The Portfolio Selection Problem: Opportunities for constrained–based metaheuristics

Student: Giacomo di Tollo
Supervisor: Andrea Roli

Dipartimento di Scienze
Università "G.D'Annunzio" Chieti–Pescara
ditollo@sci.unich.it

Abstract. The Portfolio Selection Problem is a well-known area of application for metaheuristics, but its basic formulation fails in incorporating real-world features. In this work we discuss some issues about how to enrich the model by introducing features and constraints to obtain realistic results.

Introduction

Portfolio selection is one of the most studied topics in finance: the problem (referred to as PSP), in its basic formulation, is concerned with selecting the portfolio of assets which minimize the risk, given a certain level of returns. The basic model is formulated in the seminal work by Markowitz[6], in which the formulation of the problem is given by minimizing the variance (as a risk measure) for a given level of return:

$$\min \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij} x_i x_j \tag{1}$$

$$\sum_{i=1}^{n} r_i x_i \geq r_p \quad \sum_{i=1}^{n} x_i = 1 \quad x_i \in [0,1] \quad i,j = 1 \ldots n \tag{2}$$

where σ_{ij} represents covariance between assets i and j, r_p is the expected return rate and r_i is the (actual or forecasted) return rate of asset i. Note that portfolios are modeled as sets of assets whose weight sum up to one and can assume any value in the range [0, 1].

In this formulation the problem is solvable with exact methods, but when adding additional features, it becomes untractable even for small instances. So metaheuristic approaches have been exploited to solve realistic instance of portfolio selection[2][7].

In the following we will discuss about features that improve PSP formulation by considering real world investor behavior. These features can be modeled as constraints in a CP framework and efficiently tackled by solution procedures as metaheuristics.

1 Constraints

A shortcoming of the introduced formulation is that it lacks incorporating many aspects of real-world trading: maximum size of portfolio, minimum lots, transaction costs, preferences of which assets to include in the portfolio and by how much, management costs, etc. These aspects can be formulated by introducing constraints and in the following we will introduce some of the most relevant ones.

Cardinality Constraints The number of assets in the portfolio is limited. Introducing for each asset a binary variable z ($z = 1$ if asset is in the portfolio and 0 otherwise), the constraint can be expressed as follows:

$$\sum_{i=1}^{n} z_i = k \tag{3}$$

This constraint can be defined also in inequality form, imposing that the portfolio must contain no more ($\leq$) than k assets, and can be, of course, expressed also as a global cardinality constraint in CP.

Floor and Ceiling Constraints With this constraint we impose a minimum and maximum proportion (ε_i and δ_i) allowed to be held for each asset in portfolio. In other words the portion of the portfolio for a specific asset (each asset or some of them) must be included in a fixed interval. Generally, floor constraint is used to avoid the cost of administrating tiny portions of assets, while ceiling constraints to avoid excessive exposure to a specific asset (in some case it is imposed by law).

$$\varepsilon_i z_i \leq x_i \leq \delta_i z_i \tag{4}$$

Minimum lots In the literature, investments are generally continuously divisible, so as to be represented by a real variable, while in real world securities are negotiated as multiples of minimum lots: for each asset there exists a minimum tradable lot, generally referred to as *round*. This constraint cannot be added in the continuous model since rounds are expressed in money, while in the continuous model assets-portions are chosen regardless of their absolute value. For these reasons minimum lots are generally encountered only when dealing with the integer formulation[5], in which assets are labeled by their actual value rather than their ratios to whole portfolio. In integer values, if p_j is the price of asset j and ml_j its minimum tradable quantity, the minimum lot expressed in money is given by $c_j = ml_j p_j$.

Adding those issues to the original formulation makes the problem very hard to be solved by exact methods. Hence the need for designing efficient approximate algorithms, such as metaheuristics[1].

2 Neighborhood and Repair mechanism

In order to develop and fathom powerful local search strategies a key point is to define and understand the neighborhood relationship. This is a crucial point in metaheuristics: Often in the literature the introduction of neighborhood is not grounded to explicit motivations; this can lead to a misunderstanding of the algorithm behavior or to wrong conclusion referring to the applicability of such algorithm to specific problems (or instances). The problem formulation we are discussing requires the definition of constraints of various nature to model real-world features, and when local search is dealing with constraints, the neighborhood can be implemented in the following ways:

- *all feasible* approach: each candidate solution $s\prime$, belonging to the neighborhood of a current solution s, must satisfy the constraints at any step of the search process;
- *repair* approach, in which if a non-feasible solution is found, this is suddenly forced to satisfy constraints by an embedded repair-mechanism;
- *penalties* approach: we allow moving toward non-feasible solutions, but those will be assigned a penalty in the objective function, depending on the amount of violation.

