Document downloaded from:

http://hdl.handle.net/10251/107467

This paper must be cited as:

Guijarro, F. (2018). A similarity measure for the cardinality constrained frontier in the mean-variance optimization model. Journal of the Operational Research Society. 69(6):928-945. doi:10.1057/s41274-017-0276-6

The final publication is available at

http://doi.org/10.1057/s41274-017-0276-6

Copyright Nature Publishing Group - Macmillan Publishers

Additional Information

A similarity measure for the cardinality constrained frontier in the mean-variance optimization model

Author

Francisco Guijarro

Technical University of Valencia

fraguima@upvnet.upv.es

Abstract

This paper proposes a new measure to find the cardinality constrained frontier in the mean-variance portfolio optimisation problem. In previous research, assets belonging to the cardinality constrained portfolio change according to the desired level of expected return, so that the cardinality constraint can actually be violated if the fund manager wants to satisfy clients with different return requirements. We introduce a perceptual approach in the mean–variance cardinality constrained portfolio optimization problem by considering a novel similarity measure, which compares the cardinality constrained frontier with the unconstrained mean-variance frontier. We assume that the closer the cardinality constrained frontier to the mean-variance frontier, the more appealing it is for the decision maker. This makes the assets included in the portfolio invariant to any specific level of return, through focusing not on the optimal portfolio but on the optimal frontier.

Keywords

Finance; Portfolio selection; Constrained frontier; Cardinality; Genetic algorithm

1. Introduction

Financial institutions, pension funds and private investors face the problem of investing billions of dollars in assets according to the appropriate levels of expected return and risk. The problem of asset selection is the core of portfolio theory and the seminal work of Markowitz (1952) sets the framework of the modern portfolio theory. The portfolio optimisation method proposed by Markowitz assumes a market with N different assets with known expected returns and risks (variances and covariances), where asset returns follow a multivariate normal distribution. Efficient portfolios are defined as a trade-off between risk and expected return, and these portfolios can be computed by minimizing the risk of the portfolio constrained to a desirable return, or by maximizing the return of the portfolio constrained to a certain risk.

The portfolios that set up the efficient frontier can be found by quadratic programming, and the selection process can be constrained by practical considerations, which can be written as linear constraints (Barak et al, 2013; Chang et al, 2000; Bertsimas and Shioda, 2009). Although the mean-variance model is the fundamental theory of modern portfolio theory, the direct application of this model is not of much practical use mainly due to the fact that it is simplified by means of a number of unrealistic assumptions (Lwin and Qu, 2013). Real-world investors and fund managers commonly face restrictions such as cardinality, because limiting the number of assets can simplify the management of the portfolio and reduce transaction costs. This is particularly important when some assets have a relatively small weight in the portfolio, so that we can reduce transaction costs by excluding them from the portfolio with no practical impact on its expected return and risk (Maringer and Kellerer, 2003; Canakgoz and Beasley, 2009; García et al, 2011; Cesarone et al, 2013).

Some studies have focused on solving the cardinality constrained efficient frontier (CCEF) by means of different exact techniques (Li et al, 2006; Shaw et al, 2008). Considering the constraints on the cardinality transforms the model from a quadratic to a quadratic mixed-integer problem, which has been proved to be NP-hard (Bienstock, 1995; Shaw et al, 2008; Ruiz-Torrubiano and Suarez, 2009). As the quadratic mixed-integer problem is difficult to solve optimally, many researchers have proposed the use of heuristics for CCEF computation; local search, simulated annealing, tabu search and genetic algorithms are heuristics techniques that have been successfully used in the past (Woodside-Oriakhi et al, 2011; Krink et al, 2009; Chang et al, 2009; Lwin and Qu, 2013; Maringer and Kellerer, 2003). When optimal solutions are known, heuristics find optimal or near optimal solutions in a sensible computation time, even after considering additional linear constraints such as upper and lower bounds, budget constraint, or no short sales.

A comprehensive survey of metaheuristics in portfolio selection can be found in Di Tollo and Rolli (2008) and Metaxiotis and Liagkouras (2012). In our survey of the literature below, we only include papers not reviewed by Di Tollo and Rolli (2008) and Metaxiotis and Liagkouras (2012).

Anagnostopounos and Mamanis (2011) perform a computational comparison of five multiobjective evolutionary algorithms on the mean–variance cardinality constrained portfolio optimization problem. Strength Pareto evolutionary algorithm II shows superiority over the rest of the algorithms. In second place, non-dominated sorting genetic algorithm II and e-multiobjective evolutionary algorithm have comparable performance.

Aouni et al (2013) formulate a Venture Capital Investment problem through a scenario-based and cardinality constrained stochastic Goal Programming model with a satisfaction

function. They include several objectives: the investment return, the survival rate, the intellectual capital rate, and the investment risk.

Cesarone et al (2013) reformulate the cardinality constrained model as a standard quadratic program. These results lead to an exact algorithm for solving small size problems. For larger problems, this algorithm can be relaxed to an efficient heuristic procedure that is able to find the optimal solution.

Liagkouras and Metaxiotis (2014) introduce a new Probe Guided Mutation operator and apply it to solving the cardinality constrained problem. The proposed mutation operator incorporates a fitness function evaluation mechanism that evaluates the fitness of the left-hand and right-hand regions of the parent solution. The results indicate that the proposed mutation operator clearly outperforms the classical Polynomial Mutation operator for all performance metrics.

Lwin and Qu (2013) propose a new hybrid algorithm that adopts a partially guided mutation and an elitist strategy to improve the quality of the solution. The computational results demonstrate that the proposed algorithm is both effective and efficient in solving the constrained mean-variance model.

Woodside-Oriakhi et al (2011) propose three metaheuristic algorithms based upon genetic algorithms, tabu search and simulated annealing to find the cardinality constrained efficient frontier. The results indicate that these heuristics give better quality solutions than the previous proposals, albeit at the expense of longer computation times.

We can therefore conclude that the literature on CCEF focuses on the problem of selecting the optimal portfolio for a specific level of expected return (or risk) and limiting the number of assets from N– the whole universe of assets – to k – the cardinality of the constrained portfolio –. As we detail in the following section, the particular assets involved in the portfolio can vary for different levels of return; i.e. if the constrained portfolio A has a desired expected return r_A^* and the constrained portfolio B has a desired expected return r_B^*, with $r_A^* \neq r_B^*$, both portfolios can differentiate from each other in one or more assets. Even worse, in the opinion of Chang et al. (2000), in the presence of cardinality constraints the efficient frontier may become discontinuous, where the discontinuities imply that there are certain returns which no rational investor would consider (since there exist portfolios with the same risk but with higher returns).

This situation can be considered as contradictory from the fund manager's point of view: if he or she tries to satisfy different client profiles, with different levels of desired expected return, in practice he or she would actually be considering more than k assets, regardless of the specific clients investing in k assets only. Our proposal is to compute a single constrained frontier which can satisfy all client profiles by investing in only certain k assets.

The aim of this paper is to focus on the cardinality constrained frontier, but not on a particular cardinality constrained portfolio, thus ensuring that the number of assets involved in the process will actually be constrained to k, whatever the risk profile of the particular investor. The novelty of this paper lies in its perceptual approach to the mean–variance cardinality constrained portfolio optimization problem by considering a novel similarity measure which compares the cardinality constrained frontier with the mean-variance frontier. The similarity ratio enables the computation and comparison of frontiers and the objective definition of the optimal constrained frontier. The similarity ratio is a quantitative measure that can be incorporated into the decision process, but it also allows the investor to visualize the solution that best fits his requirements. We also

propose the similarity ratio to calculate the degree of diversification of the constrained frontier with respect to the mean-variance frontier, which may be particularly relevant for the decision maker. The combinatorial problem of finding the optimal frontier constrained to k assets is solved by using a genetic algorithm, as proposed in a wide range of recent studies (Ruiz-Torrubiano and Suárez, 2009; Woodside-Oriakhi et al, 2011; Anagnostopoulos and Mamanis, 2011; Chiam et al, 2013; Sant'Anna et al, 2016).

The paper is organized as follows: Section 2 summarizes the problem of finding the CCEF and the violation of the cardinality constraint when risk profile changes. We formally introduce the perceptual similarity ratio to show the similarity of the mean-variance unconstrained frontier and the cardinality constrained frontier. Section 3 offers the results of a computational study using five real-world stock market indices. We then summarize our conclusions and suggest some possible applications of the similarity ratio to the analysis of diversification in the portfolio selection problem.

2. Statement of the problem

In this section we formulate the classical Markowitz mean-variance portfolio model that computes the minimum variance frontier (MVF). The unconstrained efficient frontier (UEF) is devised by considering only the efficient MVF portfolios, whereas the CCEF is constructed when cardinality constraints are considered on the UEF. Using the example of Chang et al (2000) we discuss the problem that arises when different decision-maker risk profiles must be attended, and then we propose a perceptual similarity measure for the optimal selection of the constrained frontier. The similarity ratio can also be useful to determine the minimum quantity of assets in portfolio to reach a reasonable level of diversification.

2.1 The minimum variance frontier and the unconstrained efficient frontier

The construction of the Markowitz MVF is based on the computation of optimal mean-variance portfolios considering both the criteria of expected return and risk. Let

N be the number of assets available to invest in,

$\mathbf{r}$ be the $N \times 1$ return vector of assets,

$\mathbf{V}$ be the $N \times N$ covariance matrix of returns,

$\mathbf{1}$ be the $N \times 1$ vector in which every entry equals 1,

r^* be the desired expected return of the portfolio.

The vector decision variable is:

$\mathbf{x}$, the $N \times 1$vector which includes the weight invested in each asset.

The MVF can be obtained by solving the following quadratic problem:

$$Min \quad \sigma^2 = \mathbf{x}^t \mathbf{V} \mathbf{x} \tag{1}$$

subject to

$$\mathbf{x}^t \mathbf{1} = 1 \tag{2}$$

$$\mathbf{x}^t \mathbf{r} = r^* \tag{3}$$

Eq. (1) minimises the risk of the portfolio whilst Eq. (2) ensures that all the money available is invested and Eq. (3) ensures the desired expected return. In this way the model (1)-(3) enables us to obtain different portfolios by adjusting the required expected return r^*, and the MVF consists of the combination of all these portfolios.

As assumed in the Markowitz mean-variance model, only a subset of the portfolios in the MVF belongs to the UEF. Efficient portfolios are the non-dominated ones, those with a higher return than the global minimum variance portfolio, which is defined as the MVF portfolio with the minimum risk. The term *unconstrained* refers to the frontier produced by considering the basic constraints only (2)-(3), in order to differentiate it from the frontier that considers additional constraints (e.g. cardinality). We simplified our approach by consciously excluding the no short sales constraint, another usual constraint formerly considered by Markowitz (1952).

2.2 The cardinality constrained frontier

Restricting the cardinality of portfolios can be addressed by adding the following constraint to model (1)-(3):

$$\mathbf{x}^t\mathbf{k} = k \tag{4}$$

where $\mathbf{k}$ is a binary vector with k ones and N-k zeros. A value of one in the i-th position denotes that the i-th asset is included in the portfolio, whereas a value of zero means that the corresponding asset is not in the portfolio.

The binary vector $\mathbf{k}$ thus specifies whether a particular asset participates in the portfolio, and the real-valued vector $\mathbf{x}$ computes the proportions of the budget invested in the assets.

Table 1 gives the data for the example of Chang et al (2000) with $N = 4$. The original data consists of return, standard deviation and correlation. We have substituted the correlation matrix for the covariance matrix, which is the one we use in both the unconstrained and constrained examples below.

Asset	Return	Standard deviation	Covariance matrix			
			1	2	3	4
1	0.004798	0.046351	0.002148	0.000168	0.000203	0.000418
2	0.000659	0.030586	0.000168	0.000935	0.000153	0.000109
3	0.003174	0.030474	0.000203	0.000153	0.000929	0.000091
4	0.001377	0.035770	0.000418	0.000109	0.000091	0.001279

Table 1. Data example from Chang et al. (2000)

Fig. 1 contains the MVF and the six cardinality constrained frontiers (CCF) constructed by considering $k = 2$ constrained portfolios from the data in Table 1. As in the case of the MVF, we can draw the CCEF by considering only the subset of portfolios in CCF

with a higher return than that the global minimum variance portfolio [1] of the corresponding CCF. This example helps us to introduce our approach based on the whole CCF, and not on a particular portfolio in the CCF following the model (1)-(4).

