Equally weighted cardinality constrained portfolio selection via factor models

Juan F. Monge

Received: 13 March 2019 / Accepted: 12 March 2020 / Published online: 17 March 2020
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract
In this work a proposal and discussion of two different 0-1 optimization models is carried out in order to solve the cardinality constrained portfolio problem by using factor models. Factor models are used to build portfolios based on tracking the market index, among other objectives, and require to estimate smaller number of parameters than the classical Markowitz model. The addition of the cardinality constraints limits the number of securities in the portfolio. Restricting the number of securities in the portfolio allows to obtain a concentrated portfolio while also limiting transaction costs. To solve this problem a new quadratic combinatorial problem is presented to obtain an equally weighted cardinality constrained portfolio. For a single factor model, some theoretical results are presented. Computational results from the 0-1 models are compared with those using a state-of-the-art Quadratic MIP solver.

Keywords Portfolio selection · Factor models · Minimum-variance portfolio · 0-1 quadratic optimization

1 Introduction

The portfolio selection problem deals with selecting a collection of financial assets and in what proportion, according to the investor’s risk preference, with the aim of obtaining the maximum expected return. The selection of assets allocated to the portfolio can be managed using different approaches: minimum risk allocation, equal weighting, risk parity, Sharpe ratio, and many others.

In the seminal work of Markowitz [26], return and risk are evaluated by means of the expected value and variance of the selected assets. Markowitz introduced the concept of an efficient frontier and showed that there is a set of optimal portfolios,
not only one. The classical Markowitz model can be formulated as a quadratic model, and the investor can find an optimal portfolio maximizing the expected return under a risk level, \( w^* = \arg w \max \{ w'\mu \text{ s.t. } w \Sigma w = \sigma^*, w'1 = 1 \} \), or minimizing the risk under a return level, \( w^* = \arg w \min \{ w'\Sigma w \text{ s.t. } w'\mu = r^*, w'1 = 1 \} \), where \( w \) denotes the vector of weights in the portfolio, \( \mu \) the vector of expected returns, and \( \Sigma \) the covariance matrix of expected returns. A significantly important portfolio is given when the constraint related to the return level is relaxed, obtaining the global minimum risk solution. This solution is important in the literature. For example, it is shown in [16] that the minimum variance portfolio is a more reliable and robust out-sample than the traditional mean variance portfolios. The minimum-variance portfolio usually performs better out sample than any other mean-variance one. In [16] the authors present non-linear models to obtain a robust portfolio estimation, the minimum variance being a particular case.

Another important portfolio is given when a tradeoff objective function return/risk is considered, \( w^* = \arg w \max \{ w'\Sigma w - \lambda w'\Sigma w \text{ s.t. } w'1 = 1 \} \), where \( \lambda \) is the risk aversion coefficient. Although this work considers the minimum variance portfolio, it is shown that the results can easily be applied to the objective functions mentioned above.

The factor model theory establishes the expected return of each asset as a linear function on the risk factors, through the parameter \( \beta \), where \( \beta \) is a measure of the risk contribution for the individual asset to the portfolio. See the seminal factor models of Sharpe [32,33] and the related Capital Asset Pricing Model (CAPM) theory.

The Markowitz mean-variance framework requires to estimate a large number of parameters. If there are \( n \) assets, it is necessary to estimate \( n \) means, \( n \) variances and \( n(n - 1)/2 \) covariances, \( O(n^2) \). The factor model requires fewer parameters to be estimated; the order is given by the number of factors \( m \), i.e. \( O(m^2) \), where \( m \) is much smaller than \( n \). Recently, some factor models have received attention in the literature. The impact of estimation error on latent factor models is measured in [4]; building indicators are also presented to measure the impact of the estimation error and the exposure of the portfolios to the factors. For other factor models, see [14,17,23,27,30].

The cardinality constrained portfolio problem is a classic problem in the literature. Some properties are presented in [13] for the efficient frontier of the cardinality constrained problem in the classical mean-variance Markowitz model, showing for example, the discontinuity of the efficient frontier. Notice also that the traditional minimization of trade-off objective function mean/risk does not provide all the efficient solutions. Different heuristics are also presented, see [36] for some other metaheuristic approaches. The exact resolution of the problem is analyzed in [12], where an exact algorithm is presented for medium sized problems, such that it provides a good approximation for larger ones.

A Lagrangian decomposition scheme is presented in [34] for the cardinality constrained portfolio problem. The covariance matrix is decomposed in two matrices, namely a diagonal matrix with the risk of each asset and another non-diagonal with the covariance among the factors. It allows us to reduce the dimensions of the quadratic problem to be solved. See in [18] another Lagrangian decomposition scheme for this problem and in [3] an alternative procedure based on solving a succession of problems into a tree search.
An approach in [24], regarding the cardinality constrained problem, refers to the investment being made in lots, such that the excess capital goes to a risk-free asset. An algebraic algorithm for solving the portfolio through formulating it as an integer model with the expected return as a linear objective function subject to linear and non-linear constraints is presented in [10]. Alternatively, mixed-integer linear programming models to selecting portfolios considering transactions costs and lots are presented in [22,25].

The simplest model in portfolio optimization requires the equally weighted constraint, i.e., the weights of the assets are identical. In [15] it is shown that none of the optimized portfolios outperforms the naive diversification when they are evaluated out-sample. A comparison of the shortfall minimization method and the naive portfolio is carried out in [21] to conclude that, while the shortfall methods perform very well in-sample, the out-sample performance is no better, and in some cases worse, than the naive model. Additionally, in [6] it is shown that the naive diversification strategy becomes optimal when the uncertainty of ambiguous assets increases; on the contrary, [37] warns that some caution be taken about [15] and the question of the add-value of portfolio optimization.

All the above works only deal with the classical Markowitz model; these papers do not integrate the cardinality constraint in factor models. To the best of our knowledge no work exists in the literature combining factor models, cardinality constraint and equally weighted constraint.

The main contribution of this work is to present a new model combining the three features, namely; factor model, cardinality constraint and equally weighted constraint. This new model can obtain high quality solutions in very little computational time; moreover, it can be helpful to practitioners as a first step in order to evaluate what best assets to consider, and using this information in a later analysis where more elaborate techniques are used. Also the work presents some theoretical results on this new 0-1 pure binary quadratic optimization model.

Standard branch and bound solver for an integer problem can be improved by the use of specialized techniques, see in [5] a computational study to solve mixed-integer quadratic problems with branch and bound techniques. However, to solve a model without a specialized algorithm or a specialized solver for portfolio optimization models would be very useful. The singularity present in the covariance matrix of the factor models allows us to take advantage over general optimization solvers such as CPLEX.

The rest of this paper is organized as follows. Section 2 deals with the main concepts of factor models and introduces the notation for the cardinality constrained minimum variance problem where the equally weighted constraint is imposed. It also presents theoretical results for this new combinatorial problem when a single factor is considered. Section 3 reports the computational results for a set of instances taken from the literature. Finally, Sect. 4 concludes and outlines future plans.

2 Factor model

For a risky asset \( i \in I \), a factor model assumes that the return rates \( r_i \) of asset \( i \) is given by \( r_i = \alpha_i + \beta_i F + \epsilon_i \), where \( F = (f_1, \ldots, f_m) \) is a column vector of random
variables called factors, with \( E(f_i) = \overline{f}_i, \alpha_i \in \mathcal{R} \) is a constant, \( \beta_i \in \mathcal{R}^m \) is a constant row vector and \( \epsilon_i \) is a (error) mean zero random variable, uncorrelated with the factors, \( E(\epsilon_i) = 0 \) and \( E(\epsilon_i \cdot f_i) = 0 \). The factors \( F \) are correlated with the covariance matrix \( \Sigma_F \). Let the notation \( \sigma_{lm} = \text{Cov}(f_l, f_m) \) and \( \sigma_{\epsilon_i}^2 = E(\epsilon_i^2) \).

