OR Spectrum
DOI 10.1007/s00291-014-0383-6

REGULAR ARTICLE

A linear risk-return model for enhanced indexation in portfolio optimization

Renato Bruni · Francesco Cesarone · Andrea Scozzari · Fabio Tardella

Abstract Enhanced indexation (EI) is the problem of selecting a portfolio that should produce excess return with respect to a given benchmark index. In this work, we propose a linear bi-objective optimization approach to EI that maximizes average excess return and minimizes underperformance over a learning period. Our model can be efficiently solved to optimality by means of standard linear programming techniques. On the theoretical side, we investigate conditions that guarantee or forbid the existence of a portfolio strictly outperforming the index. On the practical side, we support our model with extensive empirical analysis on publicly available real-world financial datasets, including comparison with previous studies, performance and diversification analysis, and verification of some of the proposed theoretical results on real data.

The third author wishes to thank the partial support received from the Spanish Ministry of Science and Technology through grant number MTM2013-46962-C2-1-P.

R. Bruni (✉)
Dip. di Ingegneria Informatica, Automatica e Gestionale, Sapienza Università di Roma, Rome, Italy
e-mail: bruni@dis.uniroma1.it

F. Cesarone
Dip. di Studi Aziendali, Università degli Studi Roma Tre, Rome, Italy
e-mail: francesco.cesarone@uniroma3.it

A. Scozzari
Facoltà di Economia, Università degli Studi Niccolò Cusano-Telematica, Rome, Italy
e-mail: andrea.scozzari@unicusano.it

F. Tardella
Dip. di Metodi e Modelli per l'Economia, il Territorio e la Finanza, Sapienza Università di Roma, Rome, Italy
e-mail: fabio.tardella@uniroma1.it

Published online: 25 November 2014

Keywords Portfolio optimization · Index tracking · Enhanced indexation · Performance analysis · Linear programming

1 Introduction

Portfolio optimization in asset management is the problem of selecting the composition of a portfolio to pursue, as far as possible, both return maximization and risk minimization. The basic index tracking (IT) problem consists in selecting a portfolio, possibly with a small number of assets, that best replicates (tracks) the performance of a given index or benchmark. This problem is usually formulated as the problem of minimizing a selected distance measure, computed over a learning period, between the index and a tracking portfolio that uses at most m out of n available assets. Extensive reviews of the literature on this problem can be found in Beasley et al. (2003) and, more recently, in Canakgoz and Beasley (2008) and Guastaroba and Speranza (2012). An evolution of this approach is called enhanced index tracking, or enhanced indexation (EI), and its purpose is to outperform, and and not just track, the given index or benchmark, see also Canakgoz and Beasley (2008) and references therein. The portfolio obtained this way is sometimes called enhanced indexation portfolio (EI portfolio), and its return in excess to that of the index is called excess return. Note that, in general, no guarantee of always reaching a positive excess return can be given, so the risk of underperforming the index always exists.

Several different approaches to the EI problem, both exact and heuristic, have been proposed in the last decade, starting from a seminal study by Beasley et al. (2003). However, in the literature, there does not seem to be a prevalent mathematical model. We now describe in some detail the more recent approaches to EI, directing the reader to the extensive overview in Canakgoz and Beasley (2008) for what concerns the earlier contributions to the problem, consisting of: Alexander and Dimitriu (2005b, a); Dose and Cincotti (2005); Konno and Hatagi (2005); Wu et al. (2007).

Canakgoz and Beasley (2008) propose a regression based model for EI, developing a two-stage mixed-integer linear programming approach, so that the use of standard and efficient solvers is possible. The first stage consists in achieving a regression slope as close to one as possible, subject to a constraint on the regression intercept (being interpreted as excess return). In the second stage, they minimize transaction costs subject to retaining the value for the slope achieved at the first stage. Computational results are presented for eight publicly available datasets from Beasley's OR-Library. Koshizuka et al. (2009) deal with EI by minimizing the mean absolute deviation between index values plus a factor alpha and the EI portfolio values by imposing a constraint on the correlation between the weights of the selected portfolio and those of the benchmark. They reformulate the problem as a convex minimization problem and test the model on the Tokyo Stock Exchange with around 1,500 assets, considering three non-overlapping time windows where the in-sample window is 3 years and the out-of-sample one is 1 year. They solve the model through an optimization package (NUOPT) but no running times are reported. Roman et al. (2013) apply a second-order stochastic dominance strategy to construct a portfolio whose return distribution dominates the one of a benchmark, see also Fábián et al. (2011); Roman et al. (2006).

Empirical studies are conducted on three datasets (FTSE 100, SP 500, and Nikkei 225) using weekly returns. The authors consider the possibility of rebalancing the portfolio composition each week, for a total of ten weeks, which represents the back-testing sample. They adopt a row generation approach for solving the problem and, for each problem instance, computational times are very encouraging. An alternative approach based on Stochastic Dominance, but aiming at approximately achieving a stronger version of Stochastic Dominance, is presented by Bruni et al. (2012), who also obtain very promising empirical and computational results on the eight publicly available datasets from Beasley's OR-Library. Meade and Beasley (2011) investigate a momentum strategy via the maximization of a modified Sortino ratio objective function: they assume that performance observed in the recent past will continue into the near future. The model is heuristically solved by means of a genetic algorithm. They use S&P Global 1200 and its subsets (Europe, UK, Japan, Australia, Canada, etc.) with their respective market indexes, finding evidence of significant momentum profits. Li et al. (2011) develop a non-linear bi-objective optimization model for EI, where the excess return is maximized and the downside standard deviation (relative to the benchmark index) is minimized. They use as decision variables the number of units of stock, thus requiring the use of integer variables. The model is solved by an evolutionary algorithm on a subset of the datasets used in Canakgoz and Beasley (2008), but no computational times are reported. Thomaidis (2012) proposes a soft computing approach to EI with prefixed investment goals (on the excess return and on the probability that the Enhanced Indexation portfolio underperforms the benchmark) and with risk constraints. The model also includes a cardinality constraint and buy-in thresholds, and it is formulated as a mixed-integer nonlinear programming problem which is solved using simulated annealing, genetic algorithms, and particle swarm optimization. Experimental results are presented for the Dow Jones Industrial Average index with 30 securities, again without reporting computational times. Guastaroba and Speranza (2012) also use a heuristic approach (called Kernel Search) for solving mixed-integer linear programming models for IT and EI that also include cardinality, buy-in, and transaction costs constraints. They evaluate the efficiency and accuracy of their heuristic comparing it with a standard exact solver. Furthermore, they perform experiments on the eight publicly available datasets from Beasley's OR-Library but they provide examples of the out-of-sample performance only on two datasets.

