08 a 11 de Outubro de 2018
Instituto Federal Fluminense
Búzios - RJ

Multi-objective optimization of portfolio selection using evolutionary algorithms: an empirical analysis involving return and skewness

Davi Riani Gotardelo[1] - davi.riani@gmail.com
Jorge F. P. G. da Motta[2] - jfgimenezjf@gmail.com
Leonardo G. da Fonseca[3] - goliatt@gmail.com
[1]Universidade Federal Rural do Rio de Janeiro, RJ, Brazil
[2]Faculdade de Engenharia (Engenharia Civil) - UFJF, MG, Brazil
[3]Programa de Pós-Graduação em Modelagem Computacional - UFJF, MG, Brazil

Abstract. *This article aims to compare the performance of 04 evolutionary algorithms in the multi-objective optimization of researched problem in the area of Finance: the portfolio selection. The multi-objective algorithms NSGAII, MOEAD, IBEA and GDE3 were used to simultaneously optimize the mean return and the portfolio skewness. The data collected in the survey were from the companies listed in the Dow Jones Index and an in sample (2010 - 2014) and out-of-sample (2015 - 2017) analysis of the optimal portfolios of each algorithm was done. Statistical tests showed that the GDE3 algorithm presented better in sample and out-of-sample performance of the optimized portfolios.*

Keywords: *Portfolio selection, evolutionary algorithm, multi-objective*

1. INTRODUCTION

The study of asset portfolio selection has been one of the most investigated topics in Finance. The maximization of return to the investor and the minimization of risk has been the target of many scientific researches in Investments. Since the initial contribution of Markowitz (1952), several researchers have sought to study methods and models applicable to the selection of portfolios.Many studies have employed evolutionary computation techniques to solve portfolios selection problems, mainly due to advances in computational optimization and computational intelligence. The objective of this work is to compare the performance of different evolutionary algorithms in the multi-objective optimization problem of portfolio selection. Among the objectives of the problem, the mean return and the skewness were the measures chosen for optimization. The algorithms chosen in this work were: NSGAII, GDE3, IBEA and SMPSO. From the collection of asset prices traded in the Dow Jones Index between 2010 and 2017, optimal portfolios were generated through each algorithm and performances were compared in-sample (2010 to 2014) and out-of-sample (2015 to 2017).This paper is structured as follows: Section 2 presents the framework theoretical approach involving the Portfolio Theory and portfolio optimization problem; Section 3 shows the computational experiments involving data collect, the features of portfolio optimization method and in sample and out-of-sample performance. Section 4 presents the conclusions and suggestions for future research.

2. PORTFOLIO OPTIMIZATION PROBLEM

In selection of a portfolio it is assumed that the investor always seeks to maximize his expected return and minimize his risk, in this case, recognized as the variance. This model was proposed by Markowitz (1952) as the Mean-Variance Model, in which the investors choose ”mean-variance” portfolios, that is, choose portfolios that minimize portfolio variance given the expected return and maximize the expected return given the variance. According to Markowitz (1952), the efficient boundary is the “geometric place formed by the infinite portfolios and their associated lines (or hyperboles), which allows us to obtain a given return with the least possible risk or a given risk with the highest possible return”.

The model can be described as below:

Be p a portfolio with n assets:

$$E[R_p] = \sum_{i=1}^{n} w_i E[R_i], \tag{1}$$

where w_i is the asset participation i in portfolio p; $[R_p]$ is the return portfolio p; $E[R_p]$ is the expected return portfolio; $E[R_i]$ is the expected return asset i;

$$Var[R_p] = \sum_{i=1}^{n} w_i^2 \sigma_i^2 + 2 \sum_{i=1}^{n} \sum_{j=i+1}^{n} w_i w_i \sigma_i \sigma_j \rho_{ij}, \tag{2}$$

where $Var[R_p]$ is the variance of portfolio returns, σ_i and σ_j are, respectively, standard deviation of asset i and portfolio p and, finally, ρ_{ij} is the correlation coefficient among assets i and j.

$$\sigma_p = \sqrt{Var[R_p]}. \tag{3}$$

The model objective is to construct the efficient boundary, which represents all optimal combinations of portfolios in the sense of return and risk. The efficient boundary can be developed by two steps:

$$\text{Maximize} \quad E[R_p] \quad \text{, subject to} \quad \sigma_p = k_1 \tag{4}$$

and

$$\text{Minimize} \quad \sigma_p \quad \text{, subject to} \quad E[R_p] = k_2 \tag{5}$$

where k_1 and k_2 are constants.