A portfolio comprised of \( n \) assets, defined by weights \( w' = (w_1, w_2, \ldots, w_n) \), can be determined by a factor model, where the return and variance are
\[
\begin{align*}
\mathbf{r} &= \sum_{i \in I} w_i \alpha_i + \sum_{i \in I} w_i \beta_i F + \sum_{i \in I} w_i \epsilon_i, \\
\mathbf{V}(\mathbf{r}) &= \sum_{i,j \in I} \sum_{l,m \in F} w_i \beta_{il} \beta_{jm} \sigma_{lm} + \sum_{i \in I} w_i^2 \sigma_{\epsilon_i}^2,
\end{align*}
\]
respectively.

2.1 Cardinality constrained minimum-variance portfolio problem with factor models (CCMVFM)

Let \( K \) be the desired number of assets in the portfolio. Consider the following decision variables:

\( x_i \), binary variable that takes value 1 if the asset \( i \) is selected and otherwise 0, \( \forall i \in I \).

\( w_i \), weight of asset \( i \) in the portfolio, \( \forall i \in I \).

The solution of CCMVFM can be given by the mixed 0-1 binary quadratic optimization problem:

\[
\begin{align*}
\text{CCMVFM} : \quad & \min_{w,x} \quad \sum_{i,j \in I} \sum_{l,m \in F} \beta_{il} \beta_{jm} \sigma_{lm} w_i w_j + \sum_{i \in I} \sigma_{\epsilon_i}^2 w_i^2 \\
\text{s.t.} & \quad \sum_{i \in I} w_i = 1, \\
& \quad \sum_{i \in I} x_i \leq K, \\
& \quad 0 \leq w_i \leq x_i, \quad \forall i \in I, \\
& \quad x_i \in \{0, 1\}, \quad \forall i \in I.
\end{align*}
\]

If the factors are uncorrelated (\( \sigma_{lm} = 0, \forall l, m \in F : l \neq m \)), then the above objective function can be expressed
\[
\begin{align*}
\min_{w,x} \quad & \sum_{l \in F} \sigma_{ll} \sum_{i,j \in I} \beta_{il} \beta_{jl} w_i w_j + \sum_{i \in I} \sigma_{\epsilon_i}^2 w_i^2.
\end{align*}
\]

The advantage of introducing uncorrelated factors is that the variance of the portfolio can be expressed as a sum of the contributions of each factor’s variance, and this issue can lead to a more efficient framework for managing portfolio diversification, see [27,28,30]. The factor-risk-parity portfolio presented in [23] is based on rotation of the principal components that are maximally independent. A comparison of the minimum variance portfolio and the equally weighted portfolio is carried out in [4], concluding that factor component of risk and the factor exposures are consistently under forecast.
2.2 Equally weighted cardinality constrained portfolio problem

An interesting simplification of \( CCMVF M \) (1) imposes the equality weighted constraint, i.e., the weight of asset \( i, w_i \), is \( 1/K \) if the asset \( i \) is selected, and otherwise, 0. The best Equally Weighted Cardinality Constrained Minimum Variance portfolio for a multi Factor Model (EWCCMVFM), i.e., the solution of problem CCMVFM where all selected assets have the same weight, is the solution of the pure binary quadratic optimization problem, to be expressed

\[
EWCCMVFM : \frac{1}{K^2} \min_x \sum_{i,j \in I} \sum_{l,m \in F} \beta_{il} \beta_{jm} \sigma_{lm} x_i x_j + \sum_{i \in I} \sigma_{i}^2 x_i^2
\]

subject to:

\[
\sum_{i \in I} x_i = K, \\
x_i \in \{0, 1\}, \quad \forall i \in I,
\]

where \( x_i \) takes value 1 if the asset \( i \) is selected, and 0 otherwise.

The rationale behind the equality in the cardinality constraint (i.e., \( \sum_{i \in I} x_i = K \)) in (2) is based on the constraint \( \sum_{i \in I} w_i = 1 \) in model (1) and the assumption that \( w_i = 1/K, \forall i \in I \). The constraint \( \sum_{i} x_i \leq K \) in model (1) does not necessarily ensure that \( K \) assets will be selected, but in practice its satisfied with equality.

Note also that \( \sum_{i \in I} \sigma_{i}^2 x_i^2 \) in the objective function in (2) can be replaced with \( \sum_{i \in I} \sigma_{i}^2 x_i \), since \( x_i \) takes the value 0 or 1.

Problem (2) can be written as \( \{ \min_x \sum_{i,j \in I} a_{ij} x_i x_j, \text{ s.t. } \sum_{i \in I} x_i = K, \ x_i \in \{0, 1\} \forall i \in I \} \), where

\[
a_{ij} = \begin{cases} 
\frac{1}{K^2} \sum_{l,m \in F} \beta_{il} \beta_{jm} \sigma_{lm} & \text{if } i \neq j, \\
\frac{1}{K^2} \sum_{l,m \in F} \beta_{il} \beta_{im} \sigma_{lm} + \sigma_{i}^2 & \text{if } i = j.
\end{cases}
\]

2.3 Equally weighted cardinality constrained minimum variance portfolio problem for a single factor model

In this subsection some properties of problem EWCCMV are studied, where only one factor is considered. For a single factor \( f \), the return rates \( r_i \) of asset \( i \in I \) is given by

\[
r_i = \alpha_i + \beta_i f + \epsilon_i,
\]

where \( E(f) = \overline{f} \) and \( Var(f) = \sigma_f^2 \).

The 0-1 quadratic model for the Equally Weighted Cardinality Constraint Minimum Variance portfolio with a Single Factor \( f \) (EWCCMVSF) can be expressed

\[
EWCCMVSF : \frac{1}{K^2} \min_x \sigma_f^2 \sum_{i,j \in I} \beta_{ij} x_i x_j + \sum_{i \in I} \sigma_{i}^2 x_i
\]

subject to:

\[
\sum_{i \in I} x_i = K, \\
x_i \in \{0, 1\}, \quad \forall i \in I.
\]

\footnote{Problem (2) can be linearized by considering variables \( u_{ij} = 1 \) if both \( x_i = 1 \) and \( x_j = 1 \), 0 otherwise, and the constraints \( u_{ij} \leq x_i, u_{ij} \leq x_j, u_{ij} \geq x_i + x_j - 1 \).}
from where

\[
\text{EWCCMV SF : } \min_{\sigma_s, \sigma_{ns}^2, x} \sigma_s^2 + \sigma_{ns}^2
\]

\[
\text{s.t. } \sum_{i \in I} x_i = K,
\]

\[
\sigma_s = \frac{1}{K} \sum_{i \in I} \beta_i x_i,
\]

\[
\sigma_{ns}^2 = \frac{1}{K^2} \sum_{i \in I} \sigma_{\epsilon_i}^2 x_i,
\]

\[
x_i \in \{0, 1\}, \ \forall i \in I.
\]

Note that \(\sigma_s\) (systematic risk) is associated with the single factor \(f\) and \(\sigma_{ns}\) (non-systematic risk) refers to the assets.

### 2.3.1 Theoretical results

Let \(A\) be the set of points on the plane, \(A = \{(\beta_i \sigma_f / K, \sigma_{\epsilon_i}^2 / K^2), \ \forall i \in I\}\) and \(K\) be the cardinality parameter.

