

Chapter 2

Related Work

Abstract In this chapter some of the fundamental concepts necessary to understand the developed work are addressed, particularly the domain relative to financial markets. Further, a substantial part of the several methodologies applied to the portfolio problematic are analyzed; throughout the first two sections, the problem related with portfolio theory and investment's analysis is presented. Subsequently, the evolutionary techniques which can be used to solve this problem are focused. Finally, Sect. 2.4 presents the connection between the presented financial domain and the evolutionary techniques, through an extended analysis on the existing solutions.

2.1 Portfolio Theory

A financial portfolio [1] consists of a group of financial assets, also called securities or investments, such as stocks, bonds, futures, CFDs, or groups of these investment vehicles known as exchange-traded-funds (ETFs). In order to one construct a portfolio, it is capital to define investment objectives that should focus on a certain and accepted degree of risk, i.e. the chance of incurring in a loss.

The core of this work is related to *portfolio management* [1], the act of deciding which assets need to be included in the portfolio, how much capital should be allocated to each kind of security and when to remove a specific investment from the holding portfolio. During this process, it is required to take into account the investor's preferences since some investors are more willing to accept a specific degree of risk than others, hoping that way to achieve better returns.

2.1.1 Diversification

As it was explained in the paragraph above, one of the fundamental goals of any investor consists on reducing his portfolio's risk. The main technique used in finance to reduce this chance of losing capital, is called diversification [1]. Diversification means that the risk needs to be spread, mixing a variety of investment vehicles, in order to minimize the loss impact of one investment in the portfolio. To understand better this concept, it is important to distinguish between two forms of risk [2]:

- *Systematic risk* the risk inherent to a market segment or the entire market which cannot be removed through diversification. Wars and economic recessions constitute an example of this kind of risk;
- *Specific risk* also known as unsystematic risk, which corresponds to the risk related to a short number of assets. Company's strikes, accidents or specific news affecting one company can be caste as unsystematic risks, which can be easily surpassed recurring to diversification.

Independently of the diversification degree of a portfolio, it is fundamental to understand that the intrinsic risk can never be shrank down to zero since there is always a form of risk (systematic) which cannot be removed. However, using a risk-managing technique, such as diversification, the specific risk can be easily reduced.

This methodology can be accomplished with the help of strategies such as the following:

- Selection of different investment vehicles, such as stocks, bonds, futures or ETFs;
- Mixing assets from different industries, countries, and sectors.

Defined this concept, it is capital to understand that diversification cannot guarantee that a losing investment is avoided. However, it can prevent loss, reducing the impact of a specific investment in the overall portfolio.

2.1.2 Management

As it was already mentioned, the goal of this work is concentrated on the automatic management of a portfolio. It is important to notice that distinct forms of management can be applied [1]:

- *Passive Management* in which the investor concentrates his objective on tracking a market index. This is related to the idea that it is not possible to beat the market index, as stated by the Random Walk Theory [3]. More concretely, a passive strategy aims only at establishing a well-diversified portfolio without trying to find under or overvalued stocks;

- *Active Management* in which the main goal of the investor consists on outperforming an investment benchmark index, buying undervalued stocks and selling overvalued ones.

Although the differences between both forms of management seem quite clear, the question active versus passive is still widely debated. Most part of the mathematical formulations used to model the aspect of portfolio optimization problem, such as the Mean–Variance Model, proposed by Markowitz [4] and which is considered as the holy grail of portfolio management theory are classified as passive management. However, as stated by Beverly Goodman [5], “passive management strives to beat (and historically does beat) the overall market by proper asset allocation and cost management.” According to him, most people wrongly think that diversification implies reducing risk while getting the market average returns. However, as stated in his article “they didn’t give (economist) Harry Markowitz the Nobel Prize for coming up with a theory that generates average returns.” For instance, Aranha and Hitoshi [6] proposed a portfolio optimization application based on the Markowitz’s model which constantly beats the index over distinct periods.

What should be understood here is the fact that passive management also tries to beat an index as the active form. At first sight, the risk and transaction costs involved when using a passive strategy are not so high when compared to an active one. Probably, most of the published articles apply this kind of approach for that reason. However, it should be noticed that an active strategy, using technical indicators, can possibly guarantee us with higher profitability levels.

In this work, both solutions are examined when coupled with evolutionary computation techniques.

For more information on the Markowitz’s model, the reader is referred to Appendix A.

2.2 Market Analysis

When defining a financial fund or portfolio our goal is to pick the best potential assets within the market in order to avoid losses and maximize our returns. There are several ways to perform a reasonable evaluation of the market and select potential profitable securities. Usually, investment analysts perform a fundamental or a technical analysis of the market. Notice that these strategies are not exclusive and both can be applied. However, a fundamental analyst tries to avoid the antagonist approach.

2.2.1 Fundamental Analysis

Fundamental Analysis [7] evaluates each security by measuring its intrinsic value through the study of overall economy, industrial conditions, and the financial situation of a specific company. When this intrinsic value is calculated it is

compared against the current security's price. If this value is inferior to the market value, then the market is possibly overvalued and the investor should sell all the company's shares, otherwise the market is undervalued and it can be extracted a potential buy signal. Summarizing the previous definition; fundamental analysis tries to understand the factors which can possibly affect market behaviour. After a rigorous probe the investor should be able to answer the following questions, and subsequently take a decision:

- Is the company's income (from business activities) growing?
- Is the company actually generating profit?
- Is the company able to repay the assets owed?

2.2.2 Technical Analysis

At the other end of the rope there is Technical Analysis [7]. A technical analyst believes that market action, namely the volume of transactions and the securities prices include all the fundamentals that can possibly affect market's price; political, economical, or psychological. Following this premise there is only the need to study those factors in order to forecast market behaviour. The applied strategies on technical analysis normally embody a set of technical indicators which try to give us a future perspective of market development according to what is visible on price charts. A technical indicator consists in a formula that is normally applied to stock's prices and volumes. The resulting values are plotted and then analysed in order to offer us a perspective on price evolution. More specifically, a technical indicator tries to capture the behaviour and investment psychology in order to determine if a stock is under or overvalued.

In order to illustrate the behaviour of such approach, the technical analyst starts by applying a simple technical indicator as the Simple Moving Average (SMA). The SMA plots per each day, the average on prices observed during the last x days. Depending on the considered data, it is also possible to employ the indicator to weekly or monthly prices. The following picture illustrates the usage of a moving average with a duration period of 12 weeks when applied to Intel weekly prices. Notice that the blue line identifies the stock price and the red line corresponds to the SMA. Observe the smoothness on the SMA line, which allow us to easily perceive the market movements (Fig. 2.1).

Regarding an indicator such as the former one, a strategy for defining buying and selling signals can be formulated:

- *Entry Signal*. Price line crosses above the SMA line;
- *Exit Signal*. Price line crosses below the SMA line.

Based on entry/exit signals and other plot characteristics different rules can be defined which allow us to score the distinct stocks within the market and subsequently pick the best securities according to the indicators employed.

