Università
Ca' Foscari
Venezia

Master's Degree Programme
Second Cycle (D.M. 270/2004)
in Economics and Finance - Finance

Final Thesis

Particle Swarm Optimization for entropy-based risk measures in portfolio selection problems:

A *mean – Entropic-VaR* application

Supervisor
Ch. Prof. Marco Corazza

Graduand
Riccardo Belli
871927

Academic Year
2018 / 2019

Introduction

Portfolio selection is a cornerstone of finance and economics. The problem consists in the minimization of a risk measure, while taking into account a series of constraints. Portfolio selection approach was introduced by Markowitz in 1952. His model was the first one to face the problem of how to efficiently invest a given amount of capital. Known also as *Modern Portfolio Theory*, Markowitz model revolutionized financial market investments. However, this model does present some limits and a set of assumptions rather utopic in the real world.

In this dissertation we will create a portfolio model trying to include some of the aspects that were avoided at that time, as for instance the presence of transaction costs and the allowance to buy or sell only determined quantities of assets. Furthermore, some model's assumptions will be modified, in particular the concept of risk measure. According to the most recent literature, in fact, only the so-called *coherent risk measures* can be employed as real financial risk measures. The most common risk measures, that have been used as alternatives to variance, are *Value-at-Risk (VaR)* – even though not being coherent – and *Expected Shortfall (ES) or Conditional Value-at-Risk (CVaR)*.
The risk measure chosen for the portfolio model developed in this work belongs to the class of entropy-based risk measures, is a coherent risk measure introduced by Ahmadi-Javid[1] and is called *Entropic Value at Risk (EVaR)*. All the above-mentioned measures of risk will be accurately described and discussed later in the dissertation.

Conversely to what proposed by Markowitz, the model developed in this work will not be based on the mean-variance criterion, but rather on *mean-entropic VaR*. Moreover, a system of characteristic Markowitz model's constraints will be

[1] *"Entropic Value-at-Risk: A New Coherent Risk Measure"*, A. Ahmadi-Javid (See Bibliography)

applied, as budget constraint and desired minimum return. We will successively introduce *mixed-integer* constraints, useful for managing transaction costs.

Since obtaining exact results from a constrained minimization problem is highly time-consuming, the solution proposed in the present dissertation will be metaheuristic-based (the concept of metaheuristic will be discussed in what follows): the metaheuristic employed, *Particle Swarm Optimization*, will not give an exact result to the problem, but a good level of approximation. This method consists in the employment of a *bio-inspired* metaheuristic algorithm able to search for an optimal solution to the problem, while not exact, exploiting the dynamics of exploration of groups of animals in the nature, like birds' flocks or shoals of fish.

The application of the model will be performed by *Matlab* and the results will be compared with the application in mean-Expected Shortfall of the same portfolio selection model.

Contents

Chapter 1

Portfolio selection problem

As well explained by (Constantinides G.M. and Malliaris A.G. 1995), in general a consumer, given a certain amount of income, typically faces two important economic decisions: the first one consists in deciding how to allocate his or her consumption among goods or services; the second one is the decision on how to invest among various assets. These two interrelated problems are known as the *consumption-saving decision* and the *portfolio selection problem*.

Portfolio selection is one of the most discussed and interesting problems in the economics and finance world. Modern portfolio theory finds his pioneer in Harry Markowitz, which developed the Mean-Variance portfolio selection model in 1952. Despite being recognized as one of the cornerstones in the portfolio selection problem, Markowitz's model proved to be too simplistic to represent the actual real world and its basic assumptions have been widely contested in recent years.

In this chapter we will give a synthetic but complete description of the Mean-Variance portfolio selection model by Harry Markowitz and a brief description of some of the models that try to overcome its limits. We will then shift our interest on the desirable characteristics for a risk measure, closing the chapter focusing on the importance of the adoption of a coherent risk measure in this environment.

1.1 Markowitz Model

Almost seventy years ago American economist Harry Markowitz developed a model which revolutionized investment the practise and became in the course of time one of the pillars of financial economics and Modern Portfolio Theory. His Mean-Variance model, rewarded with the Nobel Prize in Economics in 1990, aims at selecting a group of assets which have collectively lower risk than any single asset on its own.

1.2 Basic assumptions

As said previously, the mean-variance analysis has been challenged through the years due to the simplicity of the model with respect to the real world. The model limitations are given by some strong assumptions on which it relies:

- Investors always maximize the rate of return yielded by their investments;
- Investors are rational and risk-averse[2]: they are completely aware of all the risk underlying an investment and take positions basing their decisions on the risk, asking higher returns for accepting higher risk and coherently expecting lower return for lower levels of risk;
- Investors make their investment judgements by taking into consideration expected returns and standard deviation (risk measure) of returns of the possible assets;
- Investments have a single period horizon, meaning that at the beginning of the period *t* the investor allocates her/his wealth among different assets

[2] If faced by the decision between two identical portfolios, a *risk-averse* investor will choose the one with the lower risk.

and she/he will hold the portfolio until the period $t + \Delta t$, without considering the opportunity to reinvest the wealth in a following period;

- Under the condition of uncertainty, investors - and more in general, individuals - make decisions by maximizing the expected value of an utility function of consumption, which is assumed to be increasing and concave;

- Investors' assets are infinitely divisible. Thus, investors may decide to buy or sell a fraction of a share;

- Investors are price-takers, meaning their actions can't affect the probability distributions of returns on the available securities;

- Financial markets are frictionless. Hence, there are no transaction costs, no taxes, absence of institutional restrictions, and so on.

1.3 Markowitz portfolio selection model

The portfolio selection process may be seen as constituted by three stages:

1. The first stage consists in the identification of appropriate measures for measuring the expected return and risk;

2. The second stage establishes a criterion to identify the "best" portfolios, distinguishing between efficient portfolios and non-efficient ones;

3. In the third and last stage takes place the selection of a proper portfolio for the investor, according to her/his risk aversion. This activity is pursued by maximizing the investor's expected utility function.

1.3.1 Measures of Risk and Return

The future profitability of an asset is uncertain at the time of the purchase. This uncertainty is given by the randomness associated to return, since we do not know the future price ex-ante.

There are some statistical tools which help the investor to manage face the uncertainty of the investment:

- The mean of the single-period rate of return. It represents the profitability – expected return – of an investment;
- The variance of the single-period rate of return. It represents the risk of an investment and it is of course undesirable;
- The correlation between the return of each pair of risky assets. It represents the linear dependency between the pair of returns of the assets.

Expected value (mean) and variance of individual securities returns can be defined as follows:

Let X be a discrete random variable $X = \{(x_1, p_1), \dots, (x_i, p_i), \dots, (x_M, p_M)\}$, where x_i with $i = 1, \dots, M$ is the possible return from a given asset, and p_i, with $i = 1, \dots, M$, is the probability of occurrence of x_i, with $0 \leq p_i \leq 1$ for all i and $\sum_{i=1}^{M} p_i = 1$. Then:

$$E(X) = \sum_{i=1}^{M} x_i p_i$$

$$Var(X) = \sum_{i=1}^{M} \big(x_i - E(x)\big)^2 p_i$$

Since a portfolio is a set of two or more individual securities, we can now define its expected rate of return and its variance. Let R_p be the portfolio rate of return:

$$R_p = x_1 R_1 + \cdots + x_N R_N = \sum_{i=1}^{N} x_i R_i$$

where R_i is the random variable representing the return of the i-th asset and x_i the portion of capital in percentage invested on the same asset. Let r_i and σ_i^2 be respectively the expected rate of return and the variance of the i-th asset, with $i = 1, \dots, N$. Then the portfolio expected return and variance can be defined as follows:

$$E(R_P) = \sum_{i=1}^{N} x_i r_i \coloneqq r_P$$

$$\begin{aligned} Var(R_P) &= \sum_{i=1}^{N} x_i^2 \sigma_i^2 + 2 \sum_{i=1}^{N} \sum_{j=i+1}^{N} x_i x_j \sigma_{i,j} \\ &= \sum_{i=1}^{N} x_i^2 \sigma_i^2 + 2 \sum_{i=1}^{N} \sum_{j=i+1}^{N} x_i x_j \rho_{i,j} \sigma_i \sigma_j := \sigma_P^2 \end{aligned}$$

where $\sigma_{i,j} = \rho_{i,j}\sigma_i\sigma_j$ is the covariance and $\rho_{i,j} \in [-1, 1]$ is the linear correlation coefficient between R_i and R_j.[3]

Mean and variance of the portfolio might also be defined with the use of vectorial notation:

- $r_P = x'r$;
- $\sigma_P^2 = x'Vx$.

where V is the usual variance-covariance matrix.

1.3.2 Mean-Variance Dominance Criterion

The efficiency criterion proposed by Markowitz for the second stage is the Mean-Variance Dominance Criterion. It takes into consideration mean, that is the expected return, and variance to differentiate between efficient portfolios and inefficient ones, leading the investor to seek the lowest variance for a given expected return or the highest expected return for a given level of variance.

Definition. *Mean-Variance Dominance Criterion.* Given two random variables X and Y, respectively with mean μ_X and μ_Y and variance σ_X^2 and σ_Y^2, it is possible to state that X dominates Y with respect to the mean-variance criterion if and only if the following three conditions hold simultaneously:

1. $\mu_X \geq \mu_Y$;
2. $\sigma_X^2 \leq \sigma_Y^2$;
3. At least one of the previous inequalities is verified in narrower sense.

[3] If $\rho > 0$ the return of the two assets move in the same direction and are positive correlated; if $\rho < 0$ the return move in the opposite direction and they are negatively correlated; if $\rho = 0$ there is no relationship between the two assets.

It is clear that the criterion introduces a partial basis for comparison since it does not allow to discriminate all pairs of portfolios. In fact, it may happen for instance that when considering two efficient portfolios, we are not able to determine whether one portfolio dominates the other.

It is possible to state that the set of efficient portfolios, named *efficient frontier*, is constituted by all portfolios which, alternatively, once determined the desired level of expected return, minimize portfolio's risk or, given a certain level of risk, maximize the expected return. Rational investors' choice can fall only on a portfolio belonging to this set whereas for every inefficient portfolio there is one which, carrying the same risk, can guarantee a greater return or, equivalently, having the same expected return, guarantees a lower risk.

In the case of N assets with random returns, Markowitz formulation for the portfolio selection problem can be stated as follows[4]:

$$\underset{x}{\text{minimize}} \quad x'Vx$$

$$\text{subject to} \quad \begin{cases} x'\bar{r} = \pi \\ x'e = 1 \\ x \geq 0 \end{cases}$$

where:

- x is the N-order vector constituted by the portion of wealth $x_1, \dots, x_n$ invested in the i-th asset of the portfolio, with $i = 1, 2, \dots, n$;
- V is the N-order quadratic matrix of variances and covariances[5];

[4] This formulation is an example of quadratic program, an optimization problem constituted by a quadratic function and linear constraints.

[5] Given the symmetric nature of covariances, the matrix is as well symmetric by definition, with variances on its diagonal. We assume that the matrix is non-singular: none of the assets returns is perfectly correlated with the return of a portfolio composed by the remaining assets and none of the assets is riskless.

- $\bar{r}$ is the N-order vector composed by mean returns $r_1, \ldots, r_N$ of N assets[6];
- e is a N-order unitary vector;
- π is the level of expected return that the investor wishes.

The constraints considered by Markowitz are basic and they can be explained as follows: the first one implicates that, in the process of risk minimization, the level of expected return desired by the investor and fixed ex-ante π must be taken into consideration; the second constraint requires that the entire wealth at disposal is invested; the last ones imply that the portions of wealth invested in each asset are non-negative, in order to avoid short selling[7].

With the aim of determining a unique vector of optimal weights, the following statement should be made.

Theorem. If the variance-covariance matrix V is positive definite[8] 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.

Notice that the first two linear constraints define a convex set and, being V positive definite, also the function $x'Vx$ is convex.

To find the formula for the optimal portfolio given the constraints, we shall start from the following lagrangian function:

$$L = x'Vx - \lambda_1(x'r - \pi) - \lambda_2(x'e - 1)$$

[6] It is assumed that not all elements of r are equal. Conversely, the entire wealth would be invested in the asset with the lowest variance.
[7] Short selling is a particular investment or trading strategy which involve the sale of a security not owned by the seller. It is undertaken when the seller has the belief that the price of the security will decline or when an investor wants to hedge, placing an offsetting position to reduce risk exposure.
[8] A $N \times N$ matrix V is positive definite if $x'Vx > 0$ for any non-zero N-vector x.

where λ_1 and λ_2 are the Lagrange multipliers.

We can now set equal to zero the partial first derivatives of L and set up the system:

$$\begin{cases} \dfrac{\partial L}{\partial x} = 2x'V - \lambda_1 r' - \lambda_2 e' = 0 \\ \dfrac{\partial L}{\partial \lambda_1} = -x'r + \pi = 0 \\ \dfrac{\partial L}{\partial \lambda_2} = -x'e + 1 = 0 \end{cases}$$

With some computations it is possible to obtain the final unique solution to optimization problem:

$$x = \frac{(\gamma V^{-1} r - \beta V^{-1} e)\pi + (\alpha V^{-1} e - \beta V^{-1} r)}{\alpha\gamma - \beta^2}$$

where:

$$\alpha = r'V^{-1}r$$

$$\beta = r'V^{-1}e = e'V^{-1}r$$

$$\gamma = e'V^{-1}e$$

The efficient frontier's analytical expression varies depending on the composition of the portfolio:

- Portfolio with N > 2 risky assets: as in the case explained, the frontier expression is still represented by a parabola in the mean-variance plane, but the vertex changes depending whether a risk-free asset is considered or not.

- Portfolio with N = 2 risky assets: in this scenario the frontier expression is particularly affected not only by the possible presence of a risk-free asset, but also by the linear correlation coefficient between the two assets.

1.3.3 Portfolio Selection

In the third and last stage the proper portfolio for the investor is selected, taking into consideration the investor's risk aversion and knowing that usually all investors prefer returns to be high and/or stable, not subject to uncertainty. In order to do so, we consider the investor's expected utility function and we maximize it, seeking to obtain from one of the portfolios laying on the efficient frontier the greatest utility for the investor.

In his model, Markowitz adopted the *quadratic utility function*, described by the following equation:

$$U(R_P) = R_P - \frac{a}{2} R_P^2$$

where:

- R_P is the random variable representing the return of the portfolio;
- a is a strictly positive coefficient reflecting the investor's risk aversion: the greater is a, the greater is the investor's risk aversion.

However, the compatibility between mean-variance criterion and the theory of expected utility maximization occurs only in two limit cases: following a "subjective" approach, when the utility function of all investors has quadratic form, whereas, following an "objective" approach, when the joined probability distribution function of the N assets constituting the portfolio is a multivariate elliptical one, independently from the utility function form.

Hence, it is necessary that the efficient frontier is consistent with the maximization of the expected utility. Knowing the diversity of forms that can characterize the efficient frontier, such as the variety of possible values that the linear correlation coefficient can assume, the determination of the optimal portfolio for a specific investor can be formulated as the following constrained maximization problem:

$$maximize \quad E[U(R_P)]$$

$$subject\ to \quad \sigma_P^2 = f(R_P)$$

Concerning this scenario, one of Markowitz's most significant contribution can be considered the concept of *diversification* in the financial world.

Definition. *Diversification.* There is diversification when, allocating wealth at investor's disposal in N > 1 assets, it is possible to exploit the correlation between the assets to reduce portfolio risk under the level of the portfolio's asset with the lowest risk.

Recalling the mathematical properties of the variance, regarding the sum of non-independent random variables[9], it is possible to infer the key role played by the linear correlation coefficient. In fact, albeit in the reality of financial markets the existence of assets linked together by a linear correlation coefficient of -1 is unlikely, it is perceivable how also coefficients with less extreme values can participate in the decrease of overall portfolio risk.

[9] Given two random variables X and Y, the variance of their sum can be defined as $\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 + 2\sigma_{X,Y} = \sigma_X^2 + \sigma_Y^2 + 2\rho_{X,Y}\sigma_X\sigma_Y$, with $\sigma_{X,Y}$ being their covariance, σ_X and σ_Y being respectively the standard deviation of X and Y and $\rho_{X,Y}$ being the correlation coefficient of the two variables.

Referring to the portfolio overall mean return, being a linear combination of its single assets' expected returns, in absence of short selling – as in the case previously presented – it has to be included between the lowest and the highest of the single assets' mean returns. Whereas, with the exploitation of short selling it is possible to enlarge the invested capital with respect to the initial quantity. In terms of portfolio returns this leads to the chance of obtaining a higher return in mean in relation to the portfolio's asset with the highest return. In this case of course the greater expected return will come at the price of a higher overall portfolio risk.

The limit of Markowitz model's low diversification is that, since in the real world empiric correlations between assets are rather low, to reduce the overall variance the mean-variance criterion often tend to over-weight assets with low variances instead of exploiting negative relations between their market price trends.

1.4 Critics to Markowitz Model

As mentioned earlier, Markowitz model constitutes a fundamental contribution to the present quantitative finance on the subject of portfolio theory. However, the assumptions underlying the model which nowadays appear rather simplistic, were more suitable at the half of the last century, period in which the model was first presented and applied. Hence it seems appropriate discussing the main limits faced by the model, concerning its *basic assumptions* and the *type of constraints* employed.