**Definition 1** The addition set of \(A\), denoted by \(A(K)\), is the set of all points generated by the addition of \(K\) points from \(A\).

\[
A(K) = \left\{ \sum_{a_i \in S \subset A} a_i, \ \forall S \subset A : |S| = K \right\}
\]

**Definition 2** Convex hull of a set \(A(K)\), denoted by \(\text{conv}(A(K))\), is the set of all convex combination points generated by addition of \(K\) points from \(A\), that is:

\[
\text{conv}(A(K)) = \left\{ \sum_{i=1}^{N} x_i a_i : a_i \in A, x_i \in \mathbb{R}, 0 \leq x_i \leq 1, \sum_{i=1}^{N} x_i = K \right\}
\]

The LP relaxation of problem (4) can be expressed

\[
\text{EWCCMV SF : } \min_{\sigma_s, \sigma_{ns}^2, x} \sigma_s^2 + \sigma_{ns}^2
\]

\[
\text{s.t. } (\sigma_s, \sigma_{ns}^2) \in \text{conv}(A(K)).
\]

**Proposition 1** The optimal solution of (5) is reached at the frontier of the set \(\text{conv}(A(K))\).

**Proof of Proposition** Trivial. \(\Box\)

**Theorem 1** (Carathéodory, [9]) For \(S \subset \mathcal{R}^d\), if \(x \in \text{conv}(S)\) then \(x \in \text{conv}(T)\) for some \(T \subset S : |T| \leq d + 1\).
The Carathéodory theorem [9] establishes that any point in $\text{conv}(A(K)) \subset \mathcal{R}^2$ can be represented as a convex combination of three points of $A(K)$. Note that each point in $A(K)$ is the addition of $K$ points of $A$. The next Corollary 1 restricts the Carathéodory theorem to the frontier of set $\text{conv}(A(K))$.

**Corollary 1** The frontier of the polyhedron $\text{conv}(A(K)) \subset \mathcal{R}^2$ is formed with faces of dimension 0 and 1, then the solution of (5), $(\sigma_s^*, \sigma_{ns}^*)$, is a convex combination of two points of $A(K)$, by assuming that there are no collinear points in the frontier of $\text{conv}(A(K))$.

From Corollary 1 it follows that the solution of (5) is reached in one point of $A(K)$, or in the linear combination of two of them. One consequence of this result is that the solutions only have two or less fractional values. This property is established in Proposition 2.

**Proposition 2** The solution of the LP relaxation of problem (3) contains two variables with a fractional value, at most.

**Proof of Proposition** If the solution of (3) is reached in a vertex $v$ of $\text{conv}(A(K))$, this point is the addition of $K$ points of $A$, therefore, exist $S \subset A : |S| = K$ such that $v = \sum_{i \in S} a_i$, and $x_i = 1$ if $i \in S$.

If the solution is reached in a face of dimension 1, an edge of $\text{conv}(A(K))$, then the solution is a convex combination of two vertexes, $v_1 = \sum_{i \in S_1} a_i$ and $v_2 = \sum_{i \in S_2} a_i$, of $\text{conv}(A(K))$, noting that the two vertexes defining the edge.

Suppose that $|S_1 \cup S_2| > K + 1$, i.e., $v_1$ and $v_2$ differ in two or more points from $A$. For example, $v_1 = a_1 + a_2 + a_5 + \cdots + a_K + a_{K+1} + a_{K+2}$, and $v_2 = a_3 + a_4 + a_5 + \cdots + a_K + a_{K+1} + a_{K+2}$. The interior point $0.5v_1 + 0.5v_2 = 0.5(a_1 + a_2) + 0.5(a_3 + a_4) + a_5 + \cdots + a_K + a_{K+1} + a_{K+2}$ can also be written as $0.5(a_1 + a_3) + 0.5(a_2 + a_4) + a_5 + \cdots + a_K + a_{K+1} + a_{K+2} = 0.5(a_1 + a_3 + a_5 + \cdots + a_K + a_{K+1} + a_{K+2}) + 0.5(a_2 + a_4 + a_5 + \cdots + a_K + a_{K+1} + a_{K+2}) = 0.5z_1 + 0.5z_2$, where $z_1, z_2 \in A(K)$. If $z_1$ and $z_2$ belong to the interior of $A(K)$, then $0.5v_1 + 0.5v_2$ is an interior point, also a contradiction. If $z_1$ or $z_2$ are vertexes of $\text{conv}(A(K))$, then $v_1$, $v_2$ and $z_1$ (or $v_1$, $v_2$ and $z_2$) are collinear points, and this contradicts the assumption that there are no collinear points in the frontier of $\text{conv}(A(K))$.

Therefore, a point in the frontier of $\text{conv}(A(K))$ is a linear combination of two points of $A(K)$ at most, and these two points of $A(K)$ differ in one point from $A$, at most.

**Remark** In the multi factor model EWCCMVFM the solution is also in the frontier, but in this case the dimension of the polyhedral facets are less or equal to $|F|$, where $|F|$ is the number of factors. Notice that the solution is a combination of $|F| + 1$ points (vertexes) of $A(K)$ but now, these points (vertexes) do not have to be consecutive, consequently they can differ in more than one point from $A$. The computational experience reported in Sect. 3 shows that the resolution of the factor models requires very small computing time, as in practice the solution of the LP relaxation of EWCCMVFM problem has few fractional variables.
2.3.2 Underline matrix problem EWCCMV SF

**Definition 3** A matrix $M$ is a Monge\(^2\) matrix if for every pair of rows $i < j$ and for every pair of columns $k < l$ satisfies the Monge property $M_{ik} + M_{jl} \leq M_{il} + M_{jk}$.

**Definition 4** A matrix $M$ is called an inverse Monge matrix if it satisfies the inverse Monge property $M_{ik} + M_{jl} \geq M_{il} + M_{jk}$, $\forall i < j$, $k < l$.

Note that a symmetric Monge matrix is called a Supnick matrix.

Monge matrices have many applications in combinatorial optimization problems, see [7,31,35]. For example, the Traveling Salesman Problem (TSP) can be solved in linear time if the underlying distance matrix is a Monge matrix, see [29].

**Proposition 3** The underlying matrix in EWCCMV SF (3) is an inverse Monge matrix.

**Proof of Proposition** The objective function of EWCC MV SF can be expressed

$$
\min \frac{1}{K^2} \sum_{i,j \in I} \left( \frac{\sigma_i^2 + \sigma_j^2}{2K} + \sigma_f^2 \beta_i \beta_j \right) x_i x_j
$$

(6)

where $\sum_{i \in I} x_i = K$. Notice that (6) is given by matrix $M_{ij} = \left\{ \frac{\sigma_i^2 + \sigma_j^2}{2K} + \sigma_f^2 \beta_i \beta_j \right\}_{ij}$.

By considering the ordered set $I$, i.e, $\beta_1 \leq \beta_2 \cdots \leq \beta_n$, is easy to show that $M_{ik} + M_{jl} \geq M_{il} + M_{jk}$, for all $i < j$ and $k < l$.

$$
\begin{align*}
\left( \frac{\sigma_i^2 + \sigma_k^2}{2K} + \sigma_f^2 \beta_i \beta_k \right) + \left( \frac{\sigma_j^2 + \sigma_l^2}{2K} + \sigma_f^2 \beta_j \beta_l \right) \\
\geq \left( \frac{\sigma_i^2 + \sigma_l^2}{2K} + \sigma_f^2 \beta_i \beta_l \right) + \left( \frac{\sigma_j^2 + \sigma_k^2}{2K} + \sigma_f^2 \beta_j \beta_k \right) \\
\left( \frac{\sigma_i^2 + \sigma_k^2}{2K} + \sigma_f^2 \beta_i \beta_k \right) + \left( \frac{\sigma_j^2 + \sigma_l^2}{2K} + \sigma_f^2 \beta_j \beta_l \right) \\
- \left( \frac{\sigma_i^2 + \sigma_l^2}{2K} + \sigma_f^2 \beta_i \beta_l \right) - \left( \frac{\sigma_j^2 + \sigma_k^2}{2K} + \sigma_f^2 \beta_j \beta_k \right) \\
= \sigma_f^2 (\beta_i - \beta_j)(\beta_k - \beta_l) \geq 0
\end{align*}
$$

Therefore, finding the equally weighted cardinality constrained portfolio for a single factor model is reduced, in fact, to finding the $K$ columns/rows in the matrix

\(^2\) Monge matrices are named after the French mathematician Gaspard Monge [1746–1818].