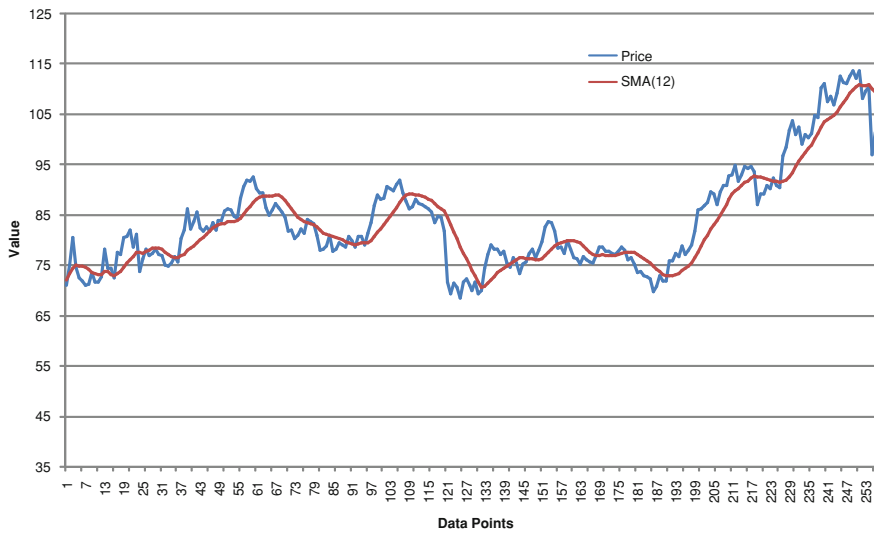


Fig. 2.1 SMA application

2.2.3 Fundamental Versus Technical

The big question which normally an inexperienced investor can formulate is; which kind of market analysis should he employ, a fundamental or a technical one?

The major drawback present on a fundamental analysis approach corresponds to the difficulty on obtaining such data and the fact that most part of this financial data is not reliable due to company's self interests. Besides, technical analysis already includes fundamental analysis because if a technical analyst believes that all factors that can possibly influence the price are already included on it, then they only need to be studied in order to evaluate the market and consequently forecast its development. Another major difference between both forms of market analysis is the time-horizon used when investing. The financial data used by a fundamental analyst is only released over long periods of time. Normally, each company announces its results following a quarterly basis, which is completely different from using daily or weekly data, such as the volume or price data employed on technical analysis.

Although both strategies seem to be on opposite sides they can coexist. Several analysts can couple fundamental data and technical analysis to provide an efficient evaluation of the market. For instance, first use fundamental analysis to pick potential profitable companies and then technical analysis for defining entry and exit signals.

2.3 Evolutionary Computation

One of the major concepts presented within this work is the subfield of artificial intelligence designated as Evolutionary Computation [8]. This methodology embodies the application of a procedure based on biological mechanisms of evolution which tries to progress iteratively to converge on an optimal solution for a combinatorial optimization problem. Normally, these evolutionary techniques involve metaheuristic optimization algorithms, i.e. algorithmic frameworks specialized in solving optimization problems. These metaheuristics, besides being based on biological evolution, can also have their groundwork on a naturally appearing phenomenon, such as Simulated Annealing (SA) and Tabu Search (TS). Since major part of this work is based on evolutionary procedures, further sub-sections will start by explaining some of these concepts, namely the Genetic Algorithms (GA) and Genetic Programming (GP).

2.3.1 Genetic Algorithms

A Genetic Algorithm corresponds [9] to a search technique used to find optimal or sub-optimal solutions to search problems. Its behaviour is inspired on evolutionary biology, by defining an initial set of random solutions, which is iteratively refined, until an optimal or a sub-optimal solution to the problem is encountered. The following diagram tries to express the behavioural process defined by the standard GA (Fig. 2.2):

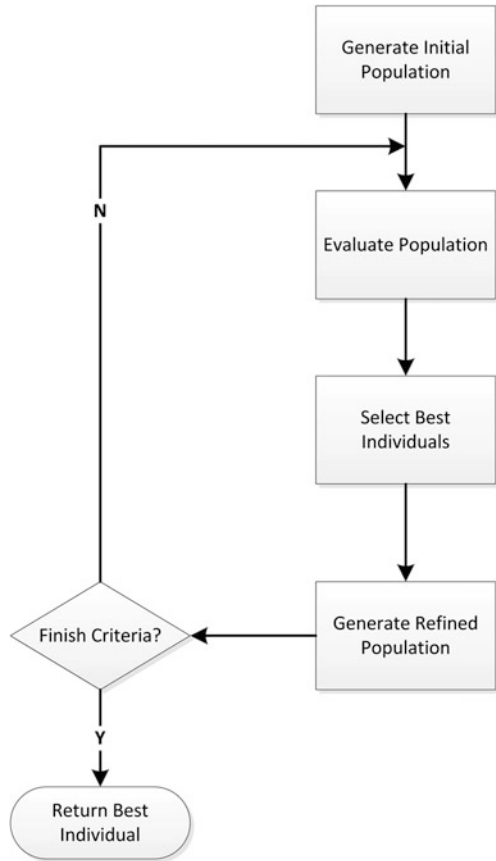
As you can see from the above figure, the algorithm proceeds as following:

- The execution starts by generating an initial population, a set of potential solutions for the problem, randomly defined;
- Following, the initial population is evaluated by a fitness function, also designated as evaluation function. Based on the values previously calculated, the best individuals are selected for reproduction. A set of operators are applied to those individuals, in order to generate a more refined population;
- If a specific finish criterion is fulfilled, for instance, the best individual has the desired fitness value, the algorithm terminates, and the best individual, i.e. the best solution is returned, otherwise this new refined population is evaluated and the same process is applied.

2.3.1.1 Individual Representation

Depending on the target problem, the first step when defining a genetic algorithm consists on specifying the representation of each solution, also designated as individual or chromosome. Normally, a chromosome is represented by a set of variables, also known as genes, depending on the considered problem.

Fig. 2.2 GA general behaviour



2.3.1.2 Initial Generation

After specifying the underlined representation it is necessary to define the initial set of individuals. According to the standard genetic algorithm, a set of random individuals is initially created, which means that random values are assigned to the variables contained within each chromosome, in order to cover a vast area of the search space.

2.3.1.3 Selection

During each iteration of the algorithm it is fundamental to pick the best individuals, i.e. solutions with the best fitness according to the evaluation function, in order to guide the search more effectively. These individuals are chosen according to a specific procedure designated as Selection. The selection operator is

responsible for selecting the chromosomes for reproduction. The fitter the chromosome, more likely it is that it will be reproduced. Several selection procedures are available [9]. On [Chap. 3](#) further details will be given on selection procedures.

2.3.1.4 Offspring Generation

After the selection operation, the picked chromosomes are combined, and subsequently generate new individuals designated as offspring. The application of this procedure is fundamental to guide the search space. This operation is normally known as Crossover, it works as a reproduction function, it combines the characteristics of two individuals, the parents, and generates a new individual, or more than one, the offspring. Like the selection procedure, several crossover operators are also available [9].

