Universitat Oberta de Catalunya

uoc.edu

Citation for published version
Calvet, L [Laura]; Kizys, R [Renatas]; Angel, J [Juan]; de Armas, J [Jesica]. (2016). A SimILS-Based Methodology for a Portfolio Optimization Problem with Stochastic Returns. Economics and Management. MS 2016. Lecture Notes in Business Information Processing, vol 254. Springer, Cham. https://doi.org/10.1007/978-3-319-40506-3_1

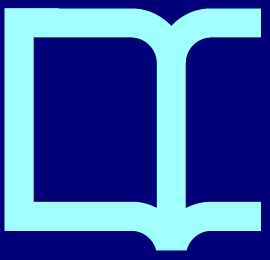
DOI
https://doi.org/10.1386/cjcs_00008_1

Handle
http://hdl.handle.net/10609/148527

Document Version
This is the Accepted Manuscript version.
The version published on the UOC's O2 Repository may differ from the final published version.

Copyright
© Duke University Press

Enquiries
If you believe this document infringes copyright, please contact the UOC's O2 Repository administrators: repositori@uoc.edu



A SimILS-based Methodology for a Portfolio Optimization Problem with Stochastic Returns

Laura Calvet[1], Renatas Kizys[2], Angel A. Juan[1], and Jesica de Armas[1]

[1] Computer Science Department, IN3 – Open University of Catalonia, Barcelona, Spain
{lcalvetl, ajuanp, jde_armasa}@uoc.edu

[2] Subject Group of Economics and Finance, University of Portsmouth, UK
renatas.kizys@port.ac.uk

Abstract. Combinatorial optimization has been a workhorse of financial and risk management, and it has spawned a large number of real-life applications. Prominent in this body of research is the mean-variance efficient frontier (MVEF) that emanates from the portfolio optimization problem (POP), pioneered by Harry Markowitz. A textbook version of POP minimizes risk for a given expected return on a portfolio of assets by setting the proportions of those assets. Most authors deal with the variability of returns by employing expected values. In contrast, we propose a simILS-based methodology (i.e., one extending the Iterated Local Search metaheuristic by integrating simulation), in which returns are modeled as random variables following specific probability distributions. Underlying simILS is the notion that the best solution for a scenario with expected values may have poor performance in a dynamic world.

Keywords: Portfolio Optimization · SimILS · Metaheuristics · Simulation.

1 Introduction

Investments play an essential role in our society through wealth creation, sustainable economic growth and ultimately improvements in welfare standards. They provide companies with the necessary funds to transform ideas and resources into profitable projects, social benefits and jobs. From the point of view of a portfolio investor, POP is a strategy of a) selection of financial assets and b) determination the optimal weights allocated to those assets that results in a desired portfolio return and an associated minimum level of risk. This combinatorial optimization problem (COP) is known as the portfolio optimization problem (POP), a milestone of modern portfolio theory, founded by Harry Markowitz [1]. Key to POP is a quadratic objective function that is a) computed by aggregating over the covariances of the constituent asset returns, and b) minimized subject to a desired rate of return. It is worth noting that other risk measures have been applied in the literature such as value-at-risk. Additionally, portfolio weights must add up to one and, in most cases, take on non-negative values. A realistic POP introduces further constraints. In particular, pre-assignment, quantity and cardinality constraints have received overwhelming attention in extant literature. The pre-assignment constraint allows the investor to pre-

select some assets, irrespective of their risk-return characteristics. The quantity constraint confines the weight allocated to an asset in the portfolio within a desired range of values. One the one hand, the upper limit (the ceiling) of the range attempts to reduce the exposure to each asset. On the other hand, the lower limit (the floor) rules out investments in negligible quantities, which may be prohibitively costly, since the transaction costs may reduce or erase the benefit. While recognising that this constraint arises as a result of the investor's discretionary decisions, it has received growing interest. For instance, [2] argue that its inclusion can lead to improved out-of-sample performance of optimization performance, can help contain portfolio volatility, boost realized portfolio performance, as well as decrease downside risk and shortfall probability. Finally, the cardinality constraint sets a minimum and maximum value for the number of assets in the portfolio. The lower bound aims to diversify the investment, i.e., allocate resources to a set of imperfectly correlated assets. Such strategy seeks to minimize the overall risk of portfolio investment. The upper bound is dictated by the evidence that marginal benefits of portfolio diversification starts to decrease after the number of assets already selected in the portfolio hits a certain threshold [3]. In addition, portfolios with a large number of assets are more costly in terms of complexity, managerial effort and the ensuing increased transaction costs. These constraints make the problem NP-hard [4].