In Equation (4) we maximize the expected return for a given standard deviation (k_1). If the schedule is repeated for several levels of different standard deviations the result will be a set of optimal portfolios that generate the efficient boundary. In the case of Equation (5), the procedure is analogous, but the standard deviation is minimized for a given level of expected return (k_2). The model assumes that the higher the return and the lower the standard deviation, better the portfolio. The relationship between return and risk, assessed at one point on the border should be chosen for each investor to attempt your interests.

2.1 Skewness in portfolio decision

The skewness and fat tail has been frequently debated in theoretical studies and articles on portfolio selection, mainly by the argument that choosing assets with greater skewness alone produces a more asymmetric portfolio and, in the case of negative skewness, values are repeated more times than the mean value (Modal values $>$ Mean values). This skewness is partially contained in the Omega view of the portfolio, since it favors larger returns to the detriment of smaller losses. But the main question is whether to privilege skewness compromises the focus of the mean-variance, that is, to evaluate whether the privilege of skewness can lead to less efficient portfolios in terms of mean-variance.

Jiang et al (1952) investigates the selection of portfolios within a structure of mean-variance skewness. They derived the composition of efficient portfolios and analyzed the properties of these efficient portfolios. They demonstrated that the required systematic skewness is achieved at the expense of the traditional mean-variance efficiency, and that a more stringent systematic skewness restriction induces a greater loss in mean-variance efficiency.

In relation to skewness composition, Canela and Colazzo (2007) work with a method that separates skewness into two components: a first, which should be avoided, derived from very high values (outliers) and the other, caused by the deviation of the tail and by the coefficient of skewness of the assets, this yes, to be sought by the investor. The authors use the polynomial goal programming to determine the ideal portfolio of industries in emerging markets and conclude that the decision to consider incorporating skewness into the portfolio decision improves portfolio performance.

Some authors have presented studies on the relation between the asymmetry of the portfolio and the asymmetry of the assets in isolation. Sun and Yan (2003) argue that there are studies that have found that ex post stock returns are positively distorted, but such skewness is only persistent for individual stocks, not for portfolios. This implies that ex post knowledge of skewness may not be useful in selecting the ex ante portfolio. According to the authors, who worked with data from American and Japanese companies, it is more significant to verify if the skewness would persist in efficiently formed portfolios through the mean-variance and also favored skewness, in a form of a polynomial goal programming method. Oliveira et al (2010) studied the influence of co-skewness and co-kurtosis, separately and jointly, in the analysis of stock prices in the Brazilian financial market and concluded that co-skewness and co-kurtosis do not improve the performance of the model pricing.

In the case of portfolio returns, negative skewness is preferable to positive skewness, since negative skewness, by definition, indicates that the tail of the left side of the probability density function is greater than that of the right side, thus favoring values higher than the mean values. Thus, a portfolio with negative skewness is more preferable to one with positive skewness because it has higher returns in its distribution. Therefore, the skewness optimization in this research is in minimizing its value. Thus, the two goals of optimization are to maximize the average return and minimize skewness. Since the algorithm works only by maximizing or minimizing, we opted to minimize the negative (-) return, which would have the same effect as maximizing.

3. COMPUTATIONAL EXPERIMENTS

3.1 Evolutionary Algorithms

According to Ragsdale (2014), one of the most interesting developments in the field of optimization was research on evolutionary algorithms. These were inspired by Darwin's theory, searching for elements of biological reproduction and applying the principle of ”survival of the most able” to get good solutions to complex problems.