Besides the crossover procedure, another function which is normally applied corresponds to the Mutation operator. The mutation procedure is fundamental within a GA in order to avoid the algorithm to concentrate on a specific search space, converging too quickly on a local maximum. A mutation operation normally corresponds to a random alteration on the genes of a specific or random chromosome. On current literature, there are several ideas on how to apply this procedure, depending on the considered chromosome representation and the meaning given to a specific parameter called Mutation Rate. The reader is referred to [10] for further details.

2.3.2 Genetic Programming

A Genetic Programming [9] procedure consists on applying a GA to write computer programs. The variables correspond to different program constructs and the algorithm tries to find the one which best achieves its goals. A simple way of viewing a genetic program can be defined as the following:

- Assume distinct numbers and several operators are available. Then the goal consists on determining the equation that best achieves a specific goal; for instance, return the maximum value as possible, given this set of operands and operators.

To solve a problem such as the presented one the reader could opt for a genetic programming where each solution or chromosome is represented as a tree structure (traditional representation for GP). Within the tree structure; a node represents an operator and each terminal node an operand. [Figure 2.3](#) provides an example of the stipulated representation.

As can be observed from the previous figure, the tree is evaluated in a recursive manner and from that the following equation can be extracted:

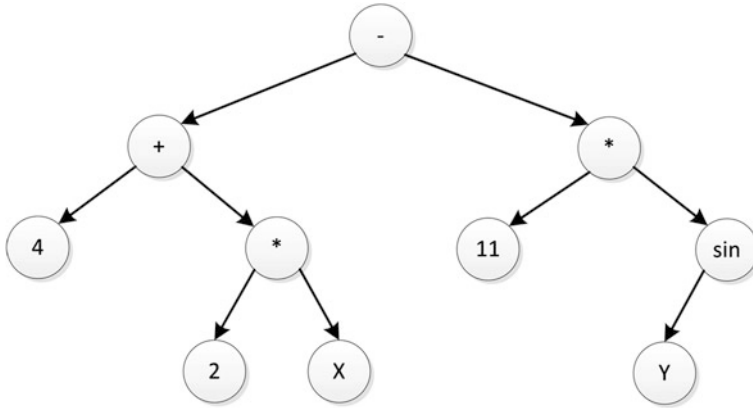


Fig. 2.3 Tree structure for GP

$$(4 + (2 * X)) - (11 * \sin(Y)) \quad (2.1)$$

Iteratively evolving the algorithm as defined under the previous section, by applying a set of genetic operators, it is possible to achieve the equation which best fits our purposes, returning the maximum value as possible, in this example.

2.4 Existing Solutions

Through this section a substantial part of the work developed in this domain is presented. The works here addressed use optimization techniques by evolving two different ways on handling this problem. The first one, given within [Sects. 2.4.1, 2.4.2 and 2.4.3](#) consists on coupling optimization algorithms with mathematical models for portfolio optimization. The procedure's goal concentrates on splitting a fixed amount of capital between different securities, each one with a specific weight within the portfolio, and majorly maintaining a passive management approach. The second strategy, given under [Sect. 2.4.4](#) involves the use of technical and fundamental analysis to define the portfolio composition, based on an active management approach.

2.4.1 Portfolio Optimization Theory

Through this section the main mathematical formulations used to model the portfolio optimization problem are addressed, more specifically, the principles employed to calculate the risk and return measures when coupling portfolio mathematical models with optimization techniques such as GAs.

2.4.1.1 Markowitz's Pioneer Work

The problem related to portfolio management suffered a major revolution during the fifties with Harry Markowitz [4]. The author is pioneer in the Modern Portfolio Theory (MTP) after analyzing the effects related with risk, correlation and diversification over the expected returns of investment portfolios.

After completing his study, Markowitz concluded that rational investors should diversify their investments, in order to reduce the respective risk and increase the expected returns. The author's assumption focus on the basis that for a well-diversified portfolio, the risk which is assumed as the average deviation from the mean, has a minor contribution to the overall portfolio risk. Instead, it is the difference (covariance) between individual investment's levels of risk that determines the global risk.

See appendix A for more details regarding the Markowitz's model.

2.4.1.2 Alternative Models

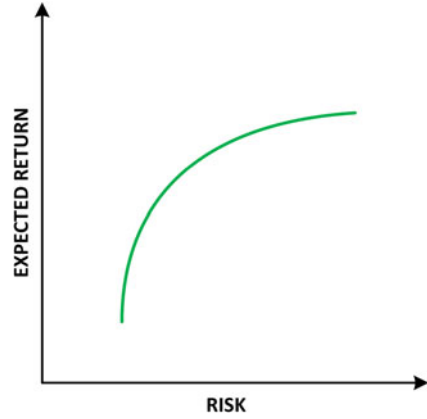
Although Markowitz's model is widely used to design the portfolio optimization problem, other models can also be considered. For instance, Black and Litterman [11] suggested a new formulation, the Black-Litterman model. In their work they propose means of estimating expected returns to achieve better-behaved portfolio models. The designed model is very similar to Markowitz's one, the main difference is concentrated on the calculation of the expected returns which generates portfolios considerably different when using the original model.

According to the authors their new design tries to rectify some of the flaws presented on Markowitz's one. They address the fact that the "expected returns are very difficult to estimate and that historical returns provide poor guides to future returns" when using the patriarch model. Also, one of the major drawbacks pointed to the original model is the time that is necessary to compute the covariance matrix from historical data and solving the resulting problem. With recent technology this problem is not an issue anymore.

Although this new model could seem a better approach according to the authors and more recent studies [12], its implementation to portfolio optimization is not so common. The main reason is due to its complexity and also because the Markowitz's one is widely used by security analysts with the respective evidence given. However, it is shown that this new model has been increasing in popularity.

Other critics pointed to the original model revealed that it fails on capturing the real essence of risk, which is the chance of incurring in a loss. Sing and Ong in one of their works [13] proposed a new method of calculating it which results on portfolios less risky than the ones generated by the Markowitz's model, when both are compared using the same risk measure. Although this new approach, normally designated as Downside Risk Framework has gained some interest by portfolio managers, the argument which states its advantage, in respect to produce less risky portfolios, cannot be of major importance when choosing one of the models since in real-life, investors are more concerned with the total return of the portfolio than with risk. More details about this comparison are addressed by Cheng and Wolverton [14].

Fig. 2.4 The efficient/pareto frontier



2.4.2 Solving Markowitz's Model

Nowadays there are several techniques which can be employed to compute an efficient combination of the portfolio's expected return and the variance between its assets, in order to follow Markowitz's maxim. More concretely, these methods concentrate their efforts on computing the Efficient Frontier (EF), a line composed of optimal portfolios. In the following, different methodologies used to calculate the EF are explored.

2.4.2.1 Quadratic Programming

Given Markowitz's model presented in appendix A, if the problem is solved as a function of R , one can obtain a set of optimal solutions which constitute the efficient frontier. This curve, also known as Pareto Frontier gives for each expected return the minimum associated risk. As stated by Markowitz, from the EF the set of all efficient portfolios can be obtained. A financial portfolio is efficient if for any given expected return there is no other portfolio with a lower risk, and for any given risk there is no other portfolio with a higher value of expected return. Figure 2.4 exemplifies the EF.