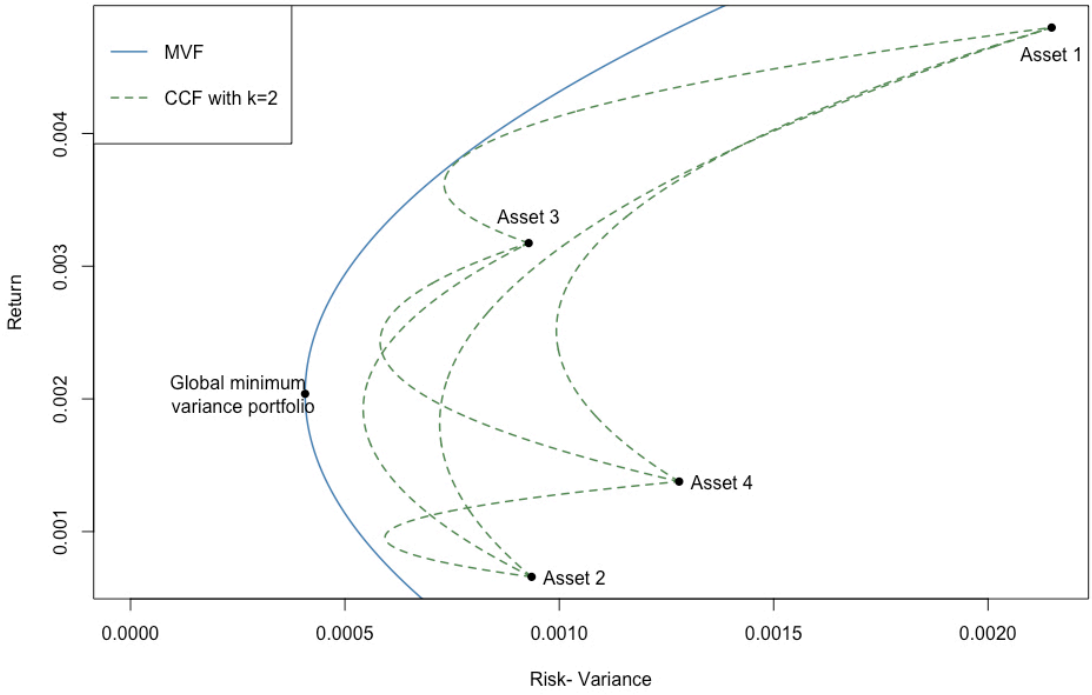


Figure 1. MVF and six different constrained frontiers with $k = 2$

If the decision maker aims at getting an expected return of 0.004, he or she would combine assets 1 and 3 and so will invest in the CCF obtained with these two assets. However, if he or she has a moderate risk profile and settles for an expected return of 0.0025, then the selected CCF should be the one combining assets 3 and 4. For an even more cautious expected return of 0.002, the considered CCF will combine assets 2 and 3. Therefore, in the usual case that the decision maker has to satisfy clients with different risk profiles, then he or she would finally invest in assets 1, 2, 3 and 4. This clearly violates the imposed cardinality constraint $k = 2$.

Summarizing, three different desired returns configure three different combinations of assets. If the decision maker is a fund manager and has to satisfy different client risk profiles, then he or she will in fact be investing in more than k assets, even though model (1)-(4) restricts the cardinality to k.

[1] The global minimum variance portfolio can be obtained by considering the model (1)-(2).

2.3 The similarity ratio approach

The following question now arises: is there any preferable CCF regardless of the desired expected return for the fund manager to invest in only k assets? Considering $k = 3$ portfolios in the above example helps us to introduce our approach.

Fig. 2 represents the MVF and four possible CCF with $k = 3$. In this example short sales are allowed, so neither the MVF nor CCF curves in the return axis are restricted to the observed profitability of the assets. Now we can see how the CCF considering assets 1, 2 and 3 – CCF(1,2,3) – dominates the other three CCF for a wide range of returns. In fact, the domination of CCF(1,2,3) in the mean-variance space goes from the return on portfolio $p1$ (0.002310) to the return on portfolio $p2$ (0.003147), and from the return on portfolio $p3$ (0.003822) to the highest return considered in Fig. 2 (0.005758).

CCF(1,3,4) is the second frontier with the widest range of return in the sense of dominance, with the best return-risk combination between the return on portfolio $p2$ and that on portfolio $p3$. However, even in this return range the improvement over CCF (1,2,3) is minimal: both frontiers are very close to each other between portfolios $p2$ and $p3$.

CCF(2,3,4) dominates the other curves in the narrow range from the return on the global minimum variance portfolio to that on portfolio $p1$. Finally, no decision maker would choose a portfolio with lower returns than the global minimum variance portfolio, so CCF(1,2,4) is not prioritized as the dominant CCF in any case.

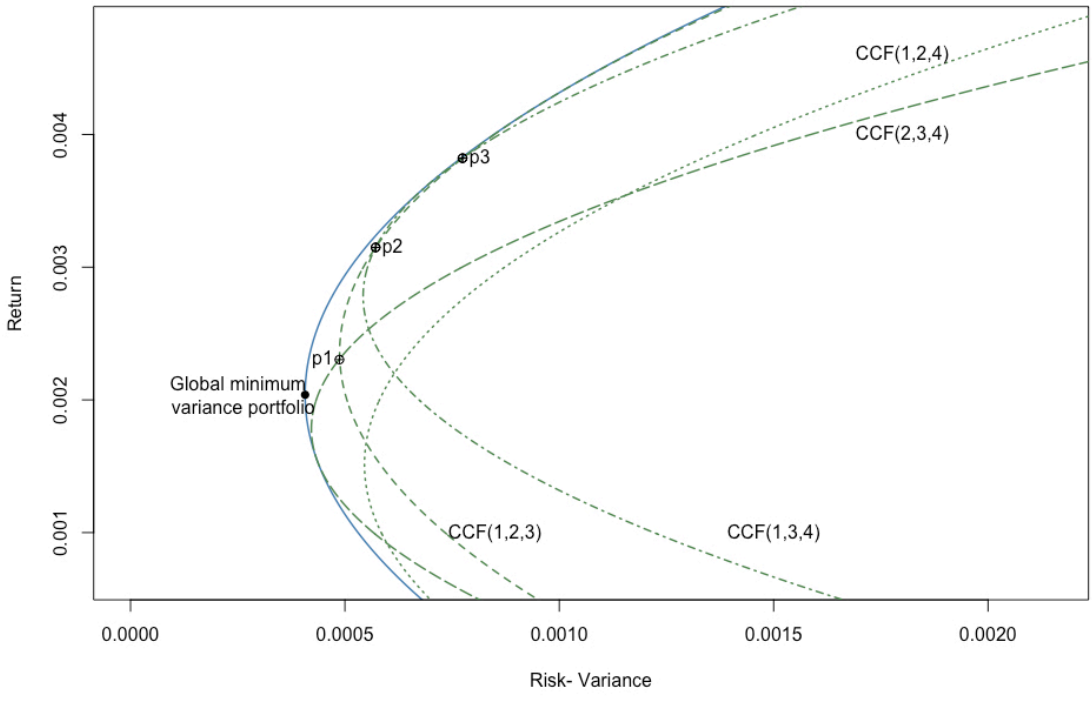


Figure 2. CCF mean-variance dominance

From a decisional point of view, the most preferable CCF is the one that fully dominates the other CCF for the whole range of returns. However, as we can observe in the example in Fig. 2, it is difficult to find a CCF that fully dominates the remaining CCF, regardless of the expected return.

For the CCF comparison we propose to limit the decision space of the return to a sensible range: between the return on the global minimum variance portfolio and the maximum return on the assets. The global minimum variance portfolio defines the minimum desired return for any rational investor, because it differentiates the efficient part of the MVF from the non-efficient part. Any MVF portfolio with a lower return than the global minimum variance portfolio will have a greater risk, so no rational investor will be interested in it. On the other hand, the asset with the maximum return can be viewed as the best achievable return if no short sales are considered. In fact, it is somewhat over-optimistic to try to obtain a portfolio with a higher return than the most profitable asset. So, both return bounds represent reasonable extremes for the return desired by the decision maker.

Bearing this in mind, the shaded section in Fig. 3 outlines the return-risk space we are considering in the decision process. Notice that the lower and the upper bounds in the return axis have equivalent bounds in the risk axis.

The lower bound for risk has now been described, and the computation of the upper bound will be explained below.

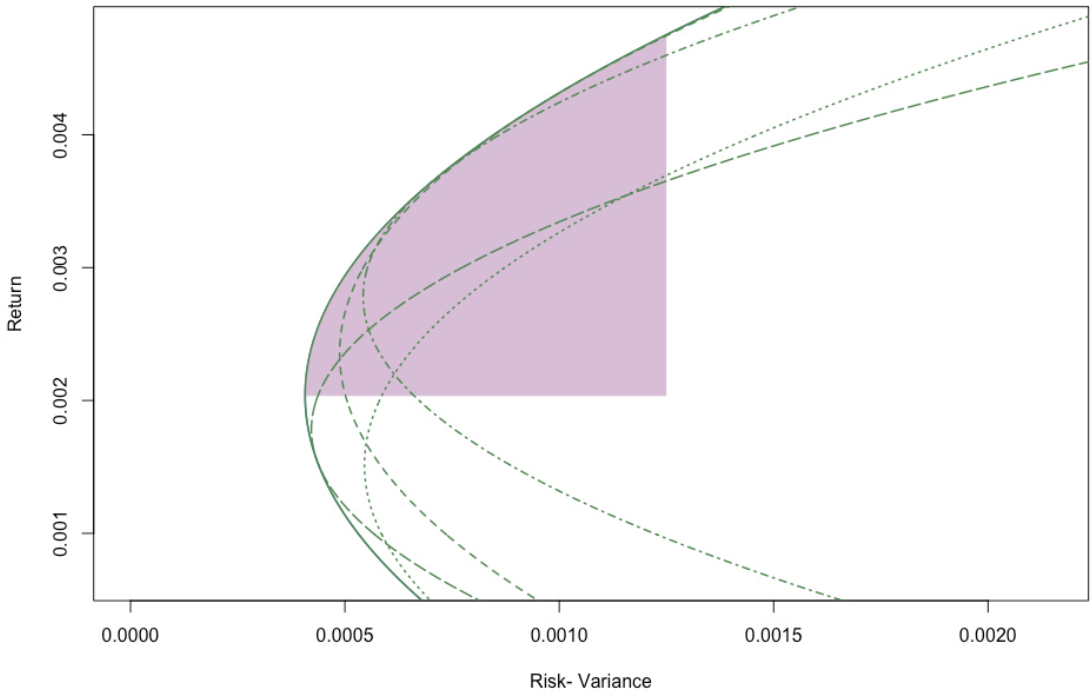


Figure 3. Mean-variance space for the decision process of selecting the optimal CCF

If we face the decision maker with this constrained space, he or she will perceive differences between the four CCF considered in the example (Fig. 4). Obviously CCF(1,2,4) will be discarded immediately because it is the most unfavourable choice. He or she must therefore compare the other three CCF in order to establish which one is "perceptually" more appealing according to both risk and return criteria. We consider it important to emphasize the term *perceptual*, because as we mentioned above, it is extremely unlikely to find a CCF that fully dominates the others in the mean-variance space.

In this situation, when several CCF are competing and the decision depends on the specific return level, the decision maker can face the problem from a perceptual perspective that jointly considers both return and risk, in a similar way to when he/she was only given the visualization of the solutions.