Optimization methods may be classified into exact methods and heuristics/metaheuristics [5]. The first group includes procedures that guarantee the optimality of a solution. However, exact methods may require making strong assumptions or large amounts of time, especially when they are used to solve real-life complex problems. Within the second group, heuristics are experience-based procedures, which usually provide near-optimal solutions in considerably less time. By contrast, metaheuristics [6] are general templates, which may solve a broad range of problems without having to be tailor-made to a particular problem and often in real time. In the literature on the portfolio optimization, linear [7] and quadratic [8] programming methods have been predominant exact methods. However, due to the complexity of these problems, metaheuristics are increasingly more employed at present [9].

Despite the non-exhaustive nature of applications of realistic POP, they have not been extensively studied. As aforementioned, a textbook version of POP underlies the empirically unsupported assumption of constant expected rate of return, a key limitation in a large and growing body of research. The main contribution of this research is to address this limitation. Indeed, since asset return is a random variable that obeys a certain probability density function, and future returns are unpredictable, the minimum desired rate of return may not be attained with certainty. More concretely, we relax the above simplifying assumption and randomize the minimum desired rate of return. The resulting problem is referred to as the *Stochastic* POP (SPOP). The solver that is constructed to solve SPOP is relatively new and is called SimILS [10]. It envisages an extension of the Iterated Local Search (ILS) metaheuristic [11] that integrates simulation techniques to address sources of uncertainty embedded in a randomized objective function or/and budget constraint. In short, while a metaheuristic searches for high-quality solutions for a deterministic version of the problem, which employs expected values of random variables, simulation techniques are applied to test them in a stochastic environment. In fact, this

approach – coined simheuristics [12] – suggests combining metaheuristics and simulation techniques. In this context, our research aims to: *(i)* derive a mathematical formulation for the Stochastic POP, *(ii)* develop a solving methodology based on an existing algorithm for the POP [13]; and *(iii)* illustrate its use by solving a benchmark instance.

This paper is organized as follows. Section 2 provides a formal description of the problem. Section 3 proposes a methodology. A computational experiment is carried out in Section 4, while the results are analyzed in Section 5. Finally, Section 6 gathers the main conclusions.

2 Description of the Problem

Let there be a set $A = \{a_1, a_2, \dots, a_n\}$ of n assets, where each asset a_i ($\forall i \in \{1,2, \dots, n\}$) is characterized by an expected return r_i. The covariance between two assets a_i and a_j ($\forall i, j \in \{1,2, \dots, n\}$) is denoted by σ_{ij}. A solution for the POP (see Fig. 1), is a vector $X = (x_1, x_2, \dots, x_n)$, where each element x_i ($0 \leq x_i \leq 1$) represents the weight of the asset a_i in the portfolio. The aim of the POP is to minimize the risk of the investment and obtain an expected return greater than a specific threshold R. A realistic version includes also the pre-selection, quantity and cardinality constraints. The pre-selection constraint dictates whether an asset a_i must be in the solution (i.e., $x_i > 0$) by means of the parameter p_i: $p_i = 1$ if a_i is pre-selected, and $p_i = 0$ otherwise. The quantity constraint specifies for each asset a_i a lower and an upper bound, ε_i and δ_i ($0 \leq \varepsilon_i \leq \delta_i \leq 1$), respectively. The cardinality constraint sets the lower and upper limits on the number of assets included in the portfolio, k_{min} and k_{max} ($1 \leq k_{min} \leq k_{max} \leq n$), respectively. One key difference between the stochastic and classical versions of POP is that the former assumes uncertain future returns on a portfolio of assets, with a certain probability of not attaining the threshold value.

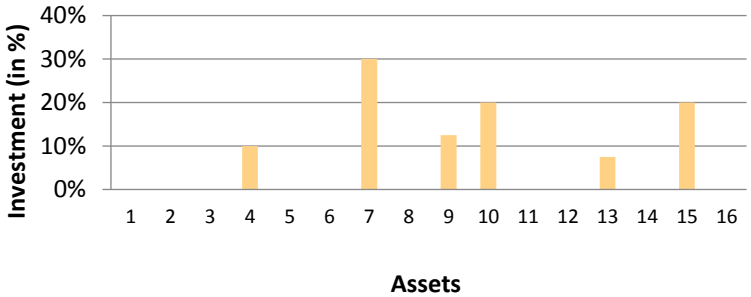


Fig. 1. Example of a solution representation

$$\min f(X) = \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij} x_i x_j$$