Given a set of assets, there are several tools capable of computing a single point in the efficient frontier or the whole curve. If the goal is to compute a single point, then a Quadratic Programming (QP) solver is sufficient, given Markowitz's model. A list of QP solvers can be found at [15].

If the objective is to compute the whole frontier, then a subtle change in Markowitz's model is necessary; the expected return RP is removed from the set of constraints and its maximization is added as a new objective. In order to calculate it, it is possible to use an active set algorithm for QP such as the Critical Line Algorithm (CLA) [2].

2.4.2.2 Modeling Real World

Although Markowitz's model became revolutionary for the portfolio selection problem, it's important to take into account that his design only forms a theoretical point of view since in the real world much more restrictions are necessary to consider, like transaction costs or industrial regulations.

At the present day, when applying a computational procedure to solve this problem in a real world, several restrictions are considered, namely cardinality constraints, buy-in thresholds, floor, ceiling, round-lots and transaction cost inclusions. In order to get a better understanding about these restrictions, the following definitions are presented:

- *Cardinality Constraints*. The maximum and minimum number of assets that a portfolio manager wants to include in the portfolio;
- *Floor and Ceiling*. The lower and upper proportion limit specified for each security;
- *Buy-in Thresholds*. Common name to design the floor constraint;
- *Round-lot*. The number of any asset included in the portfolio must be multiple of normal trading lot (100). This constraint is applied in several of the presented publications. However, nowadays, it is not applied anymore;
- *Budget Constraint*. Requires that all the capital should be invested in the portfolio.

Taking into account these more complex constraints, two different approaches can be used to find solutions for the portfolio selection problem. One uses a suitable mixed integer solver, and the other one, metaheuristics to compute the solution. Stein et al. [16] explored both techniques and defined the respective advantages and disadvantages. They concluded that using a mixed integer approach has a major drawback since exact solutions are unsuccessful when applied to large-scale problems. Despite the fact that the use of metaheuristics comes with several handicaps, such as the requirement of extensive parameter tuning which implies the realization of a variety of tests in order to find the appropriated values, most of the articles recently published focus on this methodology since they are capable of finding reasonable solutions very quickly, allow the use of alternative risk measures, and can be easily applied to different models of the problem. The most usual metaheuristics applied on the portfolio selection problem are Genetic Algorithms (GAs), Simulated Annealing (SA) and Tabu Search (TS).

2.4.3 Metaheuristics Approaches to Portfolio Optimization

During this section, several heuristic approaches to solve the portfolio optimization problematic are addressed as well the respective variants.

2.4.3.1 Single-objective Evolutionary Algorithms

Starting with GAs, the first approaches to emerge consisted on considering a single-objective optimization problem by using a trade-off function relating risk and return, instead of considering a Multi-objective Optimization Problem (MOOP) where the goal is to optimize simultaneously two conflicting objectives, in this case, minimizing risk and maximizing the return of the portfolio. This original approximation was made by Loraschy et al. [17]. Two years later the same authors proposed a distributed version of their former algorithm with much better results [18]. Instead of considering the variance as a risk measure, as it is proposed by Markowitz, they opt to use the Downside Risk approach, referred on Sect. 2.4.1.2. Their distributed version was based on an island model where a GA is used with multiple independent subpopulations running on distinct processors. From time to time, highly-fit individuals migrate between those subpopulations.

Later, in 2000, Chang et al. [19] conducted an investigation where they experimented a variety of metaheuristics, namely GAs, SA and TS. The accomplished tests on deciding which heuristic performed better were not conclusive. First, they tried to check which one was the best to approximate the efficient frontier taking into account the original Markowitz's model. The results showed that genetic algorithms were the best approach, immediately followed by simulated annealing and by last, tabu search. In the second experience, they start to enrich Markowitz's model with cardinality constraints and then applied the algorithms. This time the differences between the three heuristics were not so clear, concluding that the best approach was to run all three and combine their results. Again, the same consideration was used in respect to have a single objective which relates risk and return, and that needs to be minimized. This claim was also confirmed by Buseti [20]. However, in 2002, Schaerf [21] developed an improved version of Chang et al.'s TS algorithm, after combining different neighborhood relations. His results were contradictory with the work already done since he proved that TS performed clearly better when compared with SA.

In all the referred proposals [17–19] the portfolio's used representation was based on two distinct lists, one identifying the assets included in the portfolio (Q) and another one with the respective investment allocation (S), as defined in the following example:

$$Q = \{\text{AMZN}, \text{GOOG}\} \quad S = \{0.6, 0.9\}$$

The portfolio is composed by two distinct securities, AMZN and GOOG, and the respective investment allocation corresponds to a total of sixty percent on Amazon and ninety percent on Google.

Notice that not all portfolios considered representations correspond to feasible solutions, i.e. solutions where the considered constraints are not violated which explains why the sum of the percentage allocations is not equal to 100 %. When applying optimization methods as the mentioned ones, several considerations can be made on how to handle these infeasible solutions.

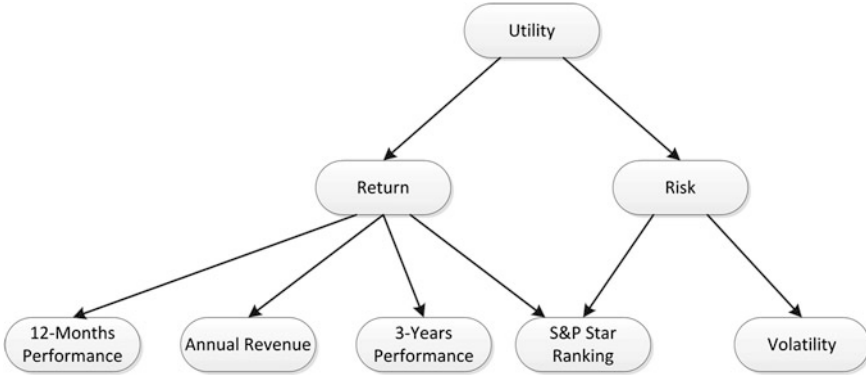


Fig. 2.5 Objective decomposition

Later, Crama and Schyns [22] developed a more sophisticated approach of the SA algorithm. They started to enrich Markowitz’s classical model with additional realistic constraints, such as floor and ceiling, turnover, trading and cardinality constraints, solving the problem via a SA algorithm. Their portfolio’s representation assumed to be the same as the former proposals. As the previous authors who approached this problem, the most difficult task was in how to handle the considered constraints. Solving this question, they concluded that the proposed method and similar ones like genetic algorithms were versatile enough, not requiring any modification, in case of considering other risk measures or arbitrary constraints.

Other approaches using completely different metaheuristics were also tried. Cura [23], for instance, used a Particle Swarm Optimization (PSO) technique. The author compared his heuristic performance with the three heuristics used by Chang et al. [19]. The results showed that none of the tested methods clearly outperformed the others, although this new model gave better results “when dealing with problem instances that demand portfolios with a low risk investment”. Similar comparisons with TS, GAs and SA were made by Fernández and Gómez [24] when using Artificial Neural Networks (ANN).