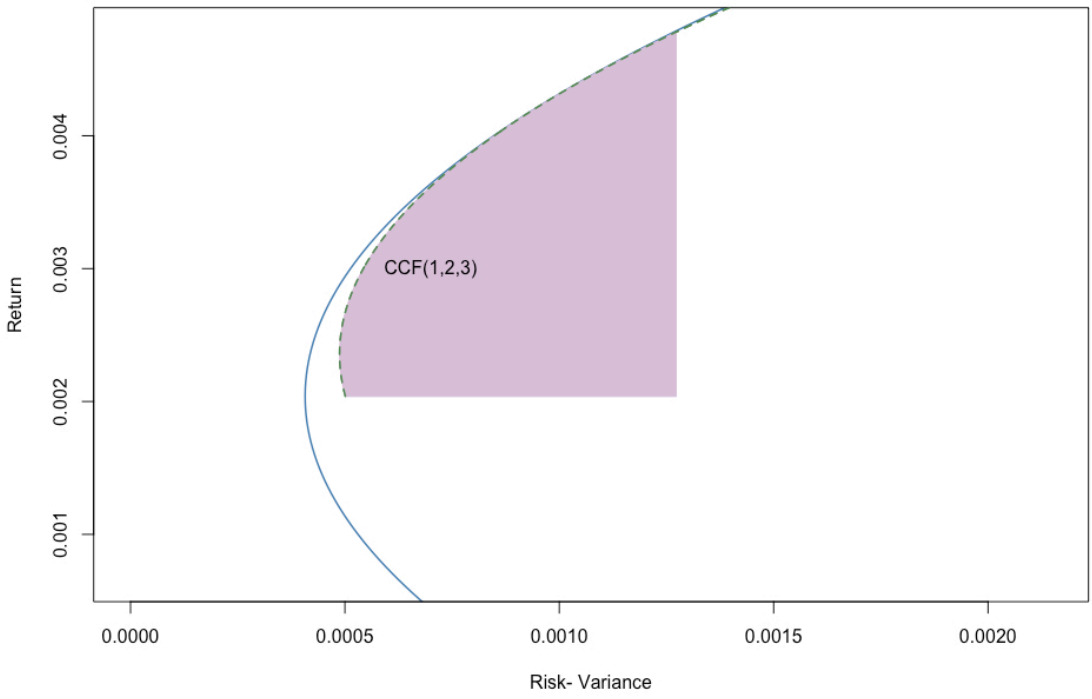
We assume for simplicity that the closer the CCF is to the MVF, the more likely it is to be selected by the decision maker. This is quite a sensible approach, because if we face an investor with two frontiers and ask him to choose one, he or she will prefer the most similar and closest to the MVF. With this in mind, we need to measure how perceptually similar each of the CCF is to the MVF, also considering that the similarity measure must be constrained to the decisional area we have depicted in Fig. 3.

We define the similarity ratio for the i-th CCF as the ratio between the CCF area and the MVF area in the considered return range.

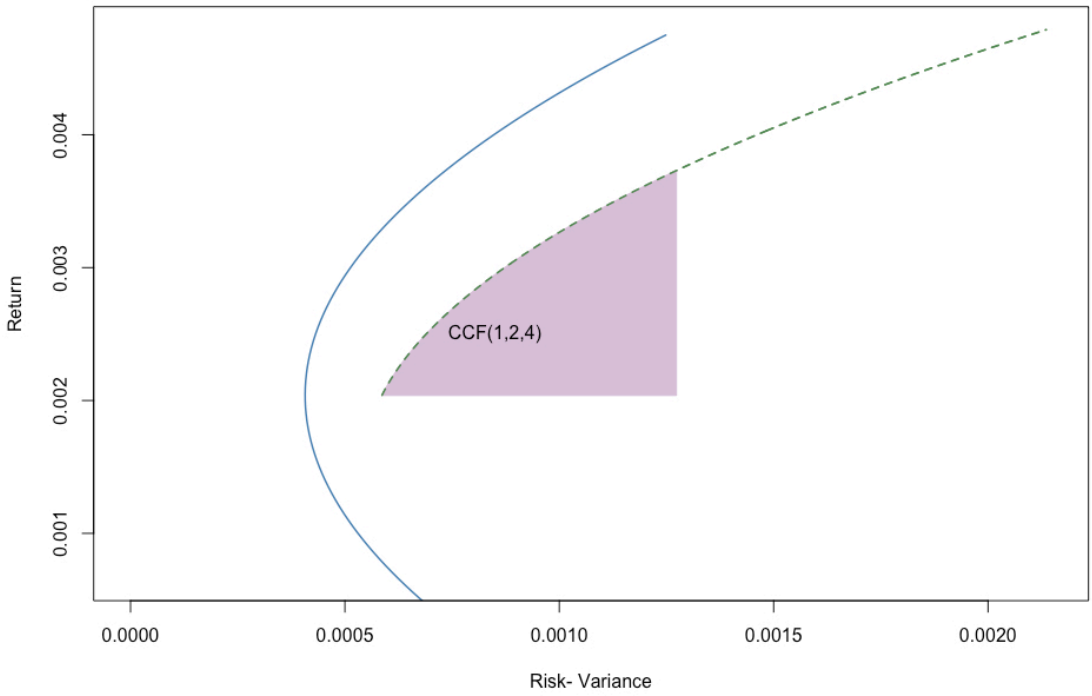
$$Similarity_i = CCF_i\ Area/MVF\ Area \quad r \in return\ range \tag{5}$$

Because the MVF area is the same regardless of the CCF, we can also rank the different CCFs by directly comparing their areas.

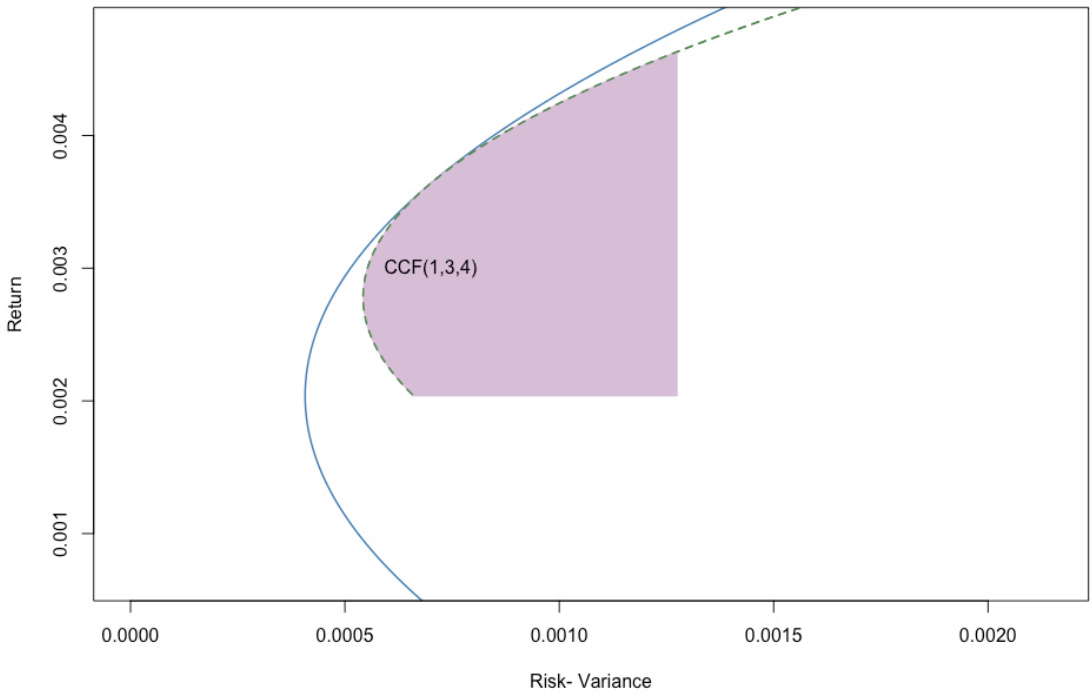
In Fig. 4 we have shaded the intersection area for the MVF and each of the CCF considered in the example, given the above established bounds for the return range. Note that CCF(1,2,4) is the least similar to the MVF. The similarity ratio for CCF(2,3,4) is slightly better than the one obtained for CCF(1,2,4), but not good enough if we compare it with the other two CCF. According to the area of CCF(1,2,3) and CCF(1,3,4), the decision maker will choose CCF(1,2,3) as the most similar to the MVF. CCF(1,3,4) can beat CCF(1,2,3) in a small range of returns, but from the perceptual point of view the area of CCF(1,2,3) is much closer to the MVF.



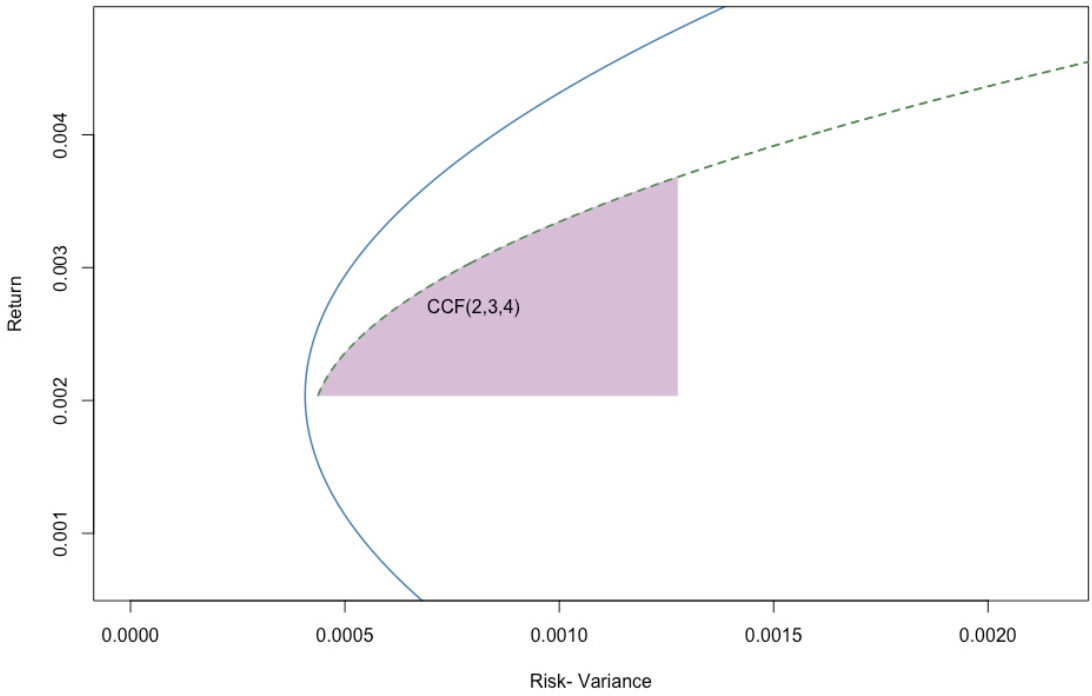
A. The MVF and the CCF with assets 1, 2 and 3 – CCF(1,2,3)



B. The MVF and the CCF with assets 1, 2 and 4 – CCF(1,2,4)



C. The MVF and the CCF with assets 1, 2 and 4 – CCF(1,3,4)



D. The MVF and the CCF with assets 1, 2 and 4 – CCF(2,3,4)

Figure 4. Comparison of the MVF and four different CCF

The area can therefore help us to transfer our subjective perceptual concept of similarity to a more objective measure, and this enables the ranking of CCFs from a decisional point of view.

Once we have defined the measure we use to compare the different CCF, we need to compute both the areas of both the MVF and CCF before estimating the similarity ratio. For this we first introduce some mathematical preliminaries.

We now briefly outline the mathematical formulation for the weights **x** of a portfolio that belongs to the MVF. It can be calculated from the model (1)-(3) by using Lagrange multipliers (Merton, 1972):

$$\frac{1}{2}\mathbf{x}^t\mathbf{V}\mathbf{x} - \lambda_1(\mathbf{x}^t\mathbf{r} - r^*) - \lambda_2(\mathbf{x}^t\mathbf{1} - 1) \tag{6}$$

Then the standard first-order conditions for a critical point are (7)-(9):

$$\mathbf{V}\mathbf{x} - \lambda_1\mathbf{r} - \lambda_2\mathbf{1} = \mathbf{0} \tag{7}$$

$$\mathbf{x}^t\mathbf{r} - r^* = 0 \tag{8}$$

$$\mathbf{x}^t\mathbf{1} - 1 = 0 \tag{9}$$

Where $\mathbf{0}$ is a column vector of zeros with dimension N, and 0 is the scalar zero.