We observe that most existing studies on the EI problem present some limitations. First of all, EI bi-objective models (or their scalarizations) based on minimizing tracking error and maximizing excess return contain a contradiction in their purposes. On one hand, the first goal penalizes both positive and negative deviations from the index while, on the other hand, one seeks to maximize the mean of positive deviations. This contradiction derives from the use of a symmetric distance measure, which in this case is not suitable for controlling the distance between the returns of the portfolio and those of the benchmark, and can be avoided using an asymmetric distance measure. Furthermore, most of the proposed models are computationally demanding, so that they cannot be practically solved to optimality for medium or large size problems, especially if cardinality or other real-word constraints are introduced. Therefore, they are solved only approximately by means of heuristics. Finally, several authors do not give detailed computational results or use datasets not publicly available, thus making computational comparison impossible.

To overcome the above limitations, we propose an innovative approach to EI based on linear bi-objective optimization. More precisely, we maximize the average excess return of the selected portfolios with respect to an index, while minimizing a downside risk measure representing the maximum underperformance with respect to the same index. As customary, we solve the bi-objective problem by converting one (the second) objective into a constraint with a parametric right-hand side. We thus obtain an efficiently solvable parametric Linear Programming model that we test on publicly available datasets. Alternatively, the bi-objective model can be solved by specialized multi-objective linear programming algorithms.

The remainder of the paper is organized as follows: in Sect. 2, we present our parametric LP for the EI problem. We also establish conditions for the existence of a portfolio strictly outperforming the benchmark when the number of assets is greater than the number of time periods. We then show that, when the number of time periods is greater than the number of assets, a classical No Arbitrage condition guarantees that there is no portfolio strictly outperforming the index, and that the only portfolio weakly outperforming the index is the one realizing the index itself. In Sect. 3, we describe the results of an extensive empirical analysis of the behavior of the proposed model. We apply our model to eight major stock markets across the world, using datasets publicly available in Beasley (1990) and already used in some similar studies, and we compare the performance of our portfolios with published results. The same analysis is conducted on two new databases, focusing on more recent periods, that we make publicly available for comparisons and further studies. We also perform a diversification analysis, and an empirical verification of some of the theoretical results on the No Arbitrage condition. Conclusions and further research projects are presented in Sect. 4.

2 A linear risk-return model

Enhanced indexation (EI) models are usually built and validated using the price values of all the assets belonging to a given market and of the corresponding benchmark index over a time interval. To simulate practical usage, a part of this interval is regarded as the past, and so it is known, and the rest is regarded as the future, supposed unknown at the time of portfolio selection. The past (called in-sample window) is used for selecting the EI portfolio, while the future (called out-of-sample window) can only be used for testing the performance of the selected portfolio. Let the in-sample window be constituted by $T+1$ time periods $0, 1, 2, \dots, T$. We use the following notation:
p_{it} is the price of the ith asset at time t, with $t = 0, \dots, T$;
b_t is the benchmark index value at time t, with $t = 0, \dots, T$;
$r_t^I = \dfrac{b_t - b_{t-1}}{b_{t-1}}$ is the benchmark index return at time t, with $t = 1, \dots, T$;
$r_{it} = \dfrac{p_{it} - p_{i(t-1)}}{p_{i(t-1)}}$ is the ith asset return at time t, with $t = 1, \dots, T$;

x is the vector whose components x_i are the fractions of a given capital invested in asset i in the EI portfolio we are selecting.

Adopting linear returns, we have that

$R_t(x) = \sum_{i=1}^{n} x_i r_{it}$ is the portfolio return at time t, with $t = 1, \ldots, T$;

$\delta_t(x) = R_t(x) - r_t^I$ is the excess return, or overperformance, of the selected portfolio w.r.t. the benchmark index at time t, with $t = 1, \ldots, T$. Note that $-\delta_t(x)$ is the underperformance of the selected portfolio w.r.t. the benchmark index at time t.

Following a classical paradigm, we would like to maximize return and, at the same time, minimize risk. Thus, we propose a linear bi-objective risk-return model where the objectives are

(a) the maximization of the (average) excess return of the selected portfolio w.r.t. the benchmark index: $\max_x \frac{1}{T} \sum_{t=1}^{T} \left(\sum_{i=1}^{n} x_i r_{it} - r_t^I \right) = \max_x \frac{1}{T} \sum_{t=1}^{T} \delta_t(x)$;

(b) the minimization of the downside risk defined as the maximum underperformance w.r.t. the same index: $\min_x \max_t -\delta_t(x)$.

Note that a negative [resp. positive] value of objective (b) corresponds to a positive [resp. negative] excess return. All efficient solutions of this bi-objective problem can be found by solving a family of single objective problems depending on a parameter K (here called risk level) that specifies the maximum allowed risk (in the sense of underperformance). Therefore, we need to solve the following problem for all values of K between two extreme risk levels $K_{\min}$ and $K_{\max}$, that will be determined later in this Section.

$$
\begin{aligned}
\phi(K) = \max_x \ & \frac{1}{T} \sum_{t=1}^{T} \delta_t(x) \\
s.t. \ & -\delta_t(x) \leq K \qquad t = 1, \ldots, T \\
& \sum_{i=1}^{n} x_i = 1 \\
& x_i \geq 0 \qquad\qquad i = 1, \ldots, n
\end{aligned}
\tag{1}
$$

2.1 No Arbitrage and the minimum risk portfolio

Solving model (1) for (small) positive values of K allows to construct portfolios that might have (small) underperformances in the in-sample window.

