Università Ca'Foscari Venezia

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

Final Thesis

Portfolio Optimization through Forward Looking Approach: an analysis of the COVID19 market crash

Supervisor
Ch. Prof. Marco Corazza

Graduand
Giacomo Pegorer
852570

Academic Year
2020 / 2021

Acknowledgments

First of all, I would like to deeply thank my supervisor, professor Marco Corazza, that helped and guided me during this intricate path with patient and useful suggestions. He has been always available to answer every question or doubt I had in the writing of the thesis.

Besides professor Corazza, I would like to thank my girlfriend Ambra, for all her love and patient. She always supported me during my academic path, encouraging me every time I felt loss. Then, my parents and my brother, who supported me in every decision I made, always wanted the best for me.

Last, but not least, I would like to sincerely thank my friend Andrea, for its incredible support and help during bad times.

Table of contents

pp

Introduction

Portfolio management and its optimization are one of the key topics of modern finance. The Modern Portfolio Theory introduce by Markowitz in 1952, revolutionized the theory and practices concerning the portfolio management at that time. Even though, in real-world scenarios, the model presents some important limitation that must be taken into consideration. As it will be explained in depth later, real-life portfolio selection is not easy as introduced by Markowitz for three main reasons. First, an appropriate measure of risk must be identified, and it has to be coherent with investors' risk attitude. It will be shown why the use of variance as a measure of risk violate the coherence properties. Second, set of constraints must be considered dealing with the construction of a portfolio. This means that constraints like the budget or the maximum number of trading stock ones must be taken into consideration. Moreover, constraints unrelated to the investor's choice, like transaction costs, can affect the final portfolio. So, dealing with coherent risk measures, financial market and investors rules and practiced lead to non-linear, non-convex and mixed-integer problems, also called NP-hard problems. In order to obtain solutions to these problems in a reasonable amount of time, researchers have developed different meta-heuristics algorithm, which are bio-inspired algorithms. This thesis will focus on the application on such algorithms introducing the implied volatility by vanilla options to prevent unexpected market crash.

So, the remainder of this research is structured as follows: first, in Section 1 provides an overview of the Modern Portfolio Theory, together with its main limits and some real-world improvements to be applied in order to render the model more realistic and effective in a real-world scenario.

Then, in Section 2 undertake the metaheuristics algorithms which have been used in this research. This Chapter introduces heuristics and metaheuristics, providing explanations of their origin, topology breakdown and operation, and explaining how they work in practice. In addition, a dedicated focus for the two algorithms used is presented.

In Section 3 the theory behind this work will be presented. An introduction of the Black-Scholes-Merton differential equation will be presented, and it will be important to understand the Black-Scholes-Merton pricing formula which a core argument when dealing with implied volatility. After the introduction of the implied volatility as a measure of future uncertainty the

Lastly, in Section 4, after the introduction of the dataset considered, PSO algorithms will be applied to obtain the optimal portfolio. Then, results obtained by the application of every algorithm to the dataset will be compared, not only by measuring expected return and variance, but also by comparing different risk-adjusted performance measures and the Maximum Drawdown in order to conduct a wider and deeper analysis.

Chapter 1

Portfolio selection problem

1.1 Markowitz model

In this chapter, an overview of the Modern Portfolio Theory started by Harry Markowitz[1] will be proposed: even if it presents some limits and it has been proposed almost 70 years ago, the model is still considered one of the keystones of the modern finance as it changed the approach to the investment world. Before its introduction investors' strategy was developed around finding the assets providing the maximum return and marginally caring about risk as it was supposed to be managed through diversification, but Markowitz demonstrated that this is not true. The overview will present the basic assumptions, limits and issues of the model, together with alternative risk measures and some improvement to it.

The portfolio selection process provided by Markowitz can be divided into three main stages:

1. Identification of a statistical tool used to measure the uncertainty of a determined investment;
2. Definition of a criterion used to select and divide, among all the possible investment, the efficient ones from the inefficient ones;
3. Choice of the optimal portfolio investment given the investor's risk aversion, maximizing her utility.

1.1.1 Assumption of the model

Even if, as just stated, the model is still considered one of the most revolutionary for the modern finance, it relies on strong assumption: investors will prefer higher returns than lower ones and the risk associated with these returns should be as less as possible. Stating this, given the main goal of the investor the maximization of the return of an investment, Markowitz is assuming that investors are rational, meaning that they will choose an asset which return will be higher than another one, or, alternatively, that will provide a lower risk given a target return.

[1] H. M. Markowitz (1952), "Portfolio Selection", Journal of Finance, Vol. 7 pp. 77-91.

Moreover, the market is considered frictionless, (e.g. no transaction costs, no taxes, no asymmetric information, etc.) and the single security can be divided indefinitely (meaning that an investor can buy whether quantity of any security she wants).

1.1.2 Formalization of the theorem

As just mentioned, the first stage aims to identify some tools useful in determining the risk-reward ratio: in its elaboration, Markowitz chose the expected value of the return as a proxy for the expected return of a security and the variance of the returns to measure uncertainty and risk.

In a very simple and straightforward way, the formalization for a single asset can be transposed to the portfolio case since its returns is defined as a linear combination of individual securities. Defining R_P as the random rate of return of a portfolio, r_i the expected return of the $i-th$ asset, σ_i^2 its variance and x_i the percentage of available wealth invested in that $i-th$ asset:

$$\mathrm{E}(R_P) = \sum_{i=1}^{n} r_i x_i = r_P \quad (1.3)$$

[2]

$$Var(R_P) = \sum_{i=1}^{n} x_i^2 \sigma_i^2 + \sum_{i=1}^{n} \sum_{j=i+1}^{n} x_i x_j \sigma_{i,j}$$

$$= \sum_{i=1}^{n} x_i^2 \sigma_i^2 + \sum_{i=1}^{n} \sum_{j=i+1}^{n} x_i x_j \sigma_i \sigma_j \rho_{i,j} = \sigma_P^2 \quad (1.4)$$

[3]

where $\sigma_{i,j}$ represents the covariance between the asset $i-th$ and $j-th$ and $\rho_{i,j}$ is the Pearson correlation coefficient[4]. As the variance is not a linear operator, measuring this for a portfolio may be a function of the covariance between the assets

[2] The formula can be also written in matrix notation as $E(r_p) = x'r$ where x' is the vector of assets' weights and r is the vector of assets' expected returns.

[3] In matrix notation the variance can be expressed as $\sigma_P^2 = x'Vx$ where x' is the vector of assets' weights and V is and N x N variance-covariance matrix.

and the correlation coefficient. Before this brilliant intuition by Markowitz, Williams in his "The Theory of Investment Value", developing the Dividend Discount Model pointed out that an investor should focus on the returns of an investment only, as its variance can be reduced through diversification due to the Jacob Bernoulli's law of large numbers[5]. But Markowitz demonstrated that the law of large numbers cannot be applied as assets are correlated each other's.

1.1.3 The Mean-Variance criterion

Once an appropriate measure for measuring return and risk has been introduced, in the second stage of the process of portfolio selection, Markowitz outlined a set of "rules" to distinguish and to divide efficient portfolios from inefficient ones: the *Mean-Variance criterion.* The M-V criterion can be defined as follow:

> *Definition.* Given two random variables *X* and *Y*, with mean μ_X and μ_y and variance σ_x^2 and σ_y^2 respectively, X dominates Y in Mean-Variance if and only if the following conditions are met together:

- $\mu_X \geq \mu_y$;
- $\sigma_x^2 \leq \sigma_y^2$;
- At least one of the prior inequalities is satisfied in a strong form.

Alternatively, the definition states that a portfolio dominates another one in M-V sense if it is impossible to obtain a greater return without increasing the variance or, as opposite, to lower the variance without giving up a certain return. Going further, plotting a curve passing through all the possible combination of assets forming the efficient portfolios, an efficient frontier will be created, describing the relationship between the expected return of a portfolio and its riskiness (variance or standard deviation). All the portfolios lying on it must be considered desirable for a rational investor. Across all the just mentioned portfolios, Markowitz' solution consists on finding the minimum variance one given a target return π set by the investor, which will depend on her risk aversion. Formalizing, Markowitz introduced the following problem:

[5] Bernoulli J., Ars Conjectandi, (Thurnisorium, Basil)

$$\min_{x} \sigma_P^2$$

$$\text{subject to} \begin{cases} r_P = \pi \\ \sum_i x_i = 1 \end{cases} \tag{1.5}$$

or, in matrix formulation:

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

$$\text{subject to} \begin{cases} x'r = \pi \\ x'e = 1 \end{cases} \tag{1.6}$$

where:

- x is an N – column vector of portfolio weights $x_1, \ldots, x_n$;
- V is the N x N variance-covariance matrix;
- r is an N – column vector of means returns for $r_1, \ldots, r_N$;
- e is a column vector of ones.

The minimization of the variance of the portfolio is subject to two constraints: the expected returns must be equal to a specific one chose by the investor and the capital available must be invested entirely. Another restriction is that $0 \geq x_i \geq 1$ for $i = i_1, i_2, \ldots, i_n$, avoiding the short-sales.

From a mathematical point of view, this is a programming problem with a quadratic objective function and linear constraints. To solve it, three assumption need to be taken: the variance-covariance matrix V needs to be nonsingular[6] and positive[7], and there must be at least two different returns across all the assets[8]. Following these conditions and assumptions the problem minimizes a convex function subject to a convex set of constraints. One of the classical approaches to

---

[6] An $n \times n$ matrix A is called nonsingular or invertible if there exists an $n \times n$ matrix B such that $AB = BA = I$, where I is the $n \times n$ *identity matrix.*

[7] An $n \times n$ matrix is said to be positive definite if the scalar $z'Mz$ is strictly positive for every non-zero column vector i of real numbers.

[8] If all the assets would have the same returns, the problem will be trivially solved investing the entire available wealth in the asset with the lowest variance.

solve linear programming problem is the Lagrange multipliers (see Markowitz, 1956), providing the following solution:

$$x = \frac{(\gamma V^{-1}r - \beta V^{-1}e)\pi + (\alpha V^{-1}e - \beta V^{-1}r)}{\alpha\gamma - \beta^2} \tag{1.7}$$

where:

- $\alpha = r^T V^{-1} r$
- $\beta = r^T V^{-1} e = e^T V^{-1} r$
- $\gamma = e^T V^{-1} e$

From the graphical point of view, the efficient frontier shape depends on the number of assets taken into consideration: as an example, for $N \geq 2$, the frontier is represented by a parabola in the M-V plane (while it is an hyperbola in the M-SD plane). Moreover, the presence of a risk-free asset modifies the shape of the frontier, moving the minimum variance portfolio on the vertical axis of the plane[9]. This condition will not occur in case of boundary constraints as it will be explained later in this section.

Figure 1 and 2 in Appendix A show the combination of different risk indicators and the presence of a risk-free asset frontiers.

1.1.4 Utility function

Given that it is possible to divide efficient portfolios from inefficient ones, real world investors have different risk propension, which lead to the third and last stage of the process: the maximization of the utility function[10] of a specific investor.

It can be proved that the expected utility framework is coherent with the Mean-Variance criterion in only one of the following mutually exclusive circumstances[11]:

---

[9] A risk-free asset is defined as an asset with a certain future return, so with $\sigma^2 = 0$. As stated in the previous reference, the portfolio on the vertical axis corresponds to the one with $\sigma_P^2 = 0$ and a certain return.

[10] The utility function U is a function on real numbers $\mathbb{R}$ defining potential wealth level, given real values. Using this function, it is possible to rank all the random wealth level given respective expected utility values associated. So, as an example, wealth x is preferred to wealth y if $E[U(x)] > E[U(y)]$.

[11] See Tobin (1958).

- the investor's utility function is a quadratic one;
- the joint probability distribution function of $R_1, \ldots, R_N$ is a multivariate elliptical one[12].

A particularly used function is the *Quadratic utility one*, defined as:

$$U(R_p) = R_p - \frac{a}{2} R_p^2 \tag{1.8}$$

where:

- R_p is a random variable describing the return of the portfolio;
- a is a positive coefficient[13] that reflects the risk aversion of the investor.

Moreover, the quadratic utility function presents a good approximation of other utility functions[14]
So, as introduced before, the portfolio maximizing the expectation of the investor's utility function will be chosen as investment. Stating this means, mathematically:

$$\begin{gathered} \max_{x} E\left[U(R_p)\right] \\ \text{subject to} \quad \sigma_p^2 = f(R_p) \end{gathered} \tag{1.9}$$

Graphically, the optimal portfolio corresponds to the tangency point between the efficient frontier and the indifference curve.

1.2 Limits of Modern Portfolio Theory

One of the main reasons behind the success of the Modern Portfolio Theory is that it provides an intuitive and not-so-complex procedure for portfolio selection. However, the model is not so used in practice by professionals, due to several limitation that investors encounter in real world investment like transaction costs[15]. In this section of the chapter an analysis of the fallacies of the model will be provided. The focus can be brought on three main topics:

---

12 An elliptical distribution is a probability distribution characterized by the property that its equi-density surfaces are ellipsoids.
13 By definition it must be strictly positive and the greater the coefficient, the higher the risk aversion.
14 See Kroll et al. (1984), Brandt and Santa-Clara (2006), Levy and Levy (2014), Markowitz (2014)
15 See, as example, Pogue (1970), Davis et al. (1990).

1. the operative choice of the target return;
2. the estimation risk related to the model's parameter;
3. the instability of mean-variance solutions.

But besides the just cited topics other issues will be introduced and discussed together with some possible solution to avoid or solve them.

1.2.1 The operative choice of the target return

As introduced, one of the key features of the MV theorem is the choice of the target return by the investor, by which it can be found the specific optimal portfolio across the efficient frontier. This characteristic can be seen from two point of view: it grants a great flexibility to the model but, on the other hand, it may be difficult for an investor to identity its "coherent" risk aversion coefficient leading to the wrong decision of choosing a too risky portfolio (with a higher variance) or to "leave money on the table".

To deal with such a problem three main procedures can be followed:

1. To select the tangency portfolio;
2. To maximize the investor's utility function;
3. To choose the maximum between $r_{1/N}$ and r_{min}

The first solution aims to make the selection process much easier with respect to ignoring the investor's estimated target return and referring to one of the core concepts of the CAPM, the fund separation theorem[16].

Following the model, the investment will be a mix between risk-free asset and portfolio of only risky assets, with the former investment to be a function of the investor's risk aversion and the latter to maximize the so-called *Sharpe ratio*[17] and to obtain a tangency portfolio between the efficient frontier and the Capital Market Line, as shown in Figure 3. The analytical expression of the tangency portfolio is the following:

$$x = \frac{V^{-1}(r - r_c e)}{e'V^{-1}(r - r_c e)} \tag{1.10}$$

---

16 See Tobin (1958)

17 The Sharpe ratio, defined as $\frac{r_p - r_f}{\sigma_p}$, is a measure of performance of an investment given its return r_p, its standard deviation σ_p and the market risk-free r_f. See Sharpe (1966).

However, it has been proved[18] that sample tangency portfolios perform poorly out-of-the sample, and they tend to present very high levels of turnover, leading to higher transaction costs.

Another possible is to directly *maximize the investor's expected utility function*, formally:

$$\begin{aligned} \min_{x} \quad & x'r - \frac{\lambda}{2}x'Vx \\ \text{s.t.} \quad & x'e = 1 \end{aligned} \tag{1.11}$$

The solution of this problem can be achieved using the Lagrangian as explained before, so:

$$x = \frac{1}{\lambda}V^{-1}\left(r - \frac{e'V^{-1}r - \lambda}{e'V^{-1}e}\right) \tag{1.12}$$

The optimal portfolio will be an efficient one, which will also lie on the highest indifference curve. Graphically, this portfolio represents the tangency point between the mean-variance efficient frontier and the highest indifference curve as shown in Figure 4.

The third way to achieve the same result is to fix the desired return to the maximum between the return expected from the equally weighted portfolio and the one expected from the global minimum variance one, as suggested by Kourtis[19]. Formally, this means:

$$\pi = max\left(r_{1/N}; r_{GMV}\right) \tag{1.13}$$

Even if this strategy seems too superficial both the portfolios have interesting properties: the equally weighted portfolio has been proved to often outperform the sample-based mean-variance optimal portfolios[20], and that it has similar out-of-

18 See, for example, DeMiguel et al. (2009)
19 Kourtis, A. (2015), A Stability Approach to Mean-Variance Optimization. Financial Review, 50, 301-330
20 See Bloomfield et al. (1977).

sample performance to the minimum-variance and the tangency portfolios obtained with Bayesian shrinkage methods[21]. In addition, it has been proved that, among 14 portfolio models based on seven empirical datasets, none of the models significantly outperformed the equally weighted portfolio in terms of Sharpe ratio, turnover or certain-equivalent return[22]. In a more recent publication, all the previous evidences have been confirmed as equal-weighted portfolio with monthly rebalancing outperforms the value- and price-weighted portfolios in terms of total mean return, certainty-equivalent return, four factor alpha and Sharpe ratio[23].

On the other side, the global minimum variance portfolio cannot be affected by the estimation errors related to the expected returns of the assets and it has been proved to yields better out-of-sample results than the tangency portfolio[24]. Many recent studies highlight the suggestion to invest into the GMV portfolio rather than on the tangency one[25]. Mathematically, it can be expressed as:

$$\min_{x} \sigma_P^2 \quad \text{s.t.} \sum_{i} x_i = 1 \tag{1.15}$$

or, in matrix formulation:

$$\min_{w} x' V x \quad \text{s.t.} \quad x' e = 1 \tag{1.16}$$

1.2.2 The estimation risk related to the model's parameter

As introduced, each security's return, variance and covariances are unknown and are usually estimated using historical data and their sample estimators: the sample mean and the sample covariance. However, mean-variance portfolios suffer the estimation errors of the parameters in a way in which the advantages of diversification are offsite. This in one of the reasons why the mean-variance

21 See Jorion (1991).
22 See DeMiguel et al. (2009) and DeMiguel, Nogales (2009).
23 See Plyakha et al. (2016).
24 See Chopra and Ziemba (1993).
25 See, e.g., Ledoit and Wolf (2003), and Jagannathan and Ma (2003).

framework is not so used by practitioners and its attractiveness is mostly confined in the academic world. The solution of such a problem is brought by using, as before, the equally-weighted portfolio (*1/N*) or the global minimum variance one (GMV) and this can be seen as one of the reasons of their better performance already explained. More on this practical limitations will be explained in Chapter 3

1.2.3 The instability of mean-variance solutions

In the previous paragraph the estimation risk problem has been introduced. One of the consequences of such risk is that it might be necessary to manage portfolio weights in a frequently rates, making them *unstable* over time. It can happen that, took a neighborhood of the optimal portfolio lying on the efficient frontier, there are many "statistical equivalent portfolios" with similar expected returns and standard deviation (or variance), but having a large or even completely different asset weights. In addition, a simple and small change in the parameters may bring to a massive portfolio revision and modification.