Making some algebraic calculations, the solution for $\mathbf{x}$ is:

$$\mathbf{x} = \mathbf{V}^{-1}[\mathbf{r} \quad \mathbf{1}]\mathbf{A}^{-1}\begin{bmatrix} r^* \\ 1 \end{bmatrix} \tag{10}$$

where $\mathbf{A}$ is a 2x2 symmetrical matrix:

$$\mathbf{A} = [\mathbf{r} \quad \mathbf{1}]^t\mathbf{V}^{-1}[\mathbf{r} \quad \mathbf{1}] = \begin{bmatrix} a & b \\ b & c \end{bmatrix} \tag{11}$$

$$a = \mathbf{r}^t\mathbf{V}^{-1}\mathbf{r}, \;\; b = \mathbf{r}^t\mathbf{V}^{-1}\mathbf{1}, c = \mathbf{1}^t\mathbf{V}^{-1}\mathbf{1} \tag{12}$$

The variance of the portfolios in the MVF can be expressed in terms of the expected return $r^*$and the scalars a, b and c:

$$\sigma^2 = \mathbf{x}^t\mathbf{V}\mathbf{x} = \left(\mathbf{V}^{-1}[\mathbf{r} \quad \mathbf{1}]\mathbf{A}^{-1}\begin{bmatrix} r^* \\ 1 \end{bmatrix}\right)^{\boldsymbol{t}} \mathbf{V}\left(\mathbf{V}^{-1}[\mathbf{r} \quad \mathbf{1}]\mathbf{A}^{-1}\begin{bmatrix} r^* \\ 1 \end{bmatrix}\right) =$$

$$([r^* \quad 1]\mathbf{A}^{-1}[\mathbf{r} \quad \mathbf{1}]^t\mathbf{V}^{-1})\mathbf{V}\left(\mathbf{V}^{-1}[\mathbf{r} \quad \mathbf{1}]\mathbf{A}^{-1}\begin{bmatrix} r^* \\ 1 \end{bmatrix}\right) =$$

$$[r^* \quad 1]\mathbf{A}^{-1}[\mathbf{r} \quad \mathbf{1}]^t\mathbf{V}^{-1}[\mathbf{r} \quad \mathbf{1}]\mathbf{A}^{-1}\begin{bmatrix} r^* \\ 1 \end{bmatrix} = [r^* \quad 1]\mathbf{A}^{-1}\mathbf{A}\mathbf{A}^{-1}\begin{bmatrix} r^* \\ 1 \end{bmatrix} =$$

$$[r^* \quad 1]\mathbf{A}^{-1}\mathbf{A}\mathbf{A}^{-1}\begin{bmatrix} r^* \\ 1 \end{bmatrix} = [r^* \quad 1]\mathbf{A}^{-1}\begin{bmatrix} r^* \\ 1 \end{bmatrix}$$

Making some changes we get (13):

$$\sigma^2 = \frac{a - 2br^* + c(r^*)^2}{ac - b^2} \tag{13}$$

Thus, Equation (13) confirms that the frontier in the mean-variance space is a parabola. The analytical form of any parabola is quadratic (14):

$$y = f(x) = Ax^2 + Bx + C \tag{14}$$

with $A \neq 0$. Making simple changes we can express A, B and C in terms of a, b and c as defined in (12):

$$A = c/(ac - b^2) \tag{15}$$

$$B = -2b/(ac - b^2) \tag{16}$$

$$C = a/(ac - b^2) \tag{17}$$

First, note that in calculating each CCF we use different values for a, b and c. These scalars depend on $\boldsymbol{R}$ and $\mathbf{V}$, and then on the assets considered in the cardinality constrained model. For example, **R** and **V** will be different for CCF(1,2,3) and CCF(1,2,4) because there is a difference in the third asset.

Secondly, we are not interested in the whole area of each CCF parabola, but are restricted to the area in the return range. The minimum value for the return is the same for all the CCF considered: the return on the global minimum variance portfolio in the MVF, 0.002038. In Fig. 4 we can see how this minimum value in the return axis translates into the bottom segment for each shadowed area at the same return value. However, the maximum value for the return differs for each CCF. We are considering the maximum asset return as the upper bound of the return range. In the case of the data in Table 4 this upper bound corresponds to the return of asset 1, *i.e.*0.004798. The portfolio in the MVF that corresponds with this return has a variance of $var_{MVF} = 0.001278$. This value depicts the right segment for all the shaded areas in Fig. 4.

So, the corresponding CCF area we are interested in can be calculated as the difference between the segment $risk = var_{MVF} = 0.001278$ and the function for each corresponding parabola (14), by taking the definite integral in the return range:

$$CCF_i\ Area = \int_{R_{min}}^{R_{max}} \left(var_{MVF} - (A_i x^2 + B_i x + C_i)\right) dx = \left[var_{MVF} x - \left(\frac{1}{3} A_i x^3 + \frac{1}{2} B_i x^2 + C_i x\right)\right]_{R_{min}}^{R_{max}} \tag{18}$$

where A_i, B_i and C_i are the corresponding A, B and C for the i-th CCF, and:

$$R_{min} = b/c \tag{19}$$

$$R_{max} = \frac{1}{c}\left(b_i + \left(b_i^2 - c_i \sqrt{a_i - var_{MVF}(a_i c_i - b_i^2)}\right)\right) \tag{20}$$

Also the values of a_i, b_i and c_i are calculated as in Equation (12), but considering the restricted versions of **R** and **V** for each CCF. As we previously stated, the value of R_{min} is the same for all the CCF considered, but the value of R_{max} differs between the CCF. R_{min} is obtained by minimizing the risk in Equation (13) with respect to the return. R_{max} is also obtained from Equation (13) by simply isolating the return.

Equation (18) is also applied to compute the MVF area, just by updating the values of A, B and C. With this, the calculation of the similarity ratio is complete and we can rank the CCF by this measure.

The pseudo-code for the similarity ratio calculation function is given in Algorithm 1. The input is a binary string that indicates the assets considered for the CCF, while the output is the similarity ratio.

```
// R0 is re return of the global minimum variance portfolio in MVF: b/c
// MVF.area is the area of the MVF, which can be computed with the similarity function
// varMVF is the variance of the projection of the asset with maximum return over the MVF
similarity = function (string) // binary string with k ones and N-k zeros
begin
    assets = which(string == 1)          // find out which assets are in the CCF
    k = length(assets)            // cardinality of the CCF
    R.CCF = R[assets]                  // R.CCF is constrained to the considered assets
    V.CCF = V[assets, assets]           // V.CCF is also constrained
    V.CCF.inv = inv(V.CCF)// inverse of V.CCF
    One = vector(1, k)                  // a k-vector of 1's
    a_i = t(R.CCF) * V.CCF.inv * R.CCF        // t is the transpose function
    b_i = t(R.CCF) * V.CCF.inv * One
    c_i = t(One) * V.CCF.inv * One
    A_i = c_i / (a_i*c_i-b_i^2)                  // CCF parabola parameters
    B_i = -2*b_i / (a_i*c_i-b_i^2)
    C_i = a_i / (a_i*c_i-b_i^2)
    Rmax = (b_i + sqrt(b_i^2 - c_i*(a_i-varMVF*(a_i*c_i-b_i^2)))) / c_i
    Rmin = R0
    CCF.area =varMVF*Rmax - A*(Rmax^3)/3 - B*(Rmax^2)/2 - C*Rmax–
              (varMVF*Rmin - A*(Rmin^3)/3 - B*(Rmin^2)/2 - C*Rmin)
    similarity.ratio = CCF.area / MVF.area
    return(similarity.ratio)
end
```

Algorithm 1. Pseudo-code for the similarity ratio calculation function

To conclude, we must point out some of the similarities and differences between the similarity ratio we propose in this paper and the hypervolume (HV) metric introduced by Ziztler and Thiele (1998). Given a finite set of $\mathcal{P}$ points, the HV metric is defined as the d -dimensional volume of the hole-free orthogonal polytope $\Pi^d = \{\mathbf{x} \in \mathbb{R}^d_{\geq 0} : \mathbf{x} \preccurlyeq \mathbf{p}\ for\ some\ \mathbf{p} \in \mathcal{P}\}$ dominated by the $\mathcal{P}$ points. This measure is used when we have a known set of finite Pareto-optimal solutions, from which we estimate the hypervolume. When the set of Pareto-optimal solutions can be established from different algorithms, the metric HV is a suitable indicator to compare the set of solutions proposed by the algorithms. The similarity ratio can be considered as a special hypervolume case, in which the polytope is defined in two dimensions, and the set $\mathcal{P}$ is defined in the infinite number of points in the CCF. The construction of the CCF is not based on a subset of Pareto-optimal solutions, but takes into account the infinite set of solutions. In addition, as can be seen in Fig. 4.A and 4.C, not all the solutions in the frontier are Pareto optimal. The

HV metric is usually used in multiobjective models, when in some cases it is not feasible to determine all the optimal solutions, but only one of their subsets. However, in the mean-variance optimization problem it is possible to approximate all the optimal solutions, since the parabolic frontier function that relates variance with profitability is precisely known. All these considerations imply that obtaining the similarity ratio does not start from the same premises as the metric HV.

2.4 A genetic algorithm for the search for optimal similarity

In the previous section we introduced a measure that enables different CCFs to be compared. However, once the decision maker has decided the number of assets in the portfolio, k, the number of portfolios to be compared can be very large. For instance, in a problem with $N = 100$ assets in the universe and a cardinality constraint $k = 10$, $\sum_{i=1}^{10} \binom{100}{i}$ similarity ratios need to be performed and compared.

As the computation time for the similarity comparison of CCFs can be very long, many authors consider genetic algorithms (GA) a good compromise between solution accuracy and computation time.

The GA is a stochastic search algorithm which is able to explore large and troublesome spaces. It is based on the ideas from natural genetics and evolutionary principle (Holland, 1975; Chang et al, 2000; Ahn et al, 2012). During the course of evolution, populations evolve according to the principles of natural selection. The fitness of each individual is evaluated and only the fittest individuals reproduce, passing their genetic information to their offspring, so that the individuals who are most successful in adapting to their environment will have a better chance of surviving and reproducing. Exploration and exploitation are two important issues; exploration is the creation of population diversity by exploring the search space, and is obtained by genetic operators, such as mutation and crossover. Mutation is a genetic operator that randomly alters the values of genes in a parent chromosome. Crossover forms new offspring from two parent chromosomes by combining part of the genetic information from each. The evolution process finishes on the basis of some convergence criteria. A maximum number of generations is usually defined. Alternatively, a GA is stopped when a sufficiently large number of generations have passed without any improvement in the best fitness value.

We now explain the configuration of our algorithm, which is similar to the one used by Chang et al (2009) and Barak et al (2013).

Chromosome: The chromosome consists of N genes, where k genes have a value of 1 and $N - k$ have zero value. The chromosomes thus differentiate between assets in the cardinality constrained frontier and those excluded from it.

Fitness functions: We use the similarity ratio of Algorithm 1 as the fitness function to evaluate the chromosomes, so our objective is to find the CCF which maximizes the similarity ratio.

Crossover: Children in our GA are produced by a uniform crossover in which two parents produce a single child. The probability of crossover between pairs of chromosomes is 0.8. If the number of the child's assets is lower than k, then the remaining assets are randomly selected. If the number of the child's assets is greater than k, then assets are randomly removed from the child until their number reaches k.

Mutation: The probability of mutation in a parent chromosome. Mutation usually occurs with a small probability, so we have set it to 0.1.

Population: The population size is set to 100.

Termination criterion: The algorithm is terminated after 500 iterations.