<u>Returns distribution:</u>

The first limit regards hypothesis underlying returns theory. The Normal density probability function generally assumed poorly describes financial assets returns in the real market. In fact, a general tendency observed in assets returns is negative asymmetry in returns distribution. A Normal distribution variable is characterized by an asymmetry equal to zero. On the other hand, asymmetry suggests that a distribution has its barycentre towards values greater than the mean, in which case asymmetry value will be positive, or towards values lower than the mean, in which case asymmetry value will be negative. A situation of negative asymmetry in returns distribution reveals that negative events often lead to greater negative returns in proportion with respect to positive returns given by favourable events. This scenario can be observed for instance in the case of bonds, where the consequence of a default[10] lead to a great negative return, while positive return is fixed.

Additional consideration that increases differences between real market and the Normal distribution assumption is kurtosis[11]. It has been shown that returns distributions observed on the market have greater values of kurtosis. Hence, extreme events (very positive or very negative returns) are more frequent with respect to theoretical returns with Normal distribution.

Expected utility function:

The assumption of a quadratic utility function has been often criticized. In particular, following the most common economic principles, as wealth increases also utility should increase, while it is evident that a quadratic utility function decreases above a certain level of wealth. This implication is of course controversial with the hypothesis of non-satiety of investors. Furthermore, other critics show that the implicit theoretical weakness in the assumption of utility function's quadratic form lies in the increasing risk aversion deriving from it. On the contrary, economic theory suggests that a decreasing risk aversion corresponding to the increase of wealth would be more suitable with economic agents' behaviour.

Measure of risk:

Being a symmetric measure, the use of variance as a risk measure has been as well disapproved since it considers positive and negative returns[12] equally weighted.

[10] Complete or partial payment of capital and interests not fulfilled.

[11] Kurtosis is an index relative to distribution's form. It measures the thickness of tails: a Normal distribution has a kurtosis value of 3, if tails are fatter the value is bigger than 3 and vice versa.

[12] In this context, positive and negative returns can also be referred as *upside potential* and *downside risk*.

However, it has been shown that investors do not treat them with the same weight. The utilization of variance as a risk measure can only be consistent in case returns present a symmetric distribution. This aspect was noticed from Markowitz himself few years after the release of his article, as we will discuss in the following paragraph.

Other assumptions:

Additional unrealistic assumptions contribute to constitute more practical limits to Markowitz model, such as:

- Friction-less market: absence of transaction costs or taxation costs. They affect portfolio performance; thus, they assume a major role in portfolio management;
- Absence of constraints deriving from the economic-politic context in which one is operating;
- Absence of constraints relative to the possibility of buying or selling an asset in a finite number.

Despite being questioned for many assumptions and limits, Markowitz model remained the most important contribution to Modern Portfolio Theory. This relevance, in conjunction with continuous debating, has led to various improvements of the model.

1.5 Improvements of Markowitz Model

With the aim of making the model proposed by Markowitz more realistic, it is possible to follow two paths. First, constraints system can be modified in order to make it more consistent with the real world; we could add for instance market friction limits as transaction and taxation costs or take into account *mixed-integer constraints*. Secondly, it is possible to operate on the objective function, hence on the measure of risk.

In this work, the attention will be prevalently addressed to the objective function and its risk measure; nonetheless, it seems useful in author's opinion to provide a brief description of mixed-integer constraints, that help portfolio selection problem to gain a more realistic form.
Mixed-integer constraints can be divided into three categories:

- Constraints relative to transaction's minimum lots, which have to be negotiated only in an integer number of units;
- Constraints relative to the maximum positive integer number of different assets which can be negotiated;
- Constraints relative to the minimum positive integer number of minimum lots of a given asset which has to be negotiated.

On one hand, the introduction of these limits provides a greater computational investigation of the programming problem; while on the other hand, it notably increases the complexity of the solution process. Indeed, the observance of mixed-integer constraints in a mathematical programming problem constitutes a problem known as *NP-complete*[13] and moreover, solving such mathematical programming problem is a *NP-hard*[14] problem.

As mentioned earlier, for the purpose of our work, we will undertake the second path which involves the objective function.

Since the beginning the use of returns' variance as risk measure inherent an investment portfolio appeared misleading. It was in fact Markowitz himself that proposed in 1959 its substitution in favour of semi-variance, formulated as follows:

[13] NP-complete are problems which are considerably burdensome to solve in terms of time requirement. In computational complexity theory, NP-complete are the most difficult problems in the NP class (non-deterministic polynomial-time problems).
[14] NP-hard are problems that are difficult at least as much as (or not less than) NP-complete.

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

While adopting this alternative, only potential losses are considered, thus returns realizations which are below their expected value. Any rational investor would not disdain to own a portfolio with returns above the expectations – hence the expected value –, so the notion of risk incorporated in variance definition does not result appropriate in this scenario. When a portfolio yields more than the expected return, it is suitable to refer to it as an opportunity, not a risk. Distinction between downside risk and upside potential appeared necessary in order to conceive risk as an element essentially negative rather than a simple dispersion measure.

Following the previous analysis, it seemed indispensable providing a definition of risk measure and identifying certain desirable characteristics.

1.5.1 Defining a risk measure

Although it is feasible to define certain desirable characteristics that a risk measure should have, the concept of risk is rather subjective[15] and it is, thus, not possible to univocally identify a measure capable of satisfying the problem of expected utility maximization – common to every investor – once the efficient frontier is determined. However, even though being affected by relativity and subjectivity, it is possible to determine some characteristics that a risk measure should have relatively to a specific set of investors: rational agents.

In order to find a risk measure that satisfies properly investors' preferences, different methodologies have been undertaken. Accordingly, a risk measure can be described with the following definition:

Definition. *Risk Measure.* A risk measure is a function ρ that assigns a non-negative numeric value to a random variable X, which can be interpreted as future return, $\rho: X \to R$.

[15] Risk is linked with the perception of uncertainty that a single investor has.

It is possible to identify the basic properties that a risk measure should satisfy in order to define function ρ as aforementioned, even though it is not enough to take into consideration. In fact, the concept of *coherent risk measure* is greatly relevant nowadays and will be discussed in the following paragraph.

We begin from the relevant and desirable characteristics that a risk measure should have:

- *Positivity:* a measure of risk associated to a random variable assumes a strictly positive value, at least null in case there is no randomness. Negative values do not make sense;
- *Linearity:* especially in the resolution of optimization problems of big dimensions, the computational complexity might be diminished linearly linking risk measure and future return. The goodness of certain risk measures is connected to the more treatable computations that come from a linear optimization problem, where risk and return are linked in a basic way;
- *Convexity:* a risk measure is convex if, given two random returns R_X and R_Y and a parameter $\vartheta \in [0;1]$, the following relation holds:

$$\rho(\vartheta R_X + (1-\vartheta)R_Y) \leq \vartheta\rho(R_X) + (1-\vartheta)\rho(R_Y).$$

  It is a property that highlights the importance of diversification, since it is a process that permits to reduce the overall portfolio risk, hence, to expose invested wealth to a minor risk.
  This property can be satisfied indirectly, satisfying the two following properties:

  - *Subadditivity:* $\rho(R_X + R_Y) \leq \rho(R_X) + \rho(R_Y)$;
  - *Positive homogeneity*: $\rho(\alpha R_X) = \alpha\rho(R_X)$, $\alpha \in \mathbb{R}^+$.

The procedure of risk measure minimization, given a certain level of expected return, has the aim of bounding the uncertainty linked to the invested capital future value rather than the increment of the latter, taking into consideration that the expected return level is established. In this way, it is possible to estimate in advance what the invested capital future value will be, with a level of uncertainty depending on returns' distribution variance. Consequently, the resulting portfolio can be defined optimal only by a risk-averse investor, which cannot consider also the concept of non-satiety.

Gathering in a unique real positive number all probability distribution's characteristics, it is clear that an important limit of risk measures is constituted by their incapacity of incorporating the whole information available in a stochastic order, which utilizes the losses' cumulative distribution function.

An important measure of *downside risk* that has been widely used in the economic and financial environment is *Value at Risk* (from now on, *VaR*). Its notion is rather simple and intuitive:

Definition. $Value\text{-}at\text{-}Risk$. Given a confidence level of $\alpha \in [0;1]$ and fixed a specific holding period, Value at Risk (VaR) indicates the maximum potential loss associated to a portfolio in $\alpha\%$ of cases during the holding period.

Originally conceived as synthetic indicator of market risk[16], this measure of risk is notably widespread in savings and credit industries. Albeit having an experienced creator – it was conceived within the American investment bank J.P. Morgan – and expressing risk in the same unit measure of invested capital (monetary terms), VaR presents various limits when the underlying losses are not distributed as a Normal. Even in this case however, the assumption of Normality – rather widespread in literature – on one hand permits to lead back the portfolio optimization problem based on VaR to a Markowitz approach, while on the other hand leads to a not

[16] Market risk is defined as risk linked to adverse movements in financial activities' prices, in goods' prices, in interest rates, in exchange rates or in option volatility.

negligible underestimate of portfolio's real VaR, since returns of portfolios containing derivatives or tools to which is associated a low rating tend to have a strongly left-asymmetric distribution (negative asymmetry).

Regarding the computational profile, this measure does not fit in a particularly convenient way to bounded optimization problems since it emerges a stochastic programming problem rather difficult to solve. Moreover, with the exception of the case in which the underlying positions' probability distributions are known, it is complex to obtain a precise measure of portfolio's VaR.

From a probabilistic standpoint, VaR with a confidence level of α is the value that satisfies the following equality:

$$P(L > VaR_{\alpha}) = 1 - \alpha$$

Where L is a generic distribution of losses. Complying with this interpretation, it is evident how being a threshold measure[17] – indeed it express the maximum potential loss with a certain level of probability – does not provide any indication on the size of losses that exceeds that threshold, thus on the nature of the profit and loss distribution's left tail (the portion exceeding VaR). The distortion presented tends to be towards lower losses, leading to a contrast with the theory of risk management, which privileges a more cautious and pessimistic behaviour in the determination of risk level associated to a portfolio.

Lastly, even though not being less relevant, another gap presented by VaR concerns the aggregation of more risk sources. The above-mentioned measure does not encourage – sometimes even prohibits – diversification, since it does not take into account events' potential economic consequences. With such behaviour, VaR does not satisfy the feature of *subadditivity* – a property that will be seen more in detail in the next paragraph – since, applying this measure, the overall portfolio risk could result even greater than the sum of the single risk sources underlying each asset. It is important to underline how the lack of subadditivity creates, in addition to the inconsistency with the diversification principle, issues

[17] VaR belongs to the set of risk measures defined *quantile-based*.

with the numeric treatability. In fact, VaR is also criticized for is inability to quantify the so-called *tail risk*, hence, its low sensibility to extreme events.

1.5.2 Concept of coherent risk measure

In order to complete the process of individuation of desirable properties that a risk measure should have and to provide a follow-up to the inadequacies of VaR measure, it is appropriate to describe the concept of coherence, as formulated by Artzner *et al.* (1999).

The writing of coherence axioms represented an attempt of translating a complex reality into a mathematical formulation that is not so restrictive to identify a unique coherent measure of risk, but it rather characterizes a class of measures. As mentioned above, in addition to the "basic" properties of a risk measure, there are other significant properties. The respect of these additional features is a necessary condition to a correct interpretation of the concept of risk associated to a financial instrument.

Definition. $Coherent\ Risk\ Measure.$ A risk measure that satisfies the four axioms of *translation invariance, subadditivity, positive homogeneity* and *monotonicity* is called coherent.

We can now list and describe the four properties that define a coherent risk measure:

- *Translation invariance:* it guarantees that investing a percentage α of the available capital in a risk-free asset[18], the overall risk associated to the portfolio contracts proportionally to the percentage α allocated in the risk-free asset:

[18] Asset that has a known future return and that does not carry any level of risk. Usually government bonds are a perfect example.

$$\rho(X+\alpha) = \rho(X) - \alpha \quad , \ \forall \ r.v. \ X, \alpha \in \mathbb{R}.$$

It implies that $\rho(X + \rho(X)) = 0$. By adding a risk-free quantity equal to $\rho(X)$ to a risky position X, we obtain a risk-free entity, coherently with the operative interpretation of ρ as minimum positive quantity to add to the initial position in order to make the instrument acceptable (thus, risk-free);

- *Subadditivity:* it represents the essence of how a risk measure should behave in the case in which the investor has to deal with a combination of assets. The risk of a portfolio should never be greater than the sum of the single risks associated to each of the assets that constitutes it. Subadditivity is strictly correlated to the concept of diversification since it can be affirmed that diversification leads to a contraction of the overall risk only if, for the risk inherent to a certain position, the following statement holds:

$$\rho(X+Y) \leq \rho(X) + \rho(Y) \quad , \ \forall \ r.v. \ X, Y.$$

- *Positive homogeneity:* it ensures that if the investment in a risky asset varies, then the riskiness associated to that investment varies proportionally. In cases in which positions dimensions directly affect risk (e.g. if positions are so large that time required to liquidate them depends on their dimensions), consequences of lack of liquidity should be considered when calculating the future net worth of a position:

$$\rho(\lambda X) = \lambda\rho(X) \quad , \ \forall \ r.v. \ X, Y \ and \ \forall \lambda \geq 0.$$

- *Monotonicity:* it underlines the preferability of an asset that systematically assures returns greater than another asset:

$$\rho(X) \leq \rho(Y) \quad , \quad \forall \ r.v. \ X,Y \ with \ X \geq Y.$$

As mentioned earlier, the concept of coherent risk measure does not define a unique risk measure, instead, it characterizes a large class of risk measures. The choice of the right measure to use within the class should be made based on some additional economic considerations.

First examples of coherent risk measures

In their *"Coherent Measures of Risk"*, where they defined the axioms of coherence, Artzner *et al.* (1999) provided also the first guidelines concerning some proposals of coherent risk measures that satisfy the axioms. They present two measures, known as *Tail Conditional Expectation (TCE)* and *Worst Conditional Expectation (WCE)*, for whom the authors demonstrated that the relation $TCE_\alpha \leq WCE_\alpha$[19] holds.

Definition. *Tail Conditional Expectation.* Tail Conditional Expectation (known also as TailVaR) is a coherent risk measure defined as:

$$TCE_\alpha(X) \stackrel{\text{def}}{=} -E[X|X \ \leq \ -VaR_\alpha(X)].$$

Definition. *Worst Conditional Expectation.* Worst Conditional Expectation is a coherent risk measure defined as:

$$WCE_\alpha(X) \stackrel{\text{def}}{=} -\inf\{E[X|A] \,|\, P[A] > \alpha\}.$$

[19] As for VaR measure, subscript α indicates the desired confidence level, which is given.

Financially, TCE and WCE try to determine "*how bad is bad*" since they focus on returns distribution's left tail – the one representing losses – and they compute the mean value subject to the fact that the losses are greater than a certain value.

The concepts of TCE and WCE represent first proposals of coherent risk measures. However, at the same time, their resemblance could erroneously let their aspects of distinction pass unnoticed. If on one hand WCE satisfies completely axioms of coherence and, nonetheless, it is widespread only in the theoretical field – since it requires knowledge of the entire underlying probability space – , on the other hand TCE is more manageable also in the application environment, even though not always satisfying axioms of coherence[20].

Conversely, as discussed before, the adoption of VaR as risk measure does not provide any indication on the size of losses beyond a threshold value, represented by the measure itself. The introduction of the axioms of coherence leads to a change in the question that we can ask ourselves with the aim of determining a more suitable risk measure, hence, coherent. More specifically, risk measures discussed in this paragraph do not observe at the maximum potential loss in the $\alpha\%$ of cases, but they rather point out the expected loss in the worst case $(1-\alpha)\%$ scenario. In other words, these measures do not concentrate on a specific threshold, which does not supply with any information besides the threshold itself, but they focus on the losses' distribution beyond the threshold value and they synthetize its features through their mean value.

The aim of creating a measure of risk that combined contemporarily the good qualities of both measures was reached through the definition of an alternative and more suitable solution represented by the measure known as *Expected Shortfall (ES)*. This index can be financially explained as the average loss considering all losses beyond a certain threshold value, VaR.

[20] TCE measure may not always respect the property of subadditivity. In fact, TCE coherence is guaranteed only restricting the analysis field on the continuous probability distribution's functions, whereas it might not be guaranteed in the general case.

Expected Shortfall can be formally defined as follows.

Definition. $\mathit{Expected\ Shortfall}$. Given a profit and loss distribution X and defined holding period and significance level $\alpha \in [0; 1]$, Expected Shortfall is defined as:

$$ES_\alpha(X) \stackrel{\text{def}}{=} -\frac{1}{\alpha}\left(E\left[X1_{(X \leq x^\alpha)}\right] - x^\alpha[P[X \leq x^\alpha] - \alpha]\right)$$

where $x^\alpha = VaR$.

The second addendum of the sum within the parenthesis can be translated as the quantity to subtract from the mean value when $X \leq x^\alpha$ has probability greater than $1 - \alpha$. Whereas, when $P(X \leq x^\alpha) = 1 - \alpha$, as it is usually the case with probability distribution's continuous functions, we obtain that the value resulting from the ES_α's formula coincides with the TCE_α's one.