Sometimes it turns out to be difficult determining which class a search method belongs to, as can be difficult to determine if a search trajectory moves only in feasible areas because of its formulation or because an implicit repair mechanism is embedded. Repair-mechanism has the effect to consider a large number of candidate solutions, but, in our opinion, it can cause loss of information and waste good partial solution features, even if it has the effect of reducing execution time.

A typical repair mechanism is explained in Streicher et al.[8], referring to a formulation with cardinality and minimum lots. This procedure takes as input a non normalized weight-vector, in which each weight represents the portion of portfolio held by an asset, and operates as follows:

1. all weights are normalized so as to sum to one. This is done by setting weights $x_i\prime = x_i / \sum_j x_j$;
2. the obtained vector is normalized so as to meet cardinality constraint: Only the k assets with largest value of $x_i\prime$ are held and then normalized to sum up to one;
3. a further normalization is required to meet minimum lots constraints: Weights of assets are forced to the previous roundlot level $x_i\prime\prime = x_i\prime - (x_i\prime \bmod c_i)$. The free portfolio amount is redistributed so as to meet minimum lots constraint buying quantities of c_is on assets with biggest $(x_i\prime \bmod c)$ until all the remainder is spent.

In this mechanism, at point 1), the repair mechanism operates normalizing all assets in the portfolio. Nevertheless, there is evidence that investors choose, for their portfolio, one highly risky asset (or a few ones) with high weight, while the remainder is partitioned in lots of small weights used to reduce risk.

In this situation, the former repair mechanism would loose important informations about the structure of portfolio. For this reason a new repair mechanism able to return a feasible solution composed of a vector $\overline{feas}$ can be defined just replacing the point 1) of the previous mechanism with the following routine:

1. Order assets in non-increasing-weights. Let o be the resulting vector[1];
2. Compute the vector $d = (o_1 - o_2), (o_2 - o_3), \ldots (o_{n-1} - o_n)$; let d_i be the i-th component in the sequence;
3. Let m be the index of the maximum element in d: This represents the maximum distance between weights of adjacent assets in the ordered array o;
4. **if** $\sum_{a=1}^{m} o_a \geq 1$ return the vector

$$\overline{feas}_i = \frac{o_i}{\sum_{j=1}^{n} o_j}$$

5. **if** $\sum_{a=1}^{m} o_a < 1$ return the vector

$$\overline{feas}_i = \begin{cases} o_i & i = 1 \ldots m \\ o_i \cdot \frac{1 - \sum_{j=1}^{m} o_j}{\sum_{l=m+1}^{n} o_l} & i = (m+1) \ldots n \end{cases}$$

3 Integer versus Continuous Formulation

The formulation we introduced (continuous fractional formulation in which weights must sum up to one) is universally used as standard approach in metaheuristics formulation: Modern Portfolio Theory relies on this formulation since it was introduced by Markowitz[6], but it represents a simplified model of real–world situations. We introduce now two issues difficult to handle with the continuous formulation.

Transaction costs Transaction costs are difficult to manage for the peculiar type of their function. As stated in Konno and Wijayanayake[3], the total costs follow a non-convex function on the size of the transaction: at the beginning it is concave up to a certain point (unit-transaction cost gradually decrease as size increase), then increases linearly to another certain point (unit-transaction costs are here constant) and then becomes convex due to the illiquidity premium (unit prices increases due to the shortage of supply). The transaction cost function is not easy to determine, but it appears to be discontinuous and it can be expressed as follows: $C = (1 + v)[f + \phi\big((b + p)s\big)] + ms$, where v is the VAT rate, f are fixed costs, b is the brokerage rate, p the illiquidity premium, s represents the transaction size, m the marketable securities tax rate and ϕ represents a subjective arbitrary function often difficult to interpolate and to define. Illiquidity premium plays an important role in this scheme and it can be introduced in different ways, but herein we consider it as an increment of the brokerage rate.

We must consider that, even if Modern Portfolio Theory states that diversified portfolio are preferable to undiversified ones, there is evidence that investors choose undiversified portfolios. This is due to the action of transaction costs, since they were not included in the original model. Considering all typologies, transaction costs tend to reduce portfolio-diversification: This is partially due

[1] For sake of readability the resulting vector will be composed of weight values and the label of the corresponding asset.

to the introduction of fixed costs, while proportional ones do not have effects because they generate only a decrease in returns rate. It is clear however that only proportional costs are suitable to be included in the continuous model, as the remainder is sensitive to the invested amount.