Regarding the use of these simple metaheuristics, another interesting publication was made by Ehrgott et al. [25]. The three authors proposed an extension to the Markowitz’s model. Instead of considering only the risk and return associated to the portfolio, they used an alternative decision criteria based on an objective hierarchy. They establish a decomposition of risk in two criteria, the volatility of an investment and an S&P investments fund ranking. The return was split in four objectives, such as 12-Month Performance, 3-Year Performance, annual revenue and also the S&P ranking. Defined this model based on a Multi Decision Criteria, they applied SA, TS and GAs to solve the resulting problem. After comparing the different heuristics, the GA approach seemed to be the most reasonable one, presenting better results. The proposed model can be defined by the diagram presented in Fig 2.5.

A similar approach was taken by Lin and Gen [26] after considering a multi-stage decision-based algorithm. They start to select 20 % of the considered assets

based on the past 3-Month Performance returns and only then apply the algorithm, in this case a genetic one. Their conviction settles on the fact that this initial process of restraining the set of considered securities can produce portfolios with higher returns.

2.4.3.2 Multi-objective Evolutionary Algorithms

Subsequently, the first approaches using Multi-objective Evolutionary Algorithms (MOEAs) start to arise, being Streichert et al. [27] the patriarchs. The authors made several tests using different solution representations and considering three real world constraints, namely cardinality constraints, buy-in thresholds and round-lots.

First, they formulated the problem to optimize, extending Markowitz's model with the mentioned constraints. Since it is a MOEA, their goal was to optimize two conflicting objectives, maximizing return and minimizing risk. Besides their different representation evaluation, they also employed a Memetic Algorithm (MA). A MA extends an EA approach by adding a new procedure on the EA process. This new step consists on performing a local search algorithm before evaluating a population in order to refine its individuals. This mechanism updates the decision variables (allocation investment percentages) so they can be inherited to the next generation which is known as the Lamarckism mechanism. In their case, the local search algorithm was applied to convert an infeasible solution to a feasible one which respects the considered constraints (cardinality, buy-in, round-lots). Their solution achieved better results when compared with one which tries to punish infeasible individuals.

Until that time, almost all the published works which were based on the use of GAs to solve the portfolio selection problem focused their genomic representation on a real-valued array [26–28] where each element represents the investment allocation on a specific asset, a binary string array [29] where each element expresses the asset allocation on a binary form, a hybrid approach [28, 30] where a real-valued array is used with a bit mask array; the value one indicates the inclusion of the asset on the portfolio and zero its absence, or recurring to the use of two distinct lists Q and S , as it was already mentioned. Streichert et al. [29] were the first to address an experiment in order to determine which representation was the most appropriated to handle the portfolio representation. They easily concluded that a hybrid representation where the investment allocation is represented by an array value clearly surpasses the other ones. In order to understand better this kind of approach, the following table is provided (Table 2.1).

The presented representation is easy to understand, a genome identifies a portfolio composed by five assets. The assets AMZN, INTL and YHOO are included on the portfolio with the respective allocations expressed by the weight's array. These values are then changed in order to maintain the model constraints, such as the budget constraint which specifies that the weight's sum is equal to one.

Table 2.1 A portfolio hybrid representation

Stock	AMZN	GOOG	INTL	MSFT	YHOO
Inclusion	1	0	1	0	1
Weight	0.32	0.17	0.02	0.44	0.12

Table 2.2 Fuzzy rules. Retrieved from [30]

Return Risk	Min	Very low	Low	Moderate	High	Very high
Max	Certain	Highly Likely	Highly likely	Likely	Likely	Probably
Very high	Highly Likely	Highly Likely	Likely	Likely	Probably	Probably
High	Highly Likely	Likely	Likely	Likely	Probably	Probably
Moderate	Likely	Likely	Likely	Probably	Unlikely	Highly unlikely
Low	Likely	Probably	Probably	Unlikely	Unlikely	Highly unlikely
Very low	Probably	Probably	Probably	Unlikely	Highly unlikely	Never

Another interesting paper related with MOEAs was made by Skolpadungket, Dahal and Harnporncha [30]. The authors investigated the performance of several multi-objective evolutionary algorithms to solve the portfolio optimization problem, considering cardinality, floor and round-lot constraints. Their experiments focus on determining which algorithm performed better among three different ones. The first one to be evaluated was the Vector Evaluated Genetic Algorithm (VEGA) which consists on an extended version of the single GA to handle multi-objectives. Secondly, they used the traditional MOEA which was specified by the Multi-objective Genetic Algorithm (MOGA). Finally, they tested advanced algorithms such as Strength Pareto Evolutionary Algorithm II (SPEA2) and a Non-dominated Sorting Genetic Algorithm II (NSGA2). Their experiments determined that SPEA2 performed better when compared with its cohorts. Another interesting point to retrieve from their findings is the fact that the simplest GA, in particularly VEGA, when extended with a fuzzy logic mechanism suffered major improvements. The authors employed the following fuzzy rules (Table 2.2).

This fuzzy mechanism specifies the probability of selection for each individual through the implementation of a fuzzy decision rule which combines two objectives, risk and return. This technique is helpful in order to facilitate the trade-off between these two measures; if the return is maximum and the risk minimum, then it is certain the selection of that individual, if the return is very low and the risk is very high, then that individual will never be selected.

Although these approaches are also based on the use of evolutionary algorithms as the previous ones listed on Sect. 2.4.3.1, the main difference between single objective EAs and MOEAs is the way the solutions are ranked. Single objective EAs are characterized by evaluate a portfolio solution through a trade-off function that relates risk and return. MOEAs try to rank solutions evaluating risk and return

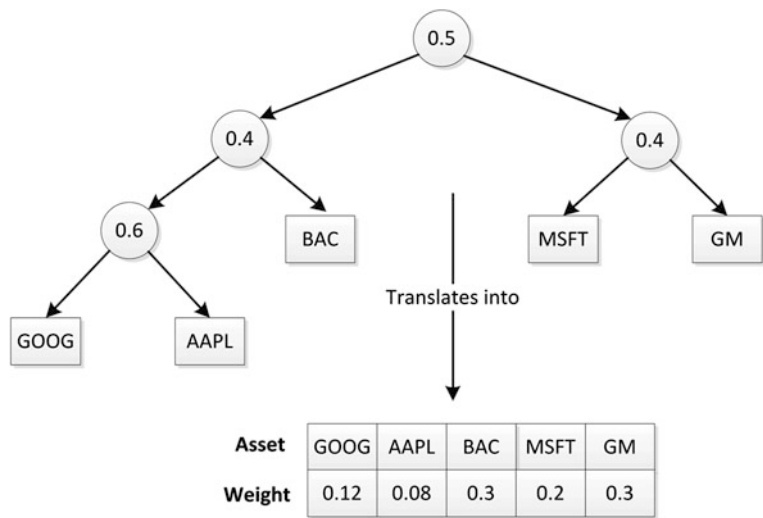


Fig. 2.6 A portfolio tree-based representation

separately as two distinct objectives to achieve. One point it should be noticed is that is still not clear if multi-objective genetic algorithms perform better than using a single objective genetic algorithm with a trade-off function to relate portfolio’s risk and return. However, the MOEAs have as major advantage the fact that is possible to generate the set of solutions, i.e. an approximation of the efficient frontier in a single run.