2.5 Transaction and tax costs

As indicated in the Introduction, constraining the cardinality of portfolios entails a significant reduction in transaction costs. The consideration of the similarity measure allows the fund manager to respect the cardinality constraint, even if she or he dealing with clients with different risk profiles, and therefore with different levels of profitability.

However, whether or not the similarity measure is used, whether it is invested in a greater or lesser number of assets, significant transaction costs must be incurred each time a portfolio is rebalanced. It is worth wondering here whether the approach of the similarity measure is really saving on transaction costs with respect to the alternative approach in which the cardinality constraint can be violated.

For this it is necessary to determine the functional form of the transaction costs. Some studies have considered that these costs vary with the traded volume (Meade & Salkin, 1990), which is reasonable in practice, since the transaction costs remain proportional to the negotiated volume if this falls within certain intervals. However, the current practice is for brokers to apply a fixed component of the cost, regardless of the volume traded, plus a variable proportional of the volume. Under this premise, Chiam et al. (2013) characterize the most common transaction cost functions in the literature:

$$Transaction\ cost = f(v) = \begin{cases} \eta, & fixed\ costs\ only \\ \lambda v, & proportional\ costs\ only \\ max[\lambda v, \eta], & proportional\ costs\ with\ lower\ limit \\ \eta + \lambda v, & proportional\ plus\ fixed\ cost \end{cases}$$

where v represents the traded volume, $0 < \lambda < 1$ and $\eta > 0$. Chiam et al. (2013) point out that in practice this function does not correspond exactly with actual market practices, where transaction costs instead often comprise a multi-tiered cost structure in which different cost functions are applied within certain ranges of the traded volume. These authors cite the example of a number of brokers in Singapore, who set their transaction costs as follows: firstly, a fixed minimum cost is required if the traded value falls below a critical value. Secondly, the variable cost decreases for higher traded values to encourage their clients to higher trading activity. The function of transaction costs that they assume in their work is the following:

$$Transaction\ cost = f(v) = \begin{cases} 25, & 0 \leq v \leq 25 \\ 0.0028v, & 25 < v \leq 50{,}000 \\ 0.0022v, & 50{,}000 < v \leq 100{,}000 \\ 0.0018v, & 100{,}000 < v \end{cases}$$

Lajili-Jarjir & Rakotondratsimba (2008) also employ a piecewise affine transaction cost function for the case of French brokers. In addition to transaction costs, they also consider the taxes incurred in buying and selling financial assets in France, which also follow a piecewise affine function. These costs arise when the volume of the transaction exceeds €7,668:

$$Tax\ cost = f(v) = \begin{cases} 0, & 0 \le v \le 7{,}668 \\ 0.003v - 23, & 7{,}668 < v \le 153{,}000 \\ 0.0015v - 23, & 153{,}000 < v \le 422{,}000 \\ 610, & 422{,}000 < v \end{cases}$$

All these examples demonstrate the behaviour of costs associated with the rebalancing of portfolios. It follows that the transaction cost per unit (transaction cost for each euro traded) decreases with the traded volume. The smaller the number of securities handled, the greater the average investment in each security, and therefore the lower the transaction cost per unit and the lower the total transaction costs. We will see in a later example how the implementation of the similarity ratio, in some cases by up to 75%, allows the number of shares effectively operated by the fund manager to be reduced. By having a number of significantly reduced securities, the average investment in assets will be higher and will benefit from a cost savings of a rebalancing of the portfolios.

3. Computational results

This section describes a number of case studies in which the proposed methodology was applied to some stock market indices frequently used in the literature. Our data source was the OR-Library (Beasley, 1990), from which we have selected five leading indices: the Hang Seng (Hong Kong), DAX 100 (Germany), FTSE 100 (UK), S&P 100 (USA) and the Nikkei 225 (Japan). These data sets contain 290 weekly price data per stock market.

The computational results given below were coded in R language with the GA package for the genetic algorithm. The size of the five problems ranged from $N = 31$ (Hang Seng) to $N = 225$ (Nikkei), which serves as a maximum value for cardinality k. The minimum value considered for k is 2. For example, in the case of the Hang Seng data we performed 28 different CCFs ($k = 2..31$). The return range goes from the return of the global minimum variance portfolio to the return of the maximum return asset. In the case of Nikkei 225 the maximum return was too small – 0.004813 –, even smaller than the highest return on the other indices, although the Nikkei database has more assets. Such a narrow return range can produce no comparable results with the other four indices, so we decided to include an upper bound for the return range of 0.015, which is in line with those considered in the other indices.

Fig. 5-14 depict different problem instances for the five stock indices. This helps us to visually perceive how the CCF identified through the maximisation of the similarity ratio are akin to the MVF. We have also constructed 500 random CCF to perceptually compare them with the one chose above (GA CCF). These random CCF have the same cardinality as the GA CCF.

These results confirm that the GA CCF that maximizes the similarity ratio is closer to the MVF than the random CCF in general terms, and when the cardinality rises the GA CCF dominates all the other 500 random CCF in the mean-variance space, as in the case of the DAX 100, FTSE 100, S&P 100 and Nikkei 225. The GA CCF for Hang Seng fully dominates the random CCF when cardinality is higher than 15 (not reported). We can therefore conclude that the similarity ratio is an appropriate approach for finding a cardinality constrained frontier that fulfils the requirements of the decision maker.

The computation times required to obtain the GA CCF were between 12.9 seconds for the Hang Seng 32 index case with $k = 5$, and 66.4 seconds for the Nikkei 225 index with

$k = 60$. It can therefore be concluded that these are reasonable times for implementing this methodology.