Such a problem is strictly linked to the transaction costs assumption: it has been proved that, even if mean-variance portfolios outperform equally-weighted in absence of transaction costs, the exact opposite happens if we consider the presence of these last ones[26].

In addition to the above criticisms, several others have been introduced and studied and a brief review of the literature will be presented.

1.2.4 Real world limitations

First, as introduced before, it has to be stated that the Modern Portfolio Theory is based on some practical unrealistic assumption: real world market is not frictionless, almost every transaction is associated with transaction costs and taxes must be paid. In addition, assets are not indivisible and often cannot event be bought at a number the investor want because of the presence of minimum or lot constraints. To cope with these issues mathematical constraints have been developed and they will be introduced later.

1.2.5 Diversification issues

The Markowitz theorem is strictly based on diversification, so adding different assets to the portfolio in order to reduce its variance, and so, the idiosyncratic risk.

[26] See DeMiguel et al. (2009) and Kourtis (2015)

Still, as the number of the assets included increases, the number of parameters to be estimated raise as well, leading to an always more complex variance-covariance matrix and increasing the estimation risk introduced before.

1.2.6 Assumption of normality

The model introduced by Markowitz describe the investor's portfolio through the first two moment of a statistical distribution, the mean and the variance. However, this condition is satisfied only if the returns follow an elliptical distribution, but it has been demonstrated that this condition doesn't hold for most securities. In particular, Cont (2001)[27] has presented a list of empirical facts emerging from the statistical analysis of asset returns, identifying the properties that commonly characterize a wide variety of markets and instruments. From this analysis, among all, two main characteristics against normality assumption has emerged: positive or negative asymmetry and heavy tails. To explain the two characteristics just mentioned the third and the fourth moment of any statistical distribution need to be introduced.

In an asymmetrical distribution, one tail can be longer than the other. The degree of asymmetry is measured by the third moment of the probability distribution, known as the skewness, and it can be computed as:

$$\text{Skewness } (X) = \frac{\sum_{i=1}^{N} \frac{(X_i - \mu_X)^3}{N}}{\sigma_X^3} \tag{1.17}$$

A Normal distribution presents a skewness equal to 0[28] but returns distributions have observed to be often positively skewed[29].

In addition to the degree of asymmetry, the other characteristics which defines a distribution is the kurtosis, the fourth moment of a distribution itself. Kurtosis gives a measure that significant deviations from the mean occur is greater than in the case of the Gaussian distribution. It can be computed as:

---

[27] Cont R., (2001), Empirical properties of asset returns: stylized facts and statistical issues, Quantitative Finance, Taylor & Francis Journals, 1(2), 223-236.

[28] In case of negative skewness, data said to be left-skewed, meaning that the left tail is longer than the right one while, as opposite, a positive skewness indicates a right-skewed data distribution, with right tail longer that the left one.

[29] Arditti, Fred D., (1971), Another Look at Mutual Fund Performance, Journal of Financial and Quantitative Analysis, 6(3), 909-912.

$$Kurtosis\ (X) = \frac{\sum_{i=1}^{N} \frac{(X_i - \mu_X)^4}{N}}{\sigma_X^4} \quad (1.18)$$

where X is a random variable defined in $\mathbb{R}$, μ_X is the mean, σ_X is the standard deviation and N is the number of observations. A Normal distribution presents a kurtosis equal to 0[30] and it has been observed that stock returns exhibit positive excess kurtosis[31].

1.2.7 Variance as a measure of risk

In the Mean-Variance model, as the name suggests, the risk involved in an investment is measured by the variance. The problem arises as the variance as a *statistical* measure and not a *financial* risk one: as it is symmetric it measures the dispersion of a security's return around its expected return, not distinguishing between positive or negative excess return. Moreover, it has been demonstrated[32] that investors treat positive and negative outcomes in a different manner[33].

Even if Markowitz introduced the variance as a measure of risk for its simplicity, he recognized its issues and proposed an alternative risk measure, *the semi-variance*, defined as:

$$Semi - Var\ (R) = \frac{1}{N} \sum_{i=1;\, r_j<\mu}^{N} \left(r_j - \mu\right)^2 \quad (1.19)$$

The fundamental characteristic of the semivariance, also called downside risk, is that it takes into account only returns below the expected return, solving the problem of the variance. However, as Markowitz underlined in his publication, due to the limited computational resources available at that time, he still chose to use the variance[34].

30 Distributions with positive excess kurtosis is called leptokurtic (heavy-tailed) while one with negative excess kurtosis is called platykurtic (light-tailed).
31 See, for example, Rao et al. (1989).
32 See Kahnemann and Tversky (1979)

34 See Markowitz (1959) and (1991)

1.2.8 Quadratic Utility assumption

As already mentioned, one basic assumption of the Modern Portfolio Theory is the assumption of the investor's quadratic utility function and, as stated before, it is consistent with the Mean-Variance criterion describing investors' preferences. However, as introduced in Borch (1963)[35] the Mean-Variance criterion can be used as an approximation, but it should be integrated with a utility function polynomial of third degree, including also the third moment of the distribution, namely the skewness, in the process. Indeed, in Hanoch and Levy (1970)[36] the authors confirmed what previously demonstrated about the quadratic utility, as the second order function requires bound to be implemented and it might not successfully explain the decision-making process of the investors. Moreover, the authors proposed the introduction of a cubic utility function which, by definition, is convex, monotonically increasing and takes into account decreasing degree of risk aversion.

1.3 Real-world model improvements

As discussed in the last section, the Modern Portfolio Theory is interesting and studied but it falls short in real-life investments as, in its original version, it includes only two constraints. To render the model more ductile and usable in real-world cases, more constraints, also mixed-integer constraints, can be added to the original, some of them used also in the application section of this research. The problem rising with the addition of further constraints is that the complexity of the problem generally increases, making the problem NP-complete[37] and its solution NP-hard[38] and it becomes necessary to use inexact techniques that will be introduced later to solve them in a reasonable amount of time.

1.3.1 Return constraint

The first constraint in the Markowitz model is the return one. It consists into requires a minimum return π to the investment. In the original theorem the

35 Borch K., 1963. Communications to the Editor — A Note on Utility and Attitudes to Risk, Management Science, 9(4), 697-700

36 Hanoch and Levy, (1970), Efficient Portfolio Selection with Quadratic and Cubic Utility, *The Journal of Business*, 43, issue 2, 181-89

37 NP-complete problems are difficult to solve as they don't admit polynomial algorithm (see Cook, 1971 and Levin, 1973)

38 NP-hard problems are optimization problems for which provably efficient algorithms do not exist, requiring exponential time to be solved using exact techniques. It can be said that they are as difficult to solve as the NP-complete problems.

constraint requires the return to be equal to π but, of course, it can be required to be at last equal to π. So, the return constraint can be expressed as:

$$\sum_{i=1}^{N} x_i r_i \geq \pi \tag{1.20}$$

where x_i denotes the share of capital invested in the $i-th$ security and r_i is the expected return of that particular $i-th$ security.

1.3.2 Budget constraint

Another constraint that Markowitz introduced in his 1952's model was the budget constraints, which requires to invest all the available wealth. Formerly, it can be expressed as:

$$\sum_{i=1}^{N} x_i = 1. \tag{1.21}$$

1.3.3 Cardinality constraint

A possible additional constraint taken into consideration is the cardinality one, which limits the number of assets for the investment. An indirect consequence of using the cardinality constraint is the easier transaction costs control as, given large dataset, the number on transaction will be so limited. The cardinality constraint can be expressed as:

$$K_d \leq \sum_{i=1}^{N} z_i \leq K_u \tag{1.22}$$

where

- K_d is an integer representing the minimum number of securities to be hold;
- K_u is an integer representing the maximum number of securities to be

hold;

- So $1 \leq K_d \leq K_u \leq N$
- z_i is a dummy variable which value can be:

$$z_i = \begin{cases} 1 & \text{if the i-th f asset } i (i = 1, \ldots, N) \text{ is to be hold,} \\ 0 & \text{otherwise} \end{cases}$$

1.3.4 Boundary constraint

As discussed previously, one of the requirements of the problem introduced by Markowitz was that at least two securities among all the selected ones have different returns. A boundary constraint forces the choice of percentage to invest in single security to be included in an interval which defines the minimum and the maximum investable. Alternatively, this means:

$$z_i \varepsilon_i \leq x_i \leq z_i \delta_i \quad \forall i \tag{1.23}$$

where ε_i and δ_i are respectively the lower and the upper bounds, and so $0 \leqq \varepsilon_i \leqq x_i \leqq \delta_i \leqq 1$.

1.3.5 Transaction costs

Buying and selling security means incurring in transaction costs, which has been proved to make the MV model inefficient if not considered[39]. Transaction costs can be considered in two different ways, the fixed ones, meaning that they are paid irrespective of the amount of cash traded, or variable which, as opposite, are related to the traded amount. During the years several researches have been published discussing common assumption which will be briefly explained.

One of the no-fixed transaction costs model is the *V-shaped variable transaction cost function* which introduced by Bertsimas and Pachamanova (2008).

During the years several research have been published discussing the optimal number of securities in a portfolio in presence of minimum transaction slot[40], fixed

[39] See Yoshimoto (1996)
[40] See Mansini and Speranza (1997)

costs of transaction[41] or concave transaction costs[42] for example. However, as this kind of constraint is outside the reach of this research, it has not been implemented for the portfolio optimization algorithm.

1.3.6 Round lot constraints

Another constraint which has been developed to deal with real world trading rules is the round lot one, which force every asset investment in portfolio to be the exact multiple of a minimum lot. Let:

$$x_i = y_i * l_i \quad i = 1, \dots, N$$

with x_i be the weights express in terms of of an investment, y_i be an integer variable and l_i the minimum lot. As it is beyond the scope of this research, this constraint has not been added to the original model.

1.3.7 Asset class constraints

In real world, investors face, in addition to the above problems, the exposure one. Exposure can be considered as the risk of losing part of the amount invested. Such a problem can be limited by investing in different asset classes like oil stocks, utility stocks, telco stocks, in other terms, by diversifying. Asset class constraints limit the proportion of the portfolio which can be invested in each class. As suggested by Cheng et al. (2000)[43] let $\Gamma_m, m = 1, \dots, M,$ be M sets of assets that are mutually exclusive, i.e. $\Gamma_\mathrm{i} \cap \Gamma_\mathrm{j} = \emptyset, \forall \mathrm{i} \neq \mathrm{j}$ and

$$L_m \leq \sum_{i \in \Gamma_m} w_i \leq U_m, \quad m = 1, \dots, M$$

with L_m and U_m be respectively the lower and the upper proportion limit.
As the assets chosen for this research are already diversified by class, this kind of constraint will not be added to the model.
In chapter 2 a complete and exhaustive presentation of the algorithm together with the implemented constraints will be presented.

41 See Brennan (1975)
42 See Konno and Wijayanayake (2001)
43 T.-J. Chang, N. Meade, J.E. Beasley, Y.M. Sharaiha, Heuristics for cardinality constrained portfolio optimisation, Computers & Operations Research, 2000, 27(13), 1271-1302

Chapter 2

Metaheuristics algorithms

In this chapter the metaheuristics for optimization Particle Swarm Optimization (PSO) and a Modified version of the Particle Swarm Optimization (MOPSO) are introduced and discussed: since they are both bio-inspired algorithm, a brief explanation of the theory behind them and an introduction to swarm intelligence will be provided. Then, the chapter will focus more in detail on the algorithms used and compared in this research.

2.1 Introduction

Optimization problems are useful in many fields, including finance: they consist into looking for the optimal value of a given set of variables or constraints. To solve them and to find optimal solution different *exact solution technique* has been developed and implemented[44] but there exist complex real-world problems which tends to be too complex and require too much computational effort to be solved: this kind of complexity may derive from characteristics like the size of the dataset, the availability of limited computational time or, as introduced before, the problem may be classified as *NP-hard*. To address such problems, approximate algorithms, namely heuristics and metaheuristics, have received great attention both from academically and practical side and have been studied in the last decades because their use a trade-off between the optimal solution, as intended for exact algorithms, and computational time needed. More specifically, the heuristics and metaheuristic methods give up a certain degree of certainty of finding the real optimal solution in order to reduce the computational time up to a reasonable one.

2.2 Heuristics and Metaheuristics

Introduced by Polya in his book "How to solve it" in 1945, the word heuristics derives from the Greek verb *heuriskein* which means "to find, to discover". In his

[44] See, for example, Aouni et al. (2005) and Peng et al. (2011).

book the author requires its students to construct an articulate though following a dictionary-like set of heuristics useful to deal with mathematical problems. As introduced before, heuristics refers to a group of algorithms implemented to solve, in a reasonable amount of time, complex optimization problems, approximating the final results in order to obtain a greater computational time. It can be clearly stated that their advantages are the easiness of implementation and their adaptability to the nature and the size of the problem.

Heuristics methods have not found a common classification among the academic researchers; however, an adequate one can be presented as follow:

- Constructive heuristics. These kind of techniques are used when the solution can be obtained selecting a subset of a given set of elements. The iterative cycle starts with an empty set, adding elements to the solution until the complete and final solution have been reached. This heuristics is composed by two different processes, the *initialization*, aiming to select the particle of the swarm to start from, and the *selection* which set the rules for the chose of the next element in the subset. Generally speaking, constructive heuristics are not time-consuming, but the complete solution might be too approximate and not good enough.
- Improvement heuristics. Differently from the constructive ones, these heuristics start from an arbitrary complete solution and then try to improve it applying small changes. Formally, at each step of the iterative cycle, the algorithm moves from the current complete solution to one of the neighbors in the search space S, defined as the set of possible solution. If the movement lead to an improvement of the complete solution, that point in the search space will become the new starting point for the next cycle. The algorithm stops when it is impossible to achieve a better complete solution, or when a specified value of the object function is obtained or if the set maximum number of iterative cycle is reached.
- Hyper-heuristics. These kind of algorithm have first been mentioned by Cowling and Soubeiga to try explaining the idea of "heuristics to chose heuristics". In order to solve the problem of choosing a heuristics to solve a problem, an algorithm able to identify the most suitable one for that particular problem. The identification process can be split into two phases: heuristic selection and move acceptance. During the first phase the heuristics is selected and applied while the move acceptance decides whether to accept or reject the solution obtained.

Apart from the just introduced classification, a particular kind of approximate algorithms emerged in the last 20 years and proved to be really effective to solve the problems just introduced, *metaheuristics* algorithms. Introduced by Glover (1986), the word metaheuristic adds the word *meta* meaning "beyond, on another level" to the word heuristic introduced before.

> *"A metaheuristic is formally defined as an iterative generation process which guides a subordinate heuristic by combining intelligently different concepts for exploring and exploiting the search space, learning strategies are used to structure information in efficiently near-optimal solutions."* [45]

Metaheuristics can be classified in many different ways, but the following ones can be considered interesting:

- *Nature inspired vs non-nature inspired*: one of the most clarifying distinction is the original inspiration of the algorithm. Nature inspired metaheuristics, like the PSO, simulates behaviors observable in nature. On the other hand, algorithms like the tabu search does not simulate such natural behaviors.

- *Trajectory based vs population based*: this classification concerns the number of solutions the algorithm can manage during its running cycle. The former is able to manage only one solution with a single agent tracking the path while the latter use a group of agents to find the best solution, the so-called global best, among all the solutions, namely local best.

- *Dynamic vs static object function*: another classification regards the way in which the algorithm uses its object function. Some metaheuristic algorithms keep the same object function defined at the beginning of the first iterative cycle fixed for all the cycles until the end of the optimization process. However, this may cause the algorithm to incur in local minima solution. To try to solve this issue, dynamic algorithm modifies the object function incorporating information taken from the past iterative cycles.

- *One vs different neighborhood structure*: most of the metaheuristic algorithms are structured to work on a unique neighborhood

[45] Osman and Laporte, 1996. Metaheuristics: A bibliography. Annals of Operational Research, 63, 513-628.

structure, so they do not modify the shape of the search space during the process. On the other hand, it exists algorithm like the Variable Neighborhood Search structured to varies and diversifies the search space.

- *Memory usage vs memory-less methods*: another important classification can be constructed by the use of adaptive memory through the search history. While the memory-less algorithms are structured to carry out a Markov process, determining the next step on the current condition, memory usage ones use techniques based on the past interactions. In particular, short-term memory algorithms use past moves, solutions and decisions to perform the next steps, while long term memory ones combine different parameters giving information on the past.

The focus of this research will be population-based algorithms, with a focus on a very promising branch of artificial intelligence, namely the swarm intelligence, which will be now introduced.

2.3 Swarm intelligence and PSO

Particle Swarm Optimization is a naturally inspired, population-based algorithm developed and presented by Eberhart and Kennedy (1995)[46] which it is heavily inspired by Heppner and Grenander (1990)[47] and Reynolds (1987)[48] scientific researches about flocks and birds looking for food. While the latter's goal was to produce a graphical simulation of the choreography created by the animals, Eberhart and Kennedy transformed it into a powerful optimization tool. One of the success of such metaheuristic algorithm was the always increasing number of NP-hard in various research ares problem and the fact that the model is derivative-free, so it doesn't need derivatives computation during the process.

Practically, PSO algorithm is based on a swarm of particles, which moves into the search space looking for the optimal[49] solution given the object function set at

[46] J. Kennedy and R. Eberhart, "Particle swarm optimization," *Proceedings of ICNN'95 - International Conference on Neural Networks*, Perth, WA, Australia, 1995, pp. 1942-1948 vol.4
[47] Heppner, F. and U. Grenander (1990). A stochastic nonlinear model for coordinated bird flocks. In S. Krasner, Ed., The Ubiquity of Chaos. AAAS Publications, Washington, DC
[48] Reynolds, C. W. (1987). Flocks, herds and schools: a distributed behavioral model. Computer Graphics, 2 1 (4):25-34.
[49] It might be good to recall that the optimal solution of an approximate techniques might be different (and so worst) from the one from one exact technique.

the beginning of the optimization process. To be more precise, the process starts with the generation of a predefined number of swarm particles in a random location of the search space. Each particle explores its surrounding area and record its best position and the swarm best position at the end of the exploration.