On the other hand, solving the same model for negative values of K yields only portfolios that strictly overperform the index in all the in-sample periods. However, requesting too small values for K may produce infeasibilty in the constraints of (1). The minimum feasible value of the risk level K in model (1) can be found by solving the problem:

$$
\begin{aligned}
K_{\min} = \min_{x,K} \ & K \\
s.t. \ & -\delta_t(x) \leq K \qquad t = 1, \ldots, T \\
& \sum_{i=1}^{n} x_i = 1 \\
& x_i \geq 0 \qquad\qquad i = 1, \ldots, n
\end{aligned}
\tag{2}
$$

The optimal solution to this problem yields the minimum risk portfolio. Note that the optimal value $K_{\min}$ of problem (2) is non-positive if the EI portfolio never underperforms the index. It can be easily guessed that this is not always possible. However, we can devise conditions under which the optimal portfolio selected in (2) is guaranteed to strictly overperform the index in all the in-sample periods, as proved in the following Theorem 1 (see also Sect. 3.5).

Let $R^I = (r_1^I, \dots, r_T^I)$ be the vector of the index returns and let $R^i = (r_{i1}, \dots, r_{iT})$ denote the vector of returns of asset i, for $i = 1, \dots, n$. We say that the index returns are realizable by a portfolio when there exists an $\hat{x}$ with $\sum_i \hat{x}_i = 1$ such that $R^I = \sum_i \hat{x}_i R^i$. We also say that the index returns are realizable by a complete portfolio when there exists an $\tilde{x}$ with $\sum_i \tilde{x}_i = 1$ such that $\tilde{x}_1 > 0, \dots, \tilde{x}_n > 0$ and $R^I = \sum_i \tilde{x}_i R^i$.

We recall that points $v^1, \dots, v^m$ in $\mathbb{R}^T$ are said to be affinely independent if $\lambda_1, \dots, \lambda_m \in \mathbb{R}$, $\sum_{i=1}^m \lambda_i = 0$, and $\sum_{i=1}^m \lambda_i v^i = 0$ imply $\lambda_1 = \lambda_2 = \dots = \lambda_m = 0$. This is equivalent to requiring that the convex hull of these points is a polytope of dimension $m - 1$. Note that $m \leq T + 1$ randomly chosen points in $\mathbb{R}^T$ are affinely independent with probability 1. We also recall that the open mapping theorem states that the images of open sets through a surjective linear mapping are open.

Theorem 1 *Assume that $T < n$, that among the vectors $R^1, \dots, R^n$ of the assets returns there are $T + 1$ affinely independent vectors, and that the index returns $R^I = (r_1^I, \dots, r_T^I)$ are realizable by a complete portfolio. Then, there exists a portfolio that strictly overperforms the index in all the in-sample periods, i.e., $\delta_t(x) = R_t(x) - r_t^I > 0$ for $t = 1, \dots, T$.*

Proof Let $\Delta = \{x \in \mathbb{R}^n : \sum_{i=1}^n x_i = 1, \ x_i \geq 0, \ i = 1 \dots, n\}$ be the standard simplex in $\mathbb{R}^n$ and let $F : \Delta \to \mathbb{R}^T$ be the linear mapping defined by $F(x) = \sum_{i=1}^n x_i R^i$. By assumption, we have that $R^I = F(\tilde{x}) = \sum_i \tilde{x}_i R^i$ for some $\tilde{x}$ in the interior of Δ. By the open mapping theorem, we then deduce that R^I belongs to the interior of $F(\Delta)$, which is the bounded polyhedron obtained as the convex hull of the points $R^1, \dots, R^n$. Since among these points there are $T + 1$ affinely independent vectors, we have that $F(\Delta)$ is a full-dimensional polyhedron, so that the ball $B(R^I, \epsilon) = \{y \in \mathbb{R}^T : \| R^I - y \| \leq \epsilon\}$ is contained in $F(\Delta)$ for some $\epsilon > 0$. Thus, in particular, the point $R_\epsilon^I = (r_1^I + \epsilon, \dots, r_T^I + \epsilon)$ belongs to $F(\Delta)$, so that there exists $(x_1', \dots, x_n') \in \Delta$ with $F(x_1', \dots, x_n') = R_\epsilon^I$. In other words, the entries of R_ϵ^I are the returns of the feasible portfolio determined by the investments $(x_1', \dots, x_n')$. This portfolio clearly outperforms the index in all the in-sample periods.

We recall that an arbitrage is a transaction that involves no negative cash flow at any probabilistic or temporal state and a positive cash flow in at least one state; in simple terms, it is the possibility of a risk-free profit at zero cost. It is well known that arbitrage opportunities, when arising, do not usually exist for long, due to the markets' natural evolution. Therefore, absence of arbitrage is a common assumption in financial markets and is a basic assumption in asset pricing theory in the context of scenario trees, see, e.g., Duffie (2010); Geyer et al. (2010); Klaassen (1998). In the framework of portfolio optimization, a similar No Arbitrage (NA) condition, see, e.g., Prisman (1986), requires that there exists no long-short portfolio $y = (y_1, \dots, y_n)$, where y_i denotes the amount of asset i purchased (if $y_i > 0$) or shorted (if $y_i < 0$), that gives a positive profit at time 0, i.e., satisfies $\sum_{i=1}^{n} y_i p_{i0} < 0$, and yields nonnegative returns for all periods, i.e., satisfies $\sum_{i=1}^{n} y_i r_{it} \geq 0$, for all $t = 1, \ldots, T$. A stronger version of the No Arbitrage condition requires in addition that every self-financing portfolio (i.e., such that $\sum_{i=1}^{n} y_i p_{i0} = 0$) that yields nonnegative returns for all periods must actually yield zero returns in all periods, i.e., $\sum_{i=1}^{n} y_i r_{it} = 0$, for all $t = 1, \ldots, T$. We note that a risk-less asset with risk-free rate ρ can be easily included in the market by setting $r_{it} = \rho$ for all $t = 1, \ldots, T$.

We now show that, under some technical assumptions, typically verified in practice by the matrix R of returns (i.e., the matrix whose columns are the vectors R^i, $i = 1, \ldots, n$), the strong No Arbitrage condition implies that the only portfolio that weakly outperforms the index in all the in-sample periods is the one realizing the index.