An equivalent representation that provide the advantage of more transparency and that permits to appreciate the simplicity of ES, can be obtained renouncing to the definition in terms of expected values. Let $F(X)$ be the probability density function[21] so that $P(X \leq x)$ and let $F^{-1}(\alpha) = \inf\{x | F(x) \geq \alpha\}$ be the inverse function of $F(X)$, it can be proved that ES can be expressed as $ES_\alpha(X) = -\frac{1}{\alpha}\int_0^\alpha F^{-1}(p)dp$.

The sample estimation of ES is obtained sorting the n possible realizations and, given a significance level, selecting the $(1 - \alpha)\%$ of the greater losses and obtaining the following result: $ES_\alpha(X) = -\frac{\sum_{i=1}^{w} x_{1:n}}{w}$, where w represents the integer part of $n(1 - \alpha)\%$, hence $w = \max\{m | m \leq n(1 - \alpha), m \in \mathbb{N}\}$.

ES is a universal risk measure, meaning that is applicable to any financial tool and to any underlying risk source. Moreover, it benefits of simplicity and completeness properties since it computes a unique number even in case of portfolios exposed to different risk sources and robustness. This is possible because, conversely to other risk measures focusing on distributions' tail, with ES, results do not vary

[21] Probability density function of a random variable X is a non-negative application $p_x(x)$ so that the probability of a set A is given by $P(X \in A) = \int_A p_x(x)dx$ for all subsets A of the sample space.

significantly when changing the confidence level of some point basis. This last aspect cannot be guaranteed by VaR, TCE or WCE.

An alternative expression of ES is the one proposed by Rockafellar and Uryasev (2002), named *Conditional Value-at-Risk (CVaR)*. Let the function associated to the loss be $z = f(x, y)$[22] with $\Psi(x, \zeta) = P\{y | f(x, y) \leq \zeta\}$, CVaR can be defined as follows:

Definition. *Conditional Value-at-Risk.* Fixed a significance level of $\alpha \in [0; 1]$, CVaR_α is equal to the expected value of the greater losses whose probability is equal to $1 - \alpha$. It is equivalent to the average of the distribution function:

$$\Psi_\alpha(x, \zeta) = \begin{cases} 0, & if\ \zeta < \zeta_\alpha(x) \\ [\Psi(x, \zeta) - \alpha]/[1 - \alpha], & if\ \zeta \geq \zeta_\alpha(x) \end{cases}$$

where $\zeta_\alpha(x)$ is the VaR_α associated to portfolio x.

Rockafellar and Uryasev (2000) themselves, besides demonstrating its coherence[23], highlighted an interesting further aspect: solving a simple convex optimization problem it is feasible to obtain separately both CVaR_α and VaR_α associated to portfolio x. It is a result of particular importance since it allows to compute CVaR_α of a position without necessarily knowing the relative VaR_α. Both risk measures can be determined simultaneously exploiting the following formula:

$$\mathrm{F}_\alpha(x, \zeta) = \zeta + \frac{1}{1 - \alpha} E\{[f(x, y) - \zeta]^+\}$$

[22] Authors express the loss associated to a portfolio in function of percentages vector x and the vector of each asset's future return y: the loss is then equal to $-x'y$.

[23] Acerbi and Tasche (2002b) let us understand that ES and CVaR are essentially two different labels employed to identify the same object, thus the expected loss in $(1 - \alpha)$ of cases.

where $[f(x,y)-\zeta]^+ = \max\{f(x,y)-\zeta; 0\}$.

From the demonstration of a relevant theorem, thanks to whom the authors proved how it is possible to determine VaR_α through a two-step approach[24], it derives that minimization of CVaR_α associated to a portfolio x is equivalent to $\mathrm{F}_\alpha(x,\zeta)$ minimization on the entire domain: $\min_{x\in X} \mathrm{CVaR}_\alpha(x) = \min_{(x,\zeta)\in X\times R} F_\alpha(x,\zeta)$.

This last result is remarkable since, with the aim of defining vector x that minimizes CVaR_α, it allows to work directly with a simple expression, convex with respect to variable ζ in $F_\alpha(x,\zeta)$, rather than managing an expression that requires the knowledge ex-ante of VaR_α's value. Lastly, the numerical analysis conducted by Rockafellar and Uryasev (2000) guarantees how such process is implicitly valid also for VaR_α minimization, being $CVaR_\alpha \geq VaR_\alpha$.

1.5.3 A new coherent risk measure: Entropic Value-at-Risk

Sometimes being coherent for a risk measure could not be enough; an important deficiency of CVaR, or Expected Shortfall, is that it cannot be computed in a reasonable time. Indeed, in most of the cases is it necessary to approximate CVaR through sampling methods. There are also other examples of coherent risk measures, as spectral risk measures[25], which cannot be efficiently computed even for simple cases. Having to face a stochastic optimization problem, as the portfolio selection one, incorporating a risk measure that has to be computed frequently, makes the need for an efficiently computable coherent risk measure more essential and relevant.

In an important paper of the Journal of Optimization Theory and Applications, A. Ahmadi-Javid (2012) introduced a new coherent risk measure called *Entropic Value-at-Risk (EVaR)*. It constitutes "*the tightest possible upper bound obtained*

[24] The approach consists of a first step which requires the definition of a set of values of ζ which minimize $\mathrm{F}_\alpha(x,\zeta)$ and a second step which identifies its left extreme in case the set contains more elements. This process results worthless if not interested in VaR's value.
[25] Acerbi C. (2002)

from Chernoff inequality[26] for the value-at-risk (VaR) as well as the conditional value-at-risk (CVaR)" Ahmadi-Javid (2012). In his work Ahmadi-Javid demonstrated that a large class of stochastic optimization problems that are computationally intractable with CVaR, is efficiently solvable when incorporating EVaR. The dual representation of EVaR is strictly linked to the Kullback-Leibler[27] divergence, also known as relative entropy.

Entropic Value-at-Risk owes its name to its connections with Value-at-Risk and relative entropy. We will describe this risk measure more in details and incorporate it in our portfolio selection model in the next chapter.

[26] Chernoff inequality will be explained in the following chapter.
[27] Kullback-Leibler divergence is a concept that will be discussed later.

Chapter 2

Entropic Value-at-Risk application

In this chapter, Entropic Value-at-Risk's implementation as a risk measure in a portfolio selection problem is discussed. It has been demonstrated [Ahmadi-Javid (2012)] that this measure, proposed by Ahmadi-Javid itself, is able to efficiently solve a broad class of stochastic optimization problems which are instead intractable with CVaR. Indeed, in recent years EVaR has been discussed in many economists and scholars' studies, often present in papers related to coherent risk measures and portfolio optimization problems[28].

EVaR will be applied to a realistic portfolio selection model, comprehensive of several constraints generally used in fund management practice, proposed by Corazza, Fasano and Gusso (2013).

2.1 Entropic VaR as a risk measure and its properties

Let the risk measure ρ be a function assigning a real value to a random variable $X \in \boldsymbol{X}$ and let $\boldsymbol{X}$ be a set of allowable random variables. Then let $(\Omega, \boldsymbol{F}, P)$ be a probability space where Ω is a set of all simple events, **F** is a σ-algebra of subsets of Ω and P is a probability measure on **F**. Furthermore, suppose that $\boldsymbol{L}$ is the set

[28] Ahmadi-Javid A. (2012c). *Application of information-type divergences to constructing multiple-priors and variational preferences.* In: Proceedings of IEEE International Symposium on Information Theory, Cambridge, MA, pp. 538-540.
Ahmadi-Javid A. (2012d). *Application of entropic value-at-risk in machine learning with corrupted input data.* In: Proceedings of 11th International Conference on Information Science, Signal Processing and their Applications (ISSPA), Montreal, QC, pp. 1104-1107.
Ahmadi-Javid A. and Pichler A. (2017); Pichler A. (2017); Delbaen F. (2018). See bibliography.

of all Borel measurable functions[29] – random variables – $X: \Omega \to \mathbb{R}$, and $\boldsymbol{X} \subseteq \boldsymbol{L}$ is a subspace including all real numbers. It is now possible to define the risk measure $\rho: \boldsymbol{X} \to \overline{\mathbb{R}}$, where $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ is the extended real line. For $p \geq 1$ let $\boldsymbol{L}_p$ be the set of all Borel measurable functions $X: \Omega \to \mathbb{R}$ for which $E(|X|^p) = \int |X|^p dP < +\infty$, $\boldsymbol{L}_\infty$ be the set of all bounded Borel measurable functions, $\boldsymbol{L}_M$ be the set of all Borel measurable functions $X: \Omega \to \mathbb{R}$ whose moment-generating function $M_X(z) = E(e^{zX})$ exists $\forall z \in \mathbb{R}$, and $\boldsymbol{L}_{M^+}$ be the set of all Borel measurable functions $X: \Omega \to \mathbb{R}$ whose moment-generating function $M_X(z)$ exists $\forall z \geq 0$. Notice that $\boldsymbol{L}_\infty \subseteq \boldsymbol{L}_M \subseteq \boldsymbol{L}_p,\ \forall p \geq 1$.

As mentioned in chapter 1, Entropic VaR can be described as the tightest possible upper bound obtained from Chernoff inequality for the VaR. Chernoff inequality [Chernoff H. (1952)] for any constant a and $X \in \boldsymbol{L}_{M^+}$ is as follows:

$$\Pr(X \geq a) \leq e^{-za} M_X(z), \qquad \forall z > 0.$$

Solving the equation $e^{-za} M_X(z) = \alpha$ with respect to a for $\alpha \in]0,1]$, we obtain

$$a_X(\alpha, z) \coloneqq z^{-1} \ln\left(\frac{M_X(z)}{\alpha}\right),$$

for which we have $\Pr(X \geq a_X(\alpha, z)) \leq \alpha$. Indeed, for each $z > 0$, $a_X(\alpha, z)$ is an upper bound for $VaR_{1-\alpha}(X)$. It is possible now to consider the best upper bound of this type as a new risk measure that bounds $VaR_{1-\alpha}(X)$ by using exponential moments.

[29] A map $f: X \to Y$ between two topological spaces is called Borel measurable if $f^{-1}(A)$ is a Borel set for any open set A. Note that the σ-algebra of Borel sets of X is the smallest σ-algebra containing the open sets. (Borel function: encyclopediaofmath.org).

Definition. *Entropic Value-at-Risk (EVaR).* The entropic value-at-risk of $X \in \boldsymbol{L}_{M^+}$ with confidence level $1-\alpha$ is defined as follows:

$$EVaR_{1-\alpha}(X) \coloneqq \inf_{z>0}\{a_X(\alpha,z)\} = \inf_{z>0}\left\{z^{-1}\ln\left(\frac{M_X(z)}{\alpha}\right)\right\}.$$

As proved in Ahmadi-Javid (2012a), EVaR is a coherent risk measure. To find its dual representation and its connection to relative entropy, we can proceed as follows:

Theorem. For every coherent risk measure $\rho: \mathbf{L}_\infty \to \mathbb{R}$ with the Fatou property[30], there exists a set of probability measures $\mathfrak{J}$ on $(\Omega, \mathbf{F})$ such that

$$\rho(X) = \sup_{Q\in\mathfrak{J}} E_Q(X).$$

The above equation is known as the dual representation or robust representation of ρ. Furthermore, the expression is an additional demonstration that this risk measure is coherent.

Lemma. *Donsker-Varadhan Variational Formula.* For any $X \in \boldsymbol{L}_\infty$,

$$\ln E_P(e^X) = \sup_{Q\ll P}\{E_Q(X) - D_{KL}(Q||P)\},$$

[30] A translation invariant supermodular mapping $\phi: L^\infty \to \mathbb{R}$ is said to satisfy the Fatou property if $\phi(X) \geq \sup \phi(X_n)$, for any sequence $(X_n)_{n\geq 1}$ of functions, uniformly bounded by 1 and converging to X in probability. *Delbaen et al. (2000).*

where $D_{KL}(Q||P) \coloneqq \int \frac{dQ}{dP}\left(\ln \frac{dQ}{dP}\right) dP$ is the *relative entropy*[31] of Q with respect to P, or the *Kullback-Leibler divergence*[32] from Q to P.

Theorem. The *dual representation of $EVaR_{1-\alpha}(X)$ for $X \in \boldsymbol{L}_\infty$* has the form:

$$EVaR_{1-\alpha}(X) = \sup_{Q \in \mathfrak{J}} E_Q(X),$$

where $\mathfrak{J} = \{Q \ll P: D_{KL}(Q||P) \leq -\ln \alpha\}$.

Entropic value-at-risk is characterized by two other important properties: the former related to another variable with same distribution, the latter linked to the comparison with VaR and CVaR.

Corollary. For $X, Y \in \boldsymbol{L}_M$, $EVaR_{1-\alpha}(X) = EVaR_{1-\alpha}(Y)$ for all $\alpha \in]0,1]$ if and only if $F_X(b) = F_Y(b)$ for all $b \in \mathbb{R}$.

The proof of this property follows from a well-known property of moment-generating functions, stating that two distributions are identical if the have the same moment-generating function. This corollary shows that $EVaR_{1-\alpha}(X)$ as a function of its parameter α characterizes the distribution of $X \in \boldsymbol{L}_M$. The initial condition $X, Y \in \boldsymbol{L}_M$ can be weakened to the existence of $M_X(b)$ and $M_Y(b)$ over the interval $b \in]-\varepsilon, +\infty]$ for a positive constant $\varepsilon > 0$.

[31] Entropy is a concept that derives from physics and allows to evaluate the level of disorder in a system. If the level of disorder grows, entropy increases as well, vice versa if it reduces, entropy decreases. Recently this measure has been reproposed in different fields as information theory, IT, biology, medicine and social sciences. In economics and finance entropy has been mostly employed as a measure of risk or as foundation of a more complex risk measure.

[32] Kullback-Leibler divergence – also known as information divergence or relative entropy – is a non-symmetric measure of the difference between two probability (P and Q) distributions. More specifically, K-L divergence from Q to P, identified as $D_{KL}(Q||P)$, is the measure of the lost information when P is used to approximate Q.

Proposition. The EVaR is an upper bound for both VaR and CVaR with the same confidence levels, i.e. for $X \in \boldsymbol{L}_{M^+}$ and every $\alpha \in]0,1]$:

$$CVaR_{1-\alpha}(X) \leq EVaR_{1-\alpha}(X).$$

Furthermore,

$$E(X) \leq EVaR_{1-\alpha}(X) \leq ess\,\sup(X),$$

where:

- $E(X) = EVaR_0(X)$;
- $ess\,\sup(X) = \lim_{\alpha \to 0} EVaR_{1-\alpha}(X).$

This statement affirms that EVaR is more risk-averse with respect to CVaR at the same confidence level. Thus, EVaR would suggest to a financial firm allocating more resources to avoid risk. This feature could make EVaR less attractive for companies which search, for instance, a greater return and are not afraid of risk; however, EVaR computational tractability results more simple, which can be important when we need to incorporate a risk measure in a stochastic optimization problem, both in terms of time and difficulty.

2.2 A realistic portfolio selection model under EVaR

Once an appropriate measure of risk is identified, with the purpose of quantifying the level of riskiness inherent in a financial investment, there is the necessity to introduce a set of constraints in order to develop a realistic portfolio selection model.

Even though it could be taken for granted, it is important to underline how the research of a portfolio which is able to minimize any measure of risk, being it represented by returns' variance or a coherent risk measure, does not lead to a solution that can be adopted in practical terms if not supplied by constraints to be followed. Referring to what described in chapter 1, the generic rational investor has to deal with the choice between conflicting objectives as return maximization and risk minimization, in the overall search of expected utility maximization.

Risk minimization without taking into consideration constraints in terms of expected return, but considering only the budget constraint[33], would lead to an optimal solution that is given by the minimum risk portfolio[34]. Without explaining in details logics and reasonings of expected utility theory, since they are not part of the aim of our work, it appears clear that this efficient solution is only one among many admissible and does not take into account the level of risk aversion of each investor, transmitted through a specific utility function.

Even considering the constraint related to the expected return, in addition to the pre-mentioned budget constraint, the solution appears too simplistic with respect to the reality of financial markets. For this reason, the model proposed in this work will take into account transaction costs, in other words those costs that the intermediary will charge to the investor once the transaction in completed and that represent, together with the tax regime, a significant financial market friction. Transaction costs will be considered indirectly through the introduction of cardinality constraints. The author believes it is appropriate to add these constraints to our model since they represent one of the major constraints categories to which a manager is subjected in daily practice.

Nevertheless, it is important to highlight that each time that a constraint is added to the model, its computational complexity increases proportionally and even more rapidly when the added constraints do not have linear and/or continuous forms. The problems that generate after the insertion of such constraints are called *NP-hard*, which, due to their difficulty in terms of computation and time, require the employment of heuristics and metaheuristics to be solved, since the

---

[33] Risk unconstrained minimization would inevitably lead to a vector of zero percentages of investment in each asset, since no portion of wealth would be allocated in an asset with an aleatory return.

[34] In the Markowitz's mean-variance approach it is called "global minimum variance portfolio", meaning the portfolio with the returns' minimum variance.

research of optimum solutions through the use of exact methods might not lead to a result.