Anagnostopoulos et al (2010) presents some features of evolutionary algorithms. According the authors, the last ability of EAs—due to their population-based nature—to handle problems having multiple objective functions has given rise to the field of evolutionary multi-objective optimization (EMO) which refers to the use of EAs to solve complex multi-objective optimization problems.The algorithms designed for this purpose are usually identified under the rubric multi-objective evolutionary algorithms (MOEAs) and they differ from their single-objective counterparts mainly in the way selection is performed. MOEAs use a non-dominated ranking and selection to guide the population towards the Pareto front, and diversity preserving techniques to avoid convergence to a single point on the front.The main advantage of MOEAs is that they generate reasonably good approximations of the non-dominated frontier in a single run and within limited computational time.

multi-objective Evolutionary Algorithms (MOEA) find multiple equally optimal solutions. In this paper, is minimized two objective functions (asymmetry and - return) simultaneously through the usage of four evolutionary strategies (NSGA-II, IBEA, MOEAD and GDE3).

According to Antoniucci (2016), GDE3, for instance are found to produce better distributions than NSGA-II and IBEA which are widely used.

In this way, it will be presented briefly each MOEA using Antoniucci (2016) work. In the next section 3.2, will be presented the obtained results.

NSGA-II - Non-dominated Sorting Algorithm II

The Non-dominated Sorting Algorithm II is one of the most popular MOEAs. First, the algorithm classifies an initial population according to the number of solutions that dominate each solution, with several fronts. ”The algorithm then sort search of these fronts according to the distance between consecutive solutions thus promoting solutions in low populated areas of the search space, before pushing them into next iteration's population”. A new iteration begins after the process of selection, crossover, mutation and merging of the new population with the previous one.

GDE3 - Generalized Differential Evolution 3

The Generalized Differential Evolution 3 algorithm uses Differential Evolution with search mechanism. ”GDE3 applies a slightly modified Differential Evolution operator to generate offspring from the current population”. After the comparison of generations and removal the dominant solutions of the population, a new iteration begins. The GDE3 resembles NSGA II in the fact that population size after the end of each iteration is reduced techniques aimed at preserving diversity. However, according to Sandra et al (2012), GDE3 modifies the crowding distance of NSGA-II in order to better deal with problems that have more than two objectives.

MOEAD - Multi-objective Evolutionary Algorithm based on Decomposition

The multi-objective Evolutionary Algorithm based on Decomposition "optimises each objective independently according to a user defined decomposition approach". In addition, it is important to note that, as each objective is independently manipulated, it is a certain diversity. This tends to produce good results.

IBEA - Indicator-Based Evolutionary Algorithm

The Indicator-Based Evolutionary Algorithm uses only one measure of quality to guide a search. First, the algorithm start a value of adequacy to the initial population. "The solution with the lowest score is deleted from the population and all the fitness scores are recomputed until the population size is lower than the user defined threshold". There is a mix of the new population with the old population after a combination and a mutation, initializing a new iteration. In addition, under the Colomine et al (2012), "attempts to incorporate practical decision-making and privileged infor-mation when searching for Pareto solutions".

3.2 Data

The data were collected by the researchers in the software Economatica and were the historical series of stock returns listed on the New York Stock Exchange (Dow Jones Index). Data are from 2010 to 2017. In total, data were collected from 30 DWJ companies.

The analysis period in sample and out of sample was divided as follows: 2010 to 2014 as in sample and 2015 to 2017 as out of sample. Initially the daily logarithmic returns of the stock prices were calculated and the optimization was done with 100 generations and 1000 iterations for each algorithm. As performance metrics in sample and out of sample the cumulative return was used. The Kolmogorov-Smirnov Test was run to verify the normality of the optimum projected portfolios. As all series do not present normality in their distribution, it was necessary to apply a non-parametric test to verify the performance difference between the algorithms. The Mann-Whitney Test was calculated to test the null hypothesis that the medians of the algorithms, by pair, are statistically equivalent, at the significance level of 5%.