\( \square \) Springer
\[
\left\{ \frac{\sigma^2_i + \sigma^2_j}{2k} + \sigma^2_i \beta_i \beta_j \right\}_{ij}
\]
with a smaller cost. Notice that for a multi-factor model, with uncorrelated factors, the matrix can be expressed as
\[
\left\{ \frac{\sigma^2_i + \sigma^2_j}{2k} + \sum_{l \in F} \sigma^2_l \beta_i \beta_j \right\}_{ij}.
\]

### 2.3.3 The complexity of problem EWCCMV SF

In spite of the characteristics of the problem presented above, Proposition 4 establishes the problem EWCCMV SF as an NP-hard. Some polynomial-time solvable cases of the cardinality constrained quadratic optimization problem are reviewed and discussed in [19].

**Proposition 4**  Problem EWCCMV SF (3) is NP-hard.

**Proof of Proposition**  The NP-Hardness of problem EWCCMV SF can be proved by the reduction method. A problem \( A \) can be reduced to problem \( B \), if any instance of problem \( A \) can be formulated as an instance of problem \( B \), and this reformulation has polynomial-time complexity, see [11]. The reduction comes from the \( K \)-\textit{Subset Sum} Problem (KSSP) [11,20].

**KSSP:**

**Input:** A set of integers \( X := \{a_1, a_2, \ldots, a_n\} \), an integer \( k \) and \( \sum k \in Z^+ \).

**Question:** Is there a subset \( \{\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_k\} \subset X \) with cardinality \( k \) and \( \sum_{i=1}^k \hat{a}_i = \overline{b} \)?

Solving KSSP has been proved to be NP-hard, see [20], among others. Note that it is equivalent to state problem KSSP by letting \( \overline{b} = 0 \) or, alternatively, restricting set \( X \) to natural numbers, adding to set \( X \) its minimum value and replacing the integer \( \overline{b} \) by \( \overline{b} + k \cdot \min \{X\} \).

It can be shown that KSSP can be reduced to problem EWCCMV SF. Considered any instance of KSSP with input being the vector \( a = \{a_1, a_2, \ldots, a_n\} \) with \( a_i \in \mathbb{Z}^+ \), \( 0 \leq k \leq n \), \( k \in \mathbb{Z}^+ \) and \( \overline{b} \). Let \( z = (z_1, z_2, \ldots, z_n) \) be the decision variables. Then solving KSSP is equivalent to solving the linear equation with binary variables,

\[
P_1 : \{ z \in \{0,1\}^n | a'z = \overline{b}, z'1 = k \}.
\]

Now, construct the following special case of problem EWCCMV SF,

\[
P_2 : \min_x \left\{ (x'a)^2 + x' \left( \frac{\overline{b}^2}{k} - 2\overline{b}a \right) : x'1 = k, x \in \{0,1\}^n \right\}.
\]

Problem \( P_2 \) can be written as \( \min_x \left\{ (x'a - \overline{b})^2 : x'1 = k, x \in \{0,1\}^n \right\} \). Obviously, \( v(P_2) \geq 0 \), where \( v(P_2) \) denotes the optimal objective value of problem \( P_2 \). In particular, \( v(P_2) = 0 \), if there exist some \( x^* \) satisfying \( a'x = \overline{b}, x \in \{0,1\}^n \) and \( x'1 = k \). That is to say, if problem \( P_2 \) with \( v(P_2) = 0 \) is solved, then its optimal solution is exactly the solution of problem \( P_1 \). On the other hand, if we have \( v(P_2) > 0 \), then there is no \( x \) satisfying problem \( P_1 \), which implies that it admits no solution. This completes the reduction from \( P_1 \) to problem \( P_2 \), which implies the NP-Hardness of problem \( P_2 \) by the reduction rule.

\( \square \)
3 Computational results

Several instances have been considered for validating the proposal made in this work. They are from the index tracking instances available at the OR-Library [2]. A full list of the test datasets in the OR-Library, for a single factor model, can be found in [http://people.brunel.ac.uk/~mastjjb/jeb/orlib/indtrackinfo.html](http://people.brunel.ac.uk/~mastjjb/jeb/orlib/indtrackinfo.html). The instances selected are indtrack 6, 7 and 8, the largest; these datasets have been used in several papers, see [1, 8, 13, 36], among others.

Each dataset contains the weekly market price for a set of assets and the market index. Additionally, four principal components have been considered for each data set for their use as factors and evaluating the factor model presented in Sect. 2, when the eigenvalues are used in descending order jointly with the associated eigenvectors.

The computational experiments were conducted in a processor Intel(R) Xeon(R) CPU E5-2650 v3 @ 2.30 GHz, 64 gigabytes of RAM. The models have been implemented in C++, where the optimizer CPLEX v12.8 is used with its default options for optimizing the related problems.

The computational experience has been divided in two parts. First, a comparison is carried out on the performance of the models that have been proposed for factor models. Next, the computational experiment has been carried out for a single factor model. All the results have been obtained insample, i.e., the solutions provided by the models have not been evaluated outsample. The main aim of this work is only to compare different formulations of the cardinality constrained portfolio problem via factor models.

3.1 Computational results for a factor models

For each dataset in the experiment, their first four principal components are computed, and then, $\beta$ and $\sigma_i^2$ for each asset in these components (factors). Note that the use of principal components as factors provide factors which are uncorrelated. It is worth pointing out that it is often acceptable to consider uncorrelated factors even if they are correlated, see [17] and the references therein.

The computational experience is performed on the models $CCMVFM$ and $EWCCMVFM$ obtaining the four solutions that are notated as follows:

- $CCMVFM$ Solution of problem Cardinality Constrained Minimum Variance via Factor Models.
- $EWCCMVFM$ Solution of problem Equally Weighted Cardinality Constrained Minimum Variance via Factor Models.
- $CCMVFM_{N:=K(EW)}$ Solution of problem $CCMVFM$ but restricted to $K$ assets; i.e., solution of problem $EWCCMVFM$.
- $CCMVFM_{N:=K(LP)}$ Solution of problem $CCMVFM$ but restricted to the $K$ largest weights from the LP relaxation solution of problem $CCMVFM$.

Let the following additional notation: problem $(\cdot)$ $\in \{ EWCCMVFM, CCMVFM_{N:=K(EW)}, CCMVFM_{N:=K(LP)} \}$. Each dataset is solved for different values of the cardinality parameter $K$. Tables 1, 2 and 3 show the computational results for the models in each dataset (the caption of each table collects the dataset name, the market and
the number of assets), where the columns for each model and cardinality are as follows: *time*, elapsed time to obtain the optimal solution or the time limit (3600 s); $\sqrt{\text{obj}}$, square root of solution value; $\%\text{dev} = 100(\sqrt{\text{obj}}(\cdot) - \sqrt{\text{obj}}(\cdot))/\sqrt{\text{obj}}(\text{CCMVFM})$, deviation of the solution value obtained by each trial from the solution value of problem *CCMVFM*; $K$, the number of assets in the solution problem *CCMVFM* and number of assets in the solution together with the number of assets that coincide with the solution problem *CCMVFM* for the solution of problem $\cdot$; $L_1$ distance of the value of the variables in the solution of problem $\cdot$ from the solution $w^*$ in problem *CCMVFM*; $SD = \sqrt{w^T\Sigma w}$, standard deviation of each solution, $\%\text{dev}_{SD}$, $\%$ deviation of $SD$ respect to the solution *CCMVFM*; and $SR$, ratio return/risk $(w^T R/\sqrt{w^T \Sigma w})$ for each solution.

**Quality evaluation of the EWCCMVFM solution** First, the very small elapsed time that is required to obtain the solution of *EWCCMVFM* can be observed. In any instance no more than 1 s is required. The solutions of *EWCCMVFM* and *CCMVFM* are very similar. In Table 3, the $\%\text{dev}$ of *EWCCMVFM* from *CCMVFM* varies from 0.74 to 3.34%. This difference comes from imposing the equally weighted constraint on the solution. Nevertheless, the equally weighted solution provides better results, in some instances, when the $SD$ and $SR$ are evaluated, which implies that both solutions are good to approximate the solution of the minimization variance problem. It is worth pointing out that the deviation $\%\text{dev}$ of the solution *CCMVFM* only varies from 0.00 to 1.37%. On the other hand, the solution *CCMVFM* is worse than the solutions for the *CCMVFM*, also when the $SD$ and the $SR$ are considered. Similar conclusions can be drawn from Tables 2 and 3.