Before the final proposal of the realistic portfolio selection model, it is important to list and describe in detail the constraints that it has been decided to consider in the model.

2.2.1 Budget and return constraints

Budget constraint and return constraint are widespread in most of portfolio selection models, since they represent the essential part of the problem. They are also included in the constrained Markowitz model.

Budget constraint assures that all available wealth is invested and can be mathematically formulated as follows:

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

where x_i represents the percentage of capital invested in the $i-th$ asset. Budget constraint can also be expressed in matrix terms as:

$$x'e = 1.$$

Return constraint establishes the determination, on the investor side, of a certain level of desired return π and guarantees that the expected return of the portfolio is not below that threshold. Hence, it permits to select a portfolio among all portfolios lying on the efficient frontier: the one which minimizes the adopted measure of risk, in our case the entropic value-at-risk. In algebric terms, portfolio mean return is equal to $\mu_p = \sum_i^N x_i r_i$, whereas in matrix terms it is defined by $x'r$. The constraint on the minimum mean return can be formulated as follows:

$$\sum_{i=1}^{N} x_i r_i \geq \pi.$$

In the specific problem that is being proposed, the perfect equality between mean return and π is not required, whereas a wider condition is imposed in order to avoid that the admissible region results empty. Indeed, any rational investor would not be dissatisfied by a portfolio that, at the same risk level, provides a return greater than expected.

From a conceptual point of view, it is possible to reach the same result if deciding to maximize return once a certain level of risk is fixed. However, considering the problem from a risk-averse investor's perspective, the concept underlying the choice to configurate the objective function that must be optimized with a measure of risk – to be minimized – instead of a measure of the return – to be maximized – reflects the investor's aversion to risk; investor who tends to focus his attention on risk rather than on return. This occurs notwithstanding these two variables are positively correlated. An example of this tendency towards risk is represented by the so-called MIFID[35] interview, which is applied by credit institutions to their clients in order to verify both their knowledge and experience referring to products and financial tools and their investment goals. Questions that constitutes the interview, deal with the risk aversion of the investor rather than with its propension to profit. Hence, it is possible to obtain a financial profile of

[35] The 2004/39/EU directive MIFID (Markets in Financial Instruments Directive) was introduced in 2007 and has the objective of increasing investor protection and guaranteeing the maximum level of transparency through mandatory information to customers.

the client which allows to guide him through the formulation of an adequate investment choice.

2.2.2 Cardinality constraints

Cardinality constraints are bounds related to the number of assets to include in the portfolio; they will also be associated with a constraint regarding the fraction of portfolio that each asset constitutes. The reason for the introduction of these constraints is to obtain an indirect control on transaction costs. The majority of models presented in literature considers, especially in the case in which the focus is on the substance of the approach rather than on the form, a system of constraints rather elementary, based on the assumption of perfect market conditions, frictionless. However, transaction costs, in addition to being significant incidence factors on the effective real performances of the portfolio, represent a constraints category that needs to be taken into consideration in the development of a portfolio model since their presence considerably influences managers activities in their daily practice.

Since the trade of assets comes with a cost, portfolio managers are led to subordinate portfolio creation to costs that has to be borne. For this reason, it is important that they operate, in first place, introducing a not too small not too big number of assets among the N constituting the basket of financial tools at their disposal. In fact, when the amount of assets selected is too large, many practical issues can befall, e.g. high dimensionality of the problem which can raise transaction costs as well. In second place, portfolio managers will also be subordinated to the selection of an investment percentage for each asset that must be not too small and not too big; these fractions of portfolio will be strictly correlated with the minimum and maximum number of assets present in the portfolio.

We can introduce and consider transaction costs in our portfolio selection problem through the employment of the following cardinality constraint:

$$K_d \leq \sum_{i=1}^{N} z_i \leq K_u,$$

where:

- $1 \leq K_d \leq K_u \leq N$;
- $z_i \in \{0, 1\}$ represent a binary variable and
  - $if\ z_i = 1, the\ i - th\ asset\ is\ included\ in\ the\ portfolio$;
  - $if\ z_i = 0, otherwise.$
- K_d is the minimum number of assets included in the portfolio;
- K_u is the maximum number of assets included in the portfolio.

Moreover, it is important to add a constraint which states that each of the selected assets cannot be a too small or too large fraction of the portfolio. This bound consists in establishing minimum and maximum fractions, respectively d_i and d_u, to allocate in each asset and can be formulated as follows:

$$z_i d_i \leq x_i \leq z_i u_i,$$

where $0 \leq d \leq u \leq 1$.

In order to guarantee the compatibility between these two constraints, parameters d and u must satisfy the following inequalities:

- $d \leq \frac{1}{K_d}$;
- $u \geq \frac{1}{K_u}$.

Now that we have described in detail the system of constraints that will be considered for the purpose of the present work, it is possible to propose the portfolio selection model developed.

2.2.3 Portfolio selection model

Let N be the number of assets from which we can choose from and, for $i = 1, \dots, N$, let $x_i \in \mathbb{R}$ be the weight of the $i-th$ asset in the portfolio, with $X^T = (x_1, \dots, x_n)$. Let $Z^T = (z_1, \dots, z_n) \in \{0,1\}^N$ be a binary vector. Moreover, consider r_i a real valued random variable for $i = 1, \dots, N$, which represents the return of the $i-th$ asset, and $\hat{r}_i = E(r_i) = \frac{1}{T}\sum_{t=1}^{T} r_{i,t}$ its expected value. Therefore, the random variable R which represents the return of the overall portfolio can be formulated as:

$$R = \sum_{i=1}^{N} x_i r_i$$

Thus, its expected value can be expressed as:

$$\hat{R} = \sum_{i=1}^{N} x_i \hat{r}_i$$

Considering our risk measure's formula, previously described, our aim is to minimize $EVaR_{1-\alpha}(R)$, subject to the system of constraints just discussed, in order to find the optimal portfolio. Translating this constrained minimization problem, our resulting portfolio selection problem can be expressed as follows:

$$
\begin{aligned}
\underset{X,Z}{\text{minimize}} \quad & EVaR_{1-\alpha}(R) \\
\textit{subject to} \quad & \hat{R} \geq \pi \\
& \sum_{i=1}^{N} x_i = 1 \\
& K_d \leq \sum_{i=1}^{N} z_i \leq K_u \\
& z_i d_i \leq x_i \leq z_i u_i, \qquad with\ i = 1, \dots, N \\
& z_i \in \{0,1\}, \qquad with\ i = 1, \dots, N
\end{aligned}
$$

Ahmadi-Javid (2012a) demonstrated, in his significant paper, that the function underlying the Entropic value-at-risk is convex. The introduction of such risk measure in our model, makes the latter a non-linear and convex mixed-integer optimization problem. Conversely to mean-variance portfolio selection problems, which are non-convex and generally admit several local solutions, our mean-EVaR portfolio selection problem facilitates the search of global solutions, rather than local minimizers. Since exact methods would be heavily time consuming, to solve the problem it has been decided to employ PSO metaheuristic, which can provide a good approximation in a reasonable amount of time. However, PSO was born as a metaheuristic technique that cannot be applied to constrained problems. Luckily, it is possible to reformulate the constraints in order to adapt the constrained optimization problem to the metaheuristic.

2.3 Reformulation of portfolio selection model for PSO

With respect to other forms of evolutive computations used in constraints management, in the opinion of Hu and Eberhart (2002), the employment of PSO presents several advantages: in addition to its simplicity in terms of number of parameters that need to be set, the amount of problems that it is able to solve is rather vast and, moreover, it does not set any restriction – regarding the potential provision of a result – neither on the objective function nor on the constraints system.

Despite the presence in literature of different methods to adapt a constrained problem to PSO algorithm, for this work it has been decided to adopt the strategy proposed by Parsopoulos and Vrahatis (2002), which consists in introducing a penalty function. This method permits to translate a constrained optimization problem into an unconstrained one by penalizing all the violations of the constraints and joining them into a particular fitness function which considers the objective function and the initial constraints. Since constraints are, in this way, integrated into it, the unique fitness function will then be optimized by applying PSO algorithm, which, as said earlier, was designed to solve unconstrained problems.

2.3.1 Penalty function

In general, the exact penalty method has the objective of solving constrained optimization problems through the exploitation of unconstrained problems. It consists in adding a term to the objective function that attaches a cost to the portfolio in case of constraints violations. Thus, solutions which comprehend constraints violations are made less favourable as potential good solutions. A parameter, ϵ, measures the magnitude of the penalty, approximating the original constrained problem with the fitness of the unconstrained one.

The general concept of this method consists in substituting a generic constrained minimization problem as the following:

$$\begin{array}{lll} minimize & f(x) & \\ subject\ to & h_j(x) = 0, & j = 1, \dots, m \\ & g_i(x) \leq 0, & i = 1, \dots, p \end{array}$$

where:

- $f(x)$ is a continuous function R^n;
- $h(x) = 0$ is a vector of m constraints expressed in equality form;
- $g(x) \leq 0$ is a vector of p constraints expressed in inequality form;

with an equivalent unconstrained minimization problem as the following:

$$\min f(x) + \epsilon P(x)$$

where:

- ϵ is a positive constant;
- $P(x)$ is a function R^n with the following characteristics:
  - $P(x)$ is continuous;
  - $P(x) \geq 0, \ \forall x \in R^n$;
  - $P(x)$ if and only if $x \in S$[36].

One of the issues of this methodology usually concerns the level of good approximation that the unconstrained problem does on the original constrained one. In general, as ϵ tends to infinite, it is necessary to check whether the solution of the unconstrained problem converges towards the constrained one. This issue should not be considered in our model, since it has been decided to employ a specific family of penalty methods, which ensures the biunivocal correspondence between the two problems' solutions.

[36] S is a vectorial subspace of R^n.

Considering the above-mentioned generic constrained minimization problem, the author decided to implement the approximation by introducing into the unconstrained optimization problem the following penalty function:

$$P(x) = \sum_{i=1}^{m} |h(x)| + \sum_{j=1}^{p} \max\,(0, g_j(x))$$

Correspondence between the original problem and the unconstrained one is guaranteed by the following important theorem.

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 $\epsilon > \max\,\{\lambda_i | \mu_j : i = 1, \dots, m\ j = 1, \dots, p\}$, x^* is also a local minimum for the unconstrained penalty objective function.

The theorem is proposed and demonstrated by Luenberger and Ye (2008). It is important to note that what previously described, together with the theorem, provides sufficient conditions in order to guarantee coincidence between unconstrained problem's and original constrained problem's solutions; however, it does not provide any information regarding the value of the penalty parameter ϵ. Such aspect will be discussed in the following paragraph.

2.3.2 Unconstrained portfolio selection model

In order to reformulate our portfolio selection model and to give a better understanding of the functioning of the selected penalty function, we start from the above-mentioned original constrained problem and we progressively transform it into the desired form.

The constrained problem is expressed as follows:

$$
\begin{aligned}
\underset{X,Z}{\text{minimize}} \quad & EVaR_{1-\alpha}(R) \\
\textit{subject to} \quad & \hat{R} \geq \pi \\
& \sum_{i=1}^{N} x_i = 1 \\
& K_d \leq \sum_{i=1}^{N} z_i \leq K_u \\
& z_i d_i \leq x_i \leq z_i u_i, \qquad \textit{with } i = 1, \dots, N \\
& z_i \in \{0, 1\}, \qquad \textit{with } i = 1, \dots, N
\end{aligned}
$$

Employing the exact penality method, it is possible to reformulate constraints as follows:

- $\hat{R} \geq \pi \quad \rightarrow \quad \max\{0, \pi - \hat{R}\} = 0;$
- $\sum_{i=1}^{N} x_i = 1 \quad \rightarrow \quad \left|\sum_{i=1}^{N} x_i - 1\right| = 0;$
- $K_d \leq \sum_{i=1}^{N} z_i \quad \rightarrow \quad \max\{0, K_d - \sum_{i=1}^{N} z_i\} = 0;$
- $\sum_{i=1}^{N} z_i \leq K_u \quad \rightarrow \quad \max\{0, \sum_{i=1}^{N} z_i - K_u\} = 0;$
- $z_i d_i \leq x_i \quad \rightarrow \quad \sum_{i=1}^{N} \max\{0, z_i d - x_i\} = 0;$
- $x_i \leq z_i u_i \quad \rightarrow \quad \sum_{i=1}^{N} \max\{0, x_i - z_i u\} = 0;$
- $z_i \in \{0, 1\} \rightarrow \sum_{i=1}^{N} |z_i(1 - z_i)| = 0.$

Then, the unconstrained optimization problem becomes:

$$\min_{X,Z,\epsilon} P(X, Z; \epsilon)$$

where:

$$P(X,Z;\epsilon) = EVaR_{1-\alpha}(R) + \frac{1}{\epsilon}[\max\left\{0, \pi - \sum_{i=1}^{N} x_i \hat{r}_i\right\} + \left|\sum_{i=1}^{N} x_i - 1\right|$$
$$+ \max\left\{0, K_d - \sum_{i=1}^{N} z_i\right\} + \max\left\{0, \sum_{i=1}^{N} z_i - K_u\right\}$$
$$+ \sum_{i=1}^{N} \max\{0, z_i d - x_i\} + \sum_{i=1}^{N} \max\{0, x_i - z_i u\} + \sum_{i=1}^{N} |z_i(1-z_i)|]$$

with ϵ as the penalty parameter.

It is possible to observe the presence of two distinguishing aspects in the application of this method, both strictly connected. In first place, we note the reformulation of constraints, differentiating between their typologies: constraints in the risk minimization problem expressed in equality form are reported in the penalty function in absolute value, while constraints in inequality form are re-written in order to have an element represented by a numeric value equal to zero, with the aim of facilitating potential violations survey. In particular, cases of double inequality have been considered into the penalty function as divided inequalities, separating the initial constraint into two parts, without altering the substance[37]. In second place, it is important to note the nature of objects that are penalized: the function penalizes not the constraints themselves, but rather their violations. Each addendum in square brackets refers to one or more constraints, but it is formulated so that it is able to take into account the potential disregard of one of the conditions set by the constraints system. In other words, in the case in which every constraint is respected, the algebric sum of elements within square

[37] For instance, instead of considering the inequality $a \leq x \leq b$, we implement separately $a - x \leq 0$ and $x - b \leq 0$.

brackets will be null, whereas, if at least one constraint is not respected, such sum will have positive value and will be added to the objective function.

Ultimately, if all constraints are observed, the constrained minimization process focuses exclusively on the measure of risk – EvaR in our case –, otherwise it will provide a greater value, given by the sum of the risk measure and the extent of the violation encountered on one or more constraints.

In conclusion, it is significant to mention that the biunivocal correspondence between solutions of the constrained minimization problem and the ones deriving from the minimization of the problem reformulated without constraints, depends on the appropriate choice of the penalty parameter ϵ.

Now that we have obtained an unconstrained objective function, we should present and describe in detail the metaheuristic algorithm that will be used in this dissertation.

Chapter 3

Particle Swarm Optimization

In this chapter *Particle Swarm Optimization (PSO)* is discussed. PSO algorithm belongs to the so-called bio-inspired metaheuristics. Some basic knowledge of heuristics, metaheuristics and swarm intelligence is provided in order to better describe the environment in which we are operating.

3.1 Heuristics and Metaheuristics

In many optimization problems, from combinatorial optimization to mathematical programming, we can find a variety of exact methods to compute the optimal solution. In some cases, however, problem's features such as real-world requirements, size or limited computational time, make exact methods not appropriate for finding the solution. In these scenarios, practitioners and researchers adopt approximate methods, which require less time and are known as *heuristics* and *metaheuristics*.

Definition. *Heuristic.* It means "to discover" or "to find" by trial and error. Heuristic methods give a suitable solution to a complex optimization problem in a reasonable time, even though not assuring the discovery of optimal solutions.

They are suitable when there is not the necessity to have the best solutions but there is the need to have sufficiently good solutions in a reasonable amount of time. A heuristic is usually an iterative algorithm, which, at each iteration, searches for the new best of the best solutions found until that instant. When the given stopping criterion is satisfied, the algorithm stops and gives an optimal solution, the best one determined from all the iterations performed. Heuristics are usually problem-specific since they are developed exploiting the properties of the problem to solve a specific problem. In order to find a more general solution to these problems, in the last years metaheuristic methods have been developed.

Definition. $Metaheuristic.$ It is a further development of heuristic methods. The term was coined by Glover in 1986: *meta* means “beyond” or “high level”, while *heuristic* “to find” by trial and error, as mentioned above. In Glover’s and Laguna’s words, a metaheuristic is a *“master strategy that guides other heuristics to produce solutions beyond those that are normally generated in a quest for local optimality”* (Laguna M. and Glover F. 1999).

In fact, guidelines and strategies provided by metaheuristics’ problem-independent techniques have the aim of developing heuristic methods that can be adapted to fit the needs of most real-world optimization problems. The ability to find “good enough” solutions in a time that is “small enough” has made metaheuristics the method of choice to solve the majority of large real-life optimization problems, both in practical applications and in academic research.

Two main features which determine the behaviour of any metaheuristic algorithm are:

- *Intensification or exploitation*: to focus the search in a local region knowing that a current good solution is found in that region;
- *Diversification or exploration:* to generate different solutions in order to explore the search space on a global scale.