From a more technical and quantitative standpoint, a swarm can be defined as a group of N particles moving in the M-dimensional search space. Each agent of the swarm is characterized by three M-dimensional vectors:

- x_j^k, the vector representing current position of the $j-th$ particle at the $k-th$ iteration.
- v_j^k, the vector representing current velocity of the $j-th$ particle at the $k-th$ iteration.
- p_j^k, the vector recording the best position explored by the $j-th$ particle up to the $k-th$ iteration. The best value recorded for the whole swarm, and represented by the fitness value, is denoted by *Pbest.*

The original PSO algorithm is quite simple, both from a conceptual and a practical point of view. Here it can be found the description of its main steps:

1. Initialization of the particle swarm with random position in the search space and velocity
2. Begin loop:
   a. Compute the value of the object function $f(x_j^k)$ in the current position x_j^k. The valuation is computed for each particle.
   b. Compare the value of the object function of the current position with the best position $Pbest$. If $f\left(x_j^k\right) > Pbest$, update the best position p_j with the current position x_j^k, otherwise the cycle continues.
   c. Identify the particle with the best global object function value and define it as $Gbest$ and its position p_g^3.
   d. Update the position and the velocity of the particles according to the following equation:

$$\begin{cases} v_j^{k+1} = v_j^k + U(0,\phi_1) \otimes \left(p_j - x_j^k\right) + U(0,\phi_2) \otimes \left(p_g - x_j^k\right) \\ x_j^{k+1} = x_j^k + v_j^{k+1} \end{cases} \tag{2.1}$$

e. If one of the criterions is met (i.e. satisfactory fitness value or limit of cycles reached), exit loop.

3. End loop.

With particular reference to equation (2.1) note that:

- v_j^k is the current velocity of the particles;
- $U(0, \phi_i)$ is a vector of random numbers uniformly distributed between $[0, \phi_i]$;
- $\otimes$ is the component-wise multiplication;
- ϕ_1 and ϕ_2 are the acceleration coefficient which are set by the author. The latter influences the behavior of the swarm through the magnitude of the force attracting every particle toward its personal best p_j and the global best p_g.
- $(p_j - x_j^k)$ is the difference between the personal best position of the particle and its current position;
- $(p_g - x_j^k)$ is the difference between the personal best position of the particle and its current position.

From equation (2.1), it can be stated that position and velocity are the results of the influence of three main component:

- v_j^k, the current velocity or "inertia component". Its role is to maintain the direction of the particle as it was in the previous iteration, avoiding unexpected direction changes.
- $U(0, \phi_1) \otimes (p_j - x_j^k)$ which is the cognitive component, can be considered as a part of the memory of the particle. It serves to attract the particles to the search space where they have obtained the best personal fitness function. As to increase the variability, this cognitive component is multiplied by a random number chose from a uniform distribution in the interval $[0, \phi_i]$.
- $U(0, \phi_2) \otimes (p_g - x_j^k)$ which is the social component, can be considered as another part of the memory. As opposite to the previous element, it considers the whole swarm and make it move to the search region in which the best global position has been encountered. Again, as to increase the variability, also this component is multiplied by a random number chose from a uniform distribution in the interval $[0, \phi_i]$.

A simple graphical representation of the update of velocity and position is reported in figure 5.

2.3.1 Parameter selection

The basic PSO introduced in this chapter has just a few parameters to be set. The first parameter is the number of particles, i.e. the size of the swarm. From a superficial analysis it might appears that the greatest the swarm, the better the results. However, the choice of the population size strongly influences the trade-off between exploration and exploitation, concepts already introduced, as a greater swarm can be useful to maintain diversity in the first iterative cycles but may be counterproductive in late stages where exploitation around the best position is needed. Moreover, as the population size increases the computational time needed to perform the task will increase. Another variable influencing the choice is the dataset size. It can be stated that swarm size needs to be set empirically however, values in the range 20 – 50 are quite common[50].

Other parameters to be fixed are the acceleration coefficients, already defined in formula (2.1), ϕ_1 and ϕ_2. Specifically, ϕ_1 check the converges of each particle to its personal best p_j^k, while ϕ_2 controls that each particle move to the global position p_g . The wrong choice of these values could make the algorithm instable and lead to an uncontrolled increase of particles velocity: small values may limit the movements of the particles, while too large values can cause the divergency of the particles. In fact, in the original paper, Eberhart and Kennedy suggested to fix the $\phi_1 = \phi_2 = 2$ but this solution did not guarantee a satisfactory stability[51]. To solve this issue, it has been suggested to keep the parameter v_j^t within the range $[-V_{max}, +V_{max}]$. Most of the time the value for $+V_{max}$ have to be set empirically on a problem-to-problem basis as to fit it perfectly the specific characteristics. Not surprisingly, a too high value of this parameter may cause the particles to ignore a good solution. On the other hand, a too low value for $-V_{max}$ limits the movement of each particle, and the optimal solution might not be even reached.

Beyond the basic formulation just introduced, several modification and improvements to cope with velocity issues have been developed to try to figure out an empirically universal way to manage it.[52]

---

50 R. Poli and J. Kennedy and T. Blackwell, Particle Swarm Optimization, 2007.

51 See, for example, R. Eberhart and Y. Shi (2001) and X. Hu, R. Eberhart, and Y. Shi (2004).

52 See, for a review, Sousa-Ferreira and Sousa (2016).

2.3.2 Modification of PSO

Even if the original version of the PSO can converge fast, providing robust results given the time needed, it may fall into local minima, especially dealing with high-dimensional optimization problems. The local minima trap practically consists into a premature convergence of the algorithm that may limit the swarm to explore other areas of the search space. To try to avoid this drawback, during the years some interesting and promising modification have been introduced to the original algorithm.

Constriction coefficient

A modified version of the PSO has been introduced by Clerc and Kennedy (2002)[53] and it is based on a modification of the velocity equation. A new parameter χ called constriction factor is introduced and the formula is modified as follow

$$\begin{cases} v_j^{k+1} = \chi\left[v_j^k + U(0,\phi_1) \otimes \left(p_j - x_j^k\right) + U(0,\phi_2) \otimes \left(p_g - x_j^k\right)\right] \\ x_j^{k+1} = x_j^k + v_j^{k+1} \end{cases} \tag{2.2}$$

with

- $\chi = \frac{2}{\phi-2+\sqrt{\phi^2-4\phi}}$;
- $\phi = \phi_1 + \phi_2$ with $\phi > 4$.

In the implementation of the constriction strategy, common value of ϕ is 4.1, given $\phi_1 = \phi_2$ and the constriction factor equal to 0.7298.

As can be seen from Equation 2.2 compared to Equation 2.1, the constriction factor is applied to all the components of the velocity equation. So, the velocity of the previous iteration v_j^k is multiplied by 0.7298, while the cognitive component $\left(p_j - x_j^k\right)$ and the social one $\left(p_g - x_j^k\right)$ are multiplied by a uniformly generated random number limited by 1.49618.

[53] Clerc, M., & Kennedy, J. (2002). The particle swarm – Explosion, stability, and convergence in a multidimensional complex space, IEEE Transactions on Evolutionary Computation, 6(1), 58–73.

Fully informed Particle Swarm

The Fully Informed Particle Swarm (FIPS) is another alternative variant of the original PSO developed by Mendes et al. (2006). In the Kennedy and Eberhart version of the PSO, a particle with k neighbors choose only one to be his source of influence, ignoring all the others. By doing this, all the information related to other neighbors remains unexploited. In the FIPS approach, every particle is fully informed by all neighbors' best position and the velocity is updated considering all the information coming from all the neighbor agents. Mathematically it is represented by

$$\begin{cases} v_j^{k+1} = \chi \left[v_j^k + \frac{1}{K_j} \sum_{n=1}^{K_j} U(0,\phi) \otimes \left(P_{nbr_n^j} - x_j^k \right) \right] \\ x_j^{k+1} = x_j^k + v_j^k \end{cases} \tag{2.3}$$

with

- K_j the number of the neighbors for particle j;
- nbr_n^j the $j's\ n-th$ neighbor.

When K_j is set equal to 2 the FIPS algorithm generates same results of the original one. However, the use of such a Fully Informed approach seems to make the algorithm less time-consuming by increasing the speed of convergence. On the other hand, is appears to be too sensitive to population topology, which is introduced in the next paragraph.

2.4 Population topology

As already discussed, PSO algorithms is based on social iterations among the particles, meaning that each particle is influenced by their neighbors. In fact, the number of iterations decreases when the neighborhood in the swarm is small and vice versa. In addition, in case of small neighborhood, the optimization process is characterized by a slower convergence, which increases the computational time needed, but it may provide better solutions. As opposite, if the neighborhood is large the convergence is faster, and the risk is that it happens too early.

Hence, considering how important are the connection between particles in the swarm and the impact they may have on the algorithm performance, researchers have developed different types of neighborhood structures. They can be divided into static and dynamic, where the first neighbors and neighborhoods remain the same for the entire optimization process: In depth, this first group includes:

- Local best. This topology, introduced by Eberhart and Kennedy in 1995, is composed by a ring structure in which every single component of the swarm communicates with K other adjacent components in the array. The main advantage of the local best structure is the capability to create multiple populations able to cooperate and give the possibility to the swarm components, to try to converge into different areas of the search space. In the original version K was equal to 2.
- Global best. In a different manner, this topology gives the possibility to all the swarm members to communicate each other's, without creating subpopulations. The algorithm identifies the best particle in the swarm and that particle will influence other members. This much deeper integration between particles gives the opportunity to faster convergence leaving much more vulnerability for local optima.
- Von Neumann. The last structure discussed in this research is that proposed by Kennedy and Mendes, which gives the possibility to each particle to interact with four of its neighbors in a rectangular lattice topology[54]. As the local best approach, the Von Neumann topology includes the parallel search when K is set equal to 4. In their work, the authors suggested that the Von Neumann topology outperforms the other topologies.

A representation of the local topology is presented in Figure 6.

As opposite, dynamic topologies constitute a much more integrated way of communication and influence between particles' swarm, as they can dynamically change their iteration method.
During the years, researchers proposed different implementation of dynamic topology structure. The main ones, chronologically, are:

- Suganthan in (1999) introduced a modified PSO including the neighborhood operator. The search cycle begins with each particles operating without any other particle's influence, and then, increasing the size of the neighborhood until all the particles are included. The idea behind this approach is to start with a local best strategy shifting to a fully connected network.

[54] For example, for a population of 20 particles, the rectangular structure is a 5 × 4 matrix, and each particle is connected to 4 other particles: the one above, the below, and at each side of it.

- Peram et al. (2003) suggested to use a weighted Euclidean distance to define the iteration partner for each member of the swarm. In details, the particle selects its neighboorhood's partner according to a so-called "fitness distance ratio" (FDR) so defined:

$$FDR = \frac{Fitness\,(x_j) - \ Fitness\,(p_i)}{|x_j - p_i|} \tag{2.4}$$

  with p_i be the *i*-neighboor of *j*-th particle.
  This kind of interactive topology have been developed to decrease the possibility of iterations between swarm members located in distant areas of the search space.

- Liang and Suganthan (2005) proposed a PSO algorithm where the population into sub-swarms of size *n*. As the swarms are randomly disjointed created, they tend to lead to an increased exploration of the search space.

- Janson and Middendorf (2005) developed a hierarchical version of the PSO based on the fitness value of each particle. In particular, each component of the swarm is attracted by its own previous best position and by the best position of the other particles, if greater in terms of performance. Doing this, best performances' particles achieve the top position of the ranking, influencing the lower positions' ones.

- Clerc (2006) introduced the TRIBES, a parameter-free PSO, where the neighborhood structure is updated based on an optimization process in response to performance feedbacks. The swarm is divided into sub-swarms with an independent structure and number of components. Then, the optimization algorithm will manage the particles of each tribe in order to improve its performance.

2.5 Improved Particle Swarm Optimization

In this section an alternative version of the Particle Swarm optimization algorithm will be proposed. Along with the PSO modification just cited, other authors tried to combine different formulation to obtain better results as Deng et al. (2012), which developed an improved version of the PSO algorithm. The authors' aim was to provide an improved algorithm in order to reach better optimal solution in a limited iterations cycle.

The first modification by these authors concerned the position of the particles in the search space. Updating particles' position may cause them to leave the search space during the iterative cycles and, as to correct this issue, the value of the new position can be set equal to the boundary value for the asset of the portfolio. However, this may cause a strong reduction in diversity characteristics and a rapid stagnation of the algorithm to the local optimum. As suggested by Paterlini and Krink (2006)[55], in order to avoid such stagnation, a reflection strategy can be applied during the initial search phase: if the value of the new position is outside the boundaries, i.e., the particle leaves the search domain, it will be reflected into the search space by

$$x_{i,j}^{t} = x_{i,j}^{t} + 2\left(x_{j}^{l} - x_{i,j}^{t}\right) \ if x_{i,j}^{t} < x_{j}^{l} \quad (2.2)$$
$$x_{i,j}^{t} = x_{i,j}^{t} - 2\left(x_{i,j}^{t} - x_{j}^{u}\right) \ if x_{i,j}^{t} > x_{j}^{u} \quad (2.3)$$

where x_j^u and x_j^l are the upper and the lower bounds of the domain space, respectively. This reflection strategy stops if no improvements is obtained after a certain number of iterations and then, the boundary values are set as by

$$x_{i,j}^{t} = x_{j}^{l} \ \ if \ \ x_{i,j}^{t} < x_{j}^{l}. \qquad x_{i,j}^{t} = x_{j}^{u} \ \ if \ \ x_{i,j}^{t} > x_{j}^{u} \quad (2.4)$$

The use of this modification has been proved to improves solution quality, allowing particles to explore a greater area escaping from the local minima at the same time.

Then, a modified version of cardinality constraint seen in (1.22) has been implemented. Defined K as the number of the desired assets to be own in the investor's portfolio, and Q a set of K, let K^{new} represent the number of assets after updating positions in portfolio. If $K^{new} < K$, then some assets must be added to Q while, on the contrary, if $K^{new} > K$, then some assets must be removed from Q until $K^{new} = K$.

Considering the case $K^{new} > K$, the methods proposed by the authors consists into delete the assets with the smallest weights in the portfolio, while in the opposite case, $K^{new} < K$, it adds a new asset $i \notin Q$ to the portfolio assigning it the minimum proportion ε_i set at the beginning. From the boundary constraint definition, it is

[55] S. Paterlini, T. Krink, Differential evolution and particle swarm optimisation in partitional clustering, Computational Statistics & Data Analysis, Volume 50, Issue 5, 2006, Pages 1220-1247.

known that x_i must satisfy $0 \leqq \varepsilon_i \leqq x_i \leqq \delta_i \leqq 1$ for $i \in Q$. Adding the new asset causes a rebalancing of the portfolio weights as to respect the budget constraint as introduced in (1.21) and it may cause that an asset weight already in portfolio, defined s_i, exceeds the limits defined before. If $s_i < \varepsilon_i$, the minimum proportional value of ε_i replaces asset s_i. If $s_i > \varepsilon_i$, the proportional share of the free portfolio is computed as

$$x_i = \varepsilon_i + \frac{s_i}{\sum_{j\in Q, s_i > \varepsilon_i} s_i}\left(1 - \sum_{j\in Q}^{\varepsilon_i}\right) \tag{2.5}$$

By using this method, the algorithm minimizes the proportional value of ε_i for the useless assets $i \in Q$ and particles converges faster to the optimal value, especially for low value of the risk aversion parameter λ, which will be explained later on.

Moving to the velocity parameter, the authors implemented the inertia weight approach, w, introduced by Shi and Eberhart (1998)[56], [57] which controls how previous velocity affects present velocity. Low values of w causes the swarm to concentrate on the local search around the current local area, while high value of w indicate the exploitation of global search for the optimal solution. As already explained, the swarm should concentrate in exploration during the first stages of the search, as the algorithm has almost no knowledge about the information in the search space. Instead, as the iterative process converges to the optimal solution, the focus should be brought to exploitation. In order to achieve this result, the proposed PSO implemented a time variant w defined as

$$w(t) = \big(w(0) - w(n_t)\big)\frac{(n_t - t)}{n_t} + w(n_t) \tag{2.6}$$

where $w(t)$ is the current inertia weight, $w(0)$ is the initial inertia weight, $w(n_t)$ is the final inertia weight, n_t is the maximum number of iterations the algorithm

[56] Y. Shi and R. Eberhart, "A modified particle swarm optimizer," 1998 IEEE International Conference on Evolutionary Computation Proceedings. IEEE World Congress on Computational Intelligence (Cat. No.98TH8360), 1998, pp. 69-73

[57] Y. Shi and R. C. Eberhart. Empirical Study of Particle Swarm Optimization. In Proceedings of the Congress on Evolutionary Computation, pages 1945–1949, Washington D.C, USA, July 1999. IEEE Service Center, Piscataway, NJ.21, 33

needs to perform and t is the current number of iterations. As suggested by Shi and Eberhart[58], and implemented by authors, $w(0) = 0.9$ and $w(n_t) = 0.4$.

Then, other authors define the modification to the acceleration coefficients. The approach of their algorithm is to use time variant coefficients introduce by Ratnaweera et al. (2004)[59], which has empirically proved to be more efficient than the fixed ones of the original definition. Mathematically, the time variant acceleration coefficients are defined as

$$c_1(t) = \left(c_{1,min} - c_{1,max}\right)\frac{t}{n_t} + c_{1,max} \tag{2.7}$$

$$c_2(t) = \left(c_{2,max} - c_{2,min}\right)\frac{t}{n_t} + c_{2,min} \tag{2.8}$$

where $c_{1,min}$ and $c_{2,min}$ are the minimum value for the acceleration coefficients, $c_{1,max}$ and $c_{2,max}$ are the maximum ones, t is the current number of iterations and n_t is the maximum number of iterations the algorithm needs to perform. In such representation of the acceleration coefficients, c_1 linearly decreases and c_2 linearly increases over time, encouraging convergence to a good optimum close to the end of the optimization process by trusting the best particle at that time. The values for the acceleration parameters, as suggested by Ratnaweera et al. (2004)[60], were defined as $c_{1,min} = c_{2,min} = 0.5$ and $c_{1,max} = c_{2,max} = 2.5$.