In summary, from the results obtained by the models, it can be deduced from this preliminary computational experiment that the solution values do not differ too much. *CCMVFM* require a high elapsed time, while *EWCCMVFM* is very fast, in fact the elapsed time can be measured in tenths of a second. *EWCCMVFM* is in practice an easy problem for an optimizer such as CPLEX, bearing in mind that the solution value of the LP relaxation is very close to the optimal one. Proposition 2 confirms this for the Single Factor problem *EWCCMVF*.

### 3.2 Computational results for the single factor model

A comparison of the model that is proposed in Sect. 2.3 is performed by considering the models *CCMVFSF* and *EWCCMVFSF* as the benchmark. The notation is as follows:

*CCMVFSF* Solution of problem Cardinality Constrained Minimum Variance for a Single Factor.

*EWCCMVFSF* Solution of problem Equally Weighted Cardinality Constrained Minimum Variance for a Single Factor.

*CCMVFSF* Solution of problem for a single factor as their counterparts in Sect. 3.1 for several factors.

*Approx-EWCCMVFSF* Solution of $K$ small values of $\sigma_i^2 \beta_i^2 + \sigma_i^2$, $\forall i \in I$, i.e., the underlying matrix in *EWCCMVFSF* as a diagonal one.
| \(N = 457\) | Model | Solution | \(SD\) | \(\%_{dev}SD\) | \(SR\) |
|---|---|---|---|---|---|
| \(K = 5\) | \(CCMV FM\) | 274.97 | 0.01501 | 5 | 0.11261 |
| \(EWCCMV FM\) | 0.07 | 0.01502 | 0.91 | 5-4 | 0.44 | 0.01662 | 5.30 | 0.08229 |
| \(CCMV FM_{N:=K(\text{EW})}\) | 0.02 | 0.01534 | 0.06 | 5-3 | 0.33 | 0.01641 | 4.00 | 0.08349 |
| \(CCMV FM_{N:=K(\text{LP})}\) | 0.02 | 0.01502 | 2.23 | 5-4 | 0.71 | 0.01723 | 9.17 | 0.07521 |
| \(K = 10\) | \(CCMV FM\) | 3600.00 | 0.01276 | 10 | 0.11329 |
| \(EWCCMV FM\) | 0.04 | 0.01285 | 0.74 | 10-10 | 0.14 | 0.01509 | 0.01520 | 0.068 | 0.11475 |
| \(CCMV FM_{N:=K(\text{EW})}\) | 0.02 | 0.01276 | 0.00 | 10-10 | 0.00 | 0.01509 | 0.00 | 0.11329 |
| \(CCMV FM_{N:=K(\text{LP})}\) | 0.02 | 0.01321 | 3.59 | 10-7 | 0.59 | 0.01655 | 9.62 | 0.10747 |
| \(K = 20\) | \(CCMV FM\) | 3600.00 | 0.01138 | 20 | 0.12653 |
| \(EWCCMV FM\) | 0.04 | 0.01166 | 2.46 | 20-17 | 0.39 | 0.01605 | 4.94 | 0.12190 |
| \(CCMV FM_{N:=K(\text{EW})}\) | 0.02 | 0.01154 | 1.37 | 20-17 | 0.26 | 0.01564 | 2.22 | 0.12077 |
| \(CCMV FM_{N:=K(\text{LP})}\) | 0.03 | 0.01187 | 4.26 | 20-16 | 0.40 | 0.01585 | 3.62 | 0.11046 |
| \(K = 30\) | \(CCMV FM\) | 3600.00 | 0.01095 | 30 | 0.13001 |
| \(EWCCMV FM\) | 0.09 | 0.01131 | 3.34 | 30-26 | 0.47 | 0.01527 | 4.41 | 0.11670 |
| \(CCMV FM_{N:=K(\text{EW})}\) | 0.01 | 0.01103 | 0.76 | 30-26 | 0.22 | 0.01556 | 1.91 | 0.11837 |
| \(CCMV FM_{N:=K(\text{LP})}\) | 0.03 | 0.01107 | 1.17 | 30-26 | 0.26 | 0.01535 | 0.52 | 0.11506 |

Time limit 3600 s
N = 457
### Table 2

| N = 1318 | Model                         | Solution | SD | %devSD | SR    |
|----------|-------------------------------|----------|----|--------|-------|
|          |                               | Time     | √obj | %dev   | K     | w − w* || 1 |
| K = 5    | **CCMV FM**                   | 3600.00  | 0.01134 | 5      | 0.01298 | 0.10413 |
|          | **EWCCMVFM**                  |          | 0.01142 | 0.67   | 5-4    | 0.12    | 0.01299 | 0.10   | 0.10526 |
|          | **CCMV FMN:=K(EW)**          | 0.05     | 0.01134 | 0.00    | 5-4    | 0.00    | 0.01298 | 0.00   | 0.10413 |
|          | **CCMV FMN:=K(LP)**          | 0.09     | 0.01280 | 12.88   | 5-2    | 1.26    | 0.01479 | 13.97  | 0.12243 |
| K = 10   | **CCMV FM**                   | 3600.00  | 0.00902 | 10     | 0.01101 | 0.12101 |
|          | **EWCCMVFM**                  | 0.13     | 0.00913 | 1.21    | 10-9   | 0.29    | 0.01070 | −2.82 | 0.15486 |
|          | **CCMV FMN:=K(EW)**          | 0.02     | 0.00903 | 0.15    | 10-9   | 0.16    | 0.01068 | −2.97 | 0.14966 |
|          | **CCMV FMN:=K(LP)**          | 0.06     | 0.00944 | 4.59    | 10-6   | 0.89    | 0.01125 | 2.16  | 0.14705 |
| K = 20   | **CCMV FM**                   | 3600.00  | 0.00725 | 20     | 0.01001 |        |
|          | **EWCCMVFM**                  | 0.24     | 0.00735 | 1.37    | 20-17  | 0.41    | 0.00996 | −0.51 | 0.14497 |
|          | **CCMV FMN:=K(EW)**          | 0.02     | 0.00727 | 0.24    | 20-17  | 0.29    | 0.00977 | −2.37 | 0.14841 |
|          | **CCMV FMN:=K(LP)**          | 0.06     | 0.00733 | 1.15    | 20-17  | 0.32    | 0.01020 | 1.97  | 0.15198 |
| K = 30   | **CCMV FM**                   | 3600.00  | 0.00653 | 30     | 0.00919 | 0.16698 |
|          | **EWCCMVFM**                  | 0.09     | 0.00663 | 1.59    | 30-28  | 0.24    | 0.00889 | −3.30 | 0.16860 |
|          | **CCMV FMN:=K(EW)**          | 0.01     | 0.00654 | 0.14    | 30-28  | 0.12    | 0.00896 | −2.56 | 0.16742 |
|          | **CCMV FMN:=K(LP)**          | 0.07     | 0.00654 | 0.27    | 30-28  | 0.11    | 0.00905 | −1.52 | 0.16372 |
| K = 40   | **CCMV FM**                   | 3600.00  | 0.00620 | 40     | 0.00883 | 0.18705 |
|          | **EWCCMVFM**                  | 0.42     | 0.00640 | 3.15    | 40-36  | 0.34    | 0.00866 | −1.94 | 0.17404 |
|          | **CCMV FMN:=K(EW)**          | 0.02     | 0.00624 | 0.62    | 40-36  | 0.15    | 0.00873 | −1.19 | 0.17241 |
|          | **CCMV FMN:=K(LP)**          | 0.06     | 0.00621 | 0.12    | 40-39  | 0.05    | 0.00871 | −1.40 | 0.18757 |
| Model                | Time  | \(\sqrt{\text{obj}}\) | %\text{dev} | \(K\)   | \(\|w - w^*\|_1\) | \(SD\) | %\text{dev}_SD | \(SR\) |
|---------------------|-------|----------------|-----------|---------|-------------------|-------|--------------|--------|
| \(K = 50\)         |       |                 |           |         |                   |       |              |        |
| \(CCMV FM\)        | 3600.00 | 0.00603         |           | 50      | 0.00865           |       |              | 0.20275|
| \(EWCCMV FM\)      | 0.22  | 0.00631         | 4.69      | 50-46   | 0.00876           | 1.27  | 0.20023      |        |
| \(CCMV FM_{N=K(EW)}\) | 0.02  | 0.00606         | 0.53      | 50-46   | 0.00868           | 0.35  | 0.19088      |        |
| \(CCMV FM_{N=K(LP)}\) | 0.06  | 0.00606         | 0.50      | 50-47   | 0.00870           | 0.59  | 0.19365      |        |