After the simulation with the four multi-objective Evolutionary Algorithms (MOEA) indicated, a qualitative and a quantitative analysis was done.

By means of the qualitative analysis, it was possible to show in Figure 1 that the MOEA was able to find better results than random simulations (black cloud), MOEA was able to be more efficient. In addition, comparing the MOEA among some algorithms have achieved better skewness solutions, while others have achieved better results for mean return. Therefore, more than talking about one algorithm is better than the other like Antoniucci (2016), it is important to note that the integration of the algorithms, in this problem with two objective functions, generates a well defined Pareto Frontier with more diversity of solutions. Next, the in-sample and out-of-sample results of the optimization obtained by each algorithm are presented.

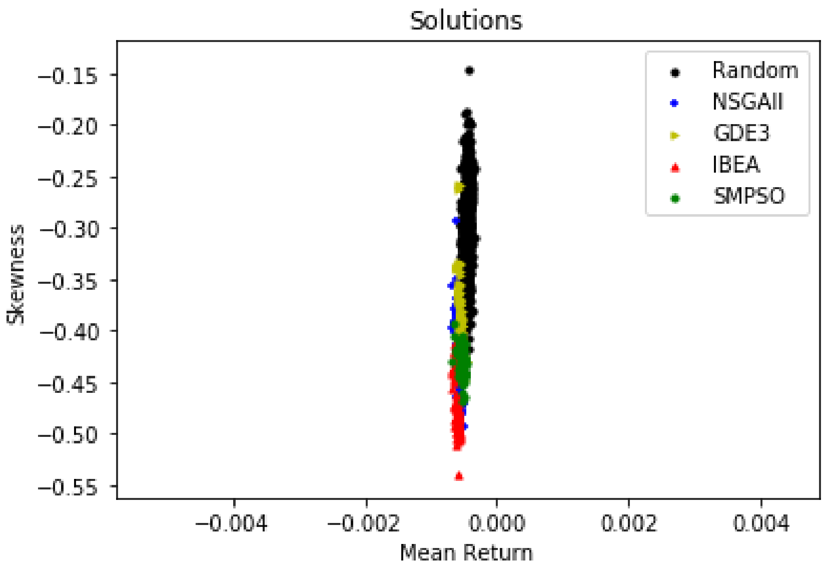

Figure 1- Overlapping Solutions

3.3 In Sample Performance

The Figure 2 shows a histogram of the returns found for each Genetic Algorithm. As can be observed visually, at first there is no normal distribution. In addition, GDE3 presented good results with higher values in the right tail of the histogram, which is positive, considering the incidence of higher values.

Boxplot cumulative return in sample was presented in Figure 3 and again shows that the GDE3 presented higher average return besides the lower amplitude of the results. This shows that the solutions found by the GDE3 could be more consistency, since the results converge for a better solution. On the other hand, it is important to highlight that the results obtained with the MOEAD were very dispersed. With regard to the boxplot it is possible to identify that the MOEAD did not show outlier. Finally, for the four algorithms used, the average value of the returns was between 72.5% and 75% when analyzed over a period of 5 years (2010 - 2014).

3.4 Out-of-Sample Performance

In every portfolio optimization test it is extremely important to test in a sample and with the results obtained in this sample, check in a second sample (out-of-sample) with the results behave. This is relevant, since in real life not always what happened in the past will re-run in the future, however, if the assets of a portfolio are allocated in an optimal way, in a future scenario this portfolio also tends to present good results . In summary, it can be stated that in-sample weights are theoretically ideal, and in the out-of-sample we check if these weights actually generate good results in the future.

In the histogram generated out-of-sample was presented in figure 4 it was possible to observe some differences with the histogram of figure 2. The GDE3 continued to present a good result for the return, however its distribution lost some of its format. The IBEA has lost some of its efficiency in determining better returns. The MOEAD and NSGAII maintained their main characteristics.