Lastly, authors define a mutation operator, based on Tripathi et al (2007)[61] which has been introduced as to improve diversity in the search space. Formally, the mutation operator is defined as

$$g_k' : \begin{cases} g_k + \Delta(t, \mathrm{UB} - g_k) & \text{if } flip = 0 \\ g_k + \Delta(t, g_k - \mathrm{LB}) & \text{if } flip = 1 \end{cases} \tag{2.9}$$

[58] In their papers the authors argued that setting $w(0)$ in the range [0.9,1.2] have less chances to fail finding the optimal solution but, among all the values in the interval, $w(0) = 0.9$ takes the least average number of iterations to find the global optimum solution. Moreover, they empirically proved that the approach performs in a satisfactory way decreasing w from 0.9 to 0.4 during the iteration.

[59] Ratnaweera, A., Halgamuge, S., & Watson, H. (2004). Self-organizing hierarchical particle swarm optimizer with time-varying acceleration coefficients. IEEE Transactions on Evolutionary Computation, 8(3), 240–255.

[60] A. Ratnaweera, S. K. Halgamuge and H. C. Watson, "Self-organizing hierarchical particle swarm optimizer with time-varying acceleration coefficients," in IEEE Transactions on Evolutionary Computation, vol. 8, no. 3, pp. 240-255, June 2004

[61] P.K. Tripathi, S. Bandyopadhyay, S.K. Pal, Multi-Objective Particle Swarm Optimization with time variant inertia and acceleration coefficients, Information Sciences, 177(22), 2007, 5033-5049,

where g_k denotes a randomly chosen variable, $flip$ represents a dummy variable taking, randomly, value 0 or 1, UB represents the upper limit of the function g_k while LB its lower limit. The function Δ is defined as

$$\Delta(t,x) = x * \left(1 - r^{\left(1 - \frac{t}{max\_t}\right)^b}\right) \tag{2.10}$$

where r is a random number generated in the interval [0,1], max_t is the maximum number of iterations, t is the current iteration number and b represents the dependence of the mutation on the iteration number. As suggested by Deb, (2001)[62] $b = 5$.

A brief overview of the modified PSO pseudo-code will now be presented, as to make it more understandable:

1. Initialization of the particle swarm with random position in the search space and velocity
2. Begin loop:
    a. Compute the value of the object function $f(x_j^k)$ in the current position x_j^k. The valuation is computed for each particle.
    b. Compare the value of the object function of the current position with the best position $Pbest$. If $f(x_j^k) > Pbest$, update the best position p_j with the current position x_j^k, otherwise the cycle continues.
    c. Identify the particle with the best global object function value and define it as $Gbest$ and its position p_g^3.
    d. Update the position and the velocity of the particles according to the following equation
    e. If one of the criterions is met (i.e. satisfactory fitness value or limit of cycles reached), exit loop.
3. End loop.

[62] K. Deb, Multi-Objective Optimization using Evolutionary Algorithms, John Wiley and Sons, USA, 2001

This modified PSO has proved to perform better than the original one but, also, against PSO-DIV by Fourie and Groenwold (2002)[63] and PSO-C by Clerc and Kennedy (2002)[64]. Moreover, it has demonstrated better performance against other kind of metaheuristics such as simulated annealing, tabu search and genetic algorithm.

[63] Fourie, P.C. & Groenwold, Albert. (2002). The particle swarm optimization algorithm in size and shape optimization. Structural and Multidisciplinary Optimization. 23. 259-267

[64] Clerc, M., & Kennedy, J. (2002). The particle swarm – Explosion, stability, and convergence in a multidimensional complex space, IEEE Transactions on Evolutionary Computation, 6(1), 58–73

Chapter 3

Options and the forward-looking approach

As introduced at the very beginning of this research, despite the seminal work by Markowitz, the optimal portfolio selection is a classical and challenging problem in finance because of the estimation errors from historical data. This section is composed by a brief explanation about option derivatives, Black-Scholes-Merton differential equation and pricing formula and two different approaches to the implementation of a forward-looking strategy.

3.1 Options

Options belong to the category of derivative instruments, i.e. financial instruments whose value depends on the value of a different type of asset called "underlying". Options can be defined as financial contracts that give to the purchaser, the right and not the obligation, to buy or to sell a specified underlying asset at a determined price, the *strike price* or *exercise price,* in a specified time in the future, the *maturity*. In depth, options can be distinguished into:

- Call option: derivative that gives the owner the right to buy an underlying asset at a specified price in the future.
- Put option: derivative that gives the owner the right to sell an underlying asset at a specified price in the future.

Moreover, options can be either American or European, where the first can be exercise at any time up to the maturity date while the latter can be exercised only on the expiration date. Options are used by investors to speculate, through holding a leveraged position in an asset at a fraction of the cost of buying it, or to hedge or reduce the risk exposure of their portfolios. They can be written on stocks, currencies, indices or futures.

Every option contract has two sided. In one side it can be found the investor who have a long position (i.e. has bought the option), while on the other there's the investor having a short position (i.e. has sold the option). So, it there exists four different options position:

- A long position in a call option;
- A long position in a put option;

- A short position in a call option.
- A short position in a put option.

Since American options can be exercised at any time, their payoff is not so easy to demonstrate as the European one. Given K to be the strike price and S_T the price of asset at maturity, the payoff for a long position in a European call options is

$$max(S_T - K, 0) \tag{3.1}$$

It can be easily noticed that the option will be exercised only if $S_T > K$. On the contrary, the payoff of a short position in a European call option is

$$-max(S_T - K, 0) = min(K - S_T, 0) \tag{3.2}$$

The payoff to a holder of a long position in a put option is

$$max(K - S_T, 0) \tag{3.3}$$

While the payoff of a short position in a put option is

$$-max(K - S_T, 0) = min(S_T - K, 0) \tag{3.4}$$

A graphical representation of these payoffs can be found in Figure 7.

Options prices on the market are determined by the results of demand and supply during the negotiation time coming from market makers[65] and financial operators. The price of an option can be derived by using two methods, the binomial tree and the Black-Scholes-Merton model which will be introduced later. In the next paragraphs the variables affecting option prices will be briefly introduced as one of them will be needed for our purposes. These six factors are:

- the current stock price;
- the strike price;
- the time of expiration;
- the volatility of the stock price;
- the risk-free rate;

[65] The market maker can be either a firm or an individual whose quotes two-sided markets in a particular security, providing bids and offers along with the market size of each. Hence, their presence is necessary to provide liquidity to the market and no delay during the operations. The market makers themselves make their profits from the bid–offer spread.

- the dividends expected to be paid.

Stock price and strike price

The mechanics of call and put option payoff has been introduced before. Starting from there, it can be easily understood the relationship between stock price, strike price and option price. In a call option, the payoff will be the amount by which the stock price exceeds the strike price. So, call options become more valuable as the stock price increases and less valuable as the strike price increase. On the contrary, for a put option the payoff on exercise is the amount by which the strike price exceeds the stock price. Hence, they become less valuable as the stock price increases and more valuable as the strike price increases.

Time to expiration

Considering the time to expiration, both put and call American options become more valuable (or at least do not decrease in value) as the time to expiration increases. Instead, European put and call options usually become more valuable as the time to expiration increases[66].

Volatility

As introduced before, volatility is a measure of uncertainty about the return provided by the stock (or the underlying in general). For the owner of the stock, the probability that the stock will perform very bad or very well offset each other. Instead, in stock options, the owner of a call option benefits from the price increase, while the downside risk of a decrease in price is limited by the presence of the option and the possibility to not exercise it (making the most the owner can lose the price of the option). On the contrary, the owner of a put option benefits from a price decrease, but the downside risk of a price increase is limited. Given these assumptions, the values of both calls and puts therefore increase as volatility increases.

To estimate volatility of a stock price, it is usually observed at a fixed intervals of time. Given:

- $n + 1$ be number of observations;

[66] This is not always the case. Considering two European call options, one with expiration day in one month, the other with expiration day in three months. If a dividend is expected to be paid in two months, a fall in the stock price is expected, making the short-life option more valuable than the long-life one.

- S_i be the stock price at the end of ith time interval;
- τ be the length of time interval in years;

and define

$$u_i = ln\left(\frac{S_i}{S_{i-1}}\right) \text{ for } i = 1,2,\dots,n \tag{3.5}$$

The estimate s of the standard deviation of u_i is

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(u_i - \bar{u})^2} \tag{3.6}$$

or

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}u_i^2 - \frac{1}{n(n-1)}\left(\sum_{i=1}^{n}u_i\right)^2} \tag{3.7}$$

with $\bar{u}$ to be the mean of u_i.

As it is known that the standard deviation of u_i is $\sigma\sqrt{T}$, the variable s is an estimate of $\sigma\sqrt{T}$ and so σ can be estimated as $\hat{\sigma}$, where

$$\hat{\sigma} = \frac{s}{\sqrt{\tau}} \tag{3.8}$$

Risk-free interest rate

Interest rates affect every aspect of market economy. In option market, as risk-free interest rates increase, investors expect a rise of the expected return from the stocks. Moreover, the present value of any future cash flow to be received by the holder of the option decreases. Combining these two phenomena, bring to the conclusion that an increase of the interest rate is linked to an increase of the value of call options and to a decrease the value of put options.

Amount of future dividends

The effects of dividends on stock price are well-known. In stock options, the reduction of stock price given the dividend payment is linked to a decrease of the value of a call option and an increase of the value of a put option. In addition, the size of the dividend influences the magnitude of this relationship.

3.2 The Black-Scholes-Merton Model

In the last paragraphs stock options have been briefly introduced together with an explanation of the factors affecting their prices. In this subchapter, one of the keystone of the modern finance will be presented as it will be needed later, for the main purpose of this research.

The Black-Scholes-Merton Model[67] represents one of the most important option pricing model. Even if it relies on not-so-realistic assumptions, it is widely used for its simplicity which is confirmed looking at the strong assumption it relies on:

- The option is European, and it can be exercised only at maturity;
- The short selling of securities is permitted;
- No dividends are paid out during the life of the option;
- There are no transaction costs or taxes;
- Securities are perfectly divisible;
- There are no riskless arbitrage opportunities;
- Security trading is continuous;
- The risk-free rate is constant and the same for all securities.

3.2.1 The Black-Scholes-Merton differential equation

Considering that the stock price, S, follows a geometric Brownian motion[68], expressed as

$$dS = \mu S dt + \sigma S dz \tag{3.5}$$

---

[67] F. Black and M. Scholes, "The Pricing of Options and Corporate Liabilities,"Journal of Political Economy", 81, 1973, 637–59; R.C. Merton, "Theory of Rational Option Pricing," Bell Journal of Economics and Management Science, 4, 1973, 141–83.
[68] This assumption is fundamental in order to avoid stock price to be negative.

Given f to be the price of a call option or any other derivative contingent on S, it has to be some function of S and t. So, from the Ito's lemma[69]

$$df = \left(\frac{\partial f}{\partial S}\mu S + \frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial S^2}\sigma^2 S^2\right)dt + \frac{\partial f}{\partial S}\sigma S dz \tag{3.6}$$

The discrete versions of equations (3.5) and (3.6) are

$$\Delta S = \mu S \Delta t + \sigma S \Delta z \tag{3.7}$$

and

$$\Delta f = \left(\frac{\partial f}{\partial S}\mu S + \frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial S^2}\sigma^2 S^2\right)\Delta t + \frac{\partial f}{\partial S}\sigma S \Delta z \tag{3.8}$$

where ΔS and Δf are the changes in f and S in the interval of time Δt. It has been demonstrated that a portfolio of a stock and a derivative can be built so that the Wiener process is eliminated. The portfolio is constructed by

-1: derivative
$\Delta f / \Delta S$: shares

In this way the holder of the portfolio is long on the number of shares $\Delta f / \Delta S$ and, as opposite, short on the derivative. Let define Π as the value of such portfolio, it means that

$$\Pi = -f + \frac{\partial f}{\partial S} S \tag{3.9}$$

Changes in portfolio values, namely $\Delta\Pi$, in the time interval Δt are given by

$$\Delta\Pi = -\Delta f + \frac{\partial f}{\partial S}\Delta S \tag{3.10}$$

Substituting equations (3.7) and (3.8) in equation (3.9) leads to

$$\Delta\Pi = \left(-\frac{\partial f}{\partial t} - \frac{1}{2}\frac{\partial^2 f}{\partial S^2}\sigma^2 S^2\right)\Delta t \tag{3.11}$$

---

[69] K. Itoˆ, "On Stochastic Differential Equations," Memoirs of the American Mathematical Society, 4 (1951): 1–51.

Equation 3.11 implies that the portfolio is risk-free, as it does not involve Δz. This assumption leads to the conclusion that the portfolio must earn the same rate of return as other short term risk-free securities. If this condition is not satisfied, an arbitrage will be possible[70]. This means that

$$\Delta\Pi = r\Pi\Delta t \tag{3.12}$$

where r is the risk-free rate. Substituting equations (3.9) and (3.11) into (3.12) yields

$$\left(\frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial S^2}\sigma^2 S^2\right)\Delta t = r\left(f - \frac{\partial f}{\partial S}S\right)\Delta t$$

So that

$$\frac{\partial f}{\partial t} + rS\frac{\partial f}{\partial S} + \frac{1}{2}\sigma^2 S^2\frac{\partial^2 f}{\partial S^2} = rf \tag{3.13}$$

Equation 3.13 is the Black-Scholes-Merton differential equation. It can be solved in many ways, corresponding to the different derivative that can be defined as with S as the underlying variable. In the next paragraphs the most known method to solve the differential equation will be introduced.

3.2.2 The Black-Scholes-Merton pricing formula

The most famous solution to the differential equation described in (3.13) are the ones for the pricing of a call or put European option. These formulas are

$$c = S_0 N(d_1) - Ke^{-rT}N(d_2) \tag{3.14}$$

and

$$p = Ke^{-rT}N(-d_2) - S_0 N(-d_1) \tag{3.15}$$

where

[70] The arbitrage is quite simple: if the portfolio earned more thatn

$$d_1 = \frac{\ln(S_0/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}$$
$$d_2 = \frac{\ln(S_0/K) + (r - \sigma^2/2)T}{\sigma\sqrt{T}} = d_1 - \sigma\sqrt{T} \tag{3.16}$$

The function $N(x)$ is the cumulative probability distribution function for a variable with a standard normal distribution meaning that it is the probability that a variable with a standard normal distribution will be less than x. The other variables are the ones introduced before. The most used approach of solving the equation is the risk-neutral valuation. Considering a European call option, its expected value $\hat{E}$ at maturity in a risk-neutral world is

$$\hat{E}[max(S_T - K, 0)] \tag{3.17}$$

From the risk-neutral world considerations, the European call price c is the expected value discounted at the risk-free interest rate

$$c = e^{-rT}\hat{E}[max(S_T - K, 0)] \tag{3.18}$$

Then, under the assumption assumed in Black-Scholes-Merton, S_T is lognormal and, from the lognormality of stock prices $\hat{E}(S_T) = S_0 e^{rT}$ and the standard deviation of $\ln S_T$ is $\sigma\sqrt{T}$. Given these considerations,

$$c = e^{-rT}[S_0 e^{rT} N(d_1) - KN(d_2)] = S_0 N(d_1) - Ke^{-rT} N(d_2) \tag{3.19}$$

where

$$d_1 = \frac{\ln\left[\hat{E}(S_T)/K\right] + \sigma^2 T/2}{\sigma\sqrt{T}} = \frac{\ln(S_0/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}$$
$$d_2 = \frac{\ln\left[\hat{E}(S_T)/K\right] - \sigma^2 T/2}{\sigma\sqrt{T}} = \frac{\ln(S_0/K) + (r - \sigma^2/2)T}{\sigma\sqrt{T}} \tag{3.20}$$

Equation (3.17) proves the Black-Scholes-Merton pricing formulas.

Since it has been proved that is it is never optimal to exercise an American call option on a non-paying-dividend stock, Equation (3.17) is used to price an American call option on a non-paying-dividend stock too. On the other hand, there no exists exact analytic formulas to price an American put option on a non-paying-dividend stock.

3.3 Implied volatility

Volatility has been introduced in 3.1 as one of the factor affecting option price. Regarding Black-Scholes-Merton pricing formula just introduced, the volatility of stock price is the only parameter that cannot be directly observed. In practice, analysts and traders work with *implied volatilities* which are volatilities implied by the option prices in the market. As it is a function of price of the underlying, the exercise price, the risk-free rate of return, the time to expiry and the price of the option, the simplest way to compute implied volatility is the iterative one: knowing all the factors of the Black-Scholes-Merton pricing formula, an initial estimate of σ can be used to check the corresponding correct price of the call option. Because c is an increasing function of σ, the iterative search will focus on lower or higher value of σ to look for the closet value of c. Other, more sophisticated methods, include the Brenner and Subrahmanyam (1988)[71] or the Newton-Rahpson[72].
Implied volatility can be used as an indicator of the market's expectations of the remaining life of the option i.e. it can be used to ascertain the market's opinion of the expected volatility of a particular security.

3.4 The forward-looking approach

One of the main issues arising from the implementation of the mean-variance framework is the estimation error based on the historical sample moments[73]. Moreover, also expected returns are difficult to estimate[74]. In order to overcome these practical limitations, researchers focused the attention the Global Minimum Variance Portfolio (GMVP) since it does not depend on the expected returns and it has been proved often provide better out-of-sample performance than a mean-variance optimized portfolio[75]. However, even considering only the GMVP, the covariance matrix estimation risk still remains, and several possible solutions have been presented, like the imposition of restriction on portfolio weights[76] or on the

---

[71] Brenner, Menachem & Subrahmanyam, Marti. (1988). A Simple Formula to Compute the Implied Standard Deviation. Financial Analysts Journal 44. 80-83.
[72] Newton-Raphson method is used to find the zeros of a real valued function $f(x) = 0$.
[73] See, for example, Best and Grauer (1991), Chopra and Ziemba (1993), and Michaud (1989)
[74] See Merton (1980)
[75] See, for example, Ledoit and Wolf (2003) and Jagannathan and Ma (2003)
[76] See DeMiguel, Garlappi, and Uppal (2009) for a review

covariance matrix[77]. But the main problem, the use of historical data which may be inconsistent, still remains.