Subject to:

$$P(\sum_{i=1}^{n} R_i x_i \geq R) \geq P_0 \quad (1)$$

$$\sum_{i=1}^{n} x_i = 1 \quad (2)$$

$$\varepsilon_i z_i \leq x_i \leq \delta_i z_i, \quad \forall i \in \{1,2,\dots,n\} \quad (3)$$

$$0 \leq \varepsilon_i \leq \delta_i \leq 1, \quad \forall i \in \{1,2,\dots,n\} \quad (4)$$

$$z_i \leq M x_i, \quad \forall i \in \{1,2,\dots,n\} \quad (5)$$

$$p_i \leq z_i, \quad \forall i \in \{1,2,\dots,n\} \quad (6)$$

$$k_{min} \leq \sum_{i=1}^{n} z_i \leq k_{max} \quad (7)$$

$$z_i \in \{0,1\}, \quad \forall i \in \{1,2,\dots,n\} \quad (8)$$

The objective function minimizes the risk of the investment. Equation (1) guarantees that the return on investment will be no smaller than the threshold R with a probability of at least P_0. Equation (2) restrains portfolio investment to the existing resources. An auxiliary variable is introduced to indicate whether the asset a_i is in the solution ($z_i = 1$ in this case, $z_i = 0$ otherwise). For each asset a_i, Equation (3) sets a lower and an upper bound (ε_i and δ_i, respectively) for x_i, in case the asset is selected (i.e., $z_i = 1$). The two bounds range from 0 and 1 (Equation 4). In Equation (5) M is a very large positive value such that $Mx_i \geq 1$ for all i if $x_i > 0$. Equation (6) defines the pre-assignment constraint, where z_i depends on the parameter p_i. If the asset a_i is pre-selected (i.e., $p_i = 1$), then it also appears in the solution (i.e., $z_i = 1$). Equation (7) describes the cardinality constraint. Finally, Equation (8) defines z_i as a binary variable.

3 Our Methodology

The proposed methodology follows a simILS approach. It is demonstrated to be successful for solving realistic COPs with sources of uncertainty [14]. It is a natural extension of ILS-based algorithms to address stochastic COPs. Specifically, we combine a solving methodology for the POP [13] with Monte Carlo simulation (MCS) techniques. The referred work describes a powerful yet simple algorithm, which includes heuristics for the selection of assets and a quadratic programming solver that allocates weights to POP. It uses memory caches to enhance the algorithm's performance. In fact, it provides high-quality solutions in real time, only within a few seconds.

Our methodology is summarized in Fig. 2 and described next.

First, the stochastic instance is transformed into deterministic by replacing random variables by their means. Second, an initial solution by means of the algorithm described in [13]. It constructs a solution by combining the pre-selected assets with high-return assets. The solution has to be feasible in the stochastic environment, i.e.,

the required return (R) has to be reached with a probability no smaller than P_0. MCS is employed to estimate this probability by means of the proportion of cases in a sample of generated scenarios where the return obtained is at least as high as R. Each scenario is created by randomly drawing a value for each return in the original instance. Third, copies of the initial solution are stored as base and best solutions. Fourth, of the above specified steps are repeated until a stopping criterion based on the elapsed time is met. First, a new solution is created by 'perturbing' the base solution. This perturbation, defined in the original algorithm, randomly erases some assets from the portfolio and introduces others. An acceptance criterion is introduced to determine whether the new solution is promising and should replace the base solution or should be discarded. It is a Demon-like acceptance criterion [15], which accepts the solutions that improve the objective function value (i.e., the risk) and those that worsen it but satisfy the following conditions; (*i*) no consecutive deteriorations take place, and (*ii*) the degradation does not exceed the value of the last improvement. The next step consists of checking the feasibility of the new solution as before, using MCS. Only if it is feasible, the new solution is copied into the base solution, and the best solution is updated (if improved). Finally, the best solution is returned.

Note that this approach presents relevant advantages: (*i*) its modularity, which enables the reuse of problem-specific procedures from the original algorithm, (*ii*) it is relatively simple to understand and implement, and (*iii*) it does not add too much time, since MCS is only used to check the feasibility of promising solutions.