Analyzing the box-plot in Figure 4, it is possible to verify that the GDE3 again presented

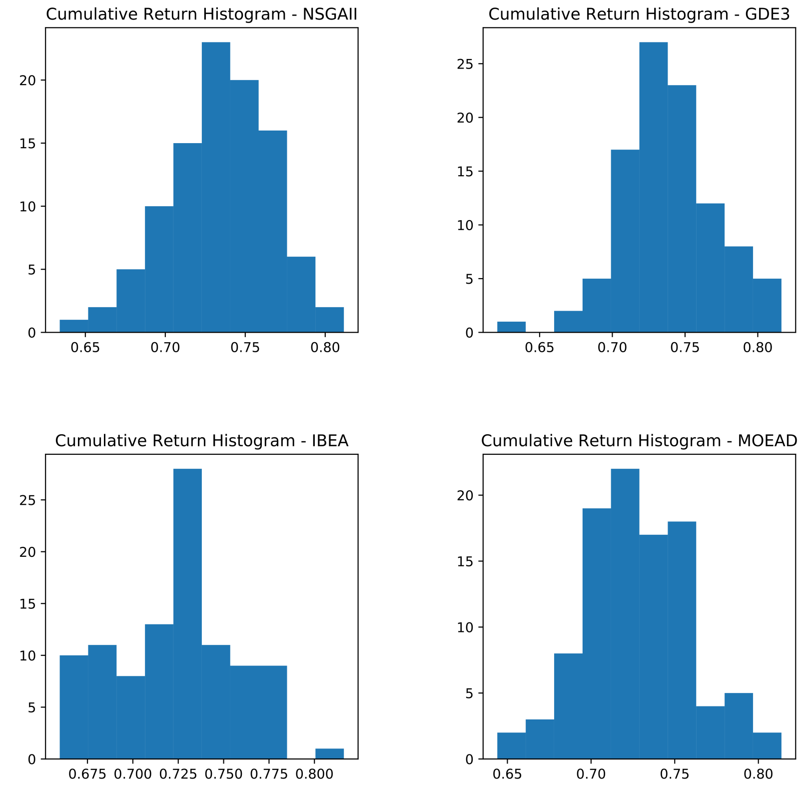

Figure 2- Histogram Cumulative Return - In Sample

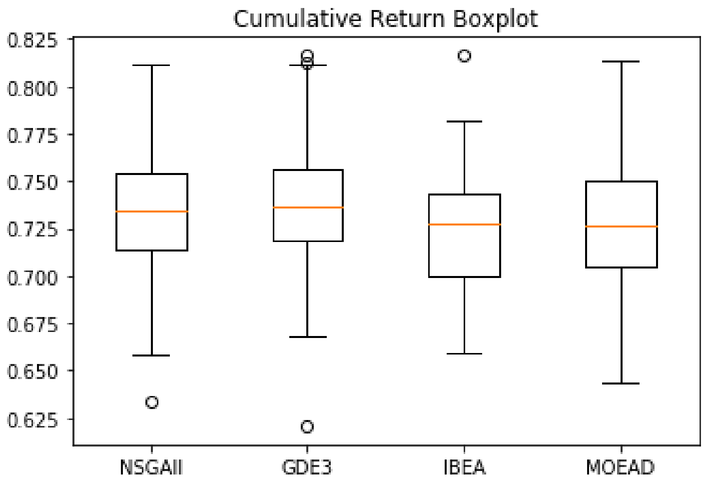

Figure 3- Boxplot Cumulative Return - In Sample

the higher mean value compared to the others and the lower amplitude. NSGAII presents the higher amplitude.