As to improve the performance of portfolios, some researchers concentrate their study on-the-so-defined “forward-looking approach”. This consists into deriving implied data extracted from option prices which should reflect the market participants’ expectations on the future performance of the underlying. Even if these information are not guaranteed, they can be used to make alternative assumption about investment and portfolio optimization and, as it will explained, allocation strategies from such approach typically outperforms historical ones.

The following subchapters will discuss two different methods of deriving moments of the statistical distribution from option-implied data.

3.4.1 Adapted Historical Covariance Model (AHCM)

The first technique to be presented is an overview on the work by DeMiguel et al. (2012) which derive the options’ covariance matrix using a partially implied technique based on the research by Buss & Vilkov (2012). This kind of approach is referred to as Adapted Historical Covariance Model (AHCM).

Firstly, as known, variance of a portfolio can be defined as

$$\sigma_{P,t}^2 = \sum_{i=1}^{N} w_{i,t}^2 \sigma_{i,t}^2 + 2 \sum_{i=1}^{N-1} \sum_{j \neq i}^{N} w_{i,t} w_{j,t} \sigma_{i,t} \sigma_{j,t} \rho_{ij,t} \qquad (3.21)$$

With $w_{i,t}$ the weight of the asset in portfolio, $\sigma_{i,t}$ the volatility of the asset i at time t and $\rho_{ij,t}$ the correlation between asset i and j. Even if one knows the weights and the volatility of all the portfolio’s components, it still remains the correlation coefficients to be estimated which represents $N \times (N-1)/2$ coefficients. To overcome this issue, Buss & Vilkov (2012) introduced a fixed proportion single state variable α_t, then DeMiguel et al. defined

$$\rho_{ij,t} - \hat{p}_{ij,t} = \alpha_t \left(1 - \rho_{ij,t}\right) \qquad (3.22)$$

[77] Ledoit and Wolf (2004)

where $\rho_{ij,t}$ is the correlation coefficient and $\hat{p}_{ij,t}$ is the expected correlation.

Solving Equation (3.22) for $\hat{p}_{ij,t}$ and substituting the results into (3.21), it leads to

$$\sigma_{P,t}^2 = \sum_{i=1} \sum_{j} w_{i,t} w_{j,t} \sigma_{i,t} \sigma_{j,t} \left(\rho_{ij,t} - \alpha_t (1 - \rho_{ij,t})\right) \tag{3.23}$$

Solving for α_t and after some rearrangements

$$\alpha_t = -\frac{\sigma_{P,t}^2 - \sum_i \sum_j w_{i,t} w_{j,t} \sigma_{i,t} \sigma_{j,t} \rho_{ij,t}}{\sum_i \sum_j w_{i,t} w_{j,t} \sigma_{i,t} \sigma_{j,t} (1 - \rho_{ij,t})} \tag{3.24}$$

which describe the composition of the α_t component. Then, substituting Equation (3.24) into (3.22) and rearranging the terms leads to the final equation of this approach

$$\hat{p}_{ij,t} = \frac{\sigma_{P,t}^2 - \sum_i \sum_j w_{i,t} w_{j,t} \sigma_{i,t} \sigma_{j,t} \rho_{ij,t}}{\sum_i \sum_j w_{i,t} w_{j,t} \sigma_{i,t} \sigma_{j,t} (1 - \rho_{ij,t})} (1 - \rho_{ij,t}) + \rho_{ij,t} \tag{3.25}$$

Equation (3.25) describes the correlation estimates computed from option-implied information which are mixed with the historical information. As the covariance matrix is defined as

$$\Sigma = D \, \Omega \, D \tag{3.26}$$

where D is a diagonal matrix of standard deviation and Ω is the correlation matrix, the implied covariance can be estimated by combining the correlation matrix with the estimated volatility matrix.

3.2 Beta Implied Covariance Model (BICM)

Among all the possible applications of the forward-looking approach to improve sample estimations, this research will focus on a Beta Implied Covariance Model developed by Kempf, Korn & Sassning (2012). In their paper the authors developed a complete family of fully implied covariance estimators from a cross-section of vanilla options. In particular, among all the components of the family of estimators, this research will keep into consideration the second moments component.

In order to develop the model, they made two strong assumptions. Firstly, the returns of the stocks in portfolio follows a generalized version of the Sharpe diagonal model[78] with time-varying coefficients

$$R_{it} = \alpha_{it} + \beta_{it} R_{mt} + \epsilon_{it} \quad \forall i = 1, \dots, N \tag{3.27}$$

where R_{it} and R_{mt} describe the returns of the $i-th$ asset and the market, respectively. α_{it} and β_{it} are the time-varying model coefficients while ϵ_{it} is a zero mean idiosyncratic error term, independent of the market return. Moreover, ϵ_{it} and ϵ_{jt} are independent for all $i \neq j$.

Equation (3.27) states that assets returns' risk can be explained by two factors, the systematic risk component which is market related and undiversifiable, and the idiosyncratic component which, as opposite, is firm-specific.

In the single index model such as the one introduced, covariances depend only on the betas and the variance of the market return

$$Cov(R_{it}, R_{jt}) = \beta_{it}\beta_{jt} Var(R_{mt}) \quad \forall i \neq j \tag{3.28}$$

Following equation (3.28), authors argued that the variance of the market can be simply derived from traded index options, so it remains to identify stocks related betas.

To solve the problem, the second assumption needs to be introduced. It is possible to derive the family of estimator by imposing a cross-sectional restriction either on the second, third, fourth, or any other higher moment, respectively. Estimators' restrictions on the different moment will now be explained.

Estimator based on second moments

The first member of the family of estimators can be derived imposing a cross-sectional restriction on the return variance. In their paper the authors introduced a

[78] W. Sharpe, A Simplified Model for Portfolio Analysis, 1963, Management Science, 9:2, 277-293

parameter c_t, with $0 \leq c_t < 1$, which represents the proportion of systematic risk for all the assets in portfolio. In addition, it is stated that c_t is time varying and it is the same for all the assets. According to the introduction of this parameter, then

$$\beta_{it}^2 \text{Var}\ (R_{mt}) = c_t \text{Var}\ (R_{it}) \tag{3.29}$$

and

$$Var\ (\epsilon_{it}) = (1 - c_t) Var\ (R_{it}) \tag{3.30}$$

The cross-sectional restriction just introduced suggests a positive relationship between beta and idiosyncratic risk, coherently with the literature and empirical evidence[79]. Moreover, it implies that high-beta stocks will have high idiosyncratic risk.

Solving the return variance of the ith assets

$$\text{Var}\ (R_{it}) = \beta_{it}^2 \text{Var}\ (R_{mt}) + (1 - c_t) \text{Var}\ (R_{it}) \tag{3.31}$$

Solving for β_{it} and rearranging

$$\beta_{it} = c_t^{1/2} \left(\frac{\text{Var}(R_{it})}{\text{Var}(R_{mt})} \right)^{1/2} \tag{3.32}$$

It is well known that the beta of the portfolio is the weighted sum of the weights of the components and the beta of the market equals one[80]. So, the parameter w_{itm} where $i = 1, \dots, N$ is introduced in the market portfolio as to derive c_t

$$\sum_{i=1}^{N} w_{itm} \beta_{it} = \sum_{i=1}^{N} w_{itm} c_t^{1/2} \left(\frac{\text{Var}(R_{it})}{\text{Var}(R_{mt})} \right)^{1/2} = 1 \tag{3.33}$$

Rearranging for c_t

$$c_t = \frac{\text{Var}(R_{mt})}{\left(\sum_{i=1}^{N} w_{itm} \text{Var}\ (R_{it})^{1/2}\right)^2} \tag{3.34}$$

79 See, for example, Fama and MacBeth (1973) or Malkiel and Xu (2002)
80 See, for example, Fama and MacBeth (1973)

From equation (3.34) it can be observed that c_t is positive and smaller than one given that not all assets in the index are perfectly correlated. Substituting c_t from (3.33) into (3.31) and substituting the resulting β_{it} in (3.34) leads to the final result

$$\mathrm{Cov}(R_{it}, R_{jt}) = \frac{\mathrm{Var}(R_{it})^{1/2}\,\mathrm{Var}(R_{jt})^{1/2}}{(\sum_{i=1}^{N} w_{itm}\,\mathrm{Var}(R_{it})^{1/2})^2}\mathrm{Var}(R_{mt}) \quad \forall i \neq j \qquad (3.35)$$

Equation (3.35) demonstrate that covariances are functions of both individual assets and the variance of the market, but no cross-moments appear. Given this result, the implied volatility from plain vanilla options on every individual asset and on the market index can be used to derive a fully-implied covariance estimate. Other fundamental characteristics of the Kempf, Korn & Sassning (2012) approach is that the matrix is guaranteed to be positive definite by construction and even all the elements composing it are restricted to be positive.

Kempf, Korn & Sassning (2012) also show that estimators can be derived from higher orders moments in a very similar way. In Appendix B the methodology used to derive such estimators is shown.

Even it has been proved to be particularly efficient and grant the possibility to construct a fully implied covariance matrix, this model faces the risk of introducing structural deficiencies in the covariance estimation.

Chapter 4

Application and discussion

This last chapter will focus the attention on the application of the portfolio selection model introduced in Chapter 1 and on the comparison between the different PSO algorithms based on the historical approach and the forward-looking one introduced in Chapter 3. Moreover, to better frame the application, a brief introduction to the financial market during the first months of 2020 will be presented. Lastly, the comparison between the results obtained will be analyzed.

4.1 Financial background

In March 2020, the World Health Organization (WHO) declared that a coronavirus outbreak (COVID-19) was a pandemic, however the impact of the disease had already heavily hit world's financial markets. Figures 8, 9 and 10 show the movement of S&P500, Nasdaq and Dow Jones Industrial Average from 01/01/2020 to 01/06/2020. It can be clearly observed the extreme downtrend caused by the raising number of cases and deaths, which are reported in Figure 11.

Figure 8 – S&P500 movement from 01/01/2020 to 01/06/2020

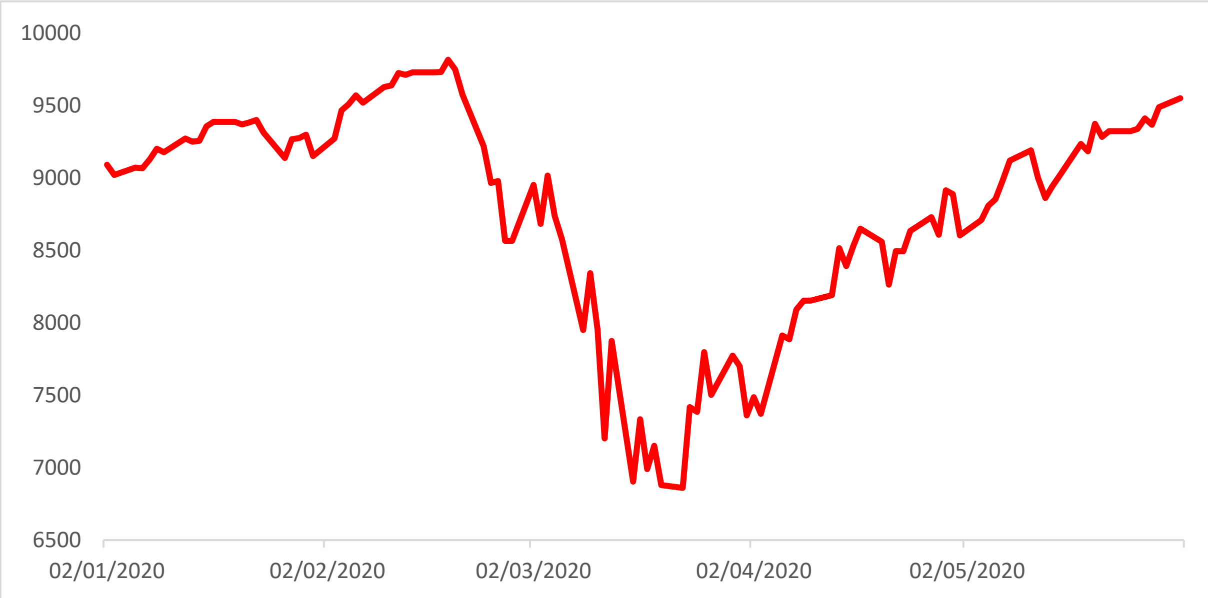

Figure 9 – NASDAQ movement from 01/01/2020 to 01/06/2020

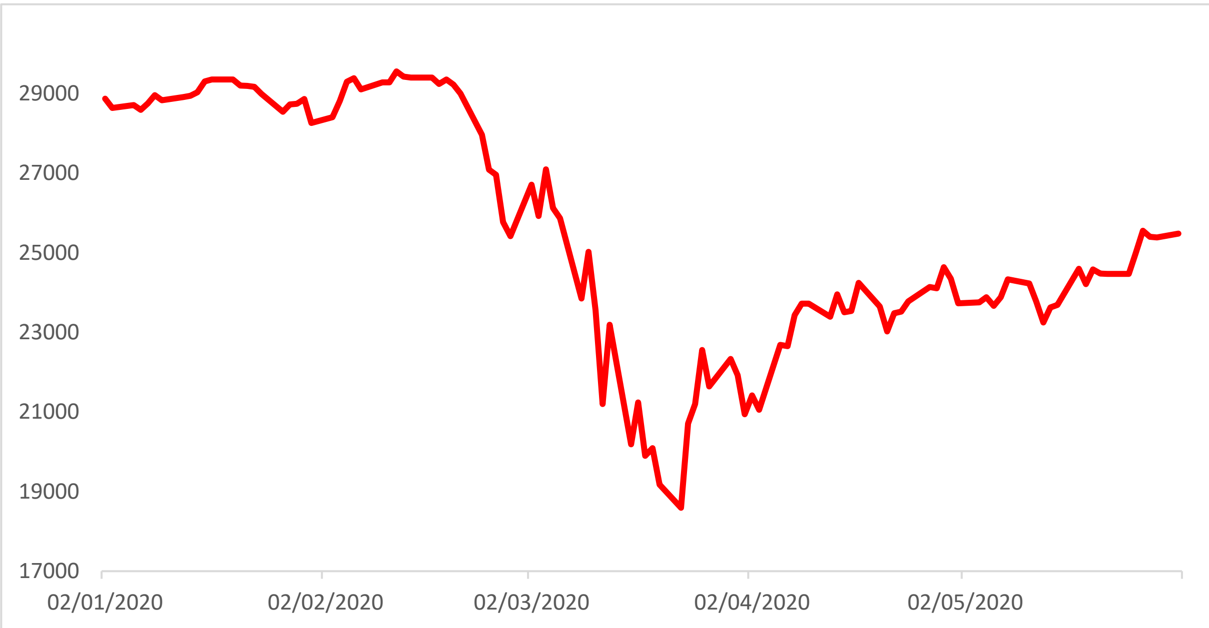

Figure 10 – Dow Jones Industrial Average movement from 01/01/2020 to 01/06/2020

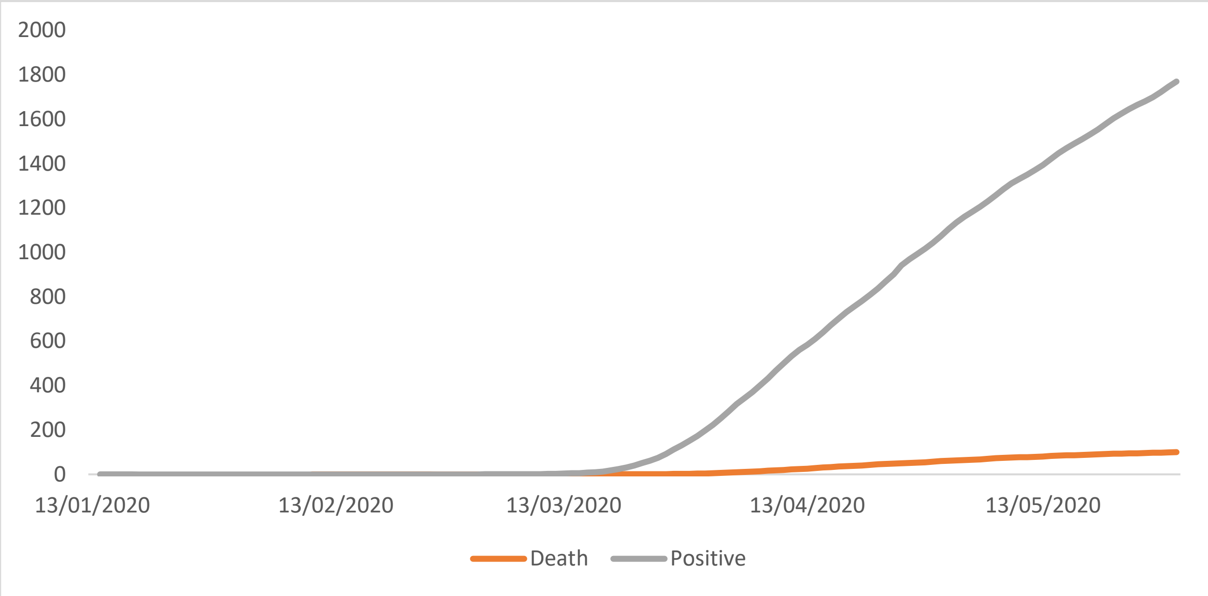

Figure 11 – Cumulative COVID positive cases and death from 01/01/2020 to 01/06/2020 (data in thousand)

Looking in depth the DJIA case, Table 4.1 shows percentage change in stock price from 01/01/2020 to 01/06/2020 by company. In Figure 12, instead, a graph about the losses occurred to all the DJIA stocks is reported.

