinc_3[it])+((1/epsilon)*vinc_4[it])

#local fitness
for l in range(particelle):
    if fitness[l,it]<Best_loc_ftn[l]:
      loc_best[l]=Pos_in[it,l,:]
    Best_loc_ftn[l]=fitness[l,it]

#global fitness
best_fit[it]=min(fitness[:,it])
index_gb[it]=np.argmin(fitness[:,it])
if best_fit[it]<best_glob_ftn:
    best_glob_ftn=best_fit[it]
global_best=Pos_in[it,int(index_gb[it])]#best portfolio
nit_m=np.argmin(best_fit)
npart=int(index_gb[nit_m])

#it computes new velocity
for n in range(particelle):
      velocità[it,n,:]=w*velocità[(it-
1),n,:]+c1*np.random.uniform()*(loc_best[n,:]-
Pos_in[it,n,:])+c2*np.random.uniform()*(global_best-Pos_in[it,n,:])

#it computes new initial position
Pos_in[it+1]=Pos_in[it]+velocità[it]

##it finds ES, constrains 1,2,3,4 of the best particle of all the
iterations
v1=vinc_1[nit_m,npart]
v2=vinc_2[nit_m,npart]
v3=vinc_3[nit_m,npart]
v4=vinc_4[nit_m,npart]
Es_m=Es_P[(nit_m,npart)]

##it generates a data frame with the participation of the best
particle
for nstock in range (nstocks):
 arrotonda[nstock]=round(global_best[nstock],3)
Pos_fin=pd.DataFrame(arrotonda.T,index=ritorni.columns,columns=["Perce
ntuali"])
Pos_fin.to_pickle("stock scelti_es_2018_5")
```

Code applied to the Entropic VaR-based portfolio selection model:

```
#import library
import pandas as pd
import numpy as np

#it imports return's data
ritorni=pd.read_pickle("ritorni_it 2014")

#it divids into in the sample and out of the sample period
ndati=len(ritorni)
n_sample=int(ndati/3)*2
ret_first=ritorni.iloc[0:n_sample,:]
ret_second=ritorni.iloc[n_sample:len(ritorni),:]
ndati_1=len(ret_first)
ndati_2=len(ret_second)
nstocks=len(ret_first.columns)

#it sets alpha
alpha=0.05
alpha_ev=0.95
n_alpha=int(ndati_1*alpha)

##Particle Swarm optimization
#number of interactions, particles, epsilon and Pi (for the whole
period)
iterazioni=250
particelle=50
epsilon=0.0001
rit_giorn=0.0002 #desired daily return
rit_voluto=np.exp(rit_giorn*ndati_1)-1 #desired return for the whole
period

#it sets the weight for velocity and constrains
w=0.7298
c1=1.49618
c2=1.49618

#loop's supports
loc_best=np.zeros((particelle,nstocks))
global_best=np.ones((1,nstocks))
Best_loc_ftn=np.ones(particelle)*(1000)
best_glob_ftn=np.ones(1)*(1000)
velocità=np.zeros([iterazioni,particelle,nstocks])
Pos_in=np.zeros([iterazioni+1,particelle,nstocks])
fitness=np.zeros([particelle,iterazioni])
ret_first=np.array(ret_first)
ritorni_st=np.zeros([nstocks,n_sample,particelle,iterazioni]).T
r_port=np.zeros([iterazioni,particelle,n_sample])
ritorno_port=np.zeros([iterazioni,particelle])
ordine=np.zeros([iterazioni,particelle,n_sample])
Ent_P=np.zeros([iterazioni,particelle])
vinc_1=np.zeros([iterazioni,particelle])
vinc_2=np.zeros([iterazioni,particelle])
vinc_2i=np.zeros([iterazioni,particelle])
vinc_3i=np.zeros([iterazioni,particelle,nstocks])
vinc_3c=np.zeros([iterazioni,particelle,nstocks])
```