Time limit 3600 s
\(N = 1318\)
Table 3  indtrack8.txt  Russel 3000 index

| N = 2151 | Model | Solution | SD | %devSD | SR |
|----------|-------|----------|----|--------|----|
|          |       | Time     | √obj | %dev  | K     | ||w − w*||_1 |
|          |       |          |      |        |       |            |
|          |       | 3600.00  | 0.1120 | 0.78  | 5     | 0.01139    | 0.09603   |
|          |       | 0.13     | 0.1129 | 0.00  | 5-5   | 0.01143    | 0.34       | 0.09534   |
|          |       | 0.02     | 0.1120 | 16.85 | 5-1   | 0.01422    | 24.88      | 0.15818   |
|          |       | 0.08     | 0.1309 | 0.01  |       |            |            |
|          |       |          |      |        |       |            |            |
|          |       | 3600.00  | 0.0893 | 0.82  | 10    | 0.01042    | 0.12100   |
|          |       | 0.10     | 0.0900 | 0.15  | 10-9  | 0.01058    | 1.49       | 0.10739   |
|          |       | 0.02     | 0.0894 | 9.20  | 10-5  | 0.01294    | 24.16      | 0.10220   |
|          |       | 0.08     | 0.0975 | 2.13  | 20-16 | 0.00922    | 0.15003    |
|          |       |          |      |        |       |            |            |
|          |       | 3600.00  | 0.0639 | 3.52  | 30    | 0.00921    | 0.16408    |
|          |       | 0.57     | 0.0651 | 1.85  | 30-26 | 0.00951    | 3.29       | 0.14650   |
|          |       | 0.02     | 0.0644 | 0.66  | 30-25 | 0.00928    | 0.81       | 0.14993   |
|          |       | 0.08     | 0.0647 | 1.16  | 30-26 | 0.00995    | 8.04       | 0.15661   |
|          |       |          |      |        |       |            |            |
|          |       | 3600.00  | 0.0600 | 1.85  | 40    | 0.00941    | 0.16969    |
|          |       | 0.24     | 0.0616 | 2.71  | 40-36 | 0.00963    | 2.38       | 0.16243   |
|          |       | 0.02     | 0.0605 | 0.79  | 40-36 | 0.00931    | 1.04       | 0.16224   |
|          |       | 0.08     | 0.0608 | 1.26  | 40-36 | 0.00993    | 5.57       | 0.15905   |
Table 3 continued

| $N = 2151$ | Model                        | Time limit 3600 s | Solution | $SD$ | $\%\text{dev}_{SD}$ | $SR$ |
|------------|-----------------------------|-------------------|----------|------|---------------------|------|
| $K = 50$   | $CCMV FM$                   | 3600.00           | $\sqrt{\text{obj}}$ 0.00578 | 50   | 0.00964             | 0.17640 |
|            | $EWCCMV FM$                 | 0.12              | $\%\text{dev}$ 3.56 | 50-44 | 0.40                | 0.00951 | -1.37 | 0.17309 |
|            | $CCMV FM_{N:=K(EW)}$        | 0.02              | $K$ 50-44 | 0.20  | 0.00918             | -4.78 | 0.17317 |
|            | $CCMV FM_{N:=K(LP)}$        | 0.08              | $||w - w^*||_1$ 0.17 | 0.14  | 0.00965             | 0.17409 |

Time limit 3600 s
$N = 2151$
Tables 4, 5 and 6 show the same information as Tables 1, 2 and 3, but for the above solutions for a single factor model.

The first observation is that the computing time for solving $CCMV SF$ (i.e., the original problem by plain use of CPLEX) is high for all the instances (1 h being the allowed computing time). On the other hand, the results in problem $EWCCMV SF$ require in all the instances less than one second, this being a consequence of Proposition 2 in Section 2.4. Moreover, the quality of the solutions obtained in problem $EWCCMV$ is high, with a deviation $%_{dev}$ of 4.35% in the worst instance, where 48 out of 50 the assets are selected in the solution of the PROBLEM $CCMV SF$ and $L_1$ is 0.31. Additionally, the solution $CCMV FSF := K(EW)$ gives a deviation $%_{dev} = 0.39\%$, see Table 5.

The solutions’ quality, compared with the best $CCMV SF$ solution obtained (within the one-hour allowed time), is also shown in the columns $SD$ and $SR$, noting that all the solutions are close to each other, except the solution Approx-$EWCCMV SF$. For example, in instance indtrack6 (Table 4) the solution Approx-$EWCCMV SF$ has a deviation $%_{dev}$ up to 21.47% from solution $CCMV SF$, but with a deviation $%_{dev}_{SD}$ up to $-18.61\%$ from the standard deviation of solution $CCMV SF$. Those results evidence that the single model factor does not significantly explain the assets risk.

Finally, in Tables 4, 5 and 6 it can be seen that when $K$ increases all the measures take similar values.

4 Conclusions

In this work two alternative models are proposed and analyzed for designing the cardinality constrained minimum-variance portfolio via factor models. Both models are pure 0-1 ones in contrast with the mixed quadratic factor model in the literature. When the the factors are uncorrelated or only a factor is considered, the computational time required to solve the pure 0-1 descends drastically. The assumption of uncorrelated factors is not very restrictive in the financial context.

Regarding the comparison of the models, the equally weighted cardinality constrained portfolio problem has provides the most promising results. In terms of computational time, all the instances require less than one second. Therefore, the models $EWCMVM FM$ and $EWCMVSF$ can be regarded as being superior to the classical factor models in terms of usability; in fact, they obtain high quality solutions with small computational time. From a practical point of view, the validity of the models presented in this work has been assessed by considering realistic datasets taken from the literature. Those models, especially the equally weighted one provide a good approximation to the cardinality constraint minimum variance portfolio problem. They can be helpful to the practitioners to evaluate which best assets to consider and in a posterior analysis to apply other more complex techniques.