AAPL	6%	**MCD**	-6%
AXP	-32%	**MMM**	-13%
BA	-127%	**MRK**	-13%
CAT	-23%	**MSFT**	13%
CSCO	0%	**NKE**	-3%
CVX	-29%	**PFE**	0%
DOW	-35%	**PG**	-5%
DIS	-26%	**TRV**	6%
GS	-18%	**UNH**	4%
HD	12%	**UTX**	-117%
IBM	-6%	**V**	2%
INTC	4%	**VZ**	-4%
JNJ	3%	**WBA**	-35%
JPM	-43%	**WMT**	5%
KO	-17%	**XOM**	-51%

Table 4.1 – Stocks performance in the analyzed period

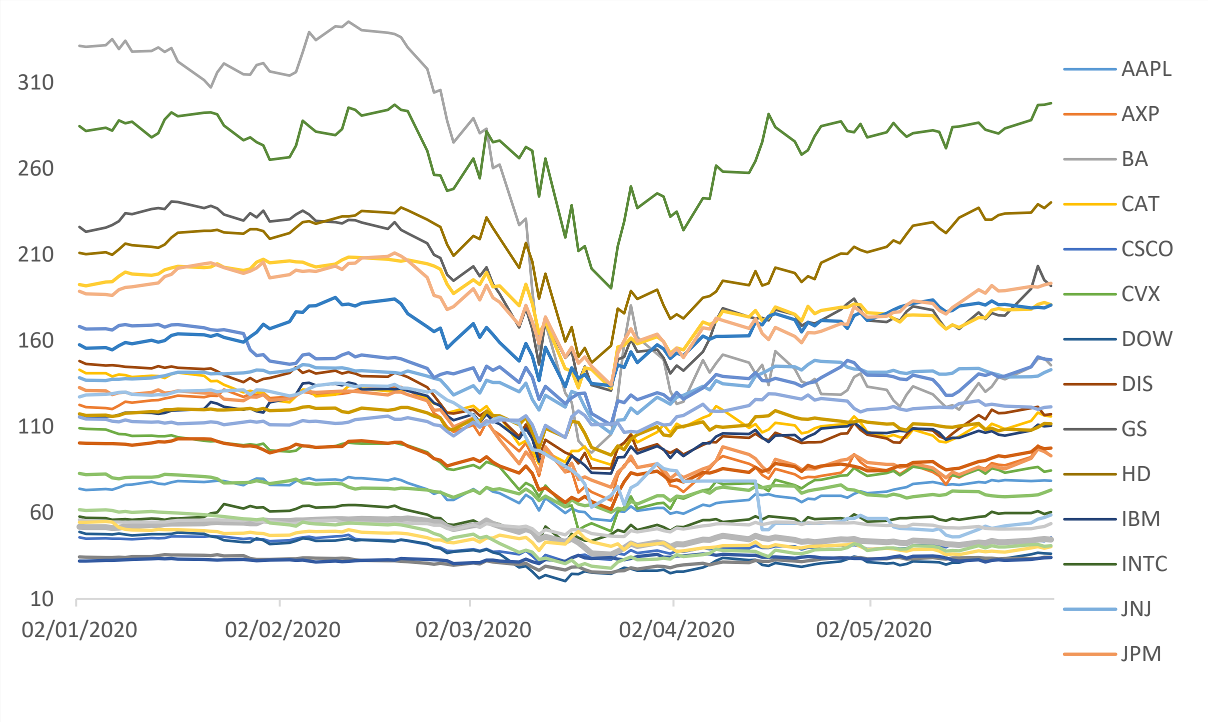

Figure 12 – Price movement of the stocks composing the DJIA where it can be clearly observed the effect of the COVID19 pandemic starting at the end of February

4.2 Preliminary information

As to assess the capabilities and the possible improvement of the forward-looking approach the application has been performed using real data, building a portfolio of eight highly diversified stocks chosen by the algorithm from the Dow Jones Industrial (DJI), which will be considered as the market portfolio.

Since the aim of this research is to try to improve portfolio performances in time of economic instability, the time window took into consideration is the one starting from 02/01/2020 to 29/05/2020 and divided into the two required subsets:

- *in-sample* one from 01/01/2020 to 28/02/2020;
- *out-of-sample* from 28/02/2020 to 30/05/2020.

To perform a precise analysis, the composition of the DJIA at the last day of the in-sample period has been chosen. The assets composing the market portfolio is presented in Figure 13. Then, the implied volatility of the option written on the stocks have been computed through the Black-Scholes-Merton pricing formula and the variance-covariance matrix has been constructed using the BCIM and Equation 3.35. It is necessary to construct such matrix because, as introduce in Equation 1.36, the aim is to minimize the variance of the portfolio through the minimization of the variance-covariance matrix. After the construction of the matrix, this has been imported into the algorithm model and substituted to the standard variance-covariance matrix.

3M Company	The Goldman Sachs Group, Inc.	Pfizer Inc.
American Express Company	The Home Depot, Inc.	The Procter & Gamble Company
Apple Inc.	Intel Corporation	The Travelers Companies, Inc.
The Boeing Company	International Business Machines Corporation	UnitedHealth Group Incorporated
Caterpillar Inc.	Johnson & Johnson	United Technologies Corporation
Chevron Corporation	JPMorgan Chase & Co.	Verizon Communications Inc.
Cisco Systems, Inc.	McDonald's Corporation	Visa Inc.
The Coca-Cola Company	Merck & Co., Inc.	Walgreens Boots Alliance, Inc.
Dow Inc.	Microsoft Corporation	Walmart Inc.
Exxon Mobil Corporation	Nike, Inc.	The Walt Disney Company

Figure 13 – List of stocks composing the DJIA in the time window taken into consideration

4.3 Parameter settings

Regarding the PSO and its implementation of the inertia weight approach, as suggested by the authors, the parameters engraving the velocity of the particles are set as follow:

- inertia weight: 0.7289;
- cognitive acceleration coefficient: ϕ_1 = 1.49681;
- social acceleration coefficient: ϕ_2 = 1.49681;
- number of iterations = 2000[81];

MOPSO parameter specific settings

As the Modified version of the PSO adds some tweaks to the original model, specific settings need to be introduced.
Indeed, the reflection strategy have been implemented using:

- Lower bound = 0.01;
- Upper bound = 0.5.

[81] The number of iterations equal to 2000 has be chosen as a value to ensures an adequate level of convergence, without making the process too much time consuming.

Regarding the inertia weight approach, as suggested by literature:

- initial inertia weight: $w(0)$ = 0.9;
- the final inertia weight: $w(n_t)$ = 0.4.

Also, for the time variant acceleration coefficient the parameters have been set following the literature with:

- $c_{1,min}$ = 0.5;
- $c_{1,max}$ = 2.5;
- $c_{2,min}$ = 0.5;
- $c_{2,max}$ = 2.5.

Lastly, the mutation operator had the following parameter settings:

- Upper bound: UB = 0.01;
- Lower bound: LB = 0.5;
- dependence of the mutation on the iteration number: b = 5.

Values of the lower bound and upper bound for both reflection strategy and mutation operator have been chosen after some preliminary run which gave the best results with such values.

As to select values of the penalty parameter ϵ, some preliminary tests have been performed. To determine the optimal value of the penalty parameter the algorithm has been run five time for every different value of ϵ. Results are reported in Table 4.2.

ϵ	PSO	PSO NO FWL	MOPSO	MOPSO FWL
1.00E-04	0.002143115	0.056755795	178.3276391	510.69539
1.00E-05	0.003768996	0.058989903	190.8943784	600.43907
1.00E-06	0.003789909	0.059573829	230.5827393	604.65748

Table 4.2 – Final value of the fitness function

As regards of number of particles, Blackwell et al (2007)[82] suggest using a value to be in range 20-50. However, in order to obtain better results in terms of fitness value the number of swarm particles have been set to 200. All the algorithms have been tested on a PC using an i5-6200u CPU with 8GB of RAM and Matlab R2020a.

4.4 Performance indicators

The analysis of the performance algorithm can be simply based on variance and

[82] Blackwell T., Kennedy J., Poli R., 2007. Particle Swarm Optimization: an overview, Swarm Intelligence, 1(1), 33–57

return. However, a portfolio may increase its total return just by increasing its variance. For this reason, risk-adjusted performance indicators have been developed, with the aim of adjust returns for the risk developed by that returns. Hence, in next paragraphs a the most used indicators will be presented, and they will be computed for each portfolio in the next chapter. In addition to the risk-adjusted indictors, the Maximum Drawdown will be introduced.

4.3.1 Sharpe ratio

Developed by the Nobel prize William F. Sharpe, the Sharpe ratio represent the most used method to compute risk-adjusted comparison. It measures the average return earned in excess to the risk-free rate per unit of risk of volatility or total risk. Mathematically, it is formulated as:

$$Sharpe\ ratio = \frac{\mathbb{E}(r_p) - r_f}{\sigma_p} \tag{4.2}$$

where:

- $\mathbb{E}(r_p)$ is the expected return of a portfolio;
- r_f is the risk-free rate;
- σ_p is the standard deviation of the return of the portfolio.

Firstly, the difference between of the risk-free component and the expected return of the portfolio (or an asset) let to estimate the over-performance associated to that risk-taking activity. Generally, high value of this ratios suggests satisfying returns with respect to the risk taken.

Being so simple to be used and understood have been the success of the spread of the ratio across the financial world as it only need to know the expected return of the portfolio and its standard deviation, given that the risk-free return is always known. However, some important shortcomings should be taken into consideration. Firstly, Sharpe ratio measures the risk in terms of volatility, assuming that returns are normally distributed. Ingersoll et al. (2007), proved that this drawback can lead to wrong investment assumption if the distribution of the returns is not the Normal one, highlighting that, is this case, Sharpe ratio is not a measure to take into consideration. Secondly, it has been shown by Ingersoll et al. (2007) that a fund manager can easily manipulate the Sharpe ratio increasing its value without raising the value of the investment. This manipulation can be obtained through option-based strategies.

4.3.2 Sortino ratio

Introduced by Sortino and Van der Meer in 1991, the Sortino ratio have been developed to overcome the drawbacks of the Sharpe ratio. In short, the aim of the authors was to not penalize volatility deriving from positive returns. According to this, the numerator of the Sortino ratio is the return in excess of a minimum acceptable return (MAR), while the denominator is the downside deviation, considering the MAR as the target return. So, the Sortino ratio can be expressed as

$$\text{Sortino ratio} = \frac{\mathbb{E}(r_p) - \tau}{\delta} \tag{4.3}$$

where:

- τ is the minimum acceptable return;
- δ is the downside deviation.

The downside deviation volatility, measuring only returns falling below the MAR, is represented as

$$\delta = \sqrt{\int_{-\infty}^{\tau} (\tau - r)^2 f(r) dr} \tag{4.4}$$

where $f(r)$ is the continuous probability distribution of the returns.

Given the mathematical formulation above, Sortino ratio can be considered as a modification of the Sharpe ratio, substituting the mean with the MAR and the standard deviation with the downside risk. It can be easily understood that, as a consequence, investors will be penalized only for the variability below the MAR. The comparison between Sharpe ratio and Sortino ratio could be useful to understand which portion of portfolio's variability can be attributed to overperformance (variability above the MAR) or underperformance (variability below the MAR).

4.4.3 Information ratio

Information ratio have been proposed by William Sharpe in 1994 as a generalization of the Sharpe ratio substituting the riskless asset to a benchmark. This new ratio is defined as the ratio of the portfolio's (or an asset) excess return over a benchmark, divided by the standard deviation of the excess return (the so-called tracking error). Specifically, it is represented as:

$$\text{Information ratio} = \frac{\mathbb{E}(r_p - r_b)}{\sigma_{p-b}} \tag{4.5}$$

where:

- r_b is the benchmark return;
- σ_{p-b} is the standard deviation of the excess return.

Hence, a high Information ratio can be translated as an investment portfolio offering greater expected return than the benchmark given a relatively low extra risk. As opposite, a low Information ratio means that an investment portfolio cannot provide enough extra return given the extra risk taken.

4.4.4 Treynor ratio

The Treynor ratio can be considered a derivation of the Sharpe ratio in which the total risk is substituted by the systematic[83] risk. As demonstrated by the CAPM[84], the systematic risk can be measured by the β coefficient, representing the slope of the regression of the portfolio's return against the returns of the market portfolio, considered as the weighted sum of every asset in a specified market[85]. Mathematically, it can be computed as

$$\beta = \frac{\text{Cov}(R_P, R_B)}{\text{Var}(R_B)} \tag{4.6}$$

where R_P is the return of the portfolio and R_B is the return of the market portfolio. Therefore, the Treynor ration can be defined as the portfolio's return in excess over the risk-free rate of return per unit of systematic risk. It is given by

$$\text{Treynor ratio } = \frac{\mathbb{E}(r_p) - r_f}{\beta(r_p, r_b)} \tag{4.7}$$

where:

- $\mathbb{E}(r_p) - r_f$ is the excess return of the investment portfolio;
- $\beta(r_p, r_b)$ is the beta of the portfolio's return r_p relative to the benchmark return r_b[86].

Evidently, the higher the ratio, the better performances on a risk-adjusted basis.

[83] Systematic risk is considered as that portion of risk inherent to the entire market, industry or market segment. Hence, it is undiversifiable, affecting the overall market and not just a particular asset.

[84] The Capital Asset Pricing Model (CAPM) developed independently by Treynor (1965), Sharpe (1964) and Linter (1965), describes the relationship between expected return and systematic risk.

[85] A β value equal to1 suggest that the performance of the individual investor is in line with the one of the market portfolio while a β greater than 1 implies that the portfolio's performance is more volatile than the benchmark. Instead, a β lower than 1 shows that the portfolio's performance is less volatile than the benchmark.

[86] It measures the sensitivity of the portfolios to the market portfolio, meaning the change in portolio's return given a change in market's return.

4.4.5 Maximum Drawdown

The Maximum Drawdown is an indicator of a downside risk over a specified time windows and it can be considered as the maximum observable loss from a peak to a trough of a portfolio. Mathematically, it is expressed as

$$MDD = \frac{\text{Trough Value - Peak Value}}{\text{Peak Value}} \tag{4.8}$$

A low value of the MDD indicates slight fluctuations in the value of the investment and, therefore, a small degree of risk, and vice versa. It can also be used as an indicator of market performance if compared to stock market index, in order to evaluate stocks' performance relative to the market.

4.5 Application

As already described, the aim of this dissertation is to understand if an option based forward looking approach applied to a metaheuristic's algorithm is capable of avoid huge portfolio losses in case of an extraordinary and sudden event, like the COVID19 pandemic of 2020. In this last section the comparison between the four algorithms will be performed. In order to have a deeper analysis of the case, it will be studied from different points of view and section will be divided in two parts: the comparison of the best portfolio obtained by the algorithms in terms of fitness, and the application of the risk-adjusted performance metrics just introduced. To try to have a wider view on the case study, the algorithm has been run three times and the best results obtained by each algorithm has been selected for the comparison.

4.5.1 Overall portfolio performance

The first and most intuitive way to compare different results has been obtained measuring the overall raw performances of portfolios in terms of expected returns and standard deviation. Results are displayed in Table 4.3.

Overall portfolios performances				
	PSO	PSO FWL	MOPSO NO FWL	MOPSO FWL
Return	-0.004682	-0.002048	0.000111	**0.000259**
Standard deviation	0.037631	0.032802	0.037787	**0.030295**

Table 4.3 – Overall portfolios performances in terms of return and standard deviation

From a simple analysis, data suggest that the FWL approach seems to perform better than the algorithms including historical approach both in terms of return and standard deviation, and so, dominating them in the mean-variance sense.

Then, to simulate a real-world case the performance of the best results in terms of fitness of each algorithm has been implemented in a 10.000€ portfolio. As it can be observed from Figure 1, one of the first things to notice is the better performance in terms of the PSO with the FWL approach with respect of the PSO with the standard variance-covariance matrix, which is coherent with the aim of this dissertation. However, focusing on the MOPSO, it is interesting to notice how they seems to perform in a similar way, but it is clear that the algorithm with the integration of the FWL approach allowed for more stable results, avoiding the negative performance spikes. Figure 15 focuses of the analysis of the portfolio value in the period from 28/02/2020 to 10/04/2020, a time period which shows higher variability than others. Looking at this part of the graph, it confirms what stated before, with the algorithm including the FWL approach which tends to perform better than its competitor.

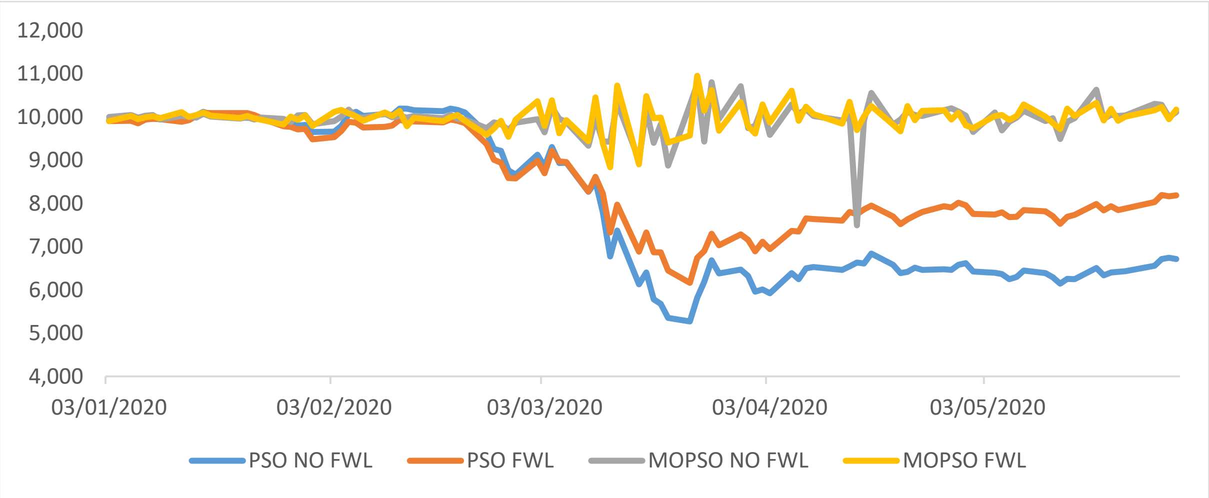

Figure 14 – Portfolio value movement in period 02/01/2020 to 29/05/2020

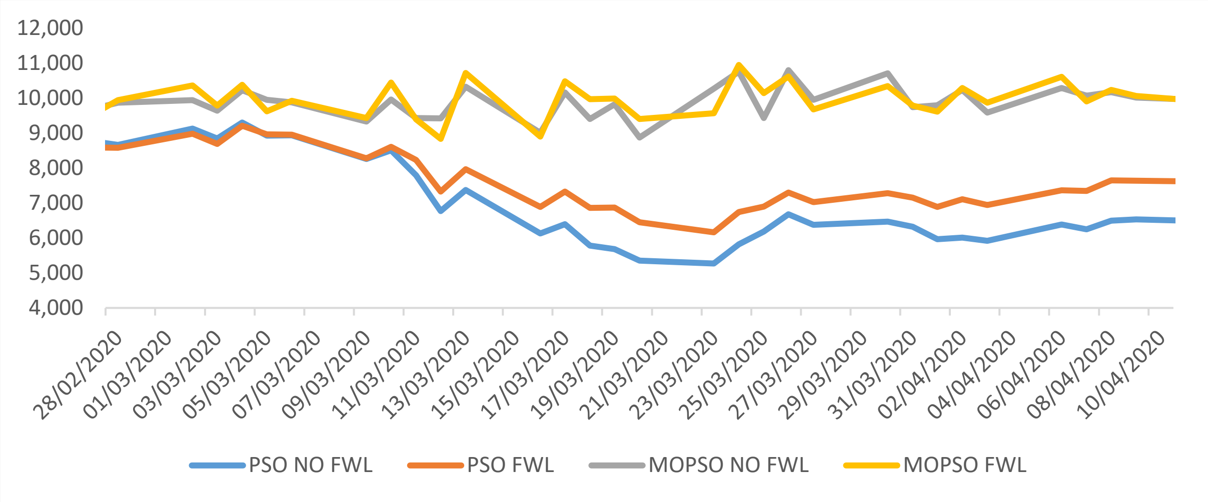

Figure 15 – Portfolio value movement from 28/02/2020 to 10/04/2020

Moving to the algorithm analysis, Table 4.4 reports the weights composing the best portfolio for each algorithm together with fitness values, constraints' violation and final value of the portfolio. It is clearly observable a general worsening of the fitness obtained by the MOPSO algorithm, in particular in the one using the FWL approach. Moreover, with the same algorithm, we note an increase in the value of fitness using the FWL approach, a symptom that the code may need further changes and refinements. Focusing on the constraint's violation, we can observe how the cardinality is always respected, while the return constraints always suffer a