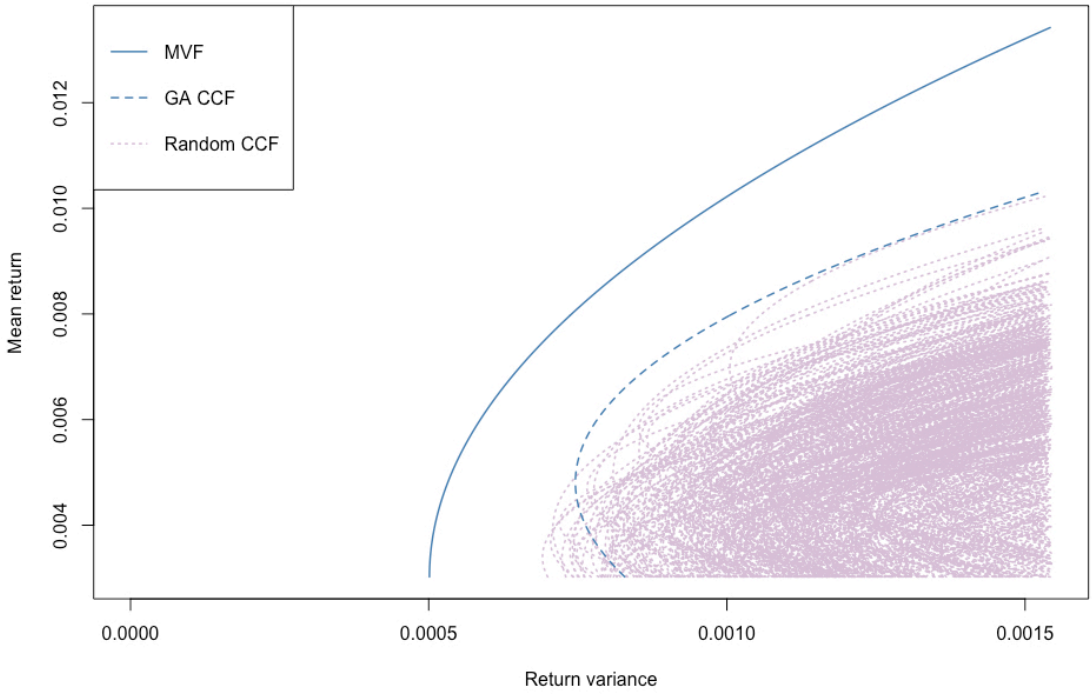


Figure 5. GA CCF and 500 random CCFs for Hang Seng, with k=5

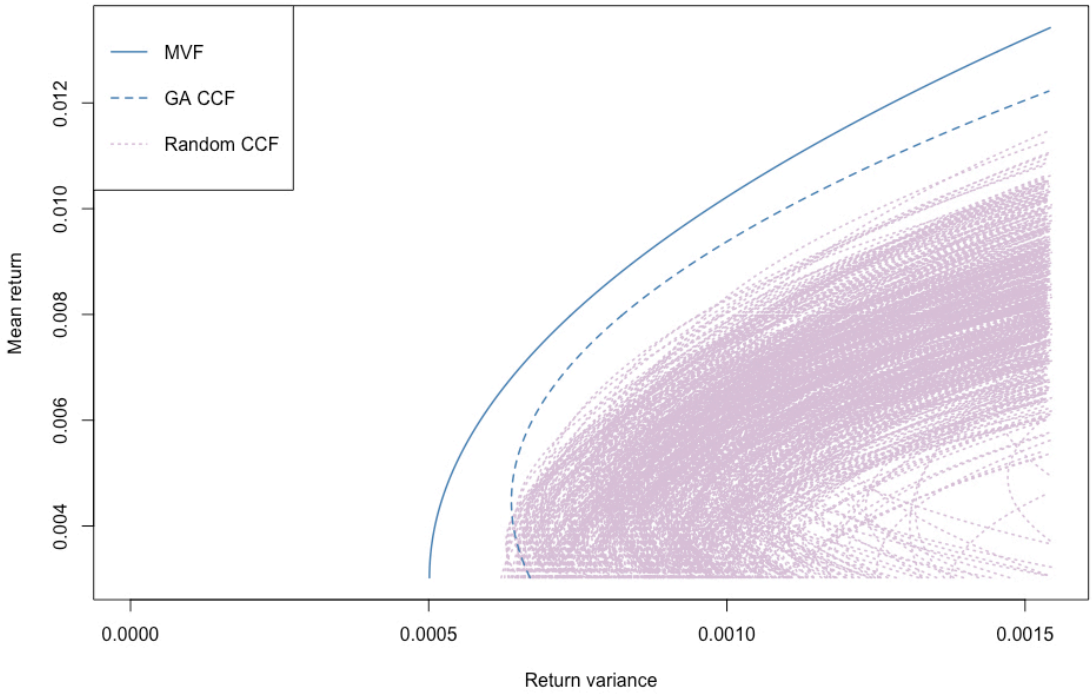


Figure 6. GA CCF and 500 random CCFs for Hang Seng, with k=10

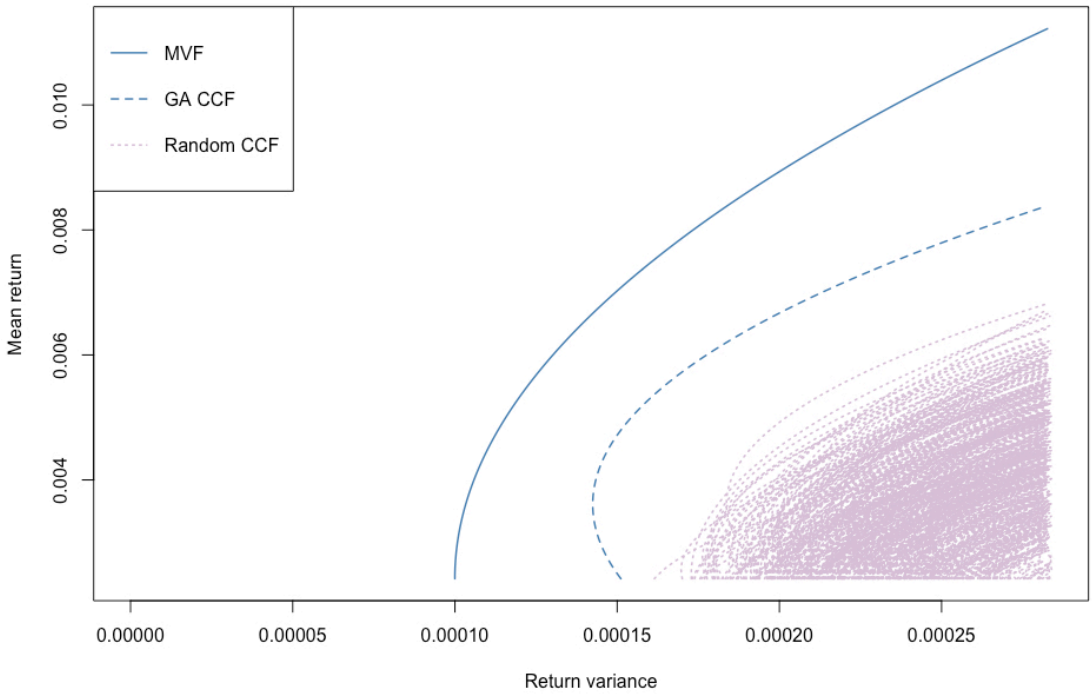


Figure 7. GA CCF and 500 random CCFs for DAX 100, with k=20

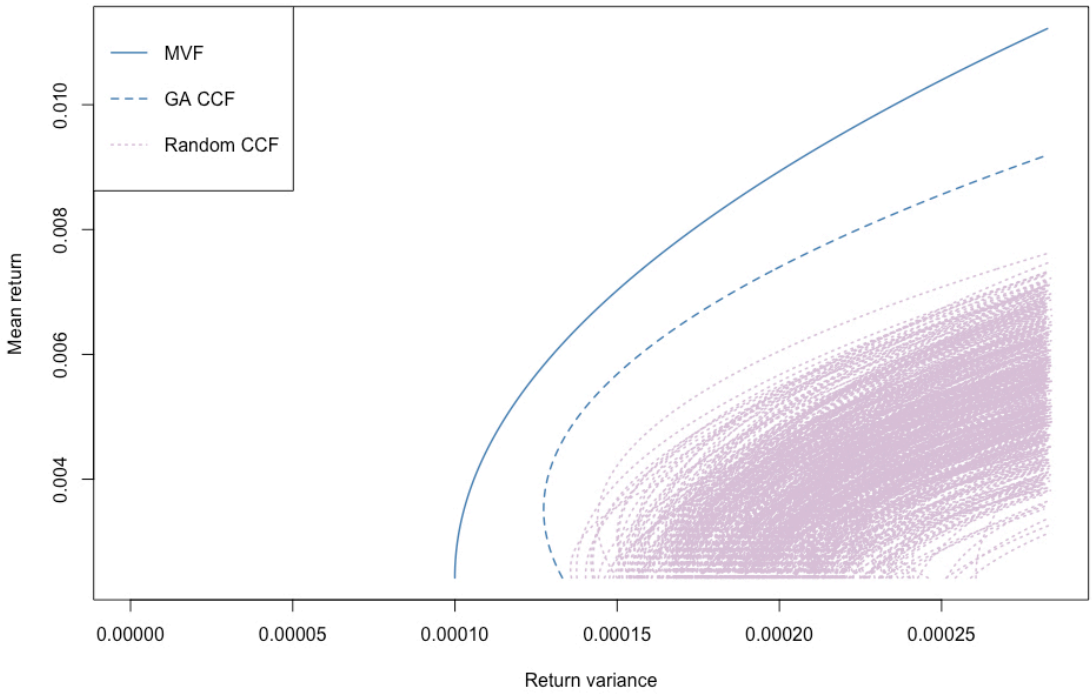


Figure 8. GA CCF and 500 random CCFs for DAX 100, with k=30

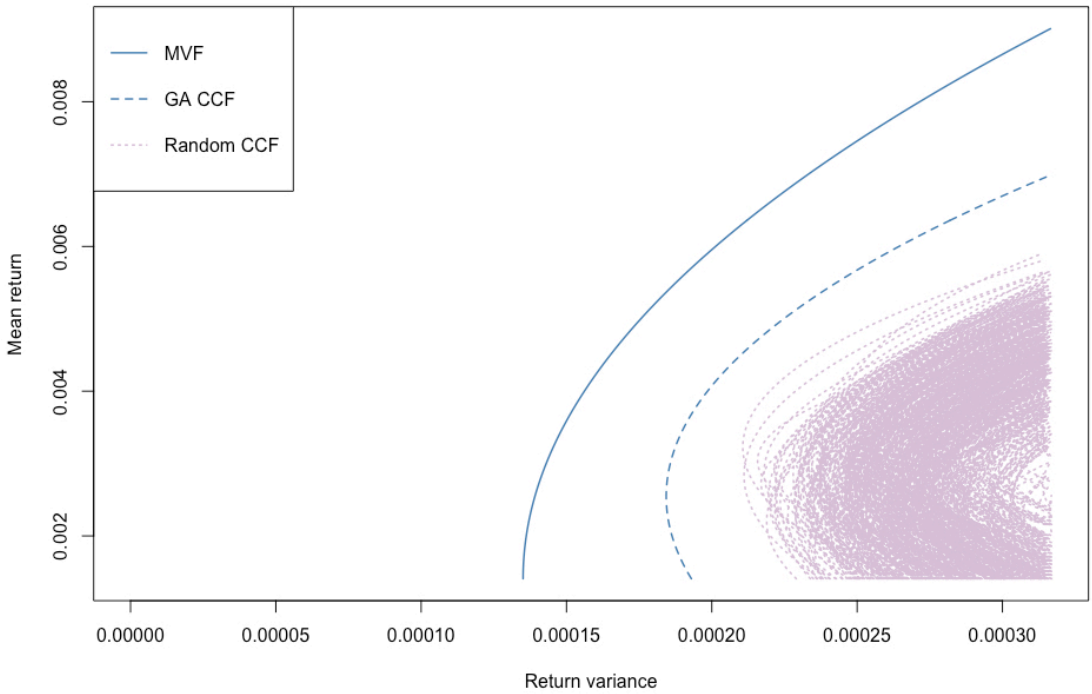


Figure 9. GA CCF and 500 random CCFs for FTSE 100, with k=20

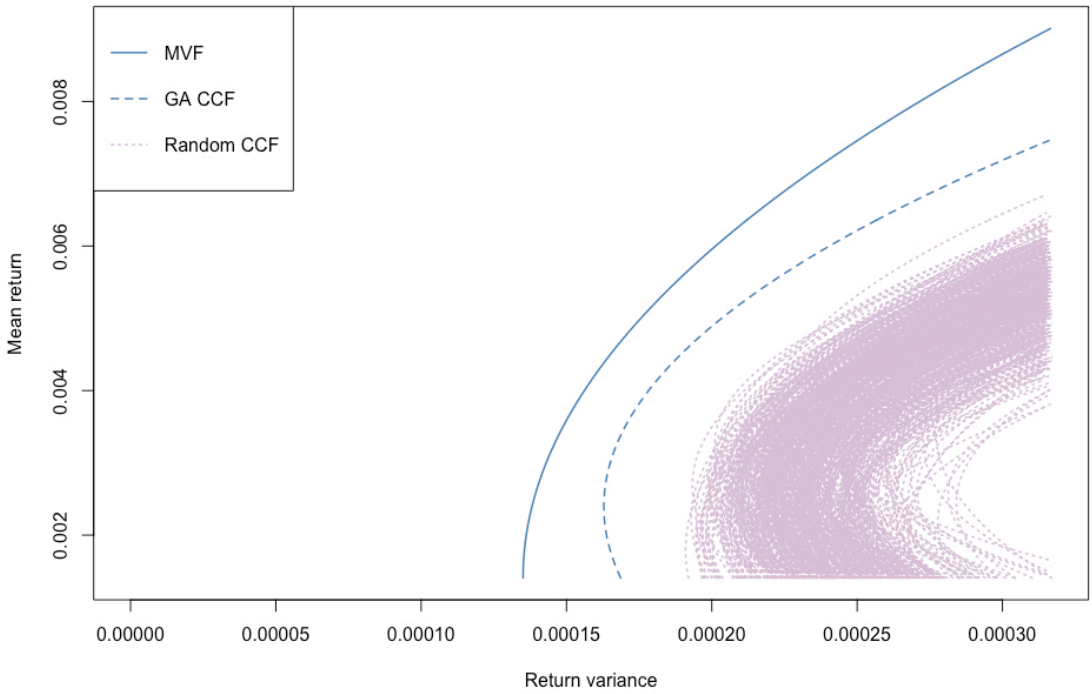


Figure 10. GA CCF and 500 random CCFs for FTSE 100, with k=30

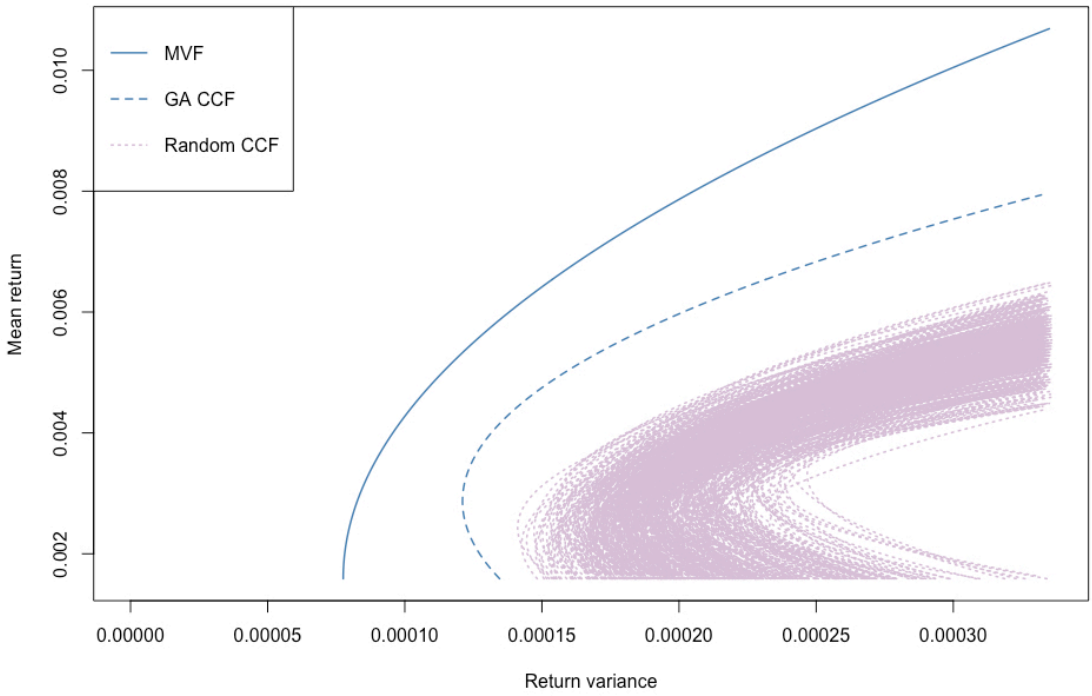


Figure 11. GA CCF and 500 random CCFs for S&P 100, with k=20

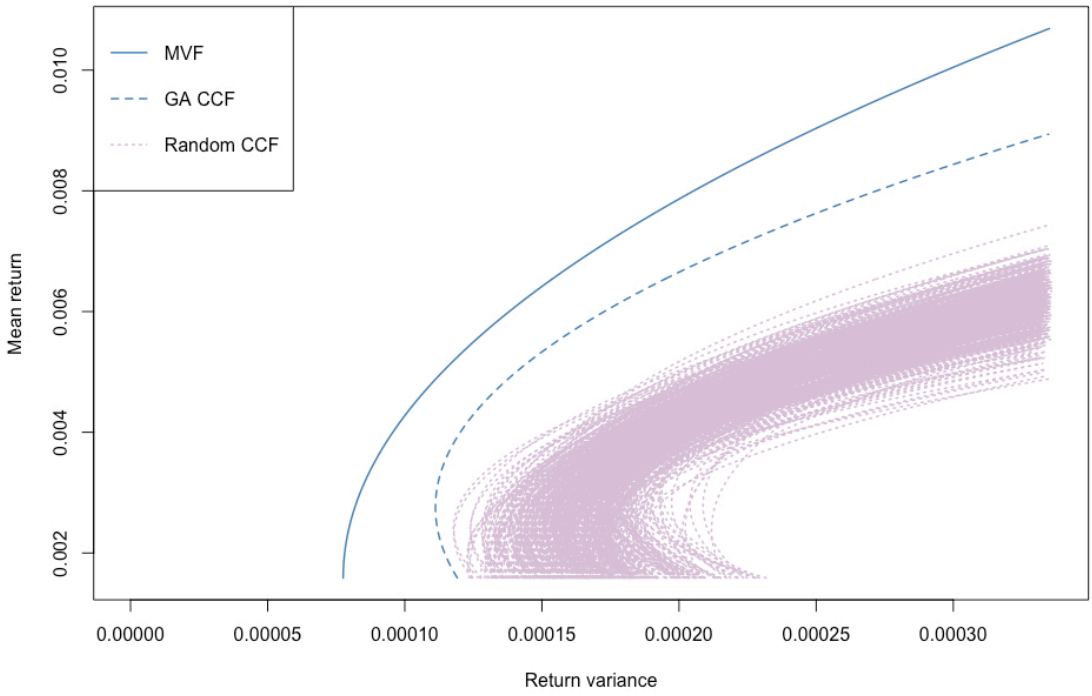


Figure 12. GA CCF and 500 random CCFs for S&P 100, with k=30

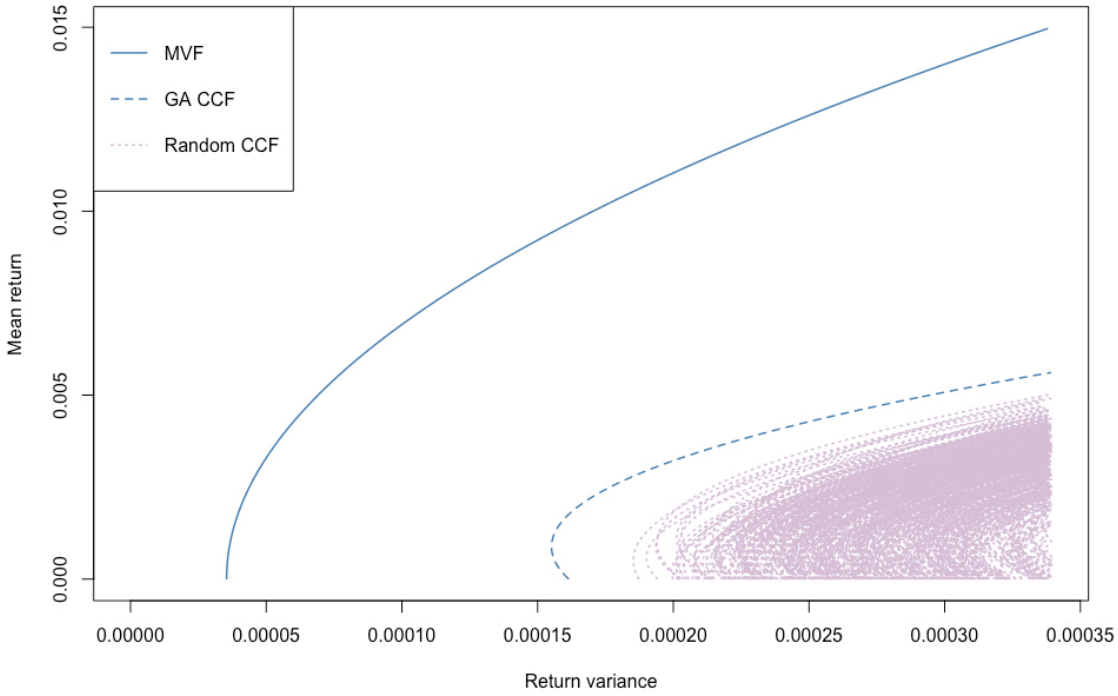


Figure 13. GA CCF and 500 random CCFs for Nikkei 225, with k=50

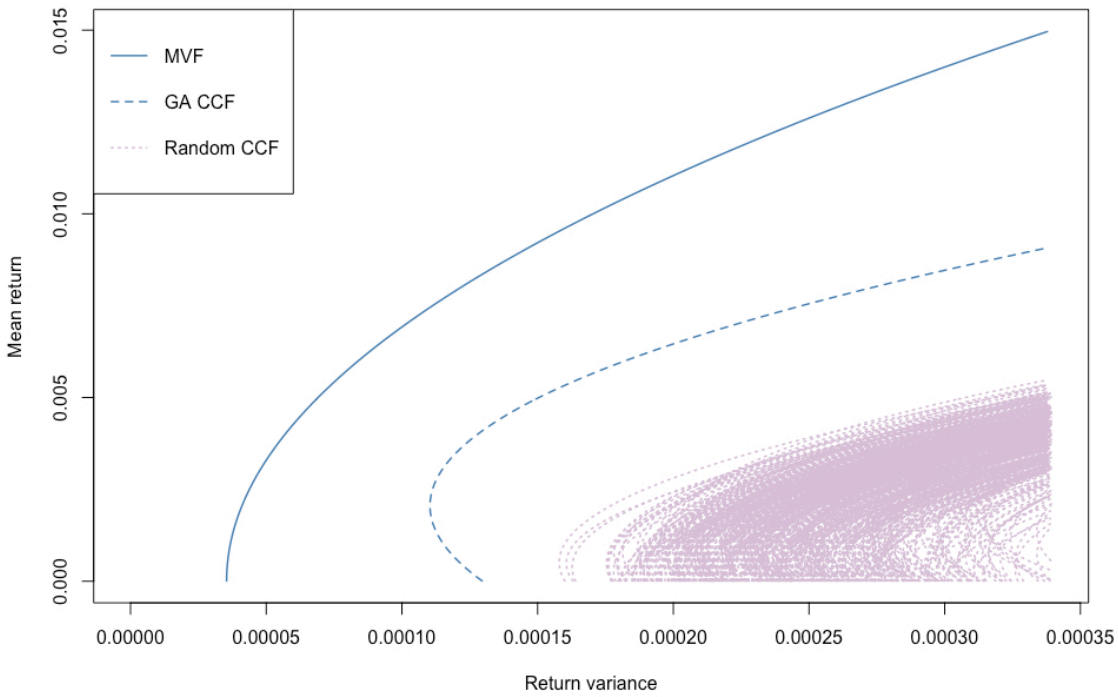


Figure 14. GA CCF and 500 random CCFs for Nikkei 225, with k=60

The use of the similarity ratio gives a considerable saving in the number of shares considered in the mean-variance portfolio optimization problem with a cardinality constraint. Fig. 15 compares the number of shares considered at the frontier obtained by the similarity ratio versus the traditional CCF approach, with no similarity ratio for the 5 stock indexes analyzed. For this, the Gain Ratio has been defined as the number of shares in the CCF with no similarity ratio divided by the number of shares in the CCF with a similarity ratio. To determine how this ratio evolves, different levels have been considered in the cardinality of the portfolios, ranging from 10% to 40% of the securities.

In general terms, we can see that the greatest saving in the number of assets is in the smaller cardinalities (10% of total assets). The most significant case is the Nikkei 225, in which the number of assets used to define the CCF with no similarity ratio is more than 4 times the number of assets needed to construct the CCF by means of the similarity ratio. In fact, the frontier obtained with the similarity ratio with 10% of the assets involves employing a total of 23 assets, while if the similarity ratio is not considered this frontier uses 94 assets. This means that the similarity ratio is able to save more than 75% of assets in constructing the frontier.