Regarding the quantitative aspect, an analysis was done in sample and another analysis was

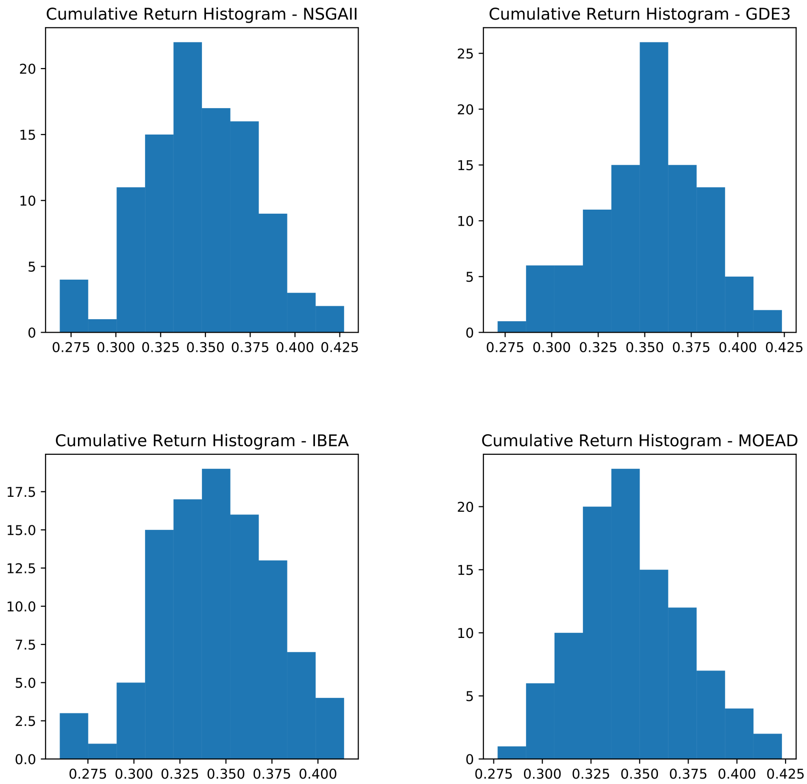

Figure 4- Histogram Cumulative Return - Out of Sample

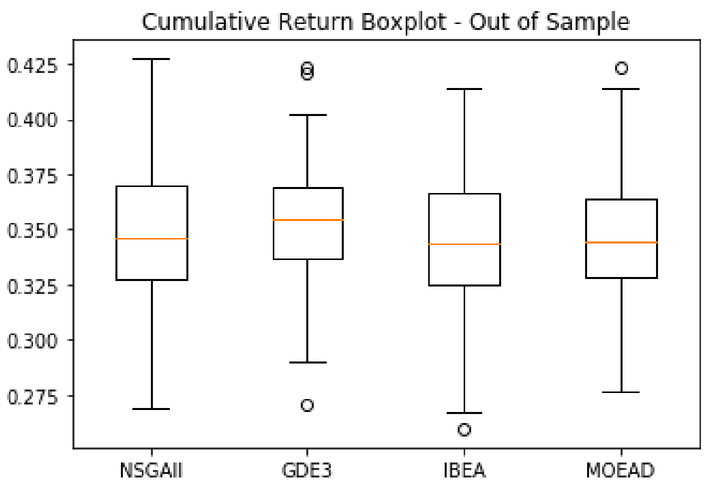

Figure 5- Box-plot Cumulative Return - Out of Sample

done outside the sample. That is, initially, a multi-objective optimization was done on a sample to determine the weights of a stock portfolio. With these weights, the returns obtained in the portfolios were analyzed in another sample (out-of-sample). After this analysis, as observed in the Figure 3, the MOEA that presented better average returns and more concentrated was the GDE3. This confirms the idea of Antoniucci presented in the previous section (3.1)

Besides that, some statistical tests were performed to understand whether the retinas generate a normal distribution. The results are shown in the table below.

Initially the KS Test was performed to verify if distributions behave as a normal distribution. The result presented in Table 01 confirms the non-normality of the distributions. Thus, a non-parametric test was performed. Thus, we can not say that mean and variance represent well the distribution behavior. Therefore, we used a nonparametric test (Mann-Whitney) between pairs of the algorithms to verify if the two populations have the same probability distribution. The Mann-Whitney Test is the non-parametric test equivalent to Student's T-Test.