Solution methods and discretization When applying solution methods (e.g. heuristics and metaheuristic strategies) to the PSP, the implementation has to be studied with particular care: Some metaheuristics for instance are designed to work in a discrete search space, while in the problem we are considering variables can assume any continuous value belonging to [0, 1] range. The basic idea for adapting the model to these techniques would be to discretize the space; this operation might not be conceptually sound since we only consider that assets weights must sum up to one, regardless of the total amount to be invested. We might decide to apply a discretization at 0.000005 intervals, without any trouble for the formulation, but it is clear that if we have to invest 1,000,000,000 euros the discretized minimum admissible lot will be 5,000 euros, while if we consider 10 euros to be invested it will amount to 0.00005 euros. This will not turn into errors or warnings, but it is clear that the meaning of the efficient frontier could be strongly misleading depending on the invested amount.

These issues lead us to face the dilemma of the integer formulation (in which assets weights are labeled by their money-value) versus the standard continuous one: After the latter was introduced, no extension was developed in order to include transaction costs and to manage ambiguity arising from discretization and total-investment. These issues seem to suggest the use of integer formulation, easily obtained by replacing equation (2) with the following:

$$\sum_{i=1}^{n} r_i x_i \geq r_p C \quad \sum_{i=1}^{n} x_i = C \quad x_i \geq 0, integer \quad i, j = 1 \dots n \tag{5}$$

where C is the invested amount. The local search approach is, in our opinion, robust w.r.t the formulation, so it is able to handle the integer version too, ensuring important advantages: Lack of necessity of discretisation, correctness of meanings of formulation, easiness in including transaction costs and rounds. The diffusion of software tools such as *Comet*[4] that enable to implement metaheuristics while preserving a CP modelling approach help us include the issues discussed so far in the analysis and development of metaheuristic for the PSP. These packages enable us to define the model (so including various constraints in the formulation), and, separately, the search strategy, so that changing the first does not trigger bad or misleading behavior in the latter. In the PSP, this is of the most importance, as constraints *must* be added to the formulation in order to obtain satisfactory results. In real-world applications these constraints can be classified in two main types:

- Hard Constraints imposed in order to make the model the most realistic as possible (these constraints must be satisfied for each category of investor in each area we are taking into account);

– Soft Constraints, imposed in order to describe preferences and behaviors of investors to whom the analysis is directed (each of this constraint must be defined to describe a specific category or area).

The second class of constraints has often been under-considered. In our opinion, efforts in introducing soft constraints can help develop solutions for several class of investors, geographic areas, regulations and so on: Metaheuristics have already been tested on this problem showing satisfactory results on the basic formulation, and in our opinion the next advance has to be made on modelling. This can be achieved by embedding in the current formulation aspects already discussed in the economic and financial literature about portfolio selection, but only a few of them has been investigated empirically. For example, no comparison amongst the different risk measures has been made on the same instance.

4 Conclusions and future works

The portfolio selection problem has been proven to be suitable for a metaheuristic approach in which the formulation is enriched by constraints used both to define the problem formulation and explain investor behavior. The versatility of these strategies enables us to add and change various aspects of the formulation without affecting the search-process. Further research is aimed at formulating an integer model in which constraints can be easily defined and included, and to use it to obtain real-world oriented results. A comparison between integer and continuous formulation will be performed in order to show differences between the resulting portfolios (if any);furthermore a comparison of different risk measures and a formulation embedding illiquidity-premium-transaction cost will be studied.

References

1. C. Blum and A. Roli. Metaheuristics in combinatorial optimization: Overview and conceptual comparison. *ACM Computing Surveys*, 35(3):268–308, September 2003.
2. T.J. Chang, N. Meade, J.E. Beasley, and Y.M. Sharaiha. Heuristics for cardinality constrained portfolio optimisation. *Comput. Oper. Res.*, 27(13):1271–1302, 2000.
3. A. Wijayanayake H. Konno. Mean-absolute deviation portfolio optimization model under transaction costs. *Journal of the Operational Research Society of Japan*, 42(4), 1999.
4. P. Van Hentenryck and L. Michel. *Constraint-Based Local Search*. The MIT Press, 2005.
5. H. Kellerer, R. Mansini, and M.G. Speranza. Selecting portfolios with fixed costs and minimum transaction lots. *Annals of Operational Research*, 99(4), 2000.
6. H. Markowitz. Portfolio selection. *Journal of Finance*, 7(1):77–91, 1952.
7. A. Schaerf. Local search techniques for constrained portfolio selectionproblems. *Comput. Econ.*, 20(3):177–190, 2002.
8. F. Streichert, H. Ulmer, and A. Zell. Evaluating a hybrid encoding and three crossover operators on the constrained portfolio selection problem. In *Proceedings of Congress on Evolutionary Computation (CEC 2004)*, 2004.