2.4.3.3 Extensions to Genetic Algorithms

More recently, extensions to the classical single-objective genetic algorithm’s approach were experimented; Aranha and Hitoshi [6] proposed a completely distinct representation of the portfolio using a tree-based structure, represented in Fig. 2.6. They conducted several experiments in order to support their choice, concluding that this new representation accelerates the evolution of a good solution. They were able to produce concise portfolios with the same utility as the ones generated when using an array-based structure. This fact brings several benefits since it permits the trading costs reduction and the increase of the portfolio’s understandability. Although this is still an early work since this representation was never proposed before, it clearly gives a good starting point on a portfolio’s representation when using genetic algorithms coupled with Markowitz’s model. However, notice that more tests must be done since the authors considered only the original Markowitz’s model without additional constraints.

The same authors [28] also tried a new approach, extending the traditional GA version with a modeling cost mechanism which can be employed to take into

consideration the previous held investments. The authors had in consideration the asset's weights held previously in the portfolio so they could take into account the costs related with buying and selling stocks that are needed to change the portfolio structure. Their goal was to minimize transaction costs by minimizing the difference between the previous held portfolio and the actual portfolio. In order to accomplish this feature, they defined the minimization of the Euclidean distance between the portfolios as a secondary objective, reached via a technique called Objective Sharing, avoiding that way the necessity of defining a MOOP. In order to maintain the consistency between the portfolios over time, they also introduced a mechanism called Population Seeding. This experience allows the possibility to get a more realistic approximation to the practical portfolio management. Although these authors were not the first on addressing the problem of considering transaction costs, the proposed approach seems to be the more realistic on how to handle this problematic. The same authors on the following year provided a more robust solution using the previous mechanism with a Memetic Algorithm [31].

In respect to extensions regarding the multi-objective evolutionary approach, Branke et al. [32] developed a system based on the combination between a multi-objective evolutionary algorithm and the Critical Line Algorithm (CLA) [2]. The considered process consists on dividing the problem into a subset of problems recurring to a MOEA. The CLA is then executed in each subset in order to produce a solution which forms a partial front designated as an envelope. Further, the EA is used to find a sequence of such envelopes which form a solution to the starting problem. Instead of representing each solution as a single-point in the efficient frontier, each solution passes to be represented as a partial front, the envelope.;

2.4.4 Technical and Fundamental Analysis in Portfolio Management

A completely different way on handling the portfolio problematic consists on performing a market evaluation based on technical and fundamental analysis, already explained in the beginning of this chapter.

It was already mentioned that technical analysis consists on studying stock charts in order to find over or undervalued stocks. Fundamental Analysis evaluates each security by measuring its intrinsic value through the study of overall economy, industry conditions and financial conditions of a specific company in order to produce a value which can be compared to the current company's price.

The first problem addressed by the exclusive use of these indicators consists on guarantying the diversification of the portfolio due to the absence of a model such as the Markowitz's one, to reduce the correlation between assets. Secondly, the risk involved in their utilization can be substantially high, and thirdly, how to decide the investment allocation percentage on each security, without doing it uniformly. Despite the presence of these problems on using such methodologies, the use of such procedures can reward us with a greater profitability since their

Table 2.3 A new portfolio representation

Stock number	1	2	...	10
Stock identification	1–300	1–300	1–300	1–300
Normalized percentage	0–1	0–1	0–1	0–1
Value added by indicators	0–1	0–1	0–1	0–1

Table 2.4 Technical rules. Based on [39]

Risk	1. Today’s price—average price of the previous 12 trading days
	2. Today’s price—average price of the previous 50 trading days
	3. Today’s price—maximum price of the previous 5 trading days
	4. Today’s price—maximum price of the previous 50 trading days
Return	5. Today’s price—minimum price of the previous 5 trading days
	6. Today’s price—minimum price of the previous 63 trading days

main goal consists on finding overvalued and undervalued stocks to produce profit. Due to its potential, it’s possible to achieve better returns, not only with the rising of security prices but also with their decline.

Liad Wagman [33] proposed a Genetic Programming (GP) approach based on the use of technical analysis. The author starts to form an initial population constituted by 1000 different portfolios, each of them composed by ten randomly stocks retrieved from an index formed with 300 distinct stocks. The investment percentage allocation for each stock is randomly assigned over these initial portfolios.

Each individual in the population, i.e. portfolio, is represented as the following (Table 2.3):

- Stock Number—Identifies the stock within the portfolio;
- Stock Identification—Identifies the stock within the index;
- Normalized Percentage—Investment allocation percentage;
- Value Added by Indicators—Percentage provided by the satisfaction of a variety of technical rules.

Each portfolio is evaluated through six technical rules responsible for generating “buy” or “not-buy” signals. These rules are generated from return and risk measures which calculation is based on the following technical indicators (Table 2.4):

Stipulated those indicators, the following rules are defined:

- Moving Average Rules (1) and (2)—Generate “buy” signals if equations (1) and (2) are greater than zero, respectively;
- Trading Range Breakout Rules (3) and (4)—Generate “buy” signals if equations (3) and (4) are greater than zero, respectively;
- Filter Rules (5) and (6)—Generate “buy” signals if today’s price has risen 1 % in respect to the minimum of previous 5 or 63 days, respectively.

All the presented values are based on [34].

These six rules are latter mapped to a percentage value, according to the respective weight. The author considered a 60 % risk value versus a 40 % return value. Each of

these percentage values is uniformly distributed over the respective rules; 15 % for each risk rule and 20 % for the two return rules. When each rule is satisfied it adds the respective weight to the overall fitness of the solution. For example, considering the past six months performance, if for stock number 6 which has a weight of 10 % in portfolio, only rule (3) is satisfied, and within the months 3, 4 and 5, then:

$$\text{Value Added by Indicators}_6 = 0.10 * \frac{(0.15 + 0.15 + 0.15)}{6} \quad (2.2)$$

The total fitness of the portfolio is calculated via the weighted average of these indicators' value, considering the normalized percentage values as the respective weights. Notice that the proposed work only aims on establishing a specific portfolio and maintain it indefinitely without having management consideration issues.

Another interesting approach was followed by Wei Yan et al. [34, 35]. The authors provide a portfolio construction system based on two distinct techniques, GP and Support Vector Machines (SVM). Both techniques are extended with a voting mechanism, and subsequent comparison is performed. Their GP application consists on a genetic programming algorithm coupled with an investment simulator. Each time an individual is evaluated through the fitness function, the investment simulator is executed. Each individual is represented by a factor model which consists on a table with 19 factors described by 18 fundamental indicators and one technical indicator, the Moving Average Convergence/Divergence (MACD). That individual is calculated considering that month's data. Further, based on this model, each market's stock is ranked. The stocks are then grouped on four market sectors and within each one they are ranked according to their expected return. The simulator then performs the following decisions:

- Top 3 stocks of each sector are bought, the bottom 3 are sold or go short;
- Sectors are equally weighted and each stock is given equal weight in the portfolio.