In an optimization problem, a good metaheuristic should provide a balance between exploitation and exploration to individuate regions with high quality solutions.

Based on exploitation and exploration, metaheuristics can be gathered in two different classes:

- *Trajectory methods:* metaheuristics that start from an initial solution and describe a trajectory in the search space: at each step of the search, a new solution – often the best with respect to the alternative ones found in its neighbourhood until that moment – replaces the current solutions. These algorithms are called *exploitation-oriented* since they allow to find locally optimal solutions, fostering intensification in the research space. The main metaheuristic families belonging to this class are Simulated Annealing (SA), Tabu Search (TA) and Variable Neighbourhood Search (VNS);

- *Population methods:* metaheuristics that consider a population of solutions. Initial populations are created randomly and then improved through an iterative process: at each step of the process the entire population, or just a part of it, is substituted by newly generated elements, which are often the best ones. Given their main feature of diversification in the search space, these algorithms are called *exploration-oriented.* The main metaheuristics families belonging to this class are Evolutionary Algorithms (EAs), Artificial Bee Colony (ABC), Ant Colony Optimization (ACO), Differential Evolution (DE) and Particle Swarm Optimization (PSO).

To conclude the definition of metaheuristic, we outline its fundamental features:

- It is a strategy that guides the search process;
- Its goal is to explore the search space in order to find the best near optimal solution;
- Metaheuristic algorithms can range from simple local search procedures to complex learning processes;
- They are approximate and usually non-deterministic;
- They might include mechanisms to avoid getting trapped in limited areas of the search space;
- They are not problem-specific;

- They might employ domain-specific knowledge[38] in the form of heuristics that are managed by the upper level strategy;
- Todays most advanced metaheuristics take into consideration search experience (incorporated in some form of memory) to guide the search.

3.2 A bio-inspired metaheuristic: PSO

Particle Swarm Optimization is a bio-inspired, population-based metaheuristic. The initial idea underlying particle swarms, as originally conceived by Kennedy and Eberhart (1995), had the aim of creating artificial intelligence (AI)[39] exploiting analogies with social interaction and individual cognitive abilities. It was initially thought to graphically simulate the choreography formed by birds' flocks in flight; however, it was soon discovered that it was also a valid tool for solving mathematical optimization problems.

Swarm intelligence systems are usually constituted by a population of agents able to accomplish simple actions, known as *particles*. Each particle:

- Represents a possible solution belonging to the set of feasible solutions;
- Interacts both with other particles and the surrounding environment;
- Is provided with a *position* (location in the search space) and a *velocity.*

The solution to an optimization problem through the use of this approach derives from the agents' social interaction: singularly, they do not possess the capacity to face the problem; while as a group (swarm) and with the deriving collaboration they are able to find a solution.

[38] Heuristics are usually problem-specific. While dealing with a domain, a certain heuristic, which works specifically in that domain, may be used.
[39] Artificial Intelligence (AI) can be defined as the science and engineering of making intelligent machines. The concept of AI is linked to computational science, which is the study of the design of intelligent agents. The latter can be described as a system that perceives its environment and takes actions to maximize its chance of success.

A swarm can be identified by three characteristics:

- *Robustness:* if for any reason an agent exits from the swarm (a solution is too far away from the others), the remainder of the group is in any case able to reach their task;
- *Flexibility:* with the employment of the same interactions rules between individuals it is possible to solve problems of different nature;
- *Self-organization:* there is no need for a supervisor that assigns the tasks; from simple rules it derives a complex and well-structured behaviour.

The analysis of swarms' behaviour, taking into account the previous three aspects, has led to the development of effective algorithms.

From a qualitative point of view, the research of the optimal solution, if discovered (being a metaheuristic, it is not always the case), is declined in the following way. A previously fixed number of particles is randomly positioned in the search space[40] and each particle is assigned with a random velocity. Each swarm member explores a specific zone of the search area, keeping track of the best position reached – the most promising – and exchanging this information with the neighbourhood particles. How much a determined position is valid as solution to the optimization problem, is measured by the so-called *fitness function*. Since information gradually spreads between all members, we can expect that at the end the whole swarm converges towards the best position globally identified between the ones explored by each single agent.

Connecting the reasoning to the animal world, a bird that locates a source of food may move away from the group to reach it, showing its individualism; alternatively, it might remain in the flock, showing a more social character.

[40] It is favourable that the particles are sufficiently spread out in the search area, in order to avoid them being trapped in local optimum points.

Translating this behaviour into mathematical terms, it is possible to describe the research strategy as a balance between an exploration phase and an exploiting one: the former is linked to the individualism of the single member that search the solution far from the swarm, the latter is related to the successful information exchanged from the other individuals.

To give a quantitative and more formal representation of the process, we can formulate the algorithm in the following way.
Let M be the number of particles – thus, the number of possible solutions to the problem – and N the number of variables of the problem, then, each of the M particles is constituted by a point in the N-dimensional space. At the k-th algorithm's step, each particle will be characterized by the following three N-dimensional vectors:

- $\boldsymbol{x_j^k}$: represents the current **position** of the j-th particle at step k;
- $\boldsymbol{v_j^k}$: represents the current **velocity** of the j-th particle at step k;
- $\boldsymbol{p_j}$**:** represents the best position ($pbest$) reached so far by the j-th particle;
- $\boldsymbol{p_g}$**:** represents the best position (*gbest*) reached globally between all particles.

Moreover, $pbest_j = f(p_j)$ denotes the fitness function's value observed in the best personal position by the j-th particle until that step. At each algorithm's iteration, current position x_j^k of each particle is considered as possible solution to the optimization problem. In case fitness function's value associated to the current position is the best one registered until that moment by the same particle, the position is saved in vector p_j and the value of the fitness function associated is known as $pbest_j$. The final aim consists in researching new positions able to improve as much as possible the value of the fitness function. Since in this work PSO is treated in general terms as tool for the solution to a generic non-constrained minimization problem, when referring to "best" values of the fitness function it is appropriate to expect that the function assumes the smallest values if facing a minimization problem; conversely, the greatest values if facing a maximization problem.

(Eberhart J., Kennedy R.C. 1995) proposed a simple algorithm executable through few rows of code. To find the solution, it requires the specification of the problem in addition to the setting of some parameters that are soon going to be discussed more in details.

The original algorithm develops as follows:

1. Initialize randomly a population of particles, in terms of initial position, initial velocity and best position visited in the search space;
2. Start of the loop (iteration):
   2.1. For each particle evaluate fitness function $f(x_j^k)$ in correspondence of the current position x_j^k;
   2.2. Compare the observed value with $pbest_j$: if the former is better, the best personal position p_j needs to be updated with the current one x_j^k;
   2.3. Identify particles in the neighbourhood with the best fitness value among all and assign it to a vector p_g called $gbest$ and defined as the vector constituted by the best global positions, not only local ones;
   2.4. Update position and velocity of all particles following the system of equations:

$$\begin{cases} v_j^{k+1} = v_j^k + U(0,\phi_1) \otimes \left(p_j^k - x_j^k\right) + U(0,\phi_2) \otimes (p_g - x_j^k) \\ x_j^{k+1} = x_j^k + v_j^{k+1} \end{cases}$$

where:

- $U(0, \phi_1), U(0, \phi_2) \in \mathbb{R}^n$ and their components are uniformly distributed respectively in intervals $[0; \phi_1]$ and $[0; \phi_2]$[41], where ϕ_1 and ϕ_2 are acceleration coefficients and will be discussed later;
- The operator $\otimes$ is the tensor product[42];
- p_g is the best position in the neighbourhood of the j-th particle;

2.5. If the given stop criterion, usually a maximum number of iterations or a determined value of the fitness function, is reached, go to step 3; otherwise update the iterations counter and go back to step 2.1;

3. End of the loop.

Referring to the first equation:

- v_j^k is the current velocity, accountable for making the particle move in the same direction it was headed before;
- $U(0, \phi_1) \otimes (p_j - x_j^k)$ is called *cognitive component*. It can be thought of as particle's memory, reason for which the particle tends to return to the areas of the search space in which it has experienced best individual fitness values. $U(0, \phi_1) \in [0, \phi_1]$ is a random number with uniform distribution;
- $U(0, \phi_2) \otimes (p_g - x_j^k)$ is called *social component*. It can be seen as the exchange of information that coming from the other particles and causes the particle to move to the best area the swarm has found until that moment. $U(0, \phi_2) \in [0, \phi_2]$ is a random number with uniform distribution.

---

[41] $\mathbb{R}^n$ is the objective function domain.

[42] The tensor product is a bilinear operator, a function that combines two vectorial spaces in the same field to produce an element of a third vectorial space, which is linear in every argument.

3.2.1 Parameter selection

When executing particle swarm algorithm, several aspects has to be taken into consideration in order to facilitate convergence and prevent dispersion – or explosion – of the swarm. These aspects include the selection of certain parameters. The original version of PSO has a restrained number of parameters that has to be determined: the first one is constituted by the *number of particles* and is determined prevalently given the level of algebraic complexity of the problem to solve. Blackwell et al. (2007) highlighted that numbers between 20 and 50 are the most widespread in practice. The other parameters that need to be set are present in the velocity update equation and are *acceleration parameters* and *maximum velocity*.

Acceleration parameters

The two acceleration parameters ϕ_1 and ϕ_2 are respectively known as *cognitive coefficient* and *social coefficient*. They influence PSO's behaviour since they determine in a relevant portion the intensity of forces that attract each particle, at each iteration, respectively towards its best position found in the past $\left(p_j - x_j^k\right)$ and towards the best position visited previously by all particles in the neighbourhood $(p_g - x_j^k)$. Changes in ϕ_1 and ϕ_2 value can make PSO more o less responsive and even unstable when particles' speed increases without control: small values restrict the movement of particles, while big values might cause the divergence of particles. It has to be highlighted that the sum of ϕ_1 and ϕ_2 should not be greater than 4 and they should not have equal value, since the weights of the personal and group coefficients differ according to the characteristics of the problem. Given the acceleration parameters' influence on the particles' velocity, the latter needs to be managed as well.

Maximum velocity

Velocity of particles is updated at each iteration of the algorithm. It is a stochastic variable and has, thus, the characteristic to create an uncontrolled trajectory, making particles undertake wider cycles in the search space. With the aim of avoiding these “explosions” or, on the contrary, a strong convergence, upper and lower bounds for velocity has been set and can be briefly described as follows:

$$v_{jn} = \begin{cases} -V_{max}, & if\ v_{jn} < -V_{max} \\ V_{max}, & if\ v_{jn} > V_{max} \\ v_{jn}, & else\ if \end{cases}$$

The general method for preventing explosions is to determine a maximum velocity parameter V_{max} to block velocity from exceeding it on each dimension n for each particle j. Usually value V_{max} is chosen empirically based on the features of the problem. It is important to underline that this constriction allows particles to fluctuate between the bounds without the tendency of the swarm to collapse towards a point or to scatter too much in the search space.

This constraint solution is object of some criticism since the bound parameter V_{max} has shown to have influence on the balance between exploration (cognitive aspect) and exploitation (social aspect). Coherently, there have been proposed some alternative solutions to improve the algorithm, that are going to be explained in the following paragraph.

3.2.2 Adjustments of PSO

As said earlier, despite being one of the most relevant method to solve an optimization problem, PSO has the shortcoming of premature convergence. Indeed, converging too fast to a certain point may cause the swarm to be trapped into a local optimum and it would not allow to explore other promising regions. Consequently, when dealing with complex problems, PSO might fail to find the global optimum. To overcome the issue of premature convergence, improving algorithm’s performance, PSO as originally stated has to be modified: in this paragraph we will describe the most relevant adjustments.

Inertia Weight Approach (IWA)

Motivated by the willing to improve the solving capacity of the metaheuristic, limiting as much as possible – until eliminating – the influence of maximum velocity parameter V_{max}, Shi and Eberhart (1998) introduced an important modification to the velocity updating equation through the insertion of a new parameter called *inertia weight* (w^k). This parameter has the objective of managing the exploration and exploitation capabilities of the swarm and to allow the swarm to converge in a more accurate and efficient way with respect to the traditional PSO update equation.

The updating system become:

$$\begin{cases} v_j^{k+1} = w^k v_j^k + U(0,\phi_1) \otimes \left(p_j^k - x_j^k\right) + U(0,\phi_2) \otimes \left(p_g - x_j^k\right) \\ x_j^{k+1} = x_j^k + v_j^{k+1} \end{cases}$$

Inertia weight may be either implemented as a fixed or dynamic value. With $w = 0$, particles' precedent velocity goes out of the equation, meaning that all particles will move without knowledge of the previous velocity at each step. With $0 < w \leq 1$ particles will tend to change direction, while conversely, with $w \geq 1$ velocity will increase over time, hardly changing direction and the swarm will diverge and scatter. Even if for fixed values of w, PSO algorithm performs well, the employment of a dynamic inertia weight parameter is more common, due to its capacity of controlling and managing exploration and exploitation abilities.

When choosing a dynamic value for the parameter, a linearly decreasing inertia weight has empirically shown good results in optimization problems. Aiming at a good balance between exploration and exploitation and ultimately at finding an optimum, parameter w's value, which is linearly decreasing in time, is set with the following equation:

$$w = w_{max} - \frac{w_{max} - w_{min}}{K} \times k$$

where:

- w_{max} and w_{min} are pre-determined maximum and minimum values of the inertia weight parameter;
- K is the given possible maximum number of iterations, chosen by the user;
- k is the current number of iterations.

The use of inertia weight parameter to control the balance between cognitive and social aspects and to ensure convergence results valid; nevertheless, once the parameter has decreased, it is not able to recover the exploration mode – hence increase again – if, for instance, the swarm needs to search in new areas. Contextually, other methods to control inertia weight behaviour have been adopted. For example, Blackwell et al. (2007) found that PSO's performance significantly improves with the employment of w using a fuzzy system. Other effective strategies, as well proposed by Blackwell et al. (2007), include the use of inertia weight with a random component or the use of an increasing inertia weight; both methods showed good results.

Constriction Factor Approach (CFA)

A widespread alternative to the introduction of inertia weight which allows to achieve the triple objective of controlling the convergence of particles, prevent explosiveness of velocity and eliminate the troublesome presence of V_{max}, is the application of a *constriction coefficient* χ on the entire velocity update equation. Hence, it is a factor that facilitates particles' convergence operating on all equation's addenda, instead of influencing only previous velocity as inertia weight. The *constriction factor approach* was proposed by Clerc (1999) and the resulting velocity update equation is as follows:

$$\begin{cases} v_j^{k+1} = \chi[v_j^k + U(0,\phi_1) \otimes \left(p_j^k - x_j^k\right) + U(0,\phi_2) \otimes (p_g - x_j^k)] \\ x_j^{k+1} = x_j^k + v_j^{k+1} \end{cases}$$

where:

- $\chi = \frac{2}{\phi - 2 + \sqrt{\phi^2 - 4\phi}}$ is the equation that determines the constriction factor;
- $\phi > 4, \quad with\ \phi = \phi_1 + \phi_2.$

Given the constriction factor equation, this approach only works for ϕ values greater than 4. Under this approach, ϕ is usually set to 4.1 with $\phi_1 = \phi_2$; consequently resulting in a constriction factor $\chi \approx 0.7298$.

Fully informed Particle Swarm

The sources of influence on particles in the classical PSO version are prevalently two: personal best and global best. This means that for the j-th particle, the remaining information coming from the other particles in its neighbourhood are not exploited. Kennedy and Mendes (2002 and 2003) have proposed some modifications on how particles interact with their neighbours, introducing a new variant of PSO, the so-called *Fully Informed Particle Swarm (FIPS)*. While in the original algorithm each PSO particle is influenced by its own previous performance and the single best position of its neighbours, the fully informed particle swarm establish that the particle is influenced by all of its neighbours. FIPS will then be formulated as follows:

$$\begin{cases} v_j^{k+1} = \chi[v_j^k + \frac{1}{M_j} \sum_{m=1}^{M} U^k \otimes (p_{nbr_m^k} - x_j^k)] \\ x_j^{k+1} = x_j^k + v_j^{k+1} \end{cases}$$

where:

- M_j is the number of neighbours for particle j;
- nbr_m is particle j's $m - th$ neighbour.

Comparing traditional PSO with FIPS, it has been shown that the latter appears to find better solutions in relatively fewer iterations than the former. However, FIPS has a main relevant shortcoming: it is much more dependent on *population topology*. This feature of the metaheuristic is the subject of the next paragraph.

3.2.3 Population topology

Population topology defines how particles interact with each other. The nature of communication between particles is affected by the adopted topology structure, hence, it significantly influences their behaviour. Possible topologies are divided into two groups: static and dynamic. In the latter case, particles belonging to the neighbourhood change at each iteration.

Considering the fact that in the traditional version of this metaheuristic, topology was conceived as static, we can notice how also this aspect of the algorithm was object of modifications and evolutions during the years. First topologies were based on proximity between particles within the search space. However, this communicative structure presented some feature of convergence less desirable, along with being computationally demanding. The following development was oriented toward a double direction, defining two general kinds of neighbourhoods:

- *Global best (gbest):* particles are influenced by the best solution found by any member of the swarm. It corresponds to a fully connected network in which each individual has access to information of all the other individuals in the swarm;
- *Local best (lbest):* each individual has access to information exchanged by its immediate neighbours, according to some swarm topology.