4 Computational Experiments

Our methodology has been implemented as a Java application. A standard personal computer, Intel Core i5 CPU at 3.2 GHz and 4 GB RAM with Windows 7 has been employed. We have experimented with a stock market database from the repository ORlib (http://people.brunel.ac.uk/~mastjjb/jeb/orlib/portinfo.html), which was proposed in [16]. It represents the market index Hang Seng (Hong Kong) measured at weekly frequency spanning the period from March 1992 to September 1997. This benchmark instance gathers expected returns r_i and standard deviations σ_i. In order to assess our simheuristic methodology, expected returns have been replaced with random variables R_i that distribute normally with mean r_i and standard deviation σ_i. The instance contains 31 assets. The expected returns average is 0.0035 (95% IC: 0.0027-0.0043) and the standard deviation is 0.002. The return standard deviations average 0.0457 (95% IC: 0.0430-0.0484) and deviate from the mean on average by 0.0073. Fig. 3 displays the probability density functions for some selected assets. Table 1 proves that there is a positive association between expected returns and standard deviation (Pearson product-moment correlation coefficient: 0.4973). In other words, investors expect a higher return for assets characterized by a higher risk. Return correlations among different assets average 0.5266 (95% IC: 0.5137-0.5395), which suggests the presence of gains from portfolio diversification.

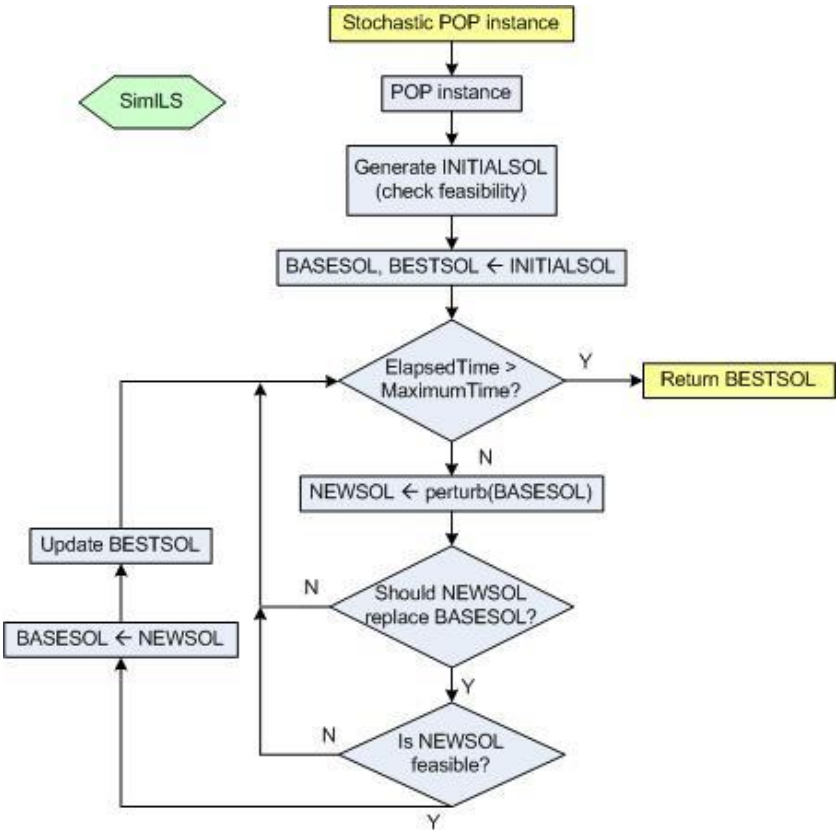


Fig. 2. Flowchart of our approach

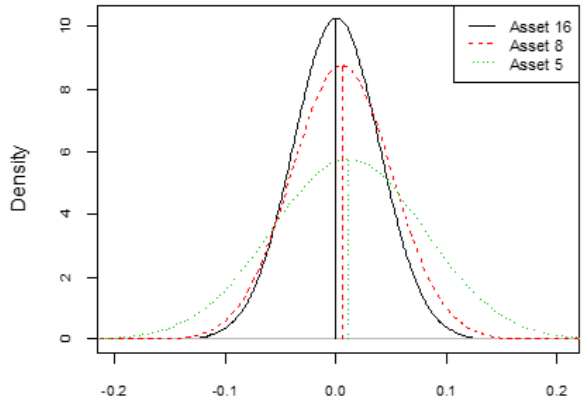


Fig. 3. Returns of selected assets following Normal distributions

Table 1. Correlation analysis between expected returns and standard deviation

Correlation coefficient	P-value	IC (95%)
0.4973	0.0044	0.1736-0.7241

Our algorithm is executed 10 times using different seeds; only the best results are shown. To minimize the computational time, the number of runs for assessing promising solutions has been set to 2000. The other parameters, including the time of the iterative procedure, have been set to the values suggested in [13].