	PSO	**PSO FWL**	**MOPSO**	**MOPSO FWL**
AAPL	0.00	0.00	0.00	0.01
AXP	0.00	0.00	0.00	0.00
BA	0.23	0.00	0.00	0.00
CAT	0.02	0.00	0.01	0.00
CSCO	0.00	0.00	0.00	0.38
CVX	0.00	0.04	0.06	0.00
DOW	0.00	0.06	0.00	0.02
DIS	0.04	0.09	0.00	0.00
GS	0.00	0.00	0.07	0.00
HD	0.00	0.00	0.05	0.00
IBM	0.03	0.00	0.00	0.00
INTC	0.00	0.06	0.00	0.00
JNJ	0.00	0.00	0.00	0.00
JPM	0.01	0.00	0.00	0.00
KO	0.00	0.03	0.00	0.19
MCD	0.00	0.00	0.00	0.06
MMM	0.00	0.12	0.04	0.03
MRK	0.00	0.04	0.00	0.00
MSFT	0.00	0.00	0.00	0.00
NKE	0.02	0.00	0.00	0.20
PFE	0.00	0.00	0.00	0.00
PG	0.25	0.19	0.00	0.03
TRV	0.00	0.00	0.17	0.00
UNH	0.30	0.30	0.01	0.02
UTX	0.00	0.00	0.41	0.00
V	0.03	0.06	0.03	0.00
VZ	0.07	0.00	0.00	0.00
WBA	0.00	0.00	0.00	0.05
WMT	0.00	0.00	0.00	0.00
XOM	0.00	0.00	0.15	0.00
Fitness value	0.002143	0.052865	178.315611	510.694503
Cardinality constraint	0.000000	0.000000	0.000000	0.000000
Return constraint	-0.003291	-0.003247	-0.004579	-0.003255
Short-sell constraint	0.000000	0.000000	0.001564	0.006788
Portfolio final value	6718.19	8190.74	10115.88	10175.14

Table 4.4 – Portfolios weights and algorithms results in terms of constraints violation and fitness

violation, even if minimal. Nevertheless, considering the period of reference and the objective of the thesis, the result is not worrying. Lastly, the short-sell constraint sees violations only in the modified PSO algorithms, confirming what was reported just above.

4.5.2 Time windows analysis

This last part of the application will move the attention to the two different time windows that have been selected for the experiment, in order to give the idea of the behavior of the algorithms in the in-sample and out-of-sample period. As the aim of the dissertation is to test the usefulness of the application a FWL approach to a portfolio, the analysis of the out-of-sample will be deeply to test the effectiveness of the modification to the approach.

In sample analysis

The first part of the analysis takes into consideration the results obtained during the in-sample period. Results of return and standard deviation are reported in Table 4.5, where the best results[87] have been highlighted.

In sample analysis				
	PSO	**PSO FWL**	**MOPSO NO FWL**	**MOPSO FWL**
Return	-0.374562	-0.397473	-0.037478	**0.018478**
Standard deviation	0.014638	0.014195	**0.009016**	0.013465

Table 4.5 In sample period results in terms of return and standard deviation

Of course, given the particular and stressful analyzed period general performances are generally bad for all the cases and the algorithms. Nevertheless, it is interesting to point out the better performance obtained by the MOPSO algorithm with respect of the standard PSO. Looking at the result, the difference in performances is significant, and it is emphasized by the better results obtained even in terms of risk.

Looking at the ratios results, shown in the Table 4.6, the MOPSO algorithm with the FWL approach performs almost always better by returning positive result while the other algorithms perform negatively. The only incoherent result can be observed looking at the Sortino ratio, which is slightly smaller in PSO. Another

[87] As in case of expected return a higher value is preferable, the best value corresponds to the highest one across all the final portfolios obtained. Viceversa, dealing with standard deviation, the best result is the smaller across the proposed.

performance indicator, besides the risk-adjusted ones just introduced is the Maximum Drawdown.

In sample analysis				
	PSO	**PSO FWL**	**MOPSO NO FWL**	**MOPSO FWL**
Sharpe Ratio	-0.255378	-0.277546	-0.038564	**0.007484**
Sortino	**-0.262470**	-0.284800	-0.323367	-0.285312
Information ratio	-0.035129	-0.050376	0.215070	**0.247380**
Treynor ratio	-0.006059	-0.007549	-0.000627	**0.000221**
Maximum Drawdown	0.482853	0.389624	0.306403	**0.169800**

Table 4.6 In-sample period results in terms of risk-adjusted indicators.

Out of sample analysis

Focusing only on the out-of-sample period the results obtained in terms of return and variance are slightly different from the overall ones. Table 4.7 reports expected return and variance of the out-of-sample period.

Out of sample analysis				
	PSO	**PSO FWL**	**MOPSO NO FWL**	**MOPSO FWL**
Return	-0.004602	-0.000762	**0.000388**	0.000363
Standard deviation	0.046159	0.039868	0.047141	**0.036743**

Table 4.7 Return and standard deviation out-of-sample results

Here the results are the opposite of those held in the in-sample period analysis. In fact, the MOPSO algorithm using the historical approach obtained better performance in terms of return but at the expense of a significantly higher standard deviation. The introduction of the FWL approach seems to have mitigated this risk, at the expense of a, albeit slightly, lower return.

The second part of the analysis focuses on the comparison between the financial performance indicators introduced before. In total return the MOPSO algorithm with FWL seems to be in line with the standard MOPSO, however, as already explained, it cannot be considered as a complete performance measure, given that it doesn't consider the risk involved. Table 4.8 summarize the comparison between the performance indicators introduced earlier.

Out of sample analysis				
	PSO	PSO FWL	MOPSO NO FWL	MOPSO FWL
Sharpe Ratio	-0.099702261	-0.019104692	0.008229819	0.009881282
Sortino	-0.083483566	0.002246977	-0.160916498	0.005371901
Information ratio	-0.063613542	0.117723005	0.172005035	0.170829564
Treynor ratio	-0.0043496	-0.000833615	0.000717606	0.000788209
Maximum Drawdown	0.433633612	0.331264401	0.306402632	0.169799569

Table 4.8 Out of sample analysis results in terms of risk-adjusted indicators

From a superficial point of view, the higher the results of these ratios correspond to a better risk-adjusted performance. However, it is interesting to point out that this could be misleading if they assume negative values[88] and so, they will not be taken into consideration in the analysis.

At a first glance, the results reported seem to confirm what already said in the previous section, with the FWL approaches performing better than the ones without this implementation. Moving closer and looking at the Sharpe ratio, it is clear the completely different performances between the standard PSO and the MOPSO algorithms, given that the former show negative values of such ratio. The comparison between the two MOPSO algorithms indicates a better risk-adjusted performance by the one including the FWL approach, coherently with the goal of this thesis.

Moving to the Sortino ratio we can observe result partially different from the analysis of the Sharpe ratio, given by the use of the downside risk instead of the variance. Practically, the main advantage of the Sortino ratio of not penalize returns above the MAR significantly impact the results: the best performance is still obtained by the MOPSO with the FWL approach, but it is followed by the PSO with the same modification applied. Then, the two algorithm without the FWL approach obtained even negative results. After the analysis of the Sortino ratio it is clear that the FWL approach heavily affected the risk-adjusted performances of the portfolio, confirming that the use of a FWL approach can be useful to obtain better results with respect of the standard historical approach.

The analysis of the Information ratio is particularly interesting, being the only risk-adjusted performance ratio which results is inconsistent with the others ratios and with the aim of the dissertation. As it is clearly observable, the best result is obtained by the MOPSO algorithm based on the historical approach. This is due to the better expected returns in out of sample as analyzed before. Anyway, the difference between the MOPSO algorithms with the historical approach and the

[88] In general, risk-adjusted indicators assume negative values in periods in which the excess return (the numerator of the ratios) is negative.

FWL one is not so important to justify a clear advantage of using the historical approach.

Analyzing the Treynor ratio seems to generally confirm what stated in the analysis of the other risk indicators. In fact, the FWL approach applied to the MOPSO algorithm slightly outperform its competitor and the same can be stated for the standard PSO algorithm.

The last analysis was conducted on the maximum drawdown and, once again, the results seem to confirm what was introduced by the study of the FWL approach applied to the variance-covariance matrix as the best results in terms of maximum drawdown is obtained by the FWL approach algorithms followed by the standard historical approach.

Conclusion

The idea this dissertation was to bring evidence of the advantages given by the implementation of a forward-looking approach based on the implied volatility applied to two metaheuristics algorithm. Based only on past data, the algorithm would not have been able to predict the arrival of an unpredictable and economically devastating event causing huge losses to security portfolio holders. Therefore, the objective is to give the possibility to a risk-adverse investor to receive anticipatory signals from the market that allow to decrease the risk of the portfolio and, consequently, the resulting losses. For this reason, a forward-looking component, i.e. a modification in the variance-covariance matrix, has been introduced in the original algorithm. Moreover, also a modified version of the original PSO has been tested to evaluate its flexibility in a forward-looking approach give that it performed always better in a historical data scenario. After the experiment trial the results have been compared together in a simple real case portfolio. Then, main financial performance indicators have been used to access different and more complete information about the results obtained. Although the modification is small, from the results of this work it seems to be able to offer partial improvements in the returns of the portfolio, through the orientation of the algorithm towards less risky assets. However, believing that the forward-looking approach is superior to one based only on historical data, further modifications to Markowitz's original work may be needed, such as introducing forward-looking estimators based on Skewness and Kurtosis such the ones introduced in the Appendix. Indeed, the use of such moments of the statistical distribution, may provide a better estimate of future behavior of the stock price and give the possibility to reach even better results in terms of variance and expected return. From the portfolio composition, further, given the results obtained a rebalancing of the assets detained in the portfolio should be accessed. Moreover, the introduction of changes to the PSO, in addition to the forward-looking approach, return more unstable results between runs and much higher fitness values Furthermore, from a computational point of view, the modification introduced in the PSO algorithms could be applied to different kind of metaheuristics such as Ant Colony Optimizations, Fireworks Algorithm or Genetic Algorithm and to more modern and advance machine learning algorithms.

In conclusion, the introduction of a forward-looking approach to the variance-covariance matrix in PSO algorithms could have had a positive impact in the performances of a portfolio during the COVID-19 financial crisis. However, the sample size, the testing period and the algorithm involved in the analysis might be too small and limited to be considered a conclusive experiment. Further research involving a cleaner and tweaked code, together with the application of a model free approach for the derivation of the implied volatility might be useful to obtain better and more price results.

Appendix A – Figures

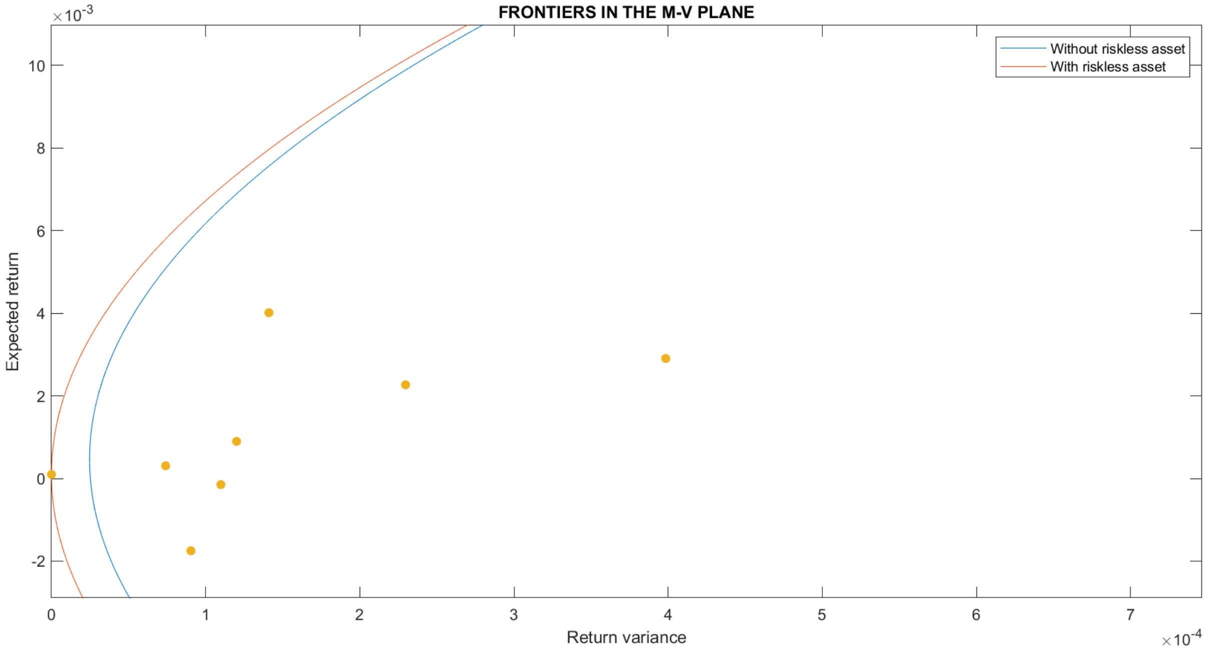

Figure 1 – Comparison between frontiers with and without risk-free asset in mean-variance plane

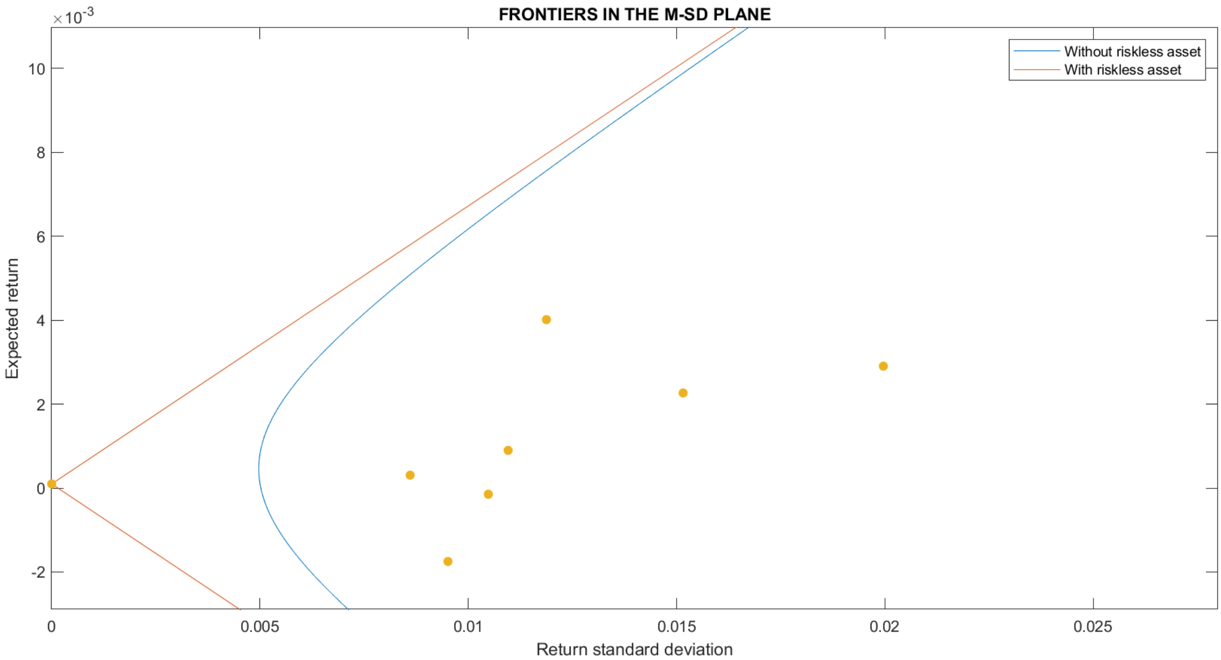

Figure 2 – Comparison between frontiers with and without risk-free asset in mean-standard deviation plane