```
vinc_3=np.zeros([iterazioni,particelle])
vinc_4i=np.zeros([iterazioni,particelle,nstocks])
vinc_4c=np.zeros([iterazioni,particelle,nstocks])
vinc_4=np.zeros([iterazioni,particelle])
best_fit=np.zeros(iterazioni)
index_gb=np.zeros(iterazioni)
arrotonda=np.zeros(nstocks)
per_min=np.ones([particelle,nstocks])*0.02
per_max=np.ones([particelle,nstocks])*0.2
t_evar=1
rit_med=np.zeros([iterazioni,particelle,nstocks])
Mx=np.zeros([iterazioni,particelle])
Ent_v=np.zeros([iterazioni,particelle])


##initialization for velocity and weight
for i in range (particelle):
  Pos_in[0,i,:]=np.random.dirichlet(np.ones(nstocks),size=1) #it
generates initial positions for all stocks, all iteractions, for the
first particle
  velocità[0,i,:]=np.random.dirichlet(np.ones(nstocks),size=1)#it
generates initial velocities for all stocks, all iteractions, for the
first particle

#it computes the daily returns for all stocks, for all particle and
for all interactions
for it in range(iterazioni):
 for p in range (particelle):
  for m in range(nstocks):
   ritorni_st[it,p,:,m]=Pos_in[it,p,m]*ret_first[:,m]
   rit_med[it,p,m]=np.exp(np.mean(ritorni_st[it,p,m,:])*(1/t_evar))
   Mx[it,p]=np.log(np.sum(rit_med))
   Ent_v[it,p]=t_evar*((Mx[it,p]/nstocks)-np.log(alpha_ev))

   #it compute minimum participation constrain
   vinc_3i[it,:,:]=per_min-abs(Pos_in[it,:,:])
   vinc_3c[it,p,m]=max(0,(vinc_3i[it,p,m]))

   #it compute maximum participation constrain
   vinc_4i[it,:,:]=abs(Pos_in[it,:,:])-per_max
   vinc_4c[it,p,m]=max(0,(vinc_4i[it,p,m]))

    #it computes the returns of each particle (portfolio) for each
interaction and the return constrain
 for part in range(particelle):
    r_port[it,part,:]=np.sum(ritorni_st[it,part,:,:],axis=1)
    r_porto=np.sum(ritorni_st[it,part,:,:],axis=1)
    ritorno_port[it,part]=np.sum(r_port[it,part,:],axis=0)
    vinc_2i[it,part]=max(0,rit_voluto-ritorno_port[it,part])

##expected shortfall portfolio and constrains
 ordine[it]=np.sort(r_port[it])
 vinc_1[it,:]=abs((np.sum(Pos_in[it],axis=1))-
np.ones(particelle))#budget constrain
 vinc_2[it]=vinc_2i[it] #return constrain
 vinc_3[it]=np.sum(vinc_3c[it],axis=1)#minimum participation constrain
 vinc_4[it]=np.sum(vinc_4c[it],axis=1)#maximum participation constrain
```

```
 ##fitness function

fitness[:,it]=Ent_v[it]+((1/epsilon)*vinc_1[it])+((1/epsilon)*vinc_2[i
t])+((1/epsilon)*vinc_3[it])+((1/epsilon)*vinc_4[it])

#local fitness
 for l in range(particelle):
    if fitness[l,it]<Best_loc_ftn[l]:
      loc_best[l]=Pos_in[it,l,:]
    Best_loc_ftn[l]=fitness[l,it]

#global fitness
 best_fit[it]=min(fitness[:,it])
 index_gb[it]=np.argmin(fitness[:,it])
 if best_fit[it]<best_glob_ftn:
    best_glob_ftn=best_fit[it]
 global_best=Pos_in[it,int(index_gb[it])]#best portfolio
 nit_m=np.argmin(best_fit)
 npart=int(index_gb[nit_m])

#it computes new velocity for each iteration
 for n in range(particelle):
      velocità[it,n,:]=w*velocità[(it-
1),n,:]+c1*np.random.uniform()*(loc_best[n,:]-
Pos_in[it,n,:])+c2*np.random.uniform()*(global_best-Pos_in[it,n,:])

    #it computes new initial position
Pos_in[it+1]=Pos_in[it]+velocità[it]

##it finds ES, constrains 1,2,3,4 of the best particle of all the
iteration
v1=vinc_1[nit_m,npart]
v2=vinc_2[nit_m,npart]
v3=vinc_3[nit_m,npart]
v4=vinc_4[nit_m,npart]
Ent_m=Ent_v[(nit_m,npart)]