Theorem 2 *Assume that the returns matrix R has full column rank, that the index returns $R^I = (r_1^I, \ldots, r_T^I)$ are realizable by a portfolio, and that the strong No Arbitrage condition holds. Then, the only portfolio that weakly outperforms the index in all the in-sample periods is the one realizing the index.*

Proof Let $\tilde{x} \in \Delta$ be a portfolio realizing the index, i.e., such that $R^I = \sum_i \tilde{x}_i R^i$ and assume that there exists a portfolio $x \in \Delta$ that outperforms the index in all the in-sample periods, i.e., such that $\sum_i x_i R^i \geq R^I$ or, equivalently, $R(x - \tilde{x}) = \sum_i (x_i - \tilde{x}_i) R^i \geq 0$. Observe that the (long-short) portfolio $y = x - \tilde{x}$ is self-financing since $\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \tilde{x}_i = 1 - 1 = 0$. Then, by the strong No Arbitrage condition, we must have $R(x - \tilde{x}) = 0$ which implies $x = \tilde{x}$ by the assumption of linear independence of the columns of R.

Note that the assumption that the index returns $R^I = (r_1^I, \ldots, r_T^I)$ are realizable by a portfolio is trivially verified if we assume, as it is often the case, that the index itself is an available asset in the market.

An immediate consequence of Theorems 1 and 2 is that arbitrage must be possible under the assumptions of Theorem 1. So, for instance, in order to assume a No Arbitrage condition one should assume that $T \geq n$. Furthermore, one can observe that, before knowing the actual in-sample values, under very mild probabilistic assumptions, obtaining a negative value for $K_{\min}$ becomes quite unlikely when increasing the number T of observations. Indeed, if we assume that any portfolio x has a positive probability ϵ of underperforming the index in any period t, then the probability of finding a portfolio that overperforms the index in all past observations is given by $(1 - \epsilon)^T$, which rapidly converges to zero as T increases. It is also straightforward to observe that the value of $K_{\min}$ is nondecreasing with respect to T, since increasing the in-sample window can never decrease the worst underperformance $K_{\min}$. Some computational experiments, described in the next section, show that for the datasets considered there seems to be a threshold $T \simeq 1.7n$ below which $K_{\min} < 0$. As a conclusion, both theory and empirical evidence show that when the number of assets in a market is much smaller than the number of observation periods, it is very unlikely to find a portfolio strictly overperforming the index in all observation periods.

2.1.1 The maximum excess return portfolio

The other extreme case in our bi-objective model consists in maximizing excess return regardless of the underperformance risk. Finding the portfolio having the maximum excess return is formulated as:

$$\begin{aligned} \delta_{\max} = \max_x \ & \frac{1}{T}\sum_{t=1}^{T}\delta_t(x) \\ s.t. \ & \sum_{i=1}^{n} x_i = 1 \\ & x_i \geq 0 \qquad i = 1, \ldots, n \end{aligned} \tag{3}$$

Let I^* be the set of all indices i^* of the assets with maximum average return, i.e., such that $\sum_{t=1}^{T} r_{i^*t} \geq \sum_{t=1}^{T} r_{it}$ for all i. Then, it is straightforward to show that the set of all solutions to problem (3) coincides with the set of all portfolios containing only assets with indices in I^*. Note that the solution of problem (3), while providing the maximum excess return, also bears the maximum risk among the efficient portfolios. Hence the maximum value $K_{\max}$ of the downside risk of the portfolios on the efficient frontier of our bi-objective model is given by

$$K_{\max} = \min_{i^* \in I^*} \max_{1 \leq t \leq T} (R_t^I - r_{i^*t}),$$

while $\delta_{\max} = \frac{1}{T}\sum_{t=1}^{T}(r_{i^*t} - R_t^I)$ is the value of the maximum average excess return with respect to the benchmark.

2.2 Properties of the efficient frontier

Recall that our model is intrinsically bi-objective, in the sense that we always pursue return maximization while controlling risk. Therefore, our EI problem does not have one or few isolated optimal solutions, but an entire set of efficient (Pareto-optimal) solutions, that we are guaranteed to find when solving to optimality our linear model (1) for the different values of the risk level K. More precisely, for every value of K between $K_{\min}$ and $K_{\max}$, the optimal solution to (1) provides a portfolio on the risk-return efficient frontier with optimal average excess return $\phi(K)$, while for $K < K_{\min}$ problem (1) is infeasible, and for $K > K_{\max}$ the optimal solution coincides with the one for $K = K_{\max}$.

The risk-return efficient frontier is thus obtained as the graph of the function $\phi(K)$ on the interval $[K_{\min}, K_{\max}]$. Some theoretical properties of the function $\phi(K)$ are easily derived from known results in parametric linear programming.

Theorem 3 *The function $\phi(K)$ is piecewise linear, concave and increasing on the interval $[K_{\min}, K_{\max}]$.*

For a proof in the general case see, e.g., Murty (1983). From this result, or from similar results in multi-objective linear programming (Löhne 2011; Ruszczynski and

Vanderbei 2003), it follows that the whole efficient frontier of our bi-objective EI model is completely determined by the breakpoints of the piecewise linear function $\phi(K)$, which are also called extreme efficient points of the bi-objective problem.

3 Empirical analysis

In this section, we test our model on some new and on some well-known publicly available datasets. We first adopt a single period approach (SP), i.e., we consider a single in-sample window and a single subsequent out-of-sample window; then, we use a rolling time window approach (RTW), i.e., we shift the in-sample window (and consequently the out-of-sample window) all over the time length of each dataset.

With the SP approach, we perform a partial comparison of our model with two recent EI techniques that use the same datasets: Canakgoz and Beasley (2008) and Guastaroba and Speranza (2012). However, the three models considered are quite different, so there is no direct correspondence between the parameters used. Nevertheless, such a comparison is performed by putting, as far as possible, these models into the same working conditions so as to provide a reasonable idea of their performances.

We observe that in the real world, the markets are in continuous evolution. Hence, it is desirable to rebalance the portfolio from time to time to take new information into account. An RTW approach allows this rebalancing, thus capturing non-stationary market conditions and is, therefore, better suited for practical application. For these reasons, extensive results on all datasets are reported with this approach.

Furthermore, we empirically test the theoretical properties discussed in Sect. 2, and we analyze the diversification of the obtained portfolios.