As mentioned in the previous section, this significant reduction in the number of securities entails a significant saving in transaction costs and in taxes, especially in the normal practice of brokers employing piecewise affine cost functions. Fig. 15 also shows how the savings in the number of assets is greater in the case of the high cardinality indexes, which are also the most attractive in cardinality constraint optimization.

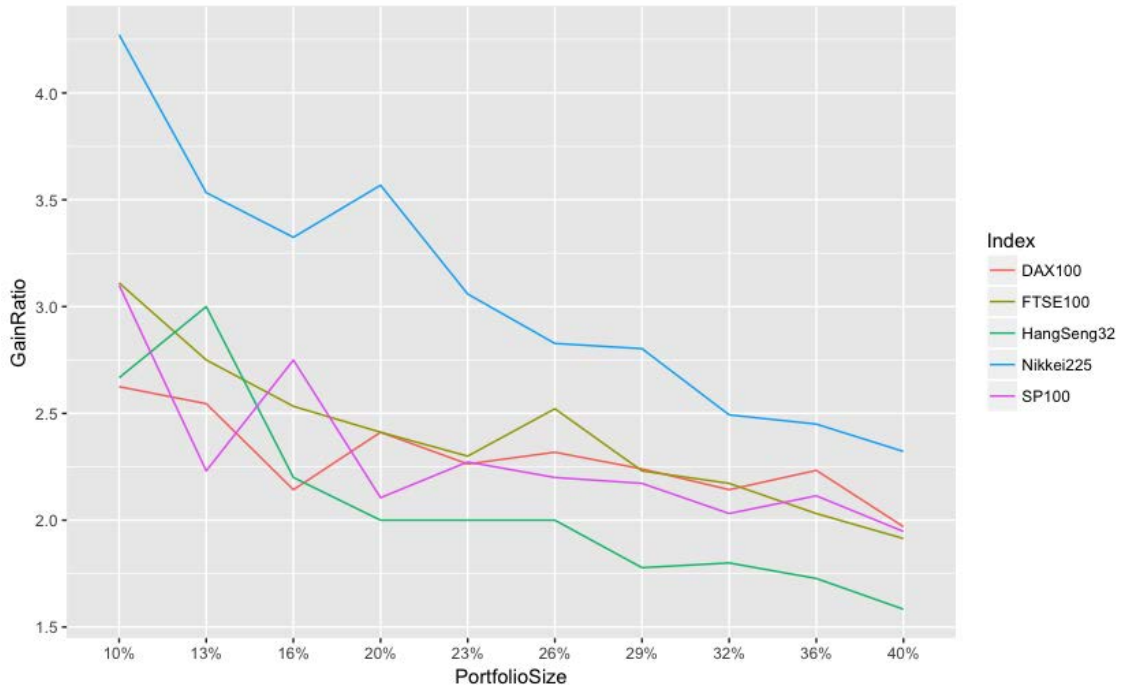


Figure 15. Savings in the number of assets when considering the similarity ratio

It is also interesting to observe how the similarity ratio evolves as the cardinality changes. This could help the decision maker to select the proper cardinality for the portfolios, in order to achieve a reasonable level of diversification with respect to the maximum level of diversification when all assets are selected – MVF –.

Fig. 16 shows the example of the S&P 100 Index and how the similarity ratio grows with the cardinality of the problem. It can be seen that as the relationship is non-linear, the diversification is accelerated for low cardinality values and grows moderately for higher cardinality values. It can therefore be concluded that the fastest diversification is obtained with the lowest cardinality values. Once a certain level of diversification has been achieved, the improvement in similarity slows down.

We also depicted the mean similarity ratio for the 500 random CCF, which mostly confirm a linear relationship between diversification and the number of assets considered in the problem. It can be seen that low cardinality problems have negative similarity ratios in the mean of the random CCFs. This is because most of the CCFs have a larger risk than the maximum risk considered in the problem, and so are on the right of the vertical segment of the maximum variance var_{MVF} (negative areas).

To summarize, Fig. 16 can help the decision maker to properly adjust cardinality by comparing the diversification of each CCF in relation to the MVF. In the case of the S&P 100, we conclude that slightly more than 50% of the diversification, as measured by the similarity ratio, is achieved with only 15 assets.