##it generates a data frame with the participations of the best
particle
for nstock in range (nstocks):
 arrotonda[nstock]=round(global_best[nstock],3)
Pos_fin=pd.DataFrame(arrotonda.T,index=ritorni.columns,columns=["Perce
ntuali"])
Pos_fin.to_pickle("stock scelti_es 2018_1")
```

Code applied to the SDR-based portfolio selection model:

```
#import library
import pandas as pd
import numpy as np

#it imports return's data
ritorni=pd.read_pickle("ritorni_it 2014")

#it divides into in the sample and out of the sample period
ndati=len(ritorni)
n_sample=int(ndati/3)*2
ret_first=ritorni.iloc[0:n_sample,:]
ret_second=ritorni.iloc[n_sample:len(ritorni),:]
ndati_1=len(ret_first)
ndati_2=len(ret_second)
nstocks=len(ret_first.columns)

##it sets alpha
alpha=0.05
n_alpha=int(ndati_1*alpha)

##Particle Swarm optimization
#number of interactions, particles, epsilon and Pi (for the whole
period)
iterazioni=250
particelle=50
epsilon=0.0001
rit_giorn=0.0002 #desired daily return
rit_voluto=np.exp(rit_giorn*ndati_1)-1 #desired return for the whole
period

#it sets the weight for velocity and constrains
w=0.7298
c1=1.49618
c2=1.49618
beta_sdr=0.2 #beta for shortfall deviation risk
p_sdr=2 #p-norm shortfall deviation

#loop's supports
loc_best=np.zeros((particelle,nstocks))
global_best=np.ones((1,nstocks))
Best_loc_ftn=np.ones(particelle)*(1000)
best_glob_ftn=np.ones(1)*(1000)
velocità=np.zeros([iterazioni,particelle,nstocks])
Pos_in=np.zeros([iterazioni+1,particelle,nstocks])
fitness=np.zeros([particelle,iterazioni])
ret_first=np.array(ret_first)
ritorni_st=np.zeros([nstocks,n_sample,particelle,iterazioni]).T
r_port=np.zeros([iterazioni,particelle,n_sample])
ritorno_port=np.zeros([iterazioni,particelle])
ordine=np.zeros([iterazioni,particelle,n_sample])
Es_P=np.zeros([iterazioni,particelle])
Sd_s=np.zeros([iterazioni,particelle,n_alpha])
Sd=np.zeros([iterazioni,particelle,n_alpha])
Sd_f=np.zeros([iterazioni,particelle,n_alpha])
Sd_fi=np.zeros([iterazioni,particelle])
conta=np.zeros([iterazioni,particelle])
vinc_1=np.zeros([iterazioni,particelle])
```

```
vinc_2=np.zeros([iterazioni,particelle])
vinc_2i=np.zeros([iterazioni,particelle])
vinc_3i=np.zeros([iterazioni,particelle,nstocks])
vinc_3c=np.zeros([iterazioni,particelle,nstocks])
vinc_3=np.zeros([iterazioni,particelle])
vinc_4i=np.zeros([iterazioni,particelle,nstocks])
vinc_4c=np.zeros([iterazioni,particelle,nstocks])
vinc_4=np.zeros([iterazioni,particelle])
best_fit=np.zeros(iterazioni)
index_gb=np.zeros(iterazioni)
arrotonda=np.zeros(nstocks)
per_min=np.ones([particelle,nstocks])*0.02
per_max=np.ones([particelle,nstocks])*0.2

##initialization for velocity and weight
for i in range (particelle):
  Pos_in[0,i,:]=np.random.dirichlet(np.ones(nstocks),size=1) #it
generates initial positions for all stocks, all iteractions, for the
first particle
  velocità[0,i,:]=np.random.dirichlet(np.ones(nstocks),size=1)#it
generates initial velocities for all stocks, all iteractions, for the
first particle

#it computes the daily returns for all stocks, for all particle and
for all interactions
for it in range(iterazioni):
 for p in range (particelle):
  for m in range(nstocks):
   ritorni_st[it,p,:,m]=Pos_in[it,p,m]*ret_first[:,m]