3.1 Data sets

We support the view of Canakgoz and Beasley (2008) that researchers should try to compare their models on a sufficiently large number of datasets, which should also be (or should be made) publicly available. This would greatly simplify the evaluation of the quality of the proposed models. For this reason, we conduct an extensive analysis on the eight real-world datasets (Beasley 1990), used in other studies on portfolio management, that are available at http://people.brunel.ac.uk/~mastjjb/jeb/orlib/indtrackinfo.html.

Those datasets consist of weakly price data from March 1992 to September 1997 (i.e., 291 historical realizations). In addition to that, we considered two recent datasets, one obtained from Yahoo Finance and consisting of weekly price data from January 2007 to May 2013; the other one from Thomson Reuters Datastream service and consisting of weekly price data from July 2005 to June 2014. We have made them publicly available on the web site http://host.uniroma3.it/docenti/cesarone/DataSets.htm

- Hang Seng (Hong Kong), file indtrack1, containing 31 assets;
- DAX 100 (Germany), file indtrack2, containing 85 assets;
- FTSE 100 (UK), file indtrack3, containing 89 assets;

- S & P 100 (USA), file indtrack4, containing 98 assets;
- Nikkei 225 (Japan), file indtrack5, containing 225 assets;
- S & P 500 (USA), file indtrack6, containing 457 assets;
- Russell 2000 (USA), file indtrack7, containing 1318 assets;
- Russell 3000 (USA), file indtrack8, containing 2151 assets.
- FTSE 100 b (UK), containing 63 assets.
- Eurostoxx 50 (EU), containing 50 assets.

All the return rates have been computed as relative variations of the quotation prices $(P_t - P_{t-1})/P_{t-1}$.

3.2 Single period performance evaluation

For the above data sets, we compute the portfolio that gives the best excess return for a given risk level in the in-sample window, and we then analyze (in terms of excess return) the performance of the chosen portfolio in the out-of-sample periods. Sample intervals are set to allow comparison with Canakgoz and Beasley (2008) and Guastaroba and Speranza (2012).

Table 1 contains the first comparison. 'Best C&B' reports the best yearly (out-of-sample) percentage Average Excess Return (AER) obtained from Table 5 of Canakgoz and Beasley (2008), and the number of assets in the corresponding portfolio. 'Fixed cardinality' reports the same information for portfolios obtained by solving model (1) on the in-sample window [1, 145] and with out-of-sample window [146, 290], as in Canakgoz and Beasley (2008), and by choosing the risk level K that produces a portfolio composed by the same number of assets of Canakgoz and Beasley (2008) (making portfolios as comparable as possible). 'Bounded cardinality' reports again the best AER obtained using model (1) on the same sample intervals, but this time choosing portfolios composed by a number of assets ranging between 5 and 10.

Observe that our portfolios provide higher AER than 'Best C&B' in half of the cases when they include the same number of assets, and in 6 out of 8 cases when we choose fewer assets. To evaluate the magnitude of the AER differences between the considered approaches, we also compute the average over all datasets of the AER (even though it has no direct financial interpretation). We observe that our EI portfolios, in particular those with few assets, have a better behavior.

Table 1 also presents average computational times for solving a single portfolio selection problem. Times are in seconds, rounded to the second decimal. Since times are the same in the cases of 'Fixed cardinality' and 'Bounded cardinality', only one time column is reported in the table. Times reported for the 'C&B' approach are taken from Canakgoz and Beasley (2008), so a direct time comparison is not possible because the procedures ran on different machines. However, they are provided to demonstrate the clear practical tractability of the proposed linear programming model.

On the other hand, Guastaroba and Speranza (2012) do not report a table, but two graphs showing out-of-sample portfolio returns on the FTSE 100 and on the S&P 100 for the portfolios obtained with their EI approach. More precisely, the graphs show the cumulative returns of the portfolios, which correspond to the values of wealth after τ periods (Rachev et al. 2004), given by

Table 1 Average excess return (AER) comparison with portfolios from Canakgoz and Beasley (2008)

	Best C&B			Our Portfolios with:				
				Fixed cardinality		Bounded cardinality		
	Selected assets	AER	Time	Selected Assets	AER	Selected assets	AER	Time
Hang Seng	10	−2.43	0.05	10	−2.51	10	−2.51	0.01
DAX 100	10	11.88	0.07	10	15.88	5	16.40	0.01
FTSE 100	10	5.01	0.07	10	10.39	5	12.59	0.01
S&P 100	10	2.30	0.05	10	11.30	5	19.39	0.01
Nikkei	10	7.81	0.15	10	4.00	7	5.62	0.01
S&P 500	40	14.78	0.35	40	11.30	8	21.84	0.01
Russell 2000	70	12.19	3.13	70	16.92	5	52.63	0.03
Russell 3000	90	22.62	6.77	90	18.59	5	69.98	0.06
Average		9.27			10.73		24.49	

$$W_\tau = W_{\tau-1}(1 + R_\tau(x)) \qquad \tau = 1, \ldots, 52$$

with initial wealth $W_0 = 1$. Thus, we present the outcome of our model for those two datasets in the same format.

Figures 1 and 2 are obtained by solving model (1) on the in-sample window [1, 104] and with out-of-sample window [105, 156], as in Guastaroba and Speranza (2012). No useful indication is given there for the choice of our risk level K. In the absence of this, we assume portfolios in Guastaroba and Speranza (2012) to be quite low-risk portfolios because they closely track the index and are obtained by imposing narrow constraints on the amount of each asset ($0.01 \leq x_i \leq 0.1$ if the ith asset is selected in the portfolio). We, therefore, show results corresponding to our minimum risk EI portfolios compared to the performance of the market index in the same period. The graphs show a slight overperformance of the EI portfolios with respect to the market index similar to that described in Guastaroba and Speranza (2012). Furthermore, in Fig. 3, we report a similar experiment (same sample windows and risk level) on the largest dataset (Russel 3000). We also observe that a good cumulative excess return is achieved in all the out-of-sample periods. We remark, however, that the best performances obtained by our model generally correspond to higher risk levels, as reported in the following subsection.