At the end of each month all the positions are closed and the profit or loss is calculated.

Although there aren't practically any published approaches using a variety of technical indicators, the referred works employ them, but in a very limited way; one only uses the Moving Average (MA) indicator and the other one a MACD indicator. Since there are many works that validate the application of technical indicators to buy or sell individual stocks, it will be interesting to deeply investigate more of those indicators in order to generate profitable portfolios. There's an infinity of technical indicators, the most widely used are described on [36]. Although it seems that if everyone uses those indicators it will get the same results, the premise is incorrect since there's a lot to explore on using them, such as the parameter specification. Also, the preferences of each investor can change. The person can opt for a more aggressive or more passive strategy, adapting the indicators to his will. Blanco et al. [37, 38] conducted an interesting study on investigating the optimization of some of these indicators using EAs.

2.4.5 Overview and Discussion

It's extremely difficult to evaluate the designed strategies in terms of profitability since most of them are applied to different market periods. Regarding the active versus passive question, an active design will try to beat the market which can probably produce higher levels of profitability when compared with a simple passive strategy, using the Markowitz's formulation. Possibly, its application conjugated with evolutionary computation is not so common due to the fact that technical and fundamental analysis requires a deep investigation on his functionality in order to one develop a solution based on its potential. Normally, it is an unfamiliar subject to most of the computational intelligence specialists which results on the employment of the widely known formulation, the Markowitz's model. When applying the notorious model, these scientists can concentrate their efforts on improving its expertise area, changing the structure and combining additional mechanisms in order to produce better and faster metaheuristics to solve this mathematical model, rather than studying other approaches which will require a deeper knowledge on economical facts.

2.5 Conclusions

From the several presented works given on this chapter, and which are briefly summarized on the following tables, it is possible to observe that most part of these solutions apply GAs to approach the portfolio problematic. Notice that a ranking involving all the different approaches presented on Table 2.5 was not performed because it is extremely hard to evaluate most part of these strategies since they are applied to different market conditions and periods. Also, the major part of these works has as principal objective the calculation of the efficient frontier in order to validate the proposed algorithm. Although the comparison is difficult, it is clear that the majority of the presented works use GAs; on several of these works where distinct optimization techniques were compared, the results showed that GAs were capable of surpassing the competitor methodologies. Based on these results, the intent of this work was to develop an application using a GA as an optimization technique.

In respect to the question active versus passive, from the previous table it is possible to observe that most part of the solutions concentrate their work on using the Markowitz's model to analyze the market, and subsequently pick the most promising stocks, according to the formulation. However, active management approaches using technical analysis can reward us with higher profitability levels, since their major intent is to beat the market, saying this; the best and most innovative way of approaching this problem is to use technical analysis to find under and overvalued stocks in the market. Given the reasons explained above and

Table 2.5 An overview over different approaches on portfolio optimization. Consult Appendix C to understand the parameters

Work	Years	Meta Heuristic	Additional features	Constraints	Portfolio analysis	Portfolio Rep.	MOOP?	Evaluation function	Data used	Training period	Testing period
[18]	1995	GA	Distributed version	-	Downside risk framework	Unknown	No	Lambda trade-off function	Prices from 100 unknown assets	Unknown	*
[19]	2000	SA	-	Floor	Markowitz	Real value array	No	Lambda trade-off function	Weekly prices from Hang Seng, DAX100, FTSE100, S&P100, Nikkei225	March 1992	*
		GA		Ceiling		structure				-	
		TS		Cardinality						Sept. 1997	
[21]	2002	HC	-	Floor	Markowitz	Hybrid structure	No	Linear combination between portfolio's variance and portfolio's return	Weekly prices from Hang Seng, Dax100, FTSE100, S&P100, Nikkei225	March 1992 - Sept. 1997	*
		TS	Token-ring procedure	Ceiling							
		SA	-	Cardinality							
[22]	2003	SA	-	Floor	Markowitz	Real value array	No	Portfolio's variance	Weekly prices from 151 US stocks	Jan. 1988 - April 1997	*
				Ceiling		structure					
				Cardinality Turnover							
[23]	2008	PSO	-	Trading	Markowitz	Hybrid structure	No	Lambda trade-off function	Weekly prices from Hang Seng, Dax100, FTSE100, S&P100, Nikkei225	March 1992 - Sept. 1997	*
				Floor							
				Ceiling							
				Cardinality							

(continued)

Table 2.5 (continued)

Work	Years	Meta	Heuristic	Additional features	Constraints	Portfolio analysis	Portfolio Rep.	MOOP?	Evaluation function	Data used	Training period	Testing period
[24]	2007	ANN	-	-	Floor Ceiling Cardinality	Markowitz	Hybrid structure	No	Lambda trade-off function	Weekly prices from Hang Seng, Dax 100, FTSE100, S&P100, Nikkei225	March 1992 - Sept. 1997	*
[26]	2007	GA	-	Multistage decision	-	Markowitz	Real value array structure	No	Sharpe ratio	40 assets from Taiwan's stock market	Unknown	*
[27]	2003	GA	-	Lamarckism mechanism	Cardinality Buy-in thresholds Round-lots	Markowitz	Binary string structure Real value array structure Hybrid real structure	Yes	Minimization portfolio's variance Maximization portfolio's return	Prices from Hang Seng	Unknown	*
[6]	2008	GA	-	-	-	Markowitz	Tree structure	No	Sharpe ratio	Prices from Nasdaq100 and Nikkei225	2003 2004 2005	Jan. 2004 Jan. 2005 Jan. 2006
[32]	2008	NSGA2	-	Critical line algorithm integration	Buy-in thresholds Cardinality 5-10-40	Markowitz	Hybrid real structure	Yes	Minimization portfolio's variance Maximization portfolio's return	Weekly prices from Hang Seng, S&P100, Nikkei225	Unknown	*

(continued)

Table 2.5 (continued)

Work	Years	Meta	Heuristic	Additional features	Constraints	Portfolio analysis	Portfolio Rep.	MOOP?	Evaluation function	Data used	Training period	Testing period
[28]	2007	GA		Objective sharing Population seeding	Multi-period consideration	Markowitz	Hybrid real structure	No	Sharpe ratio	Prices from Nasdaq 100	Nov. 2000 – Oct. 2006	Last 53 months of the indicated periods.
[31]	2009	GA		MA inclusion	Multi-period consideration	Markowitz	Tree structure	No	Sharpe ratio	Prices from Nasdaq 100, S&P500	2006 – 2007	2007 – 2008
[30]	2007	VEGA MOGA		Fuzzy	Cardinality Buy-in thresholds Round-lots	Markowitz	Hybrid real structure	Yes	Minimization of portfolio's variance Maximization of portfolio's return	Weekly prices from Hang Seng market	Unknown	*
[33]	2003	GP			Cardinality	Technical rules	Predefined structure	No	Maximization of technical rules satisfaction	Prices from Dow Jones industrial average	1979 – 1980	1980
[25]	2004	GA		MCDM	Cardinality Round-lots	Markowitz's based on risk and return decomposition	Predefined structure	Yes	Minimization of portfolio's variance Maximization of portfolio's return	Unknown	Unknown	*
[34]	2008	GP			Cardinality Multi-period consideration	1 technical rule 18 fundamental indicators	Predefined structure	Yes	Sharpe ratio	33 Malaysian stocks	Jan. 1999 – Dec. 2004	Jul. 1997 – Dec. 1998

the performed study on the developed works, it is proposed a solution based on technical analysis coupled with GAs.