The future research plan consists of extending the proposed models to the minimization of a trade-off function risk/return, where the transactions are taking into account as well as the evaluation out-sample of the performance of the models.
| $N = 457$ | Model | Solution | $SD$ | $\%dev_{SD}$ | $SR$ |
| --- | --- | --- | --- | --- | --- |
| $K = 5$ | $CCMV SF$ | 382.44 | 0.01299 | 5 | 0.02198 | 0.06554 |
| | $EWCCMV SF$ | 0.04 | 0.01313 | 1.06 | 5-5 | 0.13 | 0.02252 | 2.49 | 0.0647 |
| | $CCMV SF_{N:=K(EW)}$ | 0.06 | 0.01299 | 0.00 | 5-5 | 0.00 | 0.02198 | 0.00 | 0.06554 |
| | Approx-$EWCCMV SF$ | 0.01 | 0.01391 | 7.07 | 5-2 | 1.23 | 0.01752 | -20.30 | 0.06216 |
| $K = 10$ | $CCMV SF$ | 3600.00 | 0.00985 | 10 | 0.02185 | 0.07077 |
| | $EWCCMV SF$ | 0.03 | 0.00993 | 0.77 | 10-10 | 0.10 | 0.02227 | 1.91 | 0.07036 |
| | $CCMV SF_{N:=K(EW)}$ | 0.05 | 0.00985 | 0.00 | 10-10 | 0.00 | 0.02185 | 0.00 | 0.07077 |
| | Approx-$EWCCMV SF$ | 0.01 | 0.01039 | 5.51 | 10-7 | 0.64 | 0.01791 | -18.03 | 0.07359 |
| $K = 20$ | $CCMV SF$ | 3600.00 | 0.00775 | 20 | 0.02085 | 0.07614 |
| | $EWCCMV SF$ | 0.05 | 0.00783 | 1.01 | 20-19 | 0.20 | 0.02049 | -1.73 | 0.07752 |
| | $CCMV SF_{N:=K(EW)}$ | 0.07 | 0.00776 | 0.08 | 20-19 | 0.09 | 0.02052 | -1.57 | 0.07716 |
| | Approx-$EWCCMV SF$ | 0.01 | 0.00941 | 21.47 | 20-12 | 0.83 | 0.01697 | -18.61 | 0.09637 |
| $K = 30$ | $CCMV SF$ | 3600.00 | 0.00703 | 30 | 0.02024 | 0.07604 |
| | $EWCCMV SF$ | 0.05 | 0.00722 | 2.69 | 30-29 | 0.24 | 0.02020 | -0.19 | 0.07605 |
| | $CCMV SF_{N:=K(EW)}$ | 0.07 | 0.00706 | 0.39 | 30-29 | 0.05 | 0.02052 | 1.39 | 0.07561 |
| | Approx-$EWCCMV SF$ | 0.01 | 0.00842 | 19.73 | 30-21 | 0.66 | 0.01806 | -10.75 | 0.08418 |

Time limit 3600 s
N=457
Table 5  indtrack7.txt  Russel 2000 index

| K  | Model             | Time   | √obj    | %dev   | K     | ||w−w*||1 | SD    | %devSD | SR    |
|----|-------------------|--------|---------|--------|-------|----------|-------|--------|-------|
| 5  | CCMVSF            | 3600.00| 0.1104  | 5      | 0.10  | 0.01221  | 0.10533|
|    | EWCCMVSF         | 0.09   | 0.01111 | 0.64   | 5-5   | 0.01220  | -0.14 | 0.10625|
|    | CCMVSF_N:=K(EW)  | 0.13   | 0.01104 | 0.00   | 5-5   | 0.01221  | 0.00   | 0.10533|
|    | Approx-EWCCMVSF  | 0.03   | 0.01183 | 7.17   | 5-2   | 0.01468  | 20.23  | 0.05573|
| 10 | CCMVSF           | 3600.00| 0.00864 | 1.05   | 10-9  | 0.01110  | 3.04   | 0.12488|
|    | EWCCMVSF         | 0.15   | 0.00873 | 0.42   | 10-10 | 0.01105  | 2.55   | 0.12495|
|    | CCMVSF_N:=K(EW)  | 0.02   | 0.00868 | 0.00   | 20-20 | 0.00931  | 0.00   | 0.15070|
|    | Approx-EWCCMVSF  | 0.01   | 0.00928 | 7.37   | 10-5  | 0.01247  | 15.69  | 0.07334|
| 20 | CCMVSF           | 3600.00| 0.00690 | 20     |       | 0.00931  | 15.61  | 0.09964|
|    | EWCCMVSF         | 0.06   | 0.00702 | 1.80   | 20-20 | 0.00941  | 1.07   | 0.14976|
|    | CCMVSF_N:=K(EW)  | 0.09   | 0.00690 | 0.00   | 20-20 | 0.00931  | 0.00   | 0.15070|
|    | Approx-EWCCMVSF  | 0.00   | 0.00784 | 13.59  | 20-10 | 0.01077  | 15.61  | 0.09964|
| 30 | CCMVSF           | 3600.00| 0.00616 | 30     |       | 0.00880  | 17.744 |
|    | EWCCMVSF         | 0.06   | 0.00627 | 1.82   | 30-28 | 0.00902  | 2.48   | 0.17906|
|    | CCMVSF_N:=K(EW)  | 0.09   | 0.00617 | 0.11   | 30-28 | 0.00898  | 2.01   | 0.17503|
|    | Approx-EWCCMVSF  | 0.01   | 0.00761 | 23.54  | 30-12 | 0.01094  | 24.34  | 0.12523|
| 40 | CCMVSF           | 3600.00| 0.00577 | 40     |       | 0.00881  | 19.180 |
|    | EWCCMVSF         | 0.07   | 0.00596 | 3.29   | 40-37 | 0.00888  | 0.84   | 0.18926|
|    | CCMVSF_N:=K(EW)  | 0.10   | 0.00581 | 0.59   | 40-37 | 0.00881  | 0.03   | 0.18511|
|    | Approx-EWCCMVSF  | 0.01   | 0.00738 | 27.83  | 40-17 | 0.01123  | 27.49  | 0.13159|
| $K = 50$ | Model | Solution | Time | $\sqrt{obj}$ | $%dev$ | $K$ | $||w - w^*||_1$ | $SD$ | $%dev_{SD}$ | $SR$ |
|----------|-------|----------|------|--------------|--------|----|----------------|------|------------|------|
|          | $CCMV_{SF}$ | 3600.00  | 0.00554 | 50 | 0.00865 | 0.20516 |
|          | $EWCCMV_{SF}$ | 0.06 | 0.00578 | 4.35 | 50-48 | 0.31 | 0.00906 | 4.81 | 0.19987 |
|          | $CCMV_{SF} \N = K(EW)$ | 0.09 | 0.00557 | 0.48 | 50-48 | 0.06 | 0.00883 | 2.08 | 0.19541 |
|          | Approx-$EWCCMV_{SF}$ | 0.01 | 0.00728 | 31.31 | 50-21 | 1.18 | 0.01117 | 29.19 | 0.13479 |

Time limit 3600 s
$N = 1318$
## Table 6

| Model          | $N = 2151$ | Solution | $SD$ | $%dev_{SD}$ | $SR$ |
|----------------|------------|----------|------|-------------|------|
|                | $K = 5$    |          |      |             |      |
| $CCMV SF$      |            | 3600.00  | 0.00 | 0.01193     | 0.11668 |
| $EWCCMV SF$    |            | 0.15     | 0.01057 | 0.01192 | -0.06 | 0.11939 |
| $CCMV SF_N:K(EW)$ | 0.02     | 0.01048  | 0.00 | 0.01193 | 0.00 | 0.11668 |
| Approx-$EWCCMV SF$ | 0.02 | 0.01081 | 3.19 | 0.01368 | 14.71 | 0.08121 |
|                | $K = 10$   |          |      |             |      |
| $CCMV SF$      |            | 3600.00  | 0.01081 | 0.01188 | 0.10545 |
| $EWCCMV SF$    |            | 0.09     | 0.008645 | 0.01265 | 7.30 | 0.10659 |
| $CCMV SF_N:K(EW)$ | 0.14     | 0.00819 | 0.16 | 0.01220 | 3.52 | 0.10890 |
| Approx-$EWCCMV SF$ | 0.02 | 0.00905 | 10.77 | 0.01315 | 10.76 | 0.08141 |
|                | $K = 20$   |          |      |             |      |
| $CCMV SF$      |            | 3600.00  | 0.00635 | 0.01179 | 0.11431 |
| $EWCCMV SF$    |            | 0.08     | 0.00562 | 0.01261 | 2.89 | 0.10829 |
| $CCMV SF_N:K(EW)$ | 0.10     | 0.00555 | 0.10 | 0.01225 | -0.06 | 0.11047 |
| Approx-$EWCCMV SF$ | 0.01 | 0.00708 | 27.61 | 0.0150 | -6.16 | 0.1157 |
|                | $K = 30$   |          |      |             |      |
| $CCMV SF$      |            | 3600.00  | 0.00555 | 0.01226 | 0.11083 |
| $EWCCMV SF$    |            | 0.08     | 0.00562 | 0.01261 | 2.89 | 0.10829 |
| $CCMV SF_N:K(EW)$ | 0.10     | 0.00555 | 0.10 | 0.01225 | -0.06 | 0.11047 |
| Approx-$EWCCMV SF$ | 0.01 | 0.00708 | 27.61 | 0.0150 | -6.16 | 0.1157 |
|                | $K = 40$   |          |      |             |      |
| $CCMV SF$      |            | 3600.00  | 0.00510 | 0.01255 | 0.11566 |
| $EWCCMV SF$    |            | 0.09     | 0.00519 | 0.01281 | 2.09 | 0.10080 |
| $CCMV SF_N:K(EW)$ | 0.11     | 0.00511 | 0.18 | 0.01254 | -0.04 | 0.10347 |
| Approx-$EWCCMV SF$ | 0.01 | 0.00687 | 34.82 | 0.01147 | -8.56 | 0.11875 |
| Model | Time (s) | √obj | %dev | K | ||w − w*||₁ | SD | %devSD | SR |
|-------|----------|------|------|---|-------------|----|--------|----|
| CCMVSF | 3600.00  | 0.00480 | 50  | 0.01204 | 0.12644 |
| EWCCMV SF | 0.10 | 0.00491 | 2.23 | 50-47 | 0.27 | 0.01224 | 1.70 | 0.13159 |
| CCMVSF<sub>N:=K(EW)</sub> | 0.12 | 0.00481 | 0.12 | 50-47 | 0.09 | 0.01217 | 1.13 | 0.12576 |
| Approx-EWCCMV SF | 0.01 | 0.00687 | 43.06 | 50-16 | 1.36 | 0.01128 | -6.28 | 0.13587 |