3.3 Solving the original Bi-objective model

As known, bi-objective problems usually do not have a single optimal solution, but rather a set of non-dominated Pareto-optimal solutions constituting the so-called efficient frontier. The choice among those solutions, each of which provides a portfolio in our case, is in practice delegated to the preference of the decision maker. Depending on the level of risk aversion, a value of K can be specified in model (1) and thus

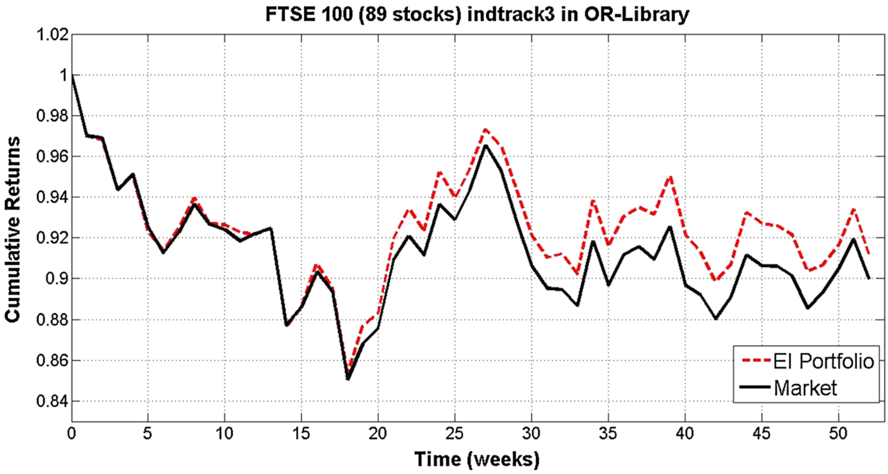


Fig. 1 Out-of-sample performance for FTSE100 (SP approach)

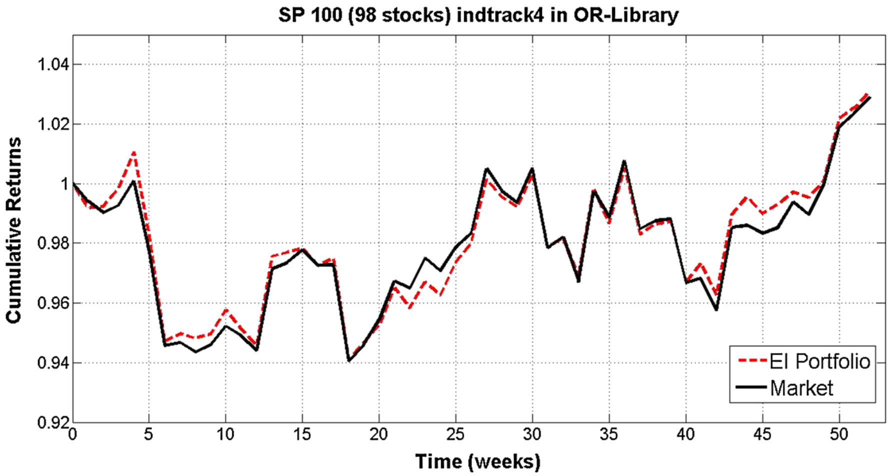


Fig. 2 Out-of-sample performance for S&P100 (SP approach)

one obtains the corresponding portfolio on the efficient frontier. By considering the in-sample window [1, 145] for each dataset of Beasley's OR-Library, we have computed the efficient frontier, that in our case is composed by a piecewise concave line having a large number of breakpoints (corresponding to extreme efficient solutions), as reported in the following Table 2. However, we should point out that the results are rather sensitive to the numerical precision adopted.

Figure 4 reports a graphical analysis of the solutions to the original bi-objective problems using two different approaches: the Benson multiobjective linear programming algorithm (Benson 1998) implemented in Bensolve (Hamel et al. 2014; Löhne 2011), for which the breakpoints of the efficient frontier are the blue circles, and a procedure based on parametric linear programming, also described in Ruszczynski and

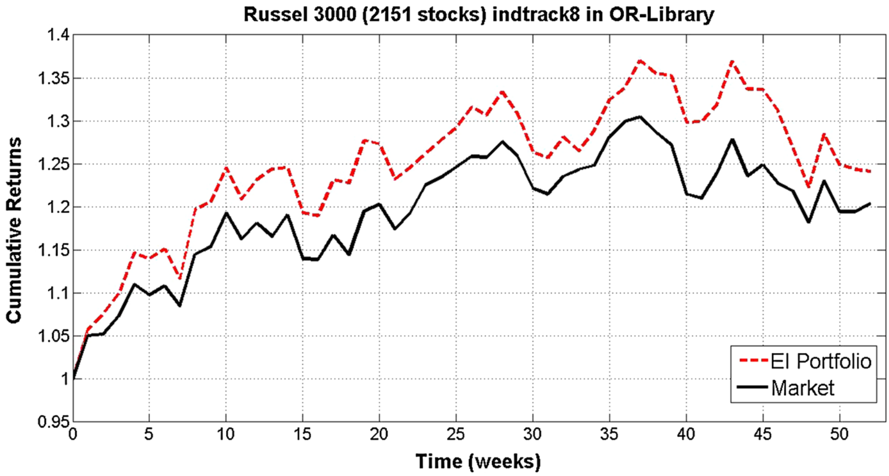


Fig. 3 Out-of-sample performance for Russel 3000 (SP approach)

Table 2 Breakpoints of the efficient frontier

	Assets	Breakpoints
Hang Seng	31	107
DAX 100	85	268
FTSE 100	89	273
S&P100	98	328
Nikkei	225	480
S&P500	457	444
Russell 2000	1318	771
Russell 3000	2151	797

Vanderbei (2003), which very efficiently finds the same breakpoints, up to numerical precision, represented in the figure as red dots. In Fig. 4, we also plotted a dashed black line representing the approximate efficient frontier obtained by repeatedly solving the scalarized LP (1) for a fixed number of equally spaced values of the RHS K between $K_{\min}$ and $K_{\max}$, thus requiring a controllable computational effort. We note that all the methods provide a good approximation of the efficient frontier. However, the first two methods also yield an implicit representation of the whole efficient frontier as the unique piecewise linear concave function passing through the extreme efficient points found.