The main topologies are the following three, with the first two belonging to the local best neighbourhood and the third one belonging to the global best neighbourhood:

- *Star topology:* in this kind of topology there is only one central particle that is connected to all the others: all the information is exchanged through it. This specific particle compares the best positions of all members of the swarm and update its position towards the particle that has found the best position so far. Afterwards, the central particle’s new position is communicated to all the members.
- *Ring topology:* each particle exchanges information to exactly two other particles, creating a single pathway, defined as a ring. With this topology the particle that finds a better solution than its two neighbours, send them the information which is going to be spread to their immediate neighbours and so on until it reaches the last particle. Following this approach, the solution is spread slowly around the pathway by all members. However, despite the slower convergence, larger areas of the search space are covered than with the previous topology.
- *Fully connected topology:* in this topology, the source of social information to a particle is the best-performing element in the whole population. All the particles are linked with each other and can receive information from every other member of the population. It has of course a global best kind of neighbourhood.

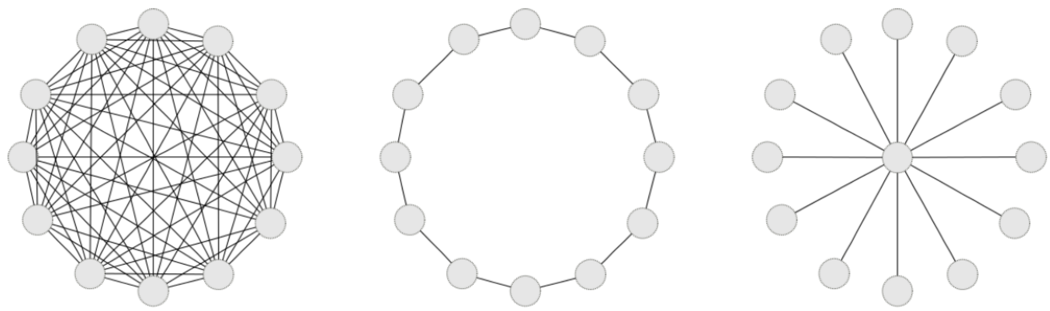

Figure 1. Fully connected or global best (left), Local best or Ring (middle) and Star (right) topology. Ugolotti et al. (2019)

In conclusion of this segment dedicated to PSO, it is considered relevant to highlight what explained by Blackwell et al. (2007) with regard to the current absence of theoretical results able to describe in a didactic way the foundations of this process. Such unavailability is mainly attributable to four reasons. In first place, PSO is constituted by a large number of elements that interact between themselves but, in spite of the relatively easy comprehension of particles' nature and of iterations, it cannot be said the same for the dynamics that involve the swarm in its totality. In second place, the concerns regard particles' memory and their ability to update it. This means that between iterations a particle could be attracted towards a new personal best, towards a global best or towards a balance of the two given by a function that depends on both aspects. The third cause is the fact that forces operating on particles have stochastic nature, impeding the employment of mathematical tools usually used in the analysis of dynamic systems. Lastly, knowing that PSO behaviour depends on the structure of the fitness function and being the choice of possible objective function very wide, to provide crosswise applicable results is highly complicated.

Chapter 4

Application on FTSE MIB and discussion

In this chapter we will apply the model introduced in chapter 2. We will consider real data, in particular daily closing prices of a set of selected assets on FTSE MIB index. The time length of historical data considered is 5 years: from July 2014 to December 2018.

4.1 Preliminary information

With the aim of assessing the capability of our portfolio selection model, we provided an application on real data. In order to keep the computation light and less time-consuming, just a set of FTSE MIB assets has been considered in our analysis. Moreover, to maintain our set representative of the index and well diversified across sectors, we have considered one asset for each sector that composes the index. FTSE MIB is the main benchmark index of Italian securities markets and it comprehends 40 companies of primary importance and with high liquidity in the different sectors.

Our set is constituted by 10 securities, which are listed in the following table.

FTSE MIB

SECTOR	SECURITY
Industrials	Atlantia
Healthcare	Diasorin

Utilities	Enel
Energy	Eni
Consumer staples	Fiat Chrysler Automobiles
Financials	Generali
Consumer discretionary	Juventus Football Club
Technology	Stmicroelectronics
Materials	Tenaris
Telecom	Telecom Italia

Table 1. Subset of FTSE MIB securities and relative sector.

In the present analysis, historical data of assets' closing prices, from 1^{st} July 2014 to 28^{th} December 2018, are considered and divided into several subsets. In particular, 3 different periods are studied, each divided into two sections: a so-called *in-sample* part of 12 months and a so-called *out-of-sample* part of 6 months. The first section of data will be employed, with the exploitation of the metaheuristic algorithm, to estimate the solution to our optimization problem; while the second one, also known as *virtual future*, will be employed to verify the accuracy of the estimation provided by the first set of data.

The general concepts behind this choice is that portfolio's risk and expected return of the out-of-sample period should follow approximately the trend of the in-sample period on which the model is based. Following the reasoning, it is assumed that the optimal percentages of investment on each asset, obtained by the in-sample-based model, produce the best portfolio model also for the near-after periods. Hence, it is possible to consider being in the last day of the in-sample period and having the will to invest a certain amount of wealth for the following 6 months, exploiting the optimal percentages of investment provided by the in-sample observations analysis. Finally, comparing the results of this artificial investment with the expectations at the end of the twelfth month, it is possible to evaluate the effectiveness of the method.

The three periods and their sections are divided as follows:

Period N.	In-sample	Out-of-sample
1	July 2014 – June 2015	July 2015 – December 2015
2	January 2016 – December 2016	January 2017 – June 2017
3	July 2017 – June 2018	July 2018 – December 2018

Table 2. Periods analysed.

As regards to returns, logarithmic returns are preferred over percentage returns. Daily log returns for each asset considered are computed from the observed daily close prices with the following formula:

$$r_{i,t} = \ln\left(\frac{p_{i,t}}{p_{i,t-1}}\right)$$

4.2 Problem specific parameters setting

First of all, before applying the model to real data, it is important to define and describe the parameters introduced as inputs in the model: both with regard to the portfolio selection problem's constraints and to the PSO algorithm.

With what concerns the portfolio model constraints, the chosen parameters for our application are the following:

- The desired minimum daily return is set equal to the portfolio mean return over the in-sample period, considering assets as equally weighted: $\pi =$

$\sum_{i=1}^{N} \hat{r}_i$, with $\hat{r}_i$ being $i-th$ asset's mean return over the in-sample period and $\hat{r}_i = \frac{1}{T}\sum_{t=1}^{T} r_{i,t}$;

- The minimum number of assets that can be held in the portfolio is: $K_d = 3$;
- The maximum number of assets that can be held in the portfolio is equal to all securities considered in this work: $K_u = 10$;
- The minimum percentage of investment in each asset is: $d = 2\%$;
- The maximum percentage of investment in each asset is: $u = 20\%$;

As mentioned in the introduction of this dissertation, results from our mean-EvaR portfolio selection model will be discussed and compared with results from the same model, but substituting EVaR risk measure with the more common Expected Shortfall (or CVaR) one. Parameters that need to be introduced for such risk measures are:

- The level of significance, common to both measures of risk, is defined by: $\alpha = 95\%$;
- The arbitrary parameter t of EVaR measure is set as advised by the author Ahmadi-Javid (2017) in its paper on portfolio optimization with the employment of EVaR: $t = 1$.

The choices concerning assets number and percentage of investment constraints are made with the intent of maintaining an appropriate minimum level of diversification within the portfolio.

4.3 PSO parameters setting

It is decided to employ inertia weight to slow down pre-mature convergence of the algorithm. Taking into consideration what suggested by literature, the following PSO parameters are considered:

- Number of particles: 50;
- Number of iterations: 1000;
- Inertia weight: $\omega = 0.7298$;
- Individual acceleration coefficient: $\phi_1 = 1.49618$;
- Social acceleration coefficient: $\phi_2 = 1.49618$;
- Penalty parameter: $\epsilon = 10^{-7}$.

The penalty parameter ϵ is relevant, since it guarantees correspondence between the original constrained optimization problem and the unconstrained one.

4.4 Application, comparison and discussion

As mentioned before, in order to give more significance to our results and comparisons, the application of our model is implemented in three different periods. In each period, the portfolio selection model is applied individually with both EvaR and ES as risk measures. The metaheuristic algorithm is run 5 times for each measure of risk; each run starting from different random positions and velocities. After the evaluation of constraints violations and the comparison between fitness values, the percentages of investment from the best run are taken as approximation of portfolio's optimal shares. Optimal results obtained in this way are then applied to logarithmic returns of the respective out-of-sample set of observations. Performances of the portfolio deriving from the mean-EVaR portfolio model will then be analysed and compared to the results computed with expected shortfall and to a benchmark. The benchmark is represented by an equally weighted portfolio composed by the same 10 assets.

Period 1

The first period considered is comprehended between 1st July 2014 and 30th December 2015. The in-sample section is constituted by the first 12 months, while the out-of-sample is represented by the last 6 months of 2015.

PSO algorithm has been launched on the in-sample observations five times: the outputs of the computations are displayed in the following table.

OUTPUTS	RUN 1	RUN 2	RUN 3	RUN 4	RUN 5
Best fitness value	8.506679	0.051333	1.092878	0.181599	2.000e+06
Num. of selected assets	7.999999	5	8.000000	8.000000	4
Budget constraint	9.008e-07	0	1.259e-08	2.405e-09	0.031330
Return constraint	0	0	0	0	0
Min. num. of assets constraint	0	0	0	0	0
Max. num. of assets constraint	0	0	0	0	0
Min. investment % constraint	2.848e-09	9.595e-20	1.879e-11	5.694e-11	1.178e-19
Max. investment % constraint	0	0	6.103e-10	2.284e-11	0.168669
z constraint	8.336e-07	4.797e-18	9.092e-08	1.308e-08	1.243e-17

Table 3. PSO outputs on mean-EVaR portfolio selection problem.

In the description of the functioning of the penalty parameter, in chapter 2, it has been explained that constraints values are equal to zero or near zero if there are no violations in the process. As we can see from Table 3, the majority of the five runs' constraints are indeed equal to zero or significantly near to zero, with the exception of Run 5 in which budget and maximum percentage of investment constraints are too relevant to be approximated to zero. As a proof of that, we can observe that the constraints violations lead to a higher and less favourable fitness value.

Having to deal with a minimization problem, it is important to choose the lowest fitness value; therefore, for this period we have selected Run 2 – with a best fitness value of 0.05 – as best solution.

The same procedure has been followed for the portfolio selection problem adopting expected shortfall as risk measure. However, with the purpose of not weighing the reading down, we just provide the best solution between the five runs. In the following table it is possible to observe the optimal weights for each asset given by the employment of the two risk measures.

Security	EVaR	ES
	Optimal weights	
Atlantia	0.200000000000000	0.120291542835261
Diasorin	0.200000000000000	0.125935424522847
Enel	0.200000000000000	0.063566196597037
Eni	0.000000000000000	0.029938581101640
FCA	-0.000000000000000	0.184483652072605
Generali	0.200000000000000	-0.000000000000000
Juventus	0.000000000000000	0.101720658347452
Stmicroelectronics	0.000000000000000	0.166996514683626
Tenaris	-0.000000000000000	0.164902854228434
Telecom	0.200000000000000	0.042164575807293
	Number of assets held in the portfolio	
	5	9

Table 4. Assets' optimal weights.

We simulated an investment with a capital of € 10.000 in the relative out-of-sample period. The investment is made on the same 10 assets three different times: one with EVaR optimal weights, one with ES optimal weights and one with equal weights (benchmark). In the following graph we can observe the behaviour of each portfolio in time.

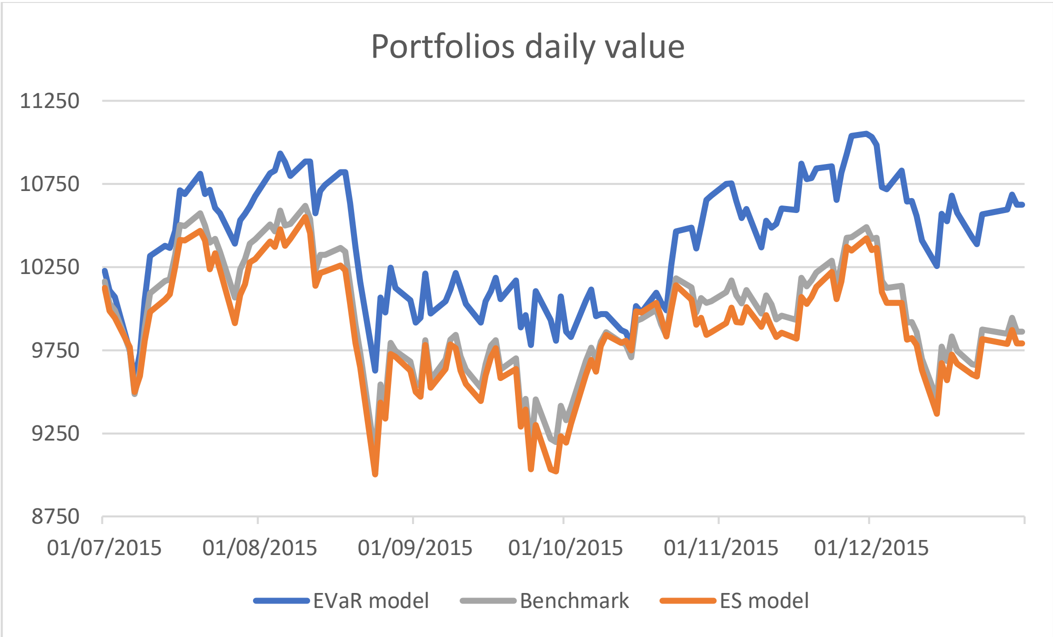

Figure 2. Period 1 – Portfolios daily values from an initial investment of €10.000.

As it can be seen, our portfolio seems to perform significantly better than both the portfolio computed with ES and the benchmark. Given that EVaR portfolio is constituted by 5 assets with respect to the 9 and 10 of ES portfolio and benchmark respectively, the first portfolio may have invested in riskier assets with a higher expected return. To study more in details the three portfolios, two important indexes are provided.

Table 5. Performance ratios.

Portfolio	*Sharpe ratio[43]	*Sortino ratio[44]
EVaR	0,03070348	0,04450178
ES	-0,006831	-0,01332989
Benchmark	-0,0098515	-0,00938271

**Risk-free rate is set equal to zero.*

Period 2

The second period considered is comprehended between 4th January 2016 and 30th June 2017. The in-sample section is constituted by the entire year 2016, while the out-of-sample is represented by the first 6 months of 2017.

PSO algorithm has been launched on the in-sample observations five times: the outputs of the computations are displayed in the following table.

Table 6. PSO outputs on mean-EVaR portfolio selection problem.

OUTPUTS	RUN 1	RUN 2	RUN 3	RUN 4	RUN 5
Best fitness value	3.802e+03	0.051389	2.206e+03	2.785e+03	0.051391
Num. of selected assets	7.999996	6.999999	5	7.999999	5.999999
Budget constraint	2.651e-07	1.887e-14	0	2.397e-08	1.075e-12
Return constraint	3.717e-04	0	2.206e-04	2.785e-04	0
Min. num. of assets constraint	0	0	0	0	0
Max. num. of assets constraint	0	0	0	0	0
Min. investment % constraint	1.158e-09	3.480e-15	8.017e-22	2.702e-11	0
Max. investment % constraint	4.237e-08	3.005e-14	2.775e-17	3.369e-09	2.602e-13
z constraint	8.265e-06	1.261e-12	4.226e-20	3.626e-08	2.422e-12

[43] Sharpe ratio is the average return gained in excess of the risk-free rate per unit of standard deviation (volatility). It is defined by the following formula: $SR = \frac{R_P - R_f}{\sigma_P}$, where $R_P$is the portfolio return, R_f is the risk-free rate and σ_P is portfolio return's standard deviation.

[44] Sortino ratio differentiates from Sharpe ratio by considering only the standard deviation of negative returns – *downside deviation* – instead of standard deviation.

We can observe that the majority of constraints values are equal to zero or really near to zero, whereas violations of the constraint on the minimum desired return are relatively significant in runs 1, 3 and 4. It is decided to consider Run 2 as best solution to the optimization problem, since it has the lowest best fitness value, slightly lower than Run 5's one. Between the runs with the lowest fitness values, Run 2's portfolio is also more diversified since it considers more assets.

In the following table we display the optimal weights for each asset given by the employment of the two risk measures.

Security	EVaR	ES
	Optimal shares	
Atlantia	-0.000000000000000	0.000000000000000
Diasorin	0.193366378398991	0.198972706272491
Enel	0.055942842714043	-0.000000000000000
Eni	0.178643610305862	0.000000000000000
FCA	0.000000000000000	0.104093925230620
Generali	0.195293341217322	0.186940704323086
Juventus	0.000000000000000	0.000000000000000
Stmicroelectronics	0.191657999722274	0.195328909026544
Tenaris	0.165073018393120	0.122130423236571
Telecom	0.020022809248410	0.192533331921531
	Number of assets held in the portfolio	
	7	6

Table 7. Assets' optimal weights.