Time limit 3600 s
N=2151
Acknowledgements This work was partly supported by the Spanish Ministry for Economy and Competitiveness, the State Research Agency and the European Regional Development Fund under Grant MTM2016-79765-P (AEI/FEDER, UE). I am grateful to anonymous reviewers for numerous suggestions to improve this paper.

References

1. Beasley, J.E., Meade, N., Chang, T.J.: An evolutionary heuristic for the index tracking problem. Eur. J. Oper. Res. 148, 621–643 (2003)
2. Beasley, J.E.: OR-Library: distributing test problems by electronic mail. J. Oper. Res. Soc. 41, 1069–1072 (1990)
3. Bertsimas, D., Shioda, R.: Algorithm for cardinality-constrained quadratic optimization. Comput. Optim. Appl. 43, 1–22 (2009)
4. Bianchi, S.W., Goldberg, L.R., Rosenberg, A.: The impact of estimation error on latent factor model forecast of portfolio risk. J Portf Manag Spec QES Issue 43, 147–156 (2017)
5. Bienstock, D.: Computational study of a family of mixed-integer quadratic programming problems. Math. Program. 74, 121–140 (1996)
6. Burak Paç, A., Pınar, M.C.: On robust portfolio and naïve diversification: mixing ambiguous and unambiguous assets. Ann. Oper. Res. 266, 223–253 (2018)
7. Burckard, R.E., Klinz, B., Rudolf, R.: Perspectives of Monge properties in optimization. Discrete Appl. Math. 70, 95–161 (1996)
8. Canakgoz, N.A., Beasley, J.E.: Mixed-integer programming approaches for index tracking and enhanced indexation. Eur. J. Oper. Res. 196, 384–399 (2009)
9. Carathéodory, C.: Über den Variabilitätsbereich der Koeffizienten von Potenzreihen, die gegebene Werte nicht annehmen. Math. Ann. 64, 95–115 (1907)
10. Castro, F., Gago, J., Hartillo, I., Puerto, J., Ucha, J.M.: An algebraic approach to integer portfolio problems. Eur. J. Oper. Res. 210, 647–659 (2011)
11. Cormen, T.H., Leiserson, C.E., Rivest, R., Stein, S.: Introduction to Algorithms. The MIT Press, Cambridge (2001)
12. Cesaroni, F., Scozzari, A., Tardeila, F.: A new method for mean-variance portfolio optimization with cardinality constraints. Ann. Oper. Res. 8, 213–234 (2013)
13. Chang, T.J., Meade, N., Beasley, J.E., Sharaiha, Y.M.: Heuristics for cardinality constrained portfolio optimisation. Comput. Oper. Res. 27, 1271–1302 (2000)
14. Deguest, R., Martellini, L., Meucci, A.L.: Risk parity and beyond—from asset allocation to risk allocation decisions. EDHEC-Risk Institute, Nice, France. Working paper (2013)
15. DeMiguel, V., Garlappi, L., Uppal, R.: Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? Rev. Financ. Stud. 18, 1219–1251 (2009)
16. DeMiguel, V., Nogales, F.J.: Portfolio selection with robust estimation. Oper. Res. 59, 560–577 (2009)
17. Driessen, J., Melenberg, B., Nijman, T.: Common factors in international bond returns. J. Int. Money Finance 22, 629–656 (2003)
18. Gao, J., Li, D.: Optimal cardinality constrained portfolio selection. Oper. Res. 61, 745–761 (2013)
19. Gao, J., Li, D.: A polynomial case of the cardinality-constrained quadratic optimization problem. J. Glob. Optim. 56, 1441–1455 (2013)
20. Garey, M.R., Johnson, D.S.: Computer and Intractability, a Guide to the Theory of NP-Completeness. W.H. Freeman Co, San Francisco (1979)
21. Haley, M.R.: Sharfall minimization and the Naive (1/N) portfolio: an out-of-sample comparison. Appl. Econ. Lett. 23, 926–929 (2016)
22. Kellerer, H., Mansini, R., Speranza, M.G.: Slecting portfolios with fixed cost and minimum transactions lots. Ann. Oper. Res. 99, 287–304 (2000)
23. Lassance, N., DeMiguel, V., Vrins, F.: Optimal portfolio diversification via independent component analysis. Available at SSRN: https://doi.org/10.2139/ssrn.3285156 (2019)
24. Li, D., Sun, X., Wang, J.: Optimal lot solution to cardinality constrained mean-variance formulation for portfolio selection. Math. Finance 16, 83–101 (2006)
25. Mansini, R., Speranza, G.: An exact approach for portfolio selection with transaction cost and rounds. IIE Trans. 37, 919–929 (2005)
26. Markowitz, H.M.: Portfolio Selection. J. Finance 7, 77–91 (1952)
27. Meucci, A.: Managing diversification. Risk 22, 74–79 (2009)
28. Meucci, A., Santangelo, A., Deguest, R.: Risk budgeting and diversification based on optimized uncorrelated factors. Risk 11, 70–75 (2015)
29. Park, J.K.: A special case of the n-vertex traveling salesman problem that can be solved in O(n) time. Inf. Process. Lett. 40, 247–254 (1991)
30. Peng, Y., Albuquerque, P.H.M., Nascimento, I.F.do, Machado, J.V.F.: Between nonlinearities, complexity, and noises: an application on portfolio selection using kernel principal component analysis. Entropy 21, 376 (2019)
31. Pferschy, U., Rudolf, R., Woeginger, G.: Monge matrices make maximization manageable. Oper. Res. Lett. 16, 245–254 (1994)
32. Sharpe, W.F.: A simplified model for portfolio analysis. Manage. Sci. 9, 277–293 (1963)
33. Sharpe, W.F.: Capital asset prices: a theory of market equilibrium under condition of risk. J. Finance 19, 425–442 (1964)
34. Shaw, D.X., Liu, S., Kopman, L.: Lagrangian relaxation procedure for cardinality—constrained portfolio optimization. Optim. Methods Softw. 23, 411–420 (2008)
35. Woeginger, G.J.: Computational problems without computation. Nieuw Archief voor Wiskunde 2, 140–147 (2003)
36. Woodside-Oriakhi, M., Lucas, C., Beasley, J.E.: Heuristic algorithms for the cardinality constrained efficient frontier. Eur. J. Oper. Res. 213, 538–550 (2011)
37. Zakamulin, V.: Superiority of optimized portfolios to naive diversification: Fact or fiction? Finance Res. Lett. 22, 122–128 (2017)

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.