   ##it computes minimum participation constrain
   vinc_3i[it,:,:]=per_min-abs(Pos_in[it,:,:])
   vinc_3c[it,p,m]=max(0,(vinc_3i[it,p,m]))

   ##it compute maximum participation constrain
   vinc_4i[it,:,:]=abs(Pos_in[it,:,:])-per_max
   vinc_4c[it,p,m]=max(0,(vinc_4i[it,p,m]))

    #compute the returns of each particle (portfolio) for each
interaction and the return constrain
 for part in range(particelle):
    r_port[it,part,:]=np.sum(ritorni_st[it,part,:,:],axis=1)
    r_porto=np.sum(ritorni_st[it,part,:,:],axis=1)
    ritorno_port[it,part]=np.sum(r_port[it,part,:],axis=0)
    vinc_2i[it,part]=max(0,rit_voluto-ritorno_port[it,part])

##expected shortfall portfolio and constrains
ordine[it]=np.sort(r_port[it])
vinc_1[it,:]=abs((np.sum(Pos_in[it],axis=1))-
np.ones(particelle))#budget constrain
vinc_2[it]=vinc_2i[it] #return constrain
vinc_3[it]=np.sum(vinc_3c[it],axis=1)#minimum participation constrain
vinc_4[it]=np.sum(vinc_4c[it],axis=1)#maximum participation constrain
Es_P[it]=np.mean(ordine[it,:,0:n_alpha],axis=1)#expected_shortfall of
all particles
Sd_s[it,:]=ordine[it,:,0:n_alpha]#shortfall deviation of all particles
for mi in range(particelle):
 for j in range(len(Sd_s)):
```

```
  Sd[it,mi] = ((Sd_s[it,mi,:] <= Es_P[it,mi])*Sd_s[it,mi,:])-
Es_P[it,mi]
  conta[it,mi]=np.sum((Sd_s[it,mi,:] <= Es_P[it,mi]))
  Sd_f[it,mi]=(Sd[it,mi]*(Sd_s[it,mi,:] <= Es_P[it,mi]))**p_sdr
  Sd_fi[it,mi]=(np.sum(Sd_f[it,mi])/conta[it,mi])**(1/p_sdr)

#fitness function
 fitness[:,it]=-
Es_P[it]+Sd_fi[it]+((1/epsilon)*vinc_1[it])+((1/epsilon)*vinc_2[it])+(
(1/epsilon)*vinc_3[it])+((1/epsilon)*vinc_4[it])

#local fitness
 for l in range(particelle):
    if fitness[l,it]<Best_loc_ftn[l]:
      loc_best[l]=Pos_in[it,l,:]
    Best_loc_ftn[l]=fitness[l,it]

#global fitness
 best_fit[it]=min(fitness[:,it])
 index_gb[it]=np.argmin(fitness[:,it])
 if best_fit[it]<best_glob_ftn:
    best_glob_ftn=best_fit[it]
 global_best=Pos_in[it,int(index_gb[it])]#best portfolio

 nit_m=np.argmin(best_fit)
 npart=int(index_gb[nit_m])

#it computes new velocity position
 for n in range(particelle):
      velocità[it,n,:]=w*velocità[(it-
1),n,:]+c1*np.random.uniform()*(loc_best[n,:]-
Pos_in[it,n,:])+c2*np.random.uniform()*(global_best-Pos_in[it,n,:])
 #it computes new initial position
Pos_in[it+1]=Pos_in[it]+velocità[it]

##it finds ES, constrains 1,2,3,4 of the best particle of all the
iteration
v1=vinc_1[nit_m,npart]
v2=vinc_2[nit_m,npart]
v3=vinc_3[nit_m,npart]
v4=vinc_4[nit_m,npart]
Es_m=Es_P[(nit_m,npart)]
SDR=-Es_P[(nit_m,npart)]+Sd_fi[(nit_m,npart)]

##it generates a data frame with the participation of the best
particle
for nstock in range (nstocks):
 arrotonda[nstock]=round(global_best[nstock],3)
Pos_fin=pd.DataFrame(arrotonda.T,index=ritorni.columns,columns=["Perce
ntuali"])
Pos_fin.to_pickle("stock scelti_sdr 2018_5")
```

And finally, the code applied for the computation of the out-of the sample results and the indexes:

```
#it imports libraries
import pandas as pd
import numpy as np
import scipy as sc
from scipy.optimize import linprog
import statsmodels.api as sm
#it imports the returns
Mib=pd.read_pickle("ritorni mib 2014")
ritorni=pd.read_pickle("ritorni_it 2014")
portafoglio=pd.read_pickle("stock scelti_ent 2014_1")
portafoglio=np.array(portafoglio.T)
Mib=np.array(Mib)

#it divides into in the sample and out of the sample period
ndati=len(ritorni)
n_sample=int(2*ndati/3)
zero=np.ones(ndati)
Mib_ols = np.array([zero,Mib[:,0]]).T
ret_first=ritorni.iloc[0:n_sample,:]
ret_first_mib=Mib[0:n_sample]
ret_second=np.array(ritorni.iloc[n_sample:len(ritorni),:])
ret_second_mib_ols=Mib_ols[n_sample:len(ritorni)]
ndati_1=len(ret_first)
ndati_2=len(ret_second)
nstocks=len(ret_first.columns)
alpha=0.05
n_alpha=int(alpha*ndati_2)
risk_free_periodo=0.01 #risk free rate for the whole period
risk_free=(np.log(risk_free_periodo+1))/ndati_2 #daily risk free rate
#it computes the returns in the out of the sample period
ritorni_oos=np.zeros([ndati_2,nstocks]) #matrix for the out of the
sample returns
for st in range (nstocks):
    ritorni_oos[:,st]=portafoglio[:,st]*ret_second[:,st]
rit_gior=np.sum(ritorni_oos,axis=1) #daily portfolio returns
rit_medio_oos=np.sum(rit_gior)/ndati_2 #average daily portfolio return
ritorno_t_oos=np.sum(rit_gior) #total portfolio return

##ratios
#standard deviation portfolio returns
sd_port=np.sqrt(np.var(rit_gior))

#sharpe ratio
sharpe_ratio=(rit_medio_oos-risk_free)/sd_port

#sortino ratio
Value_at_risk_p=(np.sort(rit_gior))[n_alpha] #value at risk portfolio
returns
Dws=np.zeros(nstocks)#vector for downside risk
Ups=np.zeros(nstocks)#vector for upside risk
for st in range(nstocks):
  Dws[st]=-min(0,rit_gior[st]-risk_free) #downside risk
  Ups[st]=max(0,rit_gior[st]-risk_free) #upside potential
Tdd=np.sqrt(((np.sum(Dws))**2)/(1/nstocks)) #target downside deviation
Sortino_ratio=(rit_medio_oos-risk_free)/Tdd
```

```
##Farinelli tibletti ratio
p=2 #gain pro
q=3 #loss aversion
Upn_ft=Ups[Ups!=0] #delete data=0
Dwn_ft=Dws[Dws!=0]#delete data=0
Ups_ft=((np.sum(Upn_ft**p))/nstocks)**(1/p)
Dws_ft=((np.sum(Dwn_ft**q))/nstocks)**(1/q)
ft_ratio=Ups_ft/Dws_ft

##treynor ratio
provaols=sm.OLS(rit_gior,ret_second_mib_ols)
results = provaols.fit()
alpha_ols=results.params[0]
beta_ols=results.params[1]
Treynor_ratio=alpha_ols/beta_ols