The same investment is simulated on this period's out-of-sample section. The following graph shows the behaviour of each portfolio in time.

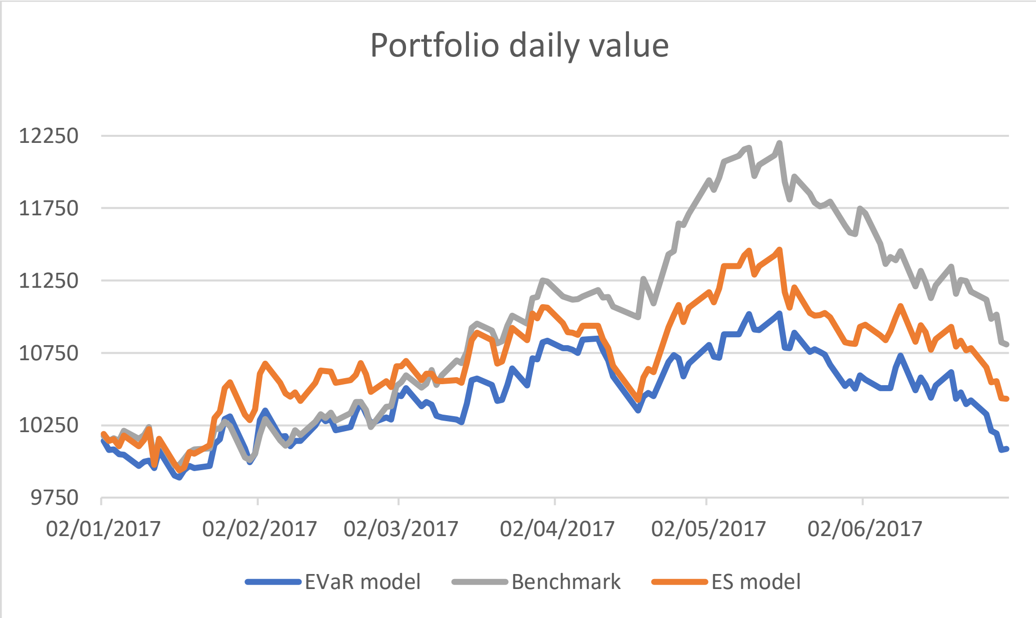

Figure 3. Period 2 – Portfolios daily values from an initial investment of €10.000.

From Figure 3 we can observe how EVaR portfolio seems to perform worse than the others. An explanation could be the caution that EVaR has as a risk measure: being an upper bound for VaR and CVaR, it should measure risk higher than the other two measures. In other words, a portfolio selection model with EVaR will choose less riskier investments with consequently lower expected returns. This could explain the general worse performance of EVaR portfolio. To obtain a clearer vision on the portfolios we are considering, we provide two important indexes: Sharpe ratio and Sortino ratio.

Table 8. Performance ratios.

Portfolio	Sharpe ratio	Sortino ratio
EVaR	0,006872349	0,009638447
ES	0,034197827	0,048292247
Benchmark	0,065733318	0,095712466

Period 3

The third period considered is comprehended between 3rd July 2017 and 28th December 2018. The in-sample section is constituted by the first 12 months, while the out-of-sample is represented by the last 6 months of 2018.

PSO algorithm has been launched on the in-sample observations five times: the outputs of the computations are displayed in the following table.

Table 9. PSO outputs on mean-EVaR portfolio selection problem.

OUTPUTS	RUN 1	RUN 2	RUN 3	RUN 4	RUN 5
Best fitness value	0.051375	1.367e+02	1.483044	0.053499	2.483e+03
Num. of selected assets	5	5.999999	7.999999	7.000000	5
Budget constraint	0	3.273e-12	1.289e-07	2.368e-11	0
Return constraint	0	1.367e-05	0	0	2.483e-04
Min. num. of assets constraint	0	0	0	0	0
Max. num. of assets constraint	0	0	0	0	0
Min. investment % constraint	6.421e-20	1.733e-13	6.835e-12	1.287e-14	3.064e-20
Max. investment % constraint	2.809e-17	1.626e-12	0	9.370e-13	1.493e-19
z constraint	4.932e-18	4.424e-11	1.384e-07	2.112e-10	2.278e-18

As the previous period, the return constraint is the one whose violations are the more significant, in this case in Run 2 and 5. Being all other constraints values near to zero, it is decided to consider Run 1 solutions as optimal weights for our portfolio.

We show these results together with the ones obtained from the mean-ES portfolio selection model in the following table.

Security	EVaR	ES
	Optimal shares	
Atlantia	0.000000000000000	-0.000000000000000

Diasorin	0.200000000000000	-0.000000000000000
Enel	0.000000000000000	0.198546963465654
Eni	0.000000000000000	0.133550476495022
FCA	0.200000000000000	0.136388654254785
Generali	0.200000000000000	0.197511079147359
Juventus	-0.000000000000000	0.000000000000000
Stmicroelectronics	0.200000000000000	0.136628773787093
Tenaris	0.000000000000000	0.197374052850480
Telecom	0.200000000000000	0.000000000000000
	Number of assets held in the portfolio	
	5	6

Table 10. Assets' optimal weights.

The investment simulation with these optimal weights lead us to the following graph, where portfolios' behaviour is described during the 6 months out-of-sample period.

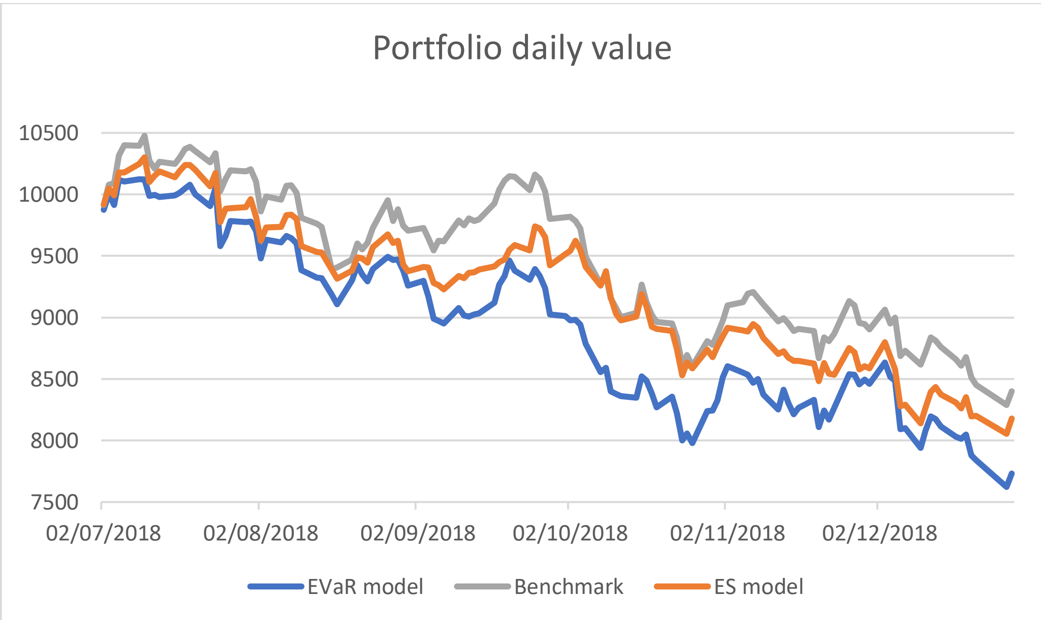

Figure 4. Period 3 – Portfolios daily values from an initial investment of €10.000.

Figure 4 shows a relevant loss for all three portfolios. In this situation, a more moderate EVaR measure should lose less as a consequence of choosing less riskier assets. However, it is observable that mean-EvaR portfolio is the one which encountered the worst loss. As explained in period 1 discussion, the cause may be related to the number of assets held in the portfolio – 5 out of 10. While in period 1 the choice of selecting only 5 specific assets resulted as a successful strategy, in this period the 5 assets selected faced more serious losses with respect to the entire population of 10 assets (benchmark).

Performance indexes are provided in the next table.

Table 11. Performance ratios.

Portfolio	Sharpe ratio	Sortino ratio
EVaR	-0,150118405	-0,1849375
ES	-0,134676284	-0,1655194
Benchmark	-0,110888982	-0,1396759

Conclusions

In the present dissertation a Particle Swarm Optimization algorithm has been applied to a complex portfolio selection model, considering Entropic Value-at-Risk as a measure of risk.

The problem was defined in chapter 2, considering several constraints that investors and fund managers could encounter in real world practice. Moreover, it has been decided to employ EVaR in defining a risk function, which needs to be minimized to obtain our best portfolio. In order to solve this constrained optimization problem in a reasonable amount of time, we relied on a bio-inspired metaheuristic algorithm, PSO, which have been discussed in chapter 3. It is worth mentioning that Particle Swarm Optimization is born as an unconstrained optimization method, hence it is usually necessary to make adaptations to deal with constrained optimization problems. We overcame this problem by employing a penalty parameter, which linearized our objective function while, at the same time, penalizing constraints violations.

In the last chapter we made the computations and applied the results to a portfolio of 10 real assets chosen from FTSE MIB index. To give more reliability to our application, we considered three distinct periods of time and implemented the algorithm several times to obtain a reasonably low fitness value and, thus, a good optimization.

Ahmadi-Javid (2012) introduced EVaR as a new coherent risk measure which can be defined as the tightest upper bound for Value-at-Risk and Conditional Value-at-Risk. We would have expected EVaR to be a more conservative and cautious measure, causing the portfolio model to choose less riskier investments and guaranteeing a more constant and less volatile return. However, we were proven wrong by the results of our work: in all three cases the EVaR model selected portfolios with a relatively small number of assets, compared to ES model and to the total number of assets (10), thus not exploiting the diversification’s property of lowering the overall portfolio risk. Moreover, portfolios computed with EVaR seemed to be riskier since they tended to over-react to market oscillations: when all three portfolios – the one computed with EVaR, the one computed with ES and the benchmark one – suffered a loss, EVaR portfolio was subject to the biggest loss; when all three portfolios obtained a profit, EVaR portfolio was the one to make the greatest profit.

In conclusion, despite the computational study showed that the EVaR approach to the portfolio selection model resulted in very different portfolios, EVaR is a coherent risk measure and was proved to own appropriate properties to be considered as a promising risk measure both from a computational and a financial standpoint. Relevant future studies should involve the introduction of new risk measures that could outperform the current measures adopted in real practice and represent better the risk borne by investors and practitioners.

Appendix

In this section we provide the developed Matlab code, adopted for this dissertation.

<u>Code applied to the EVaR-based portfolio selection model:</u>

```
clc;
format long

%% Input of the problem

% Uploading historical data and computing logarithmic returns
[prices] = importdata("file_name.xlsx");
[t,n_asset] = size(prices); % t is the period of time, n_asset is
the number of assets considered
returns = log(prices(2:end,:)./prices(1:end-1,:)); % log returns
C = 10000; % capital invested
% Differentiation between in-sample and out-of-sample
oos = 127; % number of returns out-of-sample +1
ris = returns(1:t-oos,:); %returns in-sample
ros = returns(t-oos+1:(t-1),:); %returns out-of-sample
rm_is = nanmean(ris); % In-sample mean returns
rm_oos = nanmean(ros); % Out-of-sample mean returns
% Data input
pi = mean(rm_is); % desired daily minimum return, set equal to the
portfolio mean return over the in-sample period
alpha = 0.95; % significance level of EVaR
Kl = 3; % minimum number of assets
Ku = 10; % maximum number of assets
l = ones(1,n_asset)*0.02; % minimum percentage of investment in each
asset
u = ones(1,n_asset)*0.20; % maximum percentage of investment in each
asset
t_evar = 1; % constant t linked to EVaR computation, set arbitrarily
equal to 1
% PSO parameters inizialization
P = 50; % particles number
niter = 1000; % iterations number
c1 = 1.49618; % individual acceleration coefficient
c2 = 1.49618; % social acceleration coefficient
iw = 0.7298 ; % inertia weight
vmaxx = zeros(1,n_asset);
vmaxz = zeros(1,n_asset);
```

```
epsilon = 1e-07; % parameter that penalizes violations of
constraints
% Creation of vectors useful for objective function
% Risk measure
EVaR_p = zeros(P,1); % vector of EVaR values for each particle
(portfolio)
rmp_is = zeros(P,n_asset); % P x n_asset matrix


% Constraints
constr_1 = zeros(P,1); % budget constraint
constr_2 = zeros(P,1); % desired minimum return constraint
constr_3 = zeros(P,1); % minimum number of asset constraint (z>=Kl)
constr_4 = zeros(P,1); % maximum number of asset constraint (z<=Ku)
app_5 = zeros(P,n_asset);
constr_5 = zeros(P,1); % minimum percentage of investment constraint
(x>=l)
app_6 = zeros(P,n_asset);
constr_6 = zeros(P,1); % maximum percentage of investment constraint
(x<=u)
app_7 = zeros(P,n_asset);
constr_7 = zeros(P,1); % z is either 0 or 1




%% Computation


% 1-Generation of position and velocity vectors and setting of
fitness
% function
x = rand(P,n_asset);
vx = rand(P,n_asset);
z = rand(P,n_asset);
vz = rand(P,n_asset);


f = ones(P,1)*1.0e+255; % fitness function
x1 = zeros(P,n_asset); % matrix which state if the asset is in the
portfolio (x*z)
% pb=pbest: vector with the best position of particles in previous
iterations
pbx = [x f];
pbz = z;


% g=gbest: vector with the best global position and the associated
objective function's value
gx = zeros(1,n_asset+1);
gz = zeros(1,n_asset);
```

```
% Beginning of the loop
tic; % measuring time spent in the computation
for k=1:niter
% Identifying dynamic range for maximum velocity
for i=1:n_asset
vmaxx(i) = abs(max(x(:,i))-min(x(:,i)));
vmaxz(i) = abs(max(z(:,i))-min(z(:,i)));
end


% 2-Objective function computation
for p=1:P
for i=1:n_asset
x1(p,i) = x(p,i)*z(p,i);
app_5(p,i) = max(0,l(i)*z(p,i)-x(p,i));
app_6(p,i) = max(0,x(p,i)-u(i)*z(p,i));
app_7(p,i) = abs(z(p,i)*(1-z(p,i)));
end
% Computing assets' weighted mean return for particle p
rmp_is(p,:) = x(p,:).*rm_is;
% Computation of historical EVaR for particle p
EVaR_p(p) = t_evar*(log(sum(exp(rmp_is(p,:)*t_evar^-1))/n_asset)-
log(alpha));




% Sum of investment percentages equal to 1
constr_1(p) = abs(sum(x1(p,:))-1);
% Expected return at least equal to pi
constr_2(p) = max(0,(pi-sum(x1(p,:)*rm_is')));
% Minimum number of assets Kl
constr_3(p) = max(0,Kl-sum(z(p,:)));
% Maximum number of assets Ku
constr_4(p) = max(0,sum(z(p,:))-Ku);
% Minimum percentage l
constr_5(p) = sum(app_5(p,:));
% Maximum percentage u
constr_6(p) = sum(app_6(p,:));
% z is either 0 or 1
constr_7(p) = sum(app_7(p,:));


end


% Objective funciton
f =
(EVaR_p+(1/epsilon)*(constr_1+constr_2+constr_3+constr_4+constr_5+co
nstr_6+constr_7));
```

```
% 3-Comparing objective function's value with pbest
for p=1:P
if f(p)<pbx(p,n_asset+1)
pbx(p,n_asset+1) = f(p);
for i=1:n_asset
pbx(p,i) = x1(p,i);
pbz(p,i) = z(p,i);
end
end
end




% 4-Identifying the particle with the best position (g)
[minimum,position] = min(pbx(:,n_asset+1));
gx(n_asset+1) = minimum;
for i=1:n_asset
gx(i) = pbx(position,i);
gz(i) = pbz(position,i);


end


% 5-Updating velocity and position
for p=1:P
for i=1:n_asset
vx(p,i) = iw*vx(p,i)+c1*rand*(pbx(p,i)-x(p,i))+c2*rand*(gx(i)-
x(p,i));
vz(p,i) = iw*vz(p,i)+c1*rand*(pbz(p,i)-z(p,i))+c2*rand*(gz(i)-
z(p,i));


if vx(p,i)>vmaxx(i) % maximum velocity limitation
vx(p,i) = vmaxx(i);
end
if vz(p,i)>vmaxz(i) % maximum velocity limitation
vz(p,i) = vmaxz(i);
end


x(p,i) = x(p,i)+vx(p,i);
z(p,i) = z(p,i)+vz(p,i);


end
end
converg(k,:) = gx(:,end);

% 6-Go back to Step 2 until stop condition
end
% End of loop
```

```
% Results of the optimization problem
optimum_shares = gx(1,1:n_asset)'
best_fitness = gx(1,n_asset+1)

rmp_in_sample = rm_is.*optimum_shares';%assets' weighted mean
returns in sample
rmp_out_of_sample = rm_oos.*optimum_shares'; %assets' weighted mean
returns out of sample
EVaR_oos = t_evar*(log(sum(exp(t_evar^-
1*rmp_out_of_sample(:)))/n_asset)-log(alpha)); %historical EVaR out
of sample


%% Output of the problem


%Results
n_selected_assets = sum(z(position,:))
constr_budget = constr_1(position)
constr_return = constr_2(position)
constr_Kl = constr_3(position)
constr_Ku = constr_4(position)
constr_min_share = constr_5(position)
constr_max_share = constr_6(position)
constr_z = constr_7(position)


EVaR_is = EVaR_p*C;
EVaR_in_sample = EVaR_is(position)
rm_in_sample = sum(rmp_in_sample)
rm_out_of_sample = sum(rmp_out_of_sample)
EVaR_oos = EVaR_oos*C


toc;

% Graph comparing the behaviour of the fitness function (Y axis) in
relation to the number of iterations made (X axis)
plot(converg)
```