3.4 Rolling time window evaluation

We now allow for the possibility of changing (rebalancing) the portfolio composition during the holding period. Clearly, frequent rebalances would be useful from the optimization point of view, but would also be practically infeasible because of the

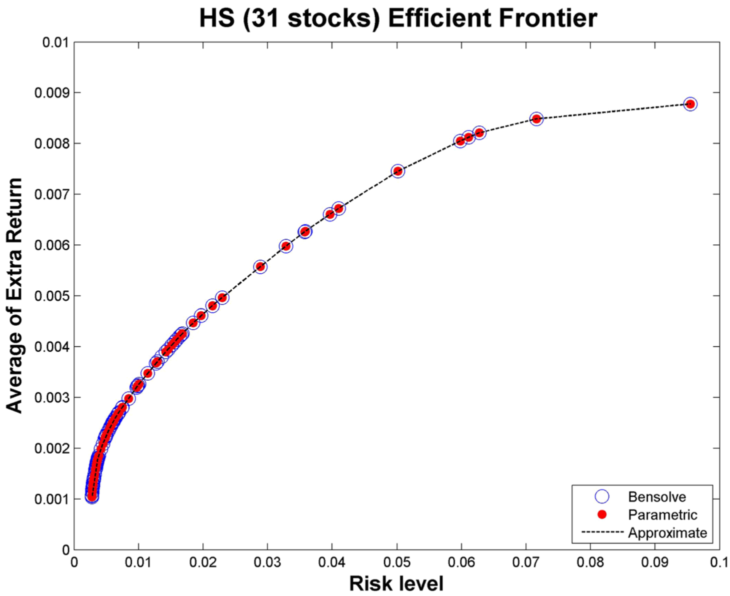


Fig. 4 Breakpoints of the efficient frontier for Hang Seng 31

implied transaction costs. Thus, although we do not explicitly take them into account in our model, we have chosen a holding period that represents a compromise between the two requirements above.

We obtain EI portfolios by solving model (1) for different values of the risk level K on in-sample windows of 200 periods repeatedly shifted all over the dataset. More precisely, for each of those in-sample windows, we evaluate the portfolio performance in the following 4 weeks (out-of-sample window), during which no rebalances are allowed. After this, we shift the mentioned in-sample window by 4 weeks to cover the out-of-sample period, we recompute the optimal portfolio w.r.t. the new in-sample window and repeat. We thus obtain 22 different EI portfolios. For instance, the first in-sample window is [1, 200] and the corresponding out-of-sample window is [201, 204], the second in-sample window is [5, 204] and the corresponding out-of-sample window is [205, 208], and so forth.

As noted in the previous section, the choice of the risk level depends on the risk aversion of the decision maker. Since it is unrealistic to analyze the performances of the efficient portfolios corresponding to a large number of risk levels, we consider only two typical decision makers: the completely risk adverse investor and a moderately risk adverse one. Therefore, we consider the following corresponding lowest and moderate risk level values:

$$K_1 = K_{\min} \qquad K_2 = K_{\min} + 1/4(K_{\max} - K_{\min}).$$

First, in Table 3, we report the average out-of-sample returns of the EI portfolios compared to the corresponding average returns of the market index. Best results for

Table 3 Out-of-sample average returns of the EI portfolios and of the market index

	Assets	$K_1(\times 10^{-2})$	$K_2(\times 10^{-2})$	Market $(\times 10^{-2})$
Hang Seng	31	0.469	**0.613**	0.456
DAX 100	85	0.567	**0.852**	0.631
FTSE 100	89	0.368	**0.486**	0.357
S&P 100	98	0.501	**0.700**	0.510
Nikkei	225	−0.049	−0.130	**−0.042**
S&P 500	457	**−0.210**	−0.893	−0.316
Russell 2000	1, 318	0.175	**0.567**	−0.004
Russell 3000	2, 151	−0.069	**0.602**	−0.297
FTSE 100 b	63	**0.282**	0.121	0.144
Eurostoxx 50	50	0.229	**0.433**	0.153

Bold values indicate best results

Table 4 Out-of-sample Sharpe ratio values of the EI portfolios and of the market index

	Assets	K_1	K_2	Market
Hang Seng	31	0.178	**0.186**	0.170
DAX 100	85	**0.314**	0.273	0.302
FTSE 100	89	**0.236**	0.221	0.222
S&P 100	98	**0.250**	0.226	0.247
Nikkei	225	–	–	–
S&P 500	457	–	–	–
Russell 2000	1, 318	0.049	**0.106**	–
Russell 3000	2, 151	–	**0.106**	–
FTSE 100 b	63	**0.094**	0.026	0.057
Eurostoxx 50	50	0.080	**0.138**	0.052

Bold values indicate best results

each dataset are marked in bold. Observe that the best of the two EI portfolios outperforms the market index in 9 out of 10 cases, and each of the two strategies K_1 and K_2 provides portfolios that outperform the market index in 7 out of 10 cases.

Then, in Tables 4 and 5, we report the outcomes of two standard performance measures: the Sharpe Ratio (Sharpe 1966, 1994) and the Rachev Ratio (Rachev et al. 2004). The Sharpe Ratio is the ratio between the expected return of a portfolio x and its standard deviation, namely $P_s = E[R(x)]/\sigma(R(x))$. However, when the expected return is negative this index has no meaning, so we report "-". The Rachev Ratio is defined as the ratio between the average of the best $\beta\%$ returns of a portfolio and that of the worst $\alpha\%$ returns. Parameters α and β have been set equal to 10. Sharpe and Rachev ratios were also selected because these synthetic indexes are somehow complementary: while the former is more focused to describe the central part of the portfolio return distribution, the latter stresses its behavior on the tails. Other commonly used performance indices, such as the Sortino ratio (Sortino and Satchell

Table 5 Out-of-sample Rachev ratio values of the EI portfolios and of the market index

	Assets	K_1	K_2	Market
Hang Seng	31	1.082	**1.280**	1.041
DAX 100	85	**1.408**	1.268	1.171
FTSE 100	89	1.233	1.065	**1.264**
S&P 100	98	1.492	**1.539**	1.510
Nikkei	225	0.932	0.847	**0.938**
S&P 500	457	**1.023**	0.910	0.920
Russell 2000	1, 318	0.919	**1.096**	0.902
Russell 3000	2, 151	**1.055**	0.933	0.889
FTSE 100 b	63	**1.070**	0.855	0.933
Eurostoxx 50	50	**1.098**	1.075	1.054

Bold values indicate best results

2001) etc., have been also computed and their results (not shown here but available upon request) turned out to be similar to the reported ones.