100 equidistant values for the required rate of return have been selected. Table 2 shows the first and the last 5 observations on the required return, risk and reliability (or probability of the return being no smaller than the required return) associated to the solution obtained by the original algorithm (i.e., considering expected values), the risk found with our methodology considering the probabilities of 0.48 and 0.52, and the gap between them. The solutions of our methodology were obtained in 4.783 seconds on average.

Table 2. Table of results

	Expected values		P_0: 0.48	P_0: 0.52	
Required return	Risk	Reliab. (%)	Risk (1)	Risk (2)	Gap (2)-(1) (%)
0.002861	0.000642	50.07	0.000642	0.000645	0.00024
0.002942	0.000643	50.28	0.000643	0.000645	0.00023
0.003023	0.000644	49.22	0.000644	0.000648	0.00047
0.003104	0.000644	49.55	0.000644	0.000646	0.00019
0.003185	0.000645	50.12	0.000645	0.000647	0.00011
0.010542	0.004194	49.02	0.004194	0.004194	0.00000
0.010622	0.004332	50.64	0.004332	0.004475	0.00014
0.010703	0.004475	49.90	0.004475	0.004475	0.00000
0.010784	0.004623	50.12	0.004623	0.004623	0.00000
0.010865	0.004776	49.67	0.004776	N/A	N/A

5 Analysis of Results

The computational results suggest that requiring returns above a given threshold with a higher probability leads to the same or higher portfolio variance. Moreover, the instance may become unsolvable.

Fig. 4 shows the differences in terms of risk between the 0.48 and 0.52 probability solutions for each of the 100 equidistant values. Solutions associated to higher probability in general have greater risk. Fig. 5 displays a multiple boxplot which depicts the distribution of returns obtained using MCS for a set of promising solutions when solving the instance for a specific required return. They are sorted according to the portfolio variance. Although no obvious differences can be identified, eyeballing suggests that the second solution have the smallest range. This case shows that it can be useful to provide a set of solutions to the decision-maker. Here, he would choose the first (risk: 0.0006444), which minimizes the risk, or the second (risk: 0.0006449), which has a slightly higher risk but with a lower return variability.

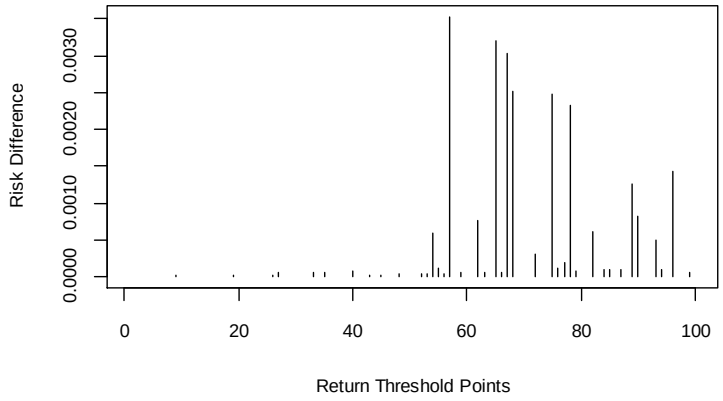


Fig. 4. Risk gaps considering two probabilities and 100 returns thresholds

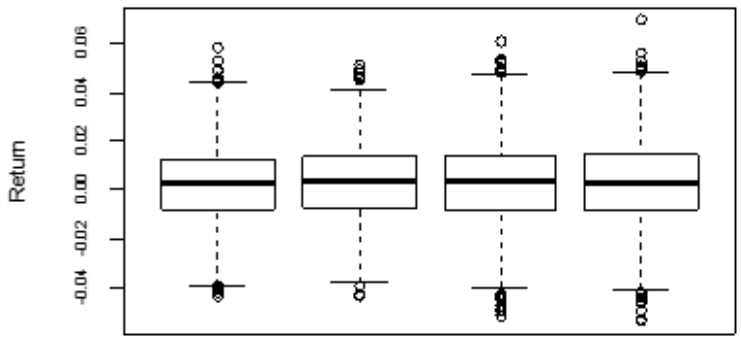


Fig. 5. Multiple Boxplot with returns distributions for several solutions

6 Conclusions

This work has addressed the Portfolio Optimization Problem (POP), which is a classic NP-hard Combinatorial Optimization Problem with plenty of applications. It consists in creating a portfolio selecting a subset of assets and setting their weights. Typically, authors solve this problem working with expected returns.