Table 2.5 summarizes the approaches given, classified according to specific parameters. For a better understanding of this table, consult Appendix C.

Besides the presented table below, under Appendix B it is possible to observe a list of commercial applications based on technical analysis and portfolio management.

References

1. Maginn JL, Tuttle DL, McLeavey DW, Pinto JE (2007) Managing investment portfolios: a dynamic process. CFA Institute Investment Series. Wiley, New Jersey
2. Markowitz HM, Todd GP, Sharpe WF (2000) Mean variance analysis in portfolio choice and capital markets. Wiley, New Jersey
3. Malkiel B (1973) A random walk down wall street. W. W. Norton & Company, New York
4. Markowitz HM (1972) Portfolio selection. *J Financ* 7:77–91
5. Goodman B (2002) Passive management: it's not an oxymoron. <http://www.thestreet.com/author/1651333/BeverlyGoodman/all.html>. Accessed 20 Aug 2009
6. Aranha C, Hitoshi I (2008) A tree-based GA representation for the portfolio optimization problem. Genetic Evolutionary Computation Conference (GECCO), Atlanta, pp 873–880
7. Achelis S (2000) Technical analysis from A to Z. McGraw-Hill, New York
8. Hamsapiya SS, Surekha P (2008) Evolutionary intelligence: an introduction to theory and applications with matlab. Springer, Heidelberg
9. Mitchell M (1999) An introduction to genetic algorithms. MIT Press, Cambridge
10. Eiben AE, Smith JE (2007) Introduction to evolutionary computing. Springer, Berlin
11. Black F, Litterman R (1992) Global portfolio optimization. *Financ Anal J* 48:28–43
12. Salomons A (2007) The Black-Litterman model hype or improvement? Dissertation, University of Groningen
13. Sing TF, Ong SE (2000) Asset allocation in a downside risk framework. *J Real Estate Portfol Manage* 6:213–223
14. Cheng P, Wolverton M (2001) MPT and the downside risk framework: a comment on two recent studies. *J Real Estate Portfol Manage* 7:125–131
15. Gould N, Toint P (2008) Quadratic programming solvers. <http://www.numerical.rl.ac.uk/qp/qp.htm>. Accessed 15 Aug 2009
16. Stein M, Branke J, Schmeck H (2005) Portfolio selection: how to integrate complex constraints. *J Financ Plan* 68–78
17. Arnone S, Loraschi A, Tettamanzi A (1993) A genetic approach to portfolio selection. *Neural Netw World III* 6:597–604
18. Loraschi A, Tettamanzi A, Tomassini M, Verda P (1995) Distributed genetic algorithms with an application to portfolio selection problems. In: Pearson D, Steele N, Albrecht R (ed) *In Artificial neural nets and genetic algorithms*, Springer-Verlag, Wien
19. Chang TJ, Meade N, Beasley JE, Sharaiha YM (2007) Heuristics for cardinality constrained portfolio optimization. *Comput Oper Res* 27:1271–1302
20. Buseti FR (2000) Metaheuristic approaches to realistic portfolio optimization. Dissertation, University of South Africa
21. Schaefer A (2002) Local search techniques for constrained portfolio selection problems. *J Comput Econ* 20:177–190
22. Crama Y, Schyns M (2003) Simulated annealing for complex portfolio selection problems. *Eur J Oper Res* 34:1177–1191

23. Cura T (2008) Particle swarm optimization approach to portfolio optimization. *Nonlinear Anal: Real World Appl* 10:2396–2406
24. Fernández A, Gómez S (2004) Portfolio selection using neural networks. *Comput Oper Res* 34:1177–1191
25. Ehrgott M, Klamroth M, Schwehm C (2004) An MCDM approach to portfolio optimization. *Eur J Oper Res* 155:752–770
26. Lin C, Gen M (2007) An effective decision-based genetic algorithm approach to multiobjective portfolio optimization problem. *Appl Math Sci* 1:201–210
27. Streichert F, Ulmer H, Zell A (2003) Evolutionary algorithms and the cardinality constrained portfolio selection problem. *Oper Res Proc* 3–5
28. Aranha C, Iba H (2007) Modelling cost into a genetic algorithm-based portfolio optimization system by seeding and objective sharing. In: Genetic and evolutionary computation conference (GECCO), London, pp 196–203
29. Streichert F, Ulmer H, Zell A (2004) Comparing discrete and continuous genotypes on the constrained portfolio selection problem. In: Genetic and evolutionary computation conference (GECCO), Seattle, pp 1239–1250
30. Skolpadungket P, Dahal K, Hampomchai N (2007) Portfolio optimization using multi-objective genetic algorithms. In: Congress evolutionary computation, Singapore, pp 516–523
31. Aranha C, Hitoshi I (2009) Using memetic algorithms to improve portfolio performance in static and dynamic trading scenarios. In: Genetic and evolutionary computation conference (GECCO), Montreal
32. Branke J, Scheckenbach B, Stein M, Deb K, Schmeck H (2008) Portfolio optimization with an envelope-based multi-objective evolutionary algorithm. *Eur J Oper Res* 199:684–693
33. Wagman L (2003) Stock portfolio evaluation: an application of genetic-programming-based technical analysis. *Gen Algorithms Gen Program*, Stanford, pp 213–220
34. Yan W, Sewell M, Clack C (2008) Learning to Optimize Profits Beats Predicting Returns—Comparing Techniques for Financial Portfolio Optimization. *Proc 10th annu conf Gen and evol computation*, 1681–1688. Atlanta, USA
35. Hassan G, Clack C (2009) Robustness of multiple objective GP stock-picking in unstable financial markets. In: Genetic and evolutionary computation conference (GECCO), Montreal, pp 1513–1520
36. Murphy JJ (1999) Technical analysis of the financial markets: a comprehensive guide to trading methods and applications. Prentice Hall Press, New York
37. Fernández-Blanco P, Bodas-Sagi D, Hidalgo JI (2008) Technical market indicators optimization using evolutionary algorithms. In: Genetic and evolutionary computation conference (GECCO), Atlanta, pp 1851–1858
38. Bodas-Sagi D, Fernández P, Hidalgo JI, Soltero F, Risco-Martín J (2009) Multiobjective optimization of technical market indicators. In: Genetic and evolutionary computation conference (GECCO), Montreal, pp 1999–2004
39. Tsang E, Li J (1999) Improving technical analysis predictions: an application of genetic programming. *Proceedings of Florida Artificial Intelligence Research Symposium*. USA

Intelligent Financial Portfolio Composition based on
Evolutionary Computation Strategies

Gorgulho, A.; Neves, R.F.M.F.; Horta, N.

2013, XI, 77 p. 30 illus., 15 illus. in color., Softcover

ISBN: 978-3-642-32988-3