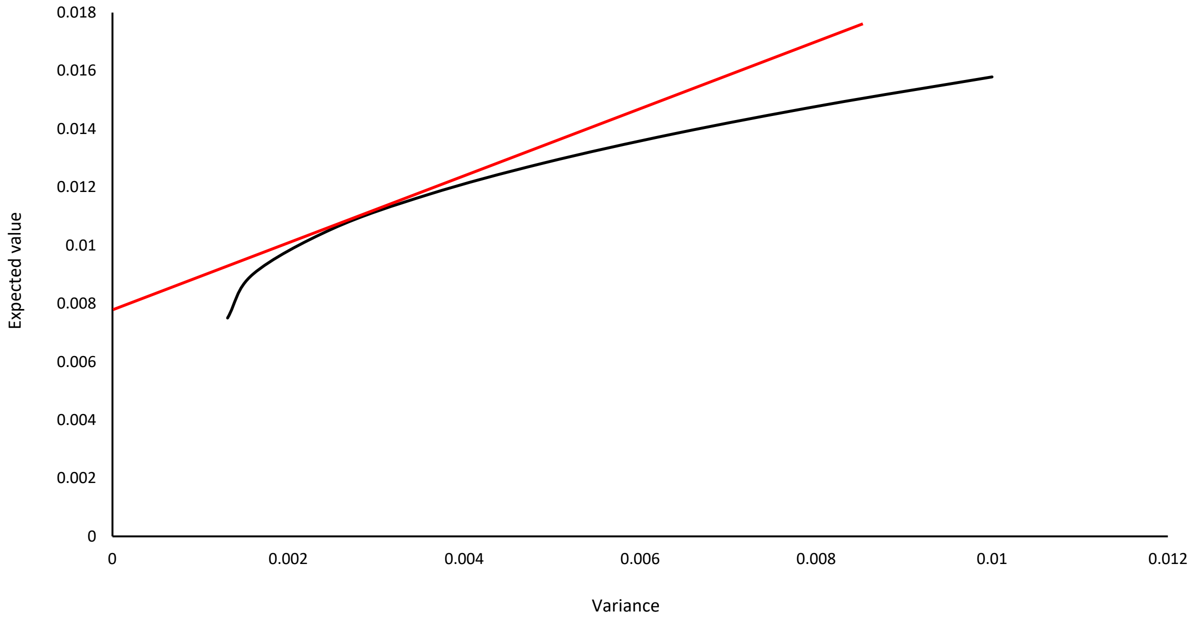

Figure 3 Tangency between the efficient frontier and the risk-free curve

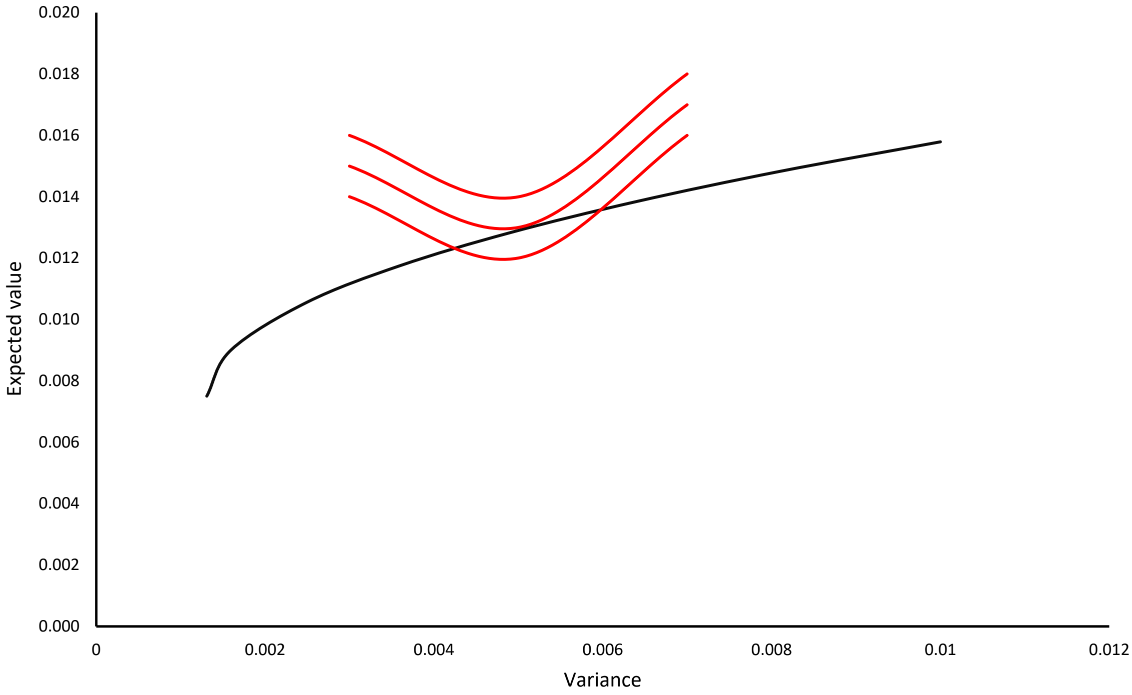

Figure 4 Different expected utility curve. The tangency between the utility curve and the efficient frontier represents the efficient portfolio for that particular investor

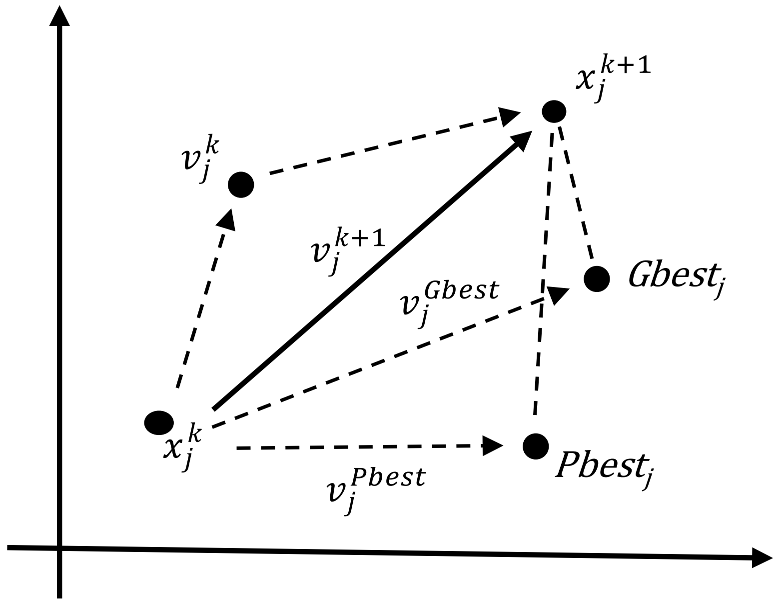

Figure 4 - Velocity and position updates for a single particle during an iteration cycle

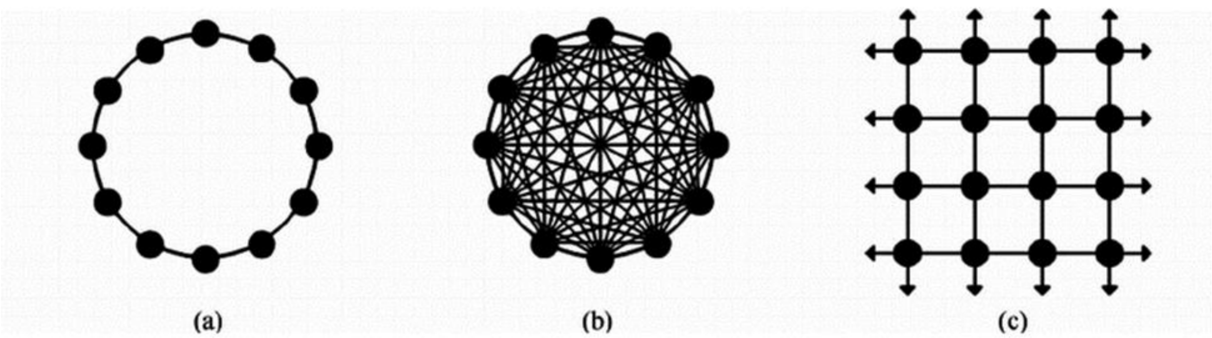

Figure 6 – Metaheuristics population topology

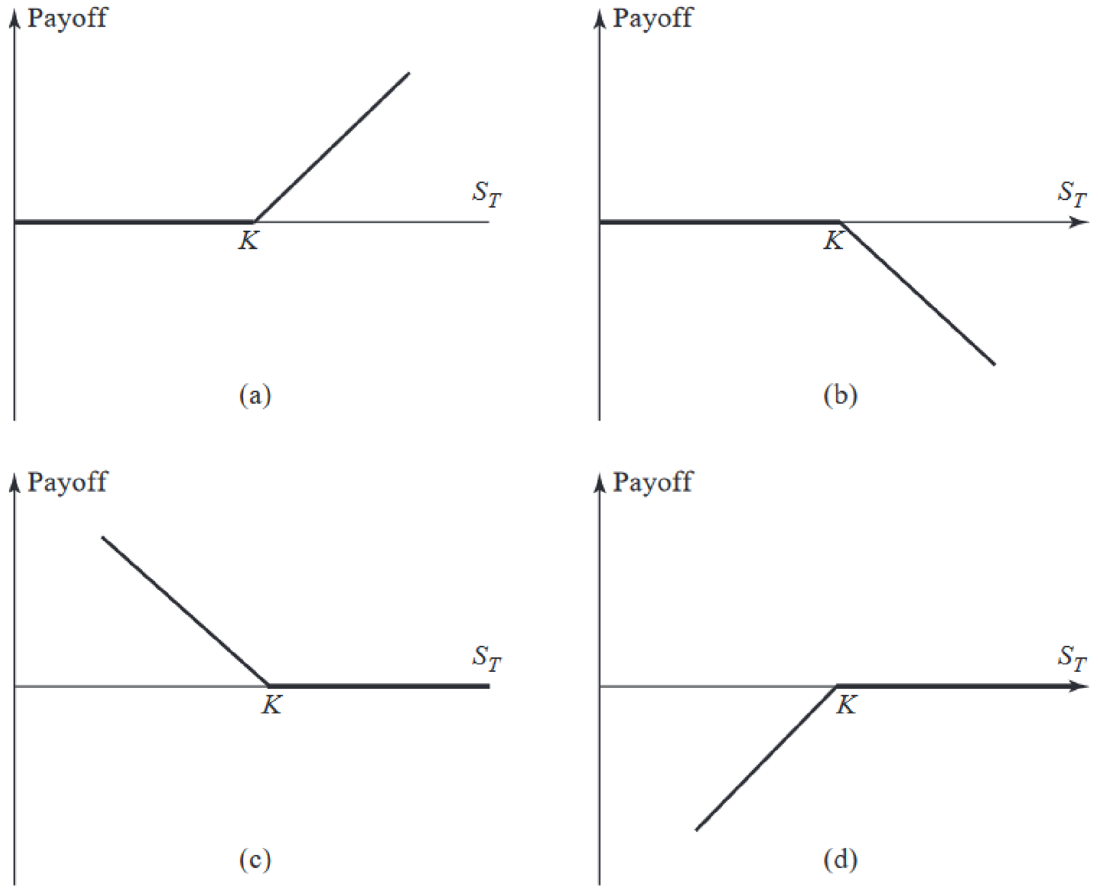

Figure 7 – Payoff of option positions

Appendix B – BICM's higher moments estimators

In chapter 3, the derivation of the estimator based on the second moment have been presented. However, authors demonstrate the derivation of the estimator's based on the third and fourth moments too. To obtain the further members of the family of fully implied estimators it is sufficient to replace the assumptions concerning about the proportion of systematic variance with a corresponding assumption about how systematic risk affects higher moments.

Estimators based on third moments

To derive the skewness-based estimator of covariance, authors assume that the proportion of systematic return skewness is equal for all the assets[89]. Then, denoting this proportion by c_t^{Skew}, the return skewnsess of of the *i*th asset is developed as

$$Skew\,(R_{it}) = \beta_{it}^3\, Skew(R_{mt}) + \left(1 - c_t^{Skew}\right) Skew(R_{it}) \tag{A.1}$$

and, solving for β_{it}

$$\beta_{it} = \left(c_t^{Skew}\right)^{1/3} \left(\frac{Skew(R_{it})}{Skew(R_{mt})}\right)^{1/3} \tag{A.2}$$

Here, again, the conditions that the market beta equals one delivers the proportion c_t^{Skew}. So, solving for c_t^{Skew} and substitute the result in Equation A.2 provide the beta coefficient which leads to the covariance equation

$$Cov(R_{it}, R_{jt}) = \frac{Skew(R_{it})^{\frac{1}{3}}\, Skew(R_{jt})^{\frac{1}{3}}}{\left(\sum_{i=1}^{N} w_{itm}\, Skew(R_{it})^{\frac{1}{3}}\right)^2} Var(R_{mt}), \qquad \forall i \neq j \tag{A.3}$$

Which represents the second member of the family of fully implied covariance estimator.

89 Note that the assumption by Chang, Christoffersen, Jacobs, and Vainberg (2012) that the proportion of systematic skewness equals 100% is a special case.

Estimators based on fourth moments

Finally, kurtosis-based covariance estimator is presented. To derive it as for the second and third moments, the assumption to be done is that the proportion of systematic kurtosis is equal for all N assets. Denoting this proportion as c_t^{Kurt}, with $0 \leq c_t^{Kurt} < 1$, the fully implied covariance estimator can be derived as

$$Cov(R_{it}, R_{jt}) = \frac{Kurt(R_{it})^{\frac{1}{4}} \, Kurt(R_{jt})^{\frac{1}{4}}}{\left(\sum_{i=1}^{N} w_{itm} \, Kurt(R_{it})^{\frac{1}{4}}\right)^2} Var(R_{mt}), \qquad \forall i \neq j \qquad \text{(A.4)}$$

References

[1] Aggarwal, R., Rao, R.P., and Hiraki, T. (1989), Skewness and kurtosis in Japanese equity returns: Empirical evidence, J. Financ. Res. 12, 253-260.

[2] Arditti, Fred D., (1971), Another Look at Mutual Fund Performance, Journal of Financial and Quantitative Analysis, 6(3), 909-912

[3] Bernoulli J., Ars Conjectandi, (Thurnisorium, Basil).

[4] Blum C., Roli A., (2003), "Metaheuristics in combinatorial optimization: overview and conceptual comparison", *ACM computing surveys*, Vol. 35, No. 3, 268-308.

[5] Cont R., (2001), Empirical properties of asset returns: stylized facts and statistical issues, Quantitative Finance, Taylor & Francis Journals, vol. 1(2), 223-236.

[6] Davis, M. H., & Norman, A. R. (1990). "Portfolio selection with transaction costs. Mathematics of Operations Research, Vol. 15, No. 4, 676–713.

[7] DeMiguel V.,Garlappi L., Uppal R., Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy?, The Review of Financial Studies, 2009 22(5) , 1915–1953

[8] Dybvig, P., & Ingersoll, J. (1982). Mean-Variance Theory in Complete Markets. The Journal of Business, Vol. 55, No. 2, 233-251.

[9] J. Tobin, Liquidity Preference as Behavior Towards Risk, The Review of Economic Studies, 1958, 25(2), 65–86

[10] Jagannathan, R. and Ma, T. (2003), Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps. The Journal of Finance, 58: 1651-1683

[11] Kahneman, Daniel & Tversky, Amos, (1979) Prospect Theory: An Analysis of Decision under Risk, Econometrica, Econometric Society, vol. 47(2), pages 263-291.

[12] Konno, H., Wijayanayake, A. Portfolio optimization problem under concave transaction costs and minimal transaction unit constraints. Math. Program. 89, 233–250 (2001)

[13] Kourtis, A., A Stability Approach to Mean-Variance Optimization. Financial Review, 2015, 50, 301-330

[14] Ledoit, Olivier and Wolf, Michael, (2003), Improved estimation of the covariance matrix of stock returns with an application to portfolio selection, Journal of Empirical Finance, 10, issue 5, p. 603-621

[15] Levin, Leonid (1973). "Универсальные задачи перебора" [Universal search problems]. Problems of Information Transmission (in Russian). 9 (3): 115–116. Translated into English by Trakhtenbrot, B. A. (1984). "A survey

of Russian approaches to perebor (brute-force searches) algorithms". Annals of the History of Computing. 6 (4): 384–400.

[16] M.J. Brennan, The Optimal Number Of Securities In A Risky Asset Portfolio When There Are Fixed Costs Of Transacting: Theory And Some Empirical Results, Journal of Financial and Quantitative Analysis 10(3) (1975) 483–496

[17] Mansini R. & Speranza, M.R., "Heuristic algorithms for the portfolio selection problem with minimum transaction lots," European Journal of Operational Research, 1999, 114(2), 219-233.

[18] Corazza, M., Fasano, G., Gusso, R.: Particle Swarm optimization with non-smooth penalty reformulation, for a complex portfolio selection problem, Applied Mathematics and Computation 224, 2013, 611-624.

[19] Marco Corazza, Daniela Favaretto, On the existence of solutions to the quadratic mixed-integer mean–variance portfolio selection problem, European Journal of Operational Research, Volume 176, Issue 3, 2007, Pages 1947-1960.

[20] Markowitz H., Portfolio Selection: Efficient Diversification of Investments. Yale University Press, 1959.

[21] Markowitz, H. (1956), "The optimization of a quadratic function subject to linear constraints", Naval Research Logistics Quarterly, Vol. 3, No. 1-2, 111–133.

[22] Miguel, Victor, Garlappi L, Nogales J., and Uppal, Raman, (2009), A Generalized Approach to Portfolio Optimization: Improving Performance by Constraining Portfolio Norms, Management Science, 55, issue 5, p. 798-812.

[23] Plyakha, Yuliya and Uppal, Raman and Vilkov, Grigory, Why Does an Equal-Weighted Portfolio Outperform Value- and Price-Weighted Portfolios? (2012).

[24] Pogue, G. A. (1970). An extension of the Markowitz portfolio selection model to include variable transactions' costs, short sales, leverage policies and taxes. The Journal of Finance, Vol. 25, No. 5, 1005–1027.

[25] Rubinstein M. (2002), "Markowitz's Portfolio Selection: a 50-year retrospective", The Journal of Finance, Vol. 57, No. 3.

[26] Stephen A. Cook. 1971. The complexity of theorem-proving procedures. In Proceedings of the third annual ACM symposium on Theory of computing (STOC '71). Association for Computing Machinery, New York, NY, USA, 151–158

[27] T.-J. Chang, N. Meade, J.E. Beasley, Y.M. Sharaiha, Heuristics for cardinality constrained portfolio optimisation, Computers & Operations Research, 2000, 27(13), 1271-1302

[28] Tobin, James (1958b). "Liquidity Preference as Behavior Towards Risk". Review of Economic Studies. 25.1 (2): 65–86.

[29] Yoshimoto A., The Mean-Variance Approach to Portfolio Optimization Subject to Transaction Costs, Journal of the Operations Research Society of Japan, 1996, 39(1), 99-117

[30] J. Kennedy and R. Eberhart, "Particle swarm optimization," *Proceedings of ICNN'95 - International Conference on Neural Networks*, Perth, WA, Australia, 1995, pp. 1942-1948 vol.4

[31] Javier Del Ser et al., Bio-inspired computation: Where we stand and what's next, Swarm and Evolutionary Computation, 2019, Volume 48, 220-250.