#it generates a data frame with returns and indexes
Tab_ratios=np.array([rit_medio_oos,ritorno_t_oos,sd_port,sharpe_ratio,
Sortino_ratio,ft_ratio,Treynor_ratio])
Tab_ratios=pd.DataFrame(Tab_ratios,columns=["Value"],index=["Daily
returns","Total return","Standard deviation","Sharpe ratio","Sortino
ratio","Farinelli Tibletti ratio","Treynor ratio"])
Tab_ratios.to_excel("Tabella Ratios_ent_2014.xlsx")
```

Bibliography

Ahmadi-Javid, *Entropic Value-at-Risk: A New Coherent Risk Measure*, Springer Science+Business Media 2011

Benati S., *The optimal portfolio problem with coherent risk measure constraints*, European Journal of Operational Research (2003)

Brutti, Righi, Ceretta ., *Shortfall deviation risk: an alternative for risk measurement*, https://ssrn.com/abstract=2539775

CFI TEAM, Digital Option, https://corporatefinanceinstitute.com/resources/knowledge/trading-investing/digital-option/

Chen J. *Modern Portfolio Theory*. Investopedia https://www.investopedia.com/terms/m/modernportfoliotheory.asp.

Chen, Z., Wang, Y.,. *Two-sided coherent risk measures and their application in realistic portfolio optimization*, Journal of Banking and Finance

Chonghui Jiang, Yongkai Ma, Yunbi An, *International portfolio selection with exchange rate risk: A behavioural portfolio theory perspective*, Journal of Banking & Finance (2013)

Corazza M, *I vincoli a variabili miste-intere nella selezione di portafoglio: una rassegna*, Dipartimento di Economia Università Ca' Foscari Venezia

Daniel Kahneman and Amos Tversky, *Prospect Theory: An Analysis of Decision under Risk*, Econometrica(1979)

Farinelli S. and Tibiletti L., *Sharpe Thinking with Asymmetrical Preferences*,2002

Farinelli S., Ferreira M., Rossello D.,Thoeny M.,Tibiletti L. *Beyond Sharpe ratio: Optimal asset allocation using different performance ratios*, Journal of Banking & Finance (2008)

Gordon J. Alexander, Alexandre M. Baptista, Shu Yanc, *Portfolio selection with mental accounts and estimation risk*, Journal of Empirical Finance (2017)

Guang-Feng Deng, Woo-Tsong Lin, Chih-Chung Lo, *Markowitz-based portfolio selection with cardinality constraints using improved particle swarm optimization*, Expert Systems with Applications (2012)

HÜBNER G., *The Generalized Treynor Ratio*, HEC Management School – University of Liège and Maastricht University

Hung-Hsi Huang, Ching-Ping Wang, *Portfolio selection and portfolio frontier with background risk*, North American Journal of Economics and Finance

José A. Soares da Fonseca, *Performance ratios for selecting international portfolios: a comparative analysis using stock market indices in the euro area*, Faculty of Economics, University of Coimbra

Mario Brandtner, *Conditional Value-at-Risk, spectral risk measures and (non-)diversification in portfolio selection problems – A comparison with mean–variance analysis,* Journal of Banking & Finance (2013)

Markowitz H, *Portfolio selection*, The Journal of Finance 1952

Peirò, A. *Skewness in financial returns*, Journal of Banking and Finance

Poli R., Kennedy J., Blackwell T., *Particle swarm optimization: An overview*, Springer Science Business Media, LLC 2007

Rachev S., Ortobelli S., Stoyanov S., Fabozzi, F.J., Biglova A., *desiderable proprieties of an ideal risk measure in portfolio theory*, International Journal of Theoretical and Applied Finance (2008)

Rachev S., Ortobelli S., Stoyanov S., Fabozzi, F.J., Biglova A., *The proper use of risk measures in portfolio theory,* International Journal of Theoretical and Applied Finance (2005)

S. V. Stoyanov, S. T. Rachev and F. J. Fabozzi, *Optimal Financial Portfolios*, Applied Mathematical Finance 2007

Salehpoor B., S. Molla-Alizadeh-Zavardehi, c*onstrained portfolio selection model at considering risk-adjusted measure by using hybrid meta-heuristic algorithm*, Applied Soft Computing Journal (2019)

Shushang Zhu , Duan Li & Shouyang Wang, *Robust portfolio selection under downside risk*

Swisher P., G.W. Kasten, *Post-Modern Portfolio Theory*, Journal of Financial Planning (2005)

Szegö, G., . *Measures of risk*, Journal of Banking and Finance

Thomas N. Rollinger and Scott T. Hoffman, *Sortino: A 'Sharper' Ratio*, published for Red Rock capital

William F. Sharpe, *A Simplified Model for Portfolio Analysis*, University of Washington

Xianming Zhu and Hongbo Wang, *A new inertia weight control strategy for particle swarm optimization* , State Key Laboratory on Integrated Optoelectronics, College of Electronic Science and Engineering Changchun.

Zymler S.,Kuhn D. and Rustem B., *Worst-Case Value-at-Risk of Non-Linear Portfolios*, Department of Computing Imperial College of Science, Technology and Medicine, London (2012)