Code applied to the ES-based portfolio selection model:

```
clc;
format long


%% Input of the problem


% Uploading historical data and computing logarithmic returns
[prices] = importdata("3rd_period.xlsx");
[t,n_asset] = size(prices); % t is the period of time, n_asset is
the number of assets considered
returns = log(prices(2:end,:)./prices(1:end-1,:)); % log returns
C = 10000; % capital invested
% Differentiation between in-sample and out-of-sample
oos = 127; % number of returns out-of-sample +1
ris = returns(1:t-oos,:); %returns in-sample
ros = returns(t-oos+1:(t-1),:); %returns out-of-sample
TT = (t-oos);
rm_is = nanmean(ris); % In-sample mean returns
rm_oos = nanmean(ros); % Out-of-sample mean returns


% Data input
pi = mean(rm_is); % desired daily minimum return, set equal to the
portfolio mean return over the in-sample period
alpha = 0.95; % significance level of EVaR
Kl = 3; % minimum number of assets
Ku = 10; % maximum number of assets
l = ones(1,n_asset)*0.02; % minimum percentage of investment in each
asset
u = ones(1,n_asset)*0.20; % maximum percentage of investment in each
asset
% PSO parameters inizialization
P = 50; % particles number
niter = 1000; % iterations number
c1 = 1.49618; % individual acceleration coefficient
c2 = 1.49618; % social acceleration coefficient
iw = 0.7298 ; % inertia weight
vmaxx = zeros(1,n_asset);
vmaxz = zeros(1,n_asset);


epsilon = 1e-07; % parameter that penalizes violations of
constraints
% Creation of vectors useful for objective function
% Risk measure
ES_port = zeros(P,1);
R_is = zeros(TT,P);
sorted_R_is = zeros(TT,P);
VaR = zeros(P,1);
```

```
% Constraints
constr_1 = zeros(P,1); % budget constraint
constr_2 = zeros(P,1); % desired minimum return constraint
constr_3 = zeros(P,1); % minimum number of asset constraint (z>=Kl)
constr_4 = zeros(P,1); % maximum number of asset constraint (z<=Ku)
app_5 = zeros(P,n_asset);
constr_5 = zeros(P,1); % minimum percentage of investment constraint
(x>=l)
app_6 = zeros(P,n_asset);
constr_6 = zeros(P,1); % maximum percentage of investment constraint
(x<=u)
app_7 = zeros(P,n_asset);
constr_7 = zeros(P,1); % z is either 0 or 1


%% Computation


% 1-Generation of position and velocity vectors and setting of
fitness
% function
x = rand(P,n_asset);
vx = rand(P,n_asset);
z = rand(P,n_asset);
vz = rand(P,n_asset);


f = ones(P,1)*1.0e+255; % fitness function
x1 = zeros(P,n_asset); % matrix which state if the asset is in the
portfolio (x*z)
% pb=pbest: vector with the best position of particles in previous
iterations
pbx = [x f];
pbz = z;


% g=gbest: vector with the best global position and the associated
objective function's value
gx = zeros(1,n_asset+1);
gz = zeros(1,n_asset);


% Beginning of the loop
tic; % measuring time spent in the computation
for k=1:niter
% Identifying dynamic range for maximum velocity
for i=1:n_asset
vmaxx(i) = abs(max(x(:,i))-min(x(:,i)));
vmaxz(i) = abs(max(z(:,i))-min(z(:,i)));
```

```
end

% 2-Objective function computation
for p=1:P
for i=1:n_asset
x1(p,i) = x(p,i)*z(p,i);
app_5(p,i) = max(0,l(i)*z(p,i)-x(p,i));
app_6(p,i) = max(0,x(p,i)-u(i)*z(p,i));
app_7(p,i) = abs(z(p,i)*(1-z(p,i)));
end
% Calculate portfolio returns for each particle (at its position)
R_is(:,p)=ris*x(p,:)'; % TTxP matrix
% Sort portfolio returns
sorted_R_is = sort(R_is); % TTxP matrix
% Store the number of returns
num_returns_is = numel(R_is(:,1));
% Calculate the index of the sorted return that will be VaR
VaR_index_is = ceil((1-alpha)*num_returns_is);
% Use the index to extract VaR from sorted returns
VaR(p) = -sorted_R_is(VaR_index_is,p);
% Calculate historical ES
ES_port(p) = -mean(sorted_R_is(1:VaR_index_is,p));



% Sum of investment percentages equal to 1
constr_1(p) = abs(sum(x1(p,:))-1);
% Expected return at least equal to pi
constr_2(p) = max(0,(pi-sum(x1(p,:)*rm_is')));
% Minimum number of assets Kl
constr_3(p) = max(0,Kl-sum(z(p,:)));
% Maximum number of assets Ku
constr_4(p) = max(0,sum(z(p,:))-Ku);
% Minimum percentage l
constr_5(p) = sum(app_5(p,:));
% Maximum percentage u
constr_6(p) = sum(app_6(p,:));
% z is either 0 or 1
constr_7(p) = sum(app_7(p,:));


end


% Objective funciton
f =
(ES_port+(1/epsilon)*(constr_1+constr_2+constr_3+constr_4+constr_5+c
onstr_6+constr_7));


% 3-Comparing objective function's value with pbest
```

```
for p=1:P
if f(p)<pbx(p,n_asset+1)
pbx(p,n_asset+1) = f(p);
for i=1:n_asset
pbx(p,i) = x1(p,i);
pbz(p,i) = z(p,i);
end
end
end




% 4-Identifying the particle with the best position (g)
[minimum,position] = min(pbx(:,n_asset+1));
gx(n_asset+1) = minimum;
for i=1:n_asset
gx(i) = pbx(position,i);
gz(i) = pbz(position,i);


end


% 5-Updating velocity and position
for p=1:P
for i=1:n_asset
vx(p,i) = iw*vx(p,i)+c1*rand*(pbx(p,i)-x(p,i))+c2*rand*(gx(i)-
x(p,i));
vz(p,i) = iw*vz(p,i)+c1*rand*(pbz(p,i)-z(p,i))+c2*rand*(gz(i)-
z(p,i));


if vx(p,i)>vmaxx(i) % maximum velocity limitation
vx(p,i) = vmaxx(i);
end
if vz(p,i)>vmaxz(i) % maximum velocity limitation
vz(p,i) = vmaxz(i);
end


x(p,i) = x(p,i)+vx(p,i);
z(p,i) = z(p,i)+vz(p,i);


end
end
converg(k,:) = gx(:,end);

% 6-Go back to Step 2 until stop condition
end
% End of loop
```

```
% Results of the optimization problem
optimum_shares = gx(1,1:n_asset)'
best_fitness = gx(1,n_asset+1)

R_oos=ros*optimum_shares;
sorted_R_oos = sort(R_oos); % TTxP matrix
num_returns_oos = numel(R_oos);
VaR_index_oos = ceil((1-alpha)*num_returns_oos);
VaR_oos = -sorted_R_oos(VaR_index_oos);
% Calculate historical ES
ES_oos = -mean(sorted_R_oos(1:VaR_index_oos));
rm_out_of_sample= gx(1,1:n_asset)*rm_oos';
rm_in_sample= gx(1,1:n_asset)*rm_is';


%% Output of the problem


%Results
n_selected_assets = sum(z(position,:))
constr_budget = constr_1(position)
constr_return = constr_2(position)
constr_Kl = constr_3(position)
constr_Ku = constr_4(position)
constr_min_share = constr_5(position)
constr_max_share = constr_6(position)
constr_z = constr_7(position)


rm_port_oos = gx(1,1:n_asset)*(mean(ros))';
ES_is = ES_port*C;
Es_in_sample = ES_is(position)
rm_in_sample = rm_in_sample
rm_out_of_sample = rm_out_of_sample
ES_oos = ES_oos*C


toc;

% Graph comparing the behaviour of the fitness function (Y axis) in
relation to the number of iterations made (X axis)
plot(converg)
```

Bibliography

Acerbi C., Bank J. "Spectral risk measures: A coherent representation of subjective risk aversion." *Finance 26*, 2002: 1443-1471.

Acerbi C., Tasche D. "Expected Shortfall: a natural coherent alternative to value-at-risk." *Economic Notes by Banca "Monte dei Paschi di Siena SpA"*, 2002: 379-388.

Acerbi D., Tasche C. "On the coherence of expected shortfall." *Journal of Banking and Finance 26* : 1487-1503.

Agnetis A. "Introduzione all'ottimizzazione non vincolata." *www.dii.unisi.it.* 2004. http://www.dii.unisi.it/~agnetis/ottnonvinc.pdf.

Ahmadi A., Javid A. and Pichler A. «An analytical study of Norms and Banach spaces induced by the entropic value-at-risk.» *Mathematics and Financial Economics. DOI: 10.1007/s11579-017-0197-9*, 2017.

Ahmadi A., Javid A. «Entropic Value-at-Risk: A New Coherent Risk Measure.» *Journal of Optimization Theory and Applications, vol. 155, Issue 3*, 2012: 1105-1123.

Artzner P., Delbaen F., Eber J.M., Heath D. «Coherent measures of risk.» In *Mathematical Finance 9*, 203-228. 2001.

Bahareh N., Zakree M., Nazri A. and Abdul S. «A survey: Particle swarm optimization based algorithms to solve premature convergence problem.» *Journal of Computer Science, vol. 10(9)*, 2014: 1758–1765.

Blum C. and Roli A. «Metaheuristics in combinatorial optimization: overview and conceptual comparison.» *ACM Computing Surveys 35*, 2003: 269-308.

Chen J. "Mean-variance analysis." *www.investopedia.com.* 2019. http: //www.investopedia.com/terms/m/meanvariance-analysis.asp.

Chen J. "Modern Portfolio Theory." *www.investopedia.com.* 2019. https://www.investopedia.com/terms/m/modernportfoliotheory.asp.

Chernoff H. «A measure of asymptotic efficiency for tests of a hypothesis based on a sum of observations, Ann. Math. Statist.» Vol. 23, Number 4, 493-507. 1952.

Clerc M. «The swarm and the queen: towards a deterministic and adaptive particle swarm optimization.» 1999. 1951–1957.

Constantinides G.M. and Malliaris A.G. "Portfolio Theory." In *Portfolio Theory, in Jarrow R.A.* Maksimovic V. and Ziemba W. T., 1995.

Corazza M., Fasano G. and Gusso R. "Particle swarm optimization with non-smooth penalty reformulatio, for complex portfolio selection problem." *Applied mathematics and computation vol. 224*, 2013: 611– 624.

Delbaen F. «Remark on the paper "Entropic value-at-risk: A new coherent risk measure.» *Barrieu P. Risk and Stochastics, World Scientific. arXiv: 1504.00640*, 2018.

Delbean F. *Coherent risk measures on general probability spaces.* Zurich: Eidgenossische Technische Hochschule, 2000.

Eberhart J., Kennedy R.C. "Particle Swarm Optimization." *Proceedings of the IEEE International Conference on Neural Networks - Conference Proceeding 4,*. 1995. 1942-1948.

Eberhart R. and Shi Y. «Empirical study of particle swarm optimization.» *in Proc. IEEE Congr. Evol. Comput. (EP), vol. 3*, 1999: 1945–1950.

Encyclopedia of Mathematics: Borel function. https://www.encyclopediaofmath.org/index.php/Borel_function.

Engelbrecht A. P. *Computational Intelligence: An introduction.* John Wiley and Sons, 2007.

Fumagalli T. "La teoria della selezione di portafoglio di Markowitz, Corso di Finanza, Università degli Studi di Bergamo." *http://www00.unibg.it.* . http://www00.unibg.it/dati/corsi/ ...(ProfessorFumagalli_BO_lezione20-La teoria della selezione di portafoglio di Markowitz.pdf).

Glover F. and Sorensen K. «Metaheuristics.» *Scholarpedia vol. 10, revision 149834*, 2015: 6532.

Grinold R. C. and Kahn N. *Active portfolio management: A Quantitative Approach for Providing Superior Returns and Controlling Risk.* Irwin Library of Investment & Finance, 1999.

Hu X. and Eberhart R. *Solving constrained nonlinear optimization problems with particle swarm optimization.* 2002.

Hu X., Eberhart R. and Shi Y. «Recent advances in particle swarm.» *in Proc. IEEE Congr. Evol. Comput, vol. 1*, 2004: 90–97.

Kamali, Samira. «"Portfolio Optimization using Particle Swarm Optimization and Genetic Algorithm".» *Journal of mathematics and computer science* 10 (2014): 85-90.

Kennedy J., Shi Y. and Eberhart R. C. *Swarm intelligence.* San Francisco: Morgan Kaufmann Publishers Inc, 2001.

Kumar G., Mohagheghi S. ,Hernandez J. C. and Del Valle Y. «Particle swarm optimization: Basic concepts, variants and applications in power systems.» *IEEE*, 2008: 171-195.

Laguna M. and Glover F. *Tabu search.* Boston, 1999.

Luenberger D. and Ye Y. "Linear and non-linear programming, 4th ed. Springer, Stanford University." *http://www.mat.uab.cat.* 2008. http://mat.uab.cat/~alseda/MasterOpt/310trialtext.pdf.

Markowitz H. M. "Portfolio selection." *Journal of Finance* 7 (1952): 77-91.

Mendes R. and Kennedy J. «Neighbourhood topologies in fullyinformed- and best-of-neighbourhood particle swarms.» *Proc. IEEE Int. Workshop Soft Computing in Industrial.* 2003: 45–50.

Mendes R. and Kennedy J. "Population structure and particle swarm performance." 2002: 1671–1676.

Mohan C. and Ozcan E. «Particle swarm optimization: Surfing the waves.» *IEEE Congress Evol. Comput., vol. 3*, 1999: 70-75.

Natarajan K., Pachamanova D., Sim M. "Incorporating asymmetric distributional information in robust value-at-risk optimization." In *Management Science 54*, by Pachamanova D., Sim M. Natarajan K., 573-585. 2008.

Parsopoulos E. and Vrahatis M.N. «Particle swarm optimization method in multiobjective problems, Proceedings ACM Symposium on Applied computing.» 603–607. 2002.

Picariello L.S. "Il modello di Markowitz e il Capital Asset Pricing Model." *www.performancetrading.it.* . http://www.performancetrading.it/Documents/LpPerformance/LpP_c_modello_markowitz.htm.

Pichler A. "A quantitative comparison of risk measures." *Annals of Operations Research n.254.* . https://doi.org/10.1007/s10479-017-2397-3.

Poli R., Kennedy J. and Blackwell T. *Particle swarm optimization.* 2007.

Quaranta A.G., Zaffaroni A. "Robust optimization of conditional value-at-risk and portfolio selection." *Journal of Banking and Finance 32*, 1999: 2046-2056.

Rachev S., Martin D.,Racheva B. and Stoyanov F. "Kurtosis - Investopedia, Stable ETL optimal portfolio and extreme risk management." *www.investopedia.com.* Edited by Physika Verlag. 2007. https://www.investopedia.com/terms/k/kurtosis.asp.

Rockafellar R.T., Uryasev S. «Optimization of conditional value-at-risk.» In *Journal of Risk*, 21-41. 1999.

Roli A. and Tollo G. «Metaheuristics for the portfolio selection problem.» *International Journal of Operations research vol. 5*, 2008: 13-35.

Rothlauf F. "Design of Modern Heuristics." *Springer-Verlag Berlin Heidelberg*, 2011.

Shi Y., Eberhart R. "A modifed particle swarm optimizer." *in Proc. Of the IEEE Congress on Evolutionary Computation*, 1998: 69–73.

Shi Y., Eberhart R. «Particle swarm optimization: Developments, applications and resources.» *in Proc. IEEE Congr. Evol. Comput, vol. 1*, 2001: 81–86.

Swagatam D., Ajith A. and Amit K. *Particle swarm optimization and differential evolution algorithms: Technical analysis, applications and hybridization perspectives,.* 2008.

Trubian M. "Complementi di ricerca operativa." *homes.di.unimi.it.* 2008. https://homes.di.unimi.it/~trubian/OttNonLineare.pdf.

Ugolotti R., Sani L. and Cagnoni S. «What Can We Learn from Multi-Objective Meta-Optimization of Evolutionary Algorithms in Continuous Domains?» *MDPI Open Access Journals, Mathematics*, 2019: 232.

Wolfram M. "Positive Definite Matrix." *http://mathworld.wolfram.com/.* . http://mathworld.wolfram.com/PositiveDefiniteMatrix.html.

Yang, X.S. "Nature-inspired Metaheuristic Algorithms." *Luniver Press, 2nd edition*, 2010.