Table 1- Kolmogorov-Smirnov Test

Algorithm	Statistic	p-value
NSGAII	0.7369	0.000
GDE3	0.7452	0.000
MOEAD	0.7402	0.000
IBEA	0.7381	0.000

Table 2 shows the Mann Whitney Test results. In the in-sample analysis, it is important to note that the GDE3 algorithm presented a significant difference in relation to the IBEA and MOEAD algorithms, since its p-value presented a value lower than the level of significance (0.05). The NSGAII algorithms also presented significant difference in relation to the IBEA and MOEAD algorithms. In the out-of-sample analysis, once again the GDE algorithm presented a significant difference in relation to the IBEA and MOEAD algorithms, and the other algorithms were equivalent. This result confirms the good results of the GDE algorithm, as demonstrated in the histograms and box-plots presented in Section 3.

Table 2- Mann Whitney Test

Algorithms	In sample		Out-of-Sample	
	Statistic	p-value	Statistic	p-value
NSGAII x GDE3	4758	0.2775	4463	0.0949
NSGAII x IBEA	4041	0.009	4682	0.2189
NSGAII x MOEAD	4319	0.048	4890	0.3945
IBEA x GDE3	3779	0.001	4150	0.0189
IBEA x MOEAD	4705	0.236	4810	0.3217
MOEAD x GDE3	4033	0.009	4293	0.0421

4. CONCLUSIONS

This work aimed to compare the performance of 04 evolutionary algorithms in the optimization of portfolio selection of assets traded on the Dow Jones Index in periods in sample (2010-2014) and out-of-sample (2015-2017). We confirm once again that evolutionary algorithms are, of course, more efficient than random solutions in the portfolio selection process.

Regarding the performance of the algorithms, the statistical tests used in this work indicate that the GDE3 algorithm presented superior performance in relation to the IBEA and MOEAD algorithms, both in the in sample and in the out of sample.
As a suggestion of future research, we suggest the inclusion of cardinality constraints in the portfolio definition, as well as the comparison of performance achieved by algorithms with market benchmarking performance, such as the naive portfolio for example.

REFERENCES

Anagnostopoulos, Konstantinos P., and Georgios Mamanis. ”A portfolio optimization model with three objectives and discrete variables.” Computers & Operations Research 37.7 (2010): 1285-1297.
Arnau Antoniucci, Guido. Analysis of the distribution of pareto optimal solutions on various multi-objective evolutionary algorithms. BS thesis. Universitat Politècnica de Catalunya, 2016.
Canela, Miguel Angel, and Eduardo Pedreira Collazo. ”Portfolio selection with skewness in emerging market industries.” Emerging Markets Review (2007), 230-250.
Duran, Feijoo E. Colomine, Carlos Cotta, and Antonio J. Fernández-Leiva. ”A comparative study of multi-objective evolutionary algorithms to optimize the selection of investment portfolios with cardinality constraints.” European Conference on the Applications of Evolutionary Computation. Springer, Berlin, Heidelberg, 2012
García, Sandra, et al. ”Time-stamped resampling for robust evolutionary portfolio optimization.” Expert Systems with Applications 39.12 (2012): 10722-10730.
Jiang, Chonghui, Yongkai Ma, and Yunbi An. ”Portfolio selection with a systematic skewness constraint.” The North American Journal of Economics and Finance (2016), 393-405.
Markowitz, Harry. ”Portfolio selection.” The journal of finance (1952), 77-91.
Oliveira, Alexandre Silva, Luis Felipe Dias Lopes, and Eduardo Botti Abbade. ”Coassimetria e cocurtose na análise dos preços das acões no mercado financeiro nacional.” Revista de Administração da UFSM (2010), 326-345.
Ragsdale, Cliff T. ”Modelagem de planilha e análise de decisão: uma introdução prática a business analytics”.São Paulo: Cengage Learning, 2014.
Sun, Qian, and Yuxing Yan. ”Skewness persistence with optimal portfolio selection.” Journal of Banking Finance (2003), 1111-1121.