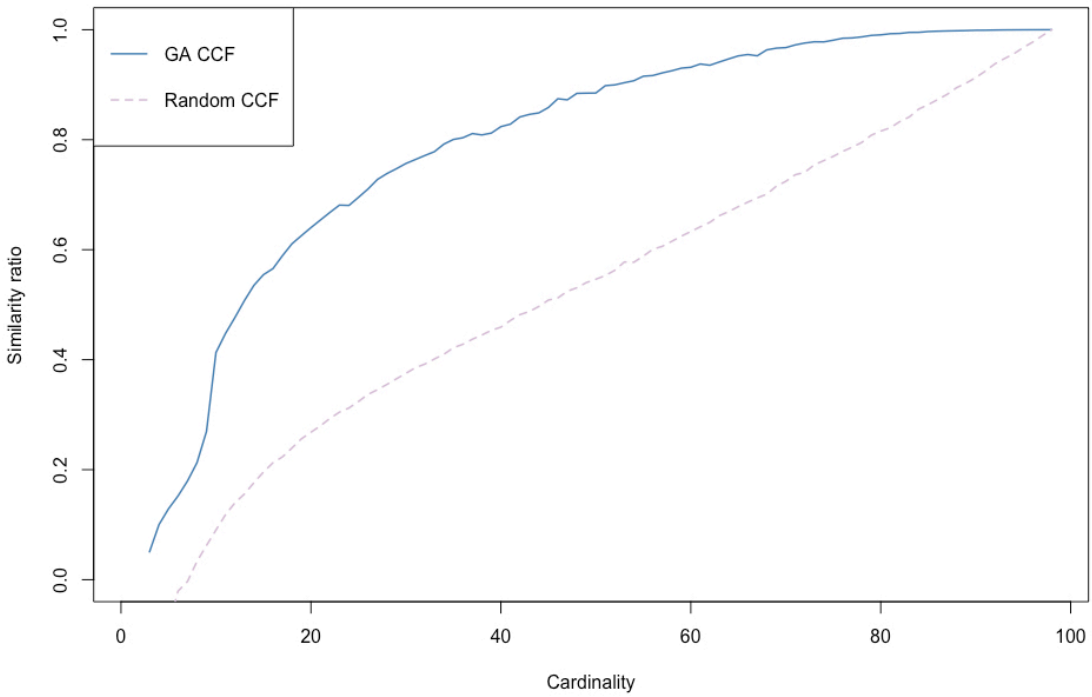


Figure 16. Similarity ratio vs. cardinality for the S&P 100 stock exchange market

The maximum difference between both curves in Fig. 16 is for cardinality 28, where the similarity ratio of the GA CCF solution is 0.738 and the mean similarity ratio for the 500 random CCF solutions is 0.355. If the decision maker is interested in maximizing this difference, then the optimal cardinality would be 28.

Finally, we also studied the persistence of assets in the CCF problem. By *persistence* we mean the fact that once a certain stock is in the CCF, it will remain in the portfolio for

higher cardinality instances of the problem, so that the composition of the CCFs is formed incrementally.

Asset number K	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
2														x			x														
3									x						x														x		
4									x					x	x														x		
5									x						x					x			x						x		
6									x			x		x	x								x						x		
7						x			x			x		x	x								x						x		
8			x			x			x			x		x	x								x						x		
9			x			x			x	x		x		x	x								x						x		
10			x			x			x	x		x		x	x					x			x						x		
11			x			x			x	x		x		x	x					x			x			x			x		
12			x			x			x	x		x		x	x				x	x			x			x			x		
13			x			x			x	x		x		x	x				x	x			x		x		x		x		
14			x			x			x	x		x		x	x				x	x			x		x		x	x	x		
15			x			x			x	x		x		x	x				x	x			x		x		x	x	x	x	
16			x			x			x	x		x		x	x		x		x	x			x		x		x	x	x	x	
17			x			x			x	x		x	x	x	x		x		x	x			x		x		x	x	x	x	
18			x			x			x	x		x	x	x	x		x		x	x			x	x	x		x	x	x	x	
19			x			x	x		x	x		x	x	x	x		x		x	x			x	x	x		x	x	x	x	
20			x			x	x	x	x	x		x	x	x	x		x		x	x			x	x	x		x	x	x	x	
21			x	x		x	x	x	x	x		x	x	x	x		x		x	x			x	x	x		x	x	x	x	
22		x	x	x		x	x	x	x	x		x	x	x	x		x		x	x			x	x	x		x	x	x	x	
23		x	x	x		x	x	x	x	x	x	x	x	x	x		x		x	x			x	x	x		x	x	x	x	
24		x	x	x		x	x	x	x	x	x	x	x	x	x		x		x	x			x	x	x	x	x	x	x	x	
25		x	x	x		x	x	x	x	x	x	x	x	x	x	x	x		x	x			x	x	x	x	x	x	x	x	
26		x	x	x		x	x	x	x	x	x	x	x	x	x	x	x		x	x		x	x	x	x	x	x	x	x	x	
27		x	x	x		x	x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x	x	x	x	x	x	x	x	
28		x	x	x		x	x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x	x	x	x	x	x	x	x	x
29		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x	x	x	x	x	x	x	x	x
30	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x	x	x	x	x	x	x	x	x
31	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x

Table 2. Persistence of the CCF composition for the Hang Seng stock exchange market

There is presumably a direct relationship between the persistence of assets and the quality of the CCF selection, so a CCF with cardinality k is probably constructed by adding some stocks to the CCF with cardinality $k-1$. Table 2 confirms this hypothesis. There we computed all GA CCF for the Hang Seng, with k from 2 to 31. We confirmed that 27 of the 30 assets remain in the portfolio once they are included in the CCF by the genetic algorithm.

4. Conclusions

This paper deals with the problem of finding the optimal cardinality constrained frontier in the mean-variance space. Previous studies concentrated on portfolio optimisation, i.e. the search for a cardinality constrained portfolio for a specific level of return. However, this can mean that the decision maker actually invests in more assets than those initially considered by the constraint, if he or she has to satisfy clients with different risk profiles. Instead of focusing on a cardinality constrained portfolio analysis, we propose an approach that perceptually focuses on the cardinality constrained frontier analysis. Through a novel similarity ratio definition, we search for the cardinality constrained frontier most similar to the mean-variance unconstrained frontier. The similarity ratio translates into an objective measure the perceptual concept of choosing the constrained frontier closest to the mean-variance frontier and use a genetic algorithm to address the combinatorial complexity introduced by the cardinality constraints.

The results obtained from five real-world stock exchange indices provide a practical example of how the similarity ratio works. When compared with a large number of random frontiers, it can be seen that the similarity ratio successfully diversifies the investment for low cardinality values. The significant reduction in the number of securities considered by the fund manager implies a reduction in transaction costs and taxes since the cost functions employed by brokers are usually piecewise affine functions. We also observed the persistence of assets included in the problem with cardinality k, which remain in the portfolio for problems with higher cardinality values.

Because the similarity ratio is defined in the mean-variance space, one could claim that variance and return are not being given equal consideration in the decision process, due to the different units calculated. In fact, the variances are smaller in absolute values than the returns in the examples given. It can also be argued that investors usually prefer standard deviation because the units are the same as with the expected return. We can overcome these two problems by considering the standard deviation instead of variance when calculating the similarity ratio, so that the proposed algorithm can be directly adapted to it.

A future related line of research will study the use of the similarity ratio for the index tracking problem. By limiting the number of stocks in the tracking portfolio, we will be able to compute the optimal tracking frontier that gives the minimum tracking error variance. We could also consider adding a third criterion (for example liquidity) to the classical return and risk, in which case the computation of the areas would be extended to volumes.

References

Ahn J J, Byun H W, Oh K J and Kim T Y (2012). Using ridge regression with genetic algorithm to enhance real estate appraisal forecasting. Expert Systems with Applications 39(9): 8369–8379.

Anagnostopoulos K P and Mamanis G (2011). The mean–variance cardinality constrained portfolio optimization problem: An experimental evaluation of five multiobjective evolutionary algorithms. Expert Systems with Applications 38(11): 14208–14217

Aouni B, Colapinto C and La Torre D (2013). A cardinality constrained stochastic goal programming model with satisfaction functions for venture capital investment decision making. Annals of Operations Research 205(1): 77–88.

Barak S, Abessi M and Modarres M (2013). Fuzzy turnover rate chance constraints portfolio model. European Journal of Operational Research 228(1): 141–147.

Beasley J E (1990). OR-Library: Distributing test problems by electronic mail. Journal of the Operational Research Society 41(11): 1069–1072.

Bertsimas D and Shioda R (2009). Algorithm for cardinality-constrained quadratic optimization. Computational Optimization and Applications 43(1): 1–22.

Bienstock D (1995). Computational study of a family of mixed integer quadratic programming problems. Mathematical Programming 74(2): 121–140.

Canakgoz N A and Beasley J E (2009). Mixed-integer programming approaches for index tracking and enhanced indexation. European Journal of Operational Research 196(1): 384–399.

Cesarone F, Scozzari A and Tardella F (2013). A new method for mean-variance portfolio optimization with cardinality constraints. Annals of Operations Research 205(1): 213–234.

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

Chang T J, Yang S C and Chang K J (2009). Portfolio optimization problems in different risk measures using genetic algorithm. Expert Systems with Applications 36(7): 10529–10537.

Chiam S C, Tan K C and Al Mamun A (2013). Dynamic index tracking via multi-objective evolutionary algorithm. Applied Soft Computing 13(7): 3392-3408.

Di Tollo G and Rolli A (2008). Metaheuristics for the portfolio selection problem. International Journal of Operations Research 5(1): 13–35.

García F, Guijarro F and Moya I (2011). The curvature of the tracking frontier: a new criterion for the partial index tracking problem. Mathematical and Computer Modelling 54(7): 1781–1784.

Holland J H (1975). Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control and artificial intelligence. The MIT Press: Cambridge.

Krink T, Mittnik S and Paterlini S (2009). Differential evolution and combinatorial search for constrained index-tracking. Annals of Operations Research 172(1): 153–176.

Lajili-Jarjir S and Rakotondratsimba Y (2006). The number of securities giving the maximum return in the presence of transaction costs. Quality and Quantity 42(5): 613–644.

Li D, Sun X and Wang J (2006). Optimal lot solution to cardinality constrained mean–variance formulation for portfolio selection. Mathematical Finance 16(1): 83–101.

Liagkouras K and Metaxiotis K (2014). A new Probe Guided Mutation operator and its application for solving the cardinality constrained portfolio optimization problem. Expert Systems with Applications 41(14): 6274–6290.

Lwin K and Qu R (2013). A hybrid algorithm for constrained portfolio selection problems. Applied Intelligence 39(2): 251–266.

Maringer D and Kellerer H (2003). Optimization of cardinality constrained portfolios with a hybrid local search algorithm. OR Spectrum 25(4): 481–495.

Markowitz H (1952). Portfolio selection. Journal of Finance 7(1): 77–91.

Meade N and Salkin G (1990). Developing and maintaining an equity index fund. The Journal of the Operational Research Society 41(7): 599–607.

Merton R C (1972). An Analytic Derivation of the Efficient Portfolio Frontier. The Journal of Financial and Quantitative Analysis 7(4): 1851–1872.

Metaxiotis K and Liagkouras K (2012). Multiobjective Evolutionary Algorithms for Portfolio Management: A comprehensive literature review. Expert Systems with Applications 39(14): 11685–11698.

Ruiz-Torrubiano R and Suarez A (2009). A hybrid optimization approach to index tracking. Annals of Operations Research 166(1): 57–71.

Sant'Anna L R, Filomena T P, Guedes P C and Borenstein D (2016). Index tracking with contolled number of assets using a hybrid heuristic combining genetic algorithm and non-linear programming. Annals of Operations Research advance online publication 3 February, doi: 10.1007/s10479-016-2111-x.

Shaw D, Liu S and Kopman L (2008). Lagrangian relaxation procedure for cardinality-constrained portfolio optimization. Optimisation Methods & Software 23(3): 411–420.

Woodside-Oriakhi M, Lucas C and Beasley J (2011). Heuristic algorithms for the cardinality constrained efficient frontier. European Journal of Operational Research 213(3): 538–550.

Zitzler E and Thiele L (1998). "Multiobjective optimization using evolutionary algorithms - A comparative case study," in Proceedings of PPSN-V, Fifth International Conference on Parallel Problem Solving from Nature. Springer Verlag, Berlin, pp. 292–301.