We have presented a mathematical formulation for the realistic POP, which considers the following constraints, commonly faced in real life: pre-assignment constraint (based on investor's preference), quantity constraint (which keeps each weight within user-specified floor and ceiling values) and cardinality constraint (providing a minimum and maximum value for the number of assets to include in the portfolio). Being a NP-hard problem, we require an approximate methodology for solving medium/high-sized instances in real time. Accordingly, we have proposed a simple methodology relying on the simILS approach. It combines an existing algorithm based on the Iterated Local Search metaheuristic for the classical version of the problem, which guides the search, with Monte Carlo simulation techniques, which checks the feasibility of promising solutions. A computational experiment employing an adapted benchmark instance is performed to illustrate its use and to analyze how the solutions change in terms of risk when varying the minimum required return and the probability of satisfying the constraint associated to this return.

Due to the stochasticity characterizing financial markets, the number of challenging versions of the POP, and their relevant applications, we plan to explore new methodologies or variants of the one presented to address problems in this field. For instance, an interesting line of research would be to consider more sources of stochasticity. Additionally, our methodology could be tested on instances describing different periods, countries, regions, sectors and asset classes.

Acknowledgments. This work has been partially supported by the Spanish Ministry of Economy and Competitiveness (TRA2013-48180-C3-P and TRA2015-71883-REDT), FEDER, and the Catalan Government (2014-CTP-00001).

References

1. Markowitz, H.M.: Portfolio selection, Journal of Finance 7, 77–91 (1952)
2. Kolm, P.N., Tütüncü, R., Fabozzi, F.J.: 60 years of portfolio optimization: Practical challenges and current trends, European Journal of Operational Research 234, 356–371 (2014)
3. Maringer, D.: Portfolio management with heuristic optimization, Springer (2005)
4. Bienstock, D.: Computational study of a family of mixed-integer quadratic programming problems, Mathematical Programming 74, 121–140 (1996)
5. Talbi, E.: Metaheuristics: from design to implementation, John Wiley & Sons (2009)
6. Boussaïd, I., Lepagnot, J., Siarry, P.: A survey on optimization metaheuristics, Information Sciences 237, 82–117 (2013)

7. Mansini, R., Ogryczak, W., Speranza, M.G.: Twenty years of linear programming based portfolio optimization, European Journal of Operational Research 234, 518–535 (2014)
8. Sawik, B.: Bi-Criteria Portfolio Optimization Models with Percentile and Symmetric Risk Measures by Mathematical Programming, Przeglad Elektrotechniczny 88, 176–180 (2012)
9. Adebiyi, A. A., Ayo, C. K.: Portfolio Selection Problem Using Generalized Differential Evolution 3, Applied Mathematical Sciences 9, 2069–2082 (2015)
10. Grasas, A., Juan, A., Lourenço, H.R.: SimILS: a simulation-based extension of the iterated local search metaheuristic for stochastic combinatorial optimization, Journal of Simulation, doi: 10.1057/jos.2014.25 (2015)
11. Lourenço, H.R., Martin, O.C., Stützle, T.: Iterated local search. In: Handbook of Metaheuristics, ed. Glover, F., Kochenberger, G., 321–353, Kluwer Academic Publishers: Norwell, MA (2003)
12. Juan, A., Faulin, J., Grasman, S., Rabe, M., Figueira, G.: A review of Simheuristics: extending metaheuristics to deal with stochastic optimization problems, Operations Research Perspectives 2, 62–72 (2015)
13. Kizys, R., Juan, A., Sawik, B., Linares, A.: Solving the Portfolio Optimization Problem under realistic constraints. In: Proceedings of the ICRA6/Risk 2015 Int. Conference, 457–464, Barcelona, Spain (2015)
14. Bianchi, L., Marco, D., Gambardella, L.M., Gutjahr, W.J.: A survey on metaheuristics for stochastic combinatorial optimization, Natural Computing 8, 239–287 (2009)
15. Talbi, E., Metaheuristics: from design to implementation, John Wiley & Sons (2009)
16. Chang, T.-J., Meade, N., Beasley, J.E., Sharaiha, Y.M.: Heuristics for cardinality constrained portfolio optimization, Computers and Operations Research 27, 1271–1302 (2000)