Table 4 reports the Sharpe ratio values for the EI portfolios and for the market index. Best results for each datasets are marked in bold. In this case, results obtained by the EI portfolios are always better than the benchmark. Table 5 reports the Rachev Ratio values for the EI portfolios and for the market index. Best results for each datasets are marked in bold. In 8 out of 10 cases, the EI portfolios outperform the market index.

In addition to the above performance evaluation, we also analyze the capability of tracking the market index for the EI portfolios found with our model. This is done by evaluating, for each dataset, the correlation and the average difference between the out-of-sample returns of the portfolios and of the market index, as reported in Table 6. The generally high-correlation values and the very low-average differences show that the EI portfolios are able to replicate the index trend while they try to overperform it. Note that the downside deviations w.r.t. the market index are usually smaller than the upside ones, as it is indeed our purpose, and as required by model (1).

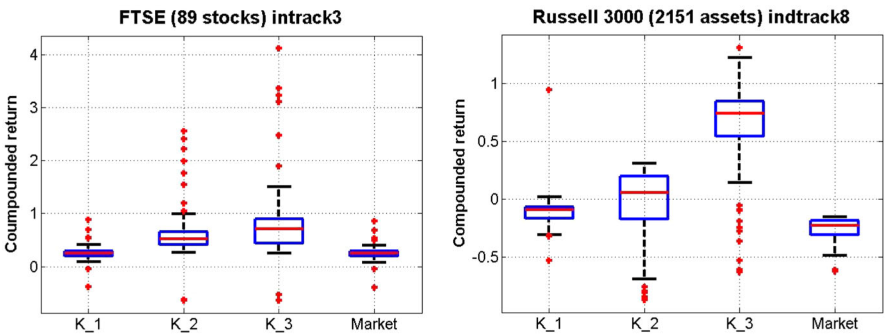
To better understand the behavior of our model, we compute the yearly compounded out-of-sample return CR_τ (after τ periods) of the 22 EI portfolios in the following way:

$$CR_\tau = \left[\prod_{t=1}^{\tau}(1 + R_t(x))\right]^{\frac{52}{\tau}} - 1 \qquad \tau = 1, \ldots, 88$$

where $R_t(x)$ is the t th value of the 88 weekly out-of-sample returns (4 values for each of the 22 out-of-sample windows) of the EI portfolios. The following analysis is performed considering 3 different risk levels: the minimum and moderate levels K_1 and K_2, defined as above, and a medium level $K_3 = K_{\min} + 1/2(K_{\max} - K_{\min})$. As an example, we provide the box plots of results for the FTSE 100 (Fig. 5a) and for the Russell 3000 (Fig. 5b) datasets. In the figures, each box represents the yearly compounded return distribution; the central mark is the median and the edges are

Table 6 Out-of-sample index tracking capability

	Assets	K_1		K_2	
		Correlation (%)	Ave. diff.	Correlation (%)	Ave. diff.
Hang Seng	31	99.6	0.00013	82.6	0.00157
DAX 100	85	75.6	−0.00064	36.4	0.00221
FTSE 100	89	98.9	0.00011	77.8	0.00129
S&P100	98	99.5	−0.00009	76.8	0.00190
Nikkei	225	99.6	−0.00091	70.3	−0.00172
S&P500	457	98.3	0.00106	79.1	−0.00577
Russell 2000	1, 318	95.4	0.00179	69.5	0.00571
Russell 3000	2, 151	97.0	0.00228	68.4	0.00899
FTSE 100 b	63	81.9	0.00138	62.7	−0.00023
Eurostoxx 50	50	99.0	0.00076	79.5	0.00280



(a) FTSE 100 **(b)** Russell 3000

Fig. 5 Box plot of yearly compounded return

the 25th and the 75th percentiles, the whiskers correspond to approximately ±2.7 times the standard deviation, and the outliers are represented individually. The yearly compounded return distribution of the EI portfolios with minimum risk level tends to be similar to that of the market index, while for higher risk levels the EI portfolios seem preferable. Note that this happens also for the FTSE 100, where the market index was preferable according to the Rachev Ratio.

3.5 Analysis of minimum risk portfolios

To analyze the theoretical results presented in Sect. 2, and the corresponding assumptions, we compute the minimum risk level (or maximum allowed underperformance) $K_{\min}$ for the well-known Beasley's datasets described above. We then examine the sign of $K_{\min}$ with respect to the value of the ratios between the number T of in-sample observations and the number n of assets. Specifically, $n = \{31, 85, 89, 98, 255, 457,$

1318, 2151}, while T ranges from 10 to 290, which is the maximum number of available observations. Table 7 reports the values of $K_{\min}$ considering the market index as a benchmark. We observe that:

1. $K_{\min}$ is negative when T is smaller, comparable or not much larger than n (approximately $T < 1.7n$);
2. $K_{\min}$ is positive for larger values of T w.r.t. n (the bottom left corner).

The first point agrees with the results of Sect. 2, because it shows that, under the assumptions of Theorem 1, for $T < n$ a minimum risk portfolio strictly outperforming the market index always exists (i.e., arbitrage is possible). On the other hand, for T sufficiently greater than n the returns matrix R is expected to have full column rank. In this case, under the strong No Arbitrage condition, no portfolio that strictly outperforms the index can exist, and if the market index is a realizable portfolio, then it is the optimal portfolio. If, on the contrary, the market index is not realizable, the optimal portfolio necessarily underperfoms the index. Point 2 above shows that the latter case holds here.

We now repeat the experiment using the equally-weighted (or uniform) portfolio as benchmark, namely $R^I_t = \sum_{i=1}^n r_{it}/n$, which is a feasible solution of model (1) and hence a realizable portfolio. Table 8 reports the values of $K_{\min}$ obtained in this case. In all instances, if $T < 1.7n$ then $K_{\min} < 0$. On the other hand, when T is sufficiently large with respect to n (i.e., approximately $T > 1.7n$), $K_{\min}$ becomes zero. This outcome is fully consistent with the results of Sect. 2, at least for the datasets where this can be tested (only the first four datasets satisfy $T > 1.7n$).

3.6 Diversification analysis and cardinality constraints

The diversification of a portfolio is a fundamental property in portfolio management. Diversification should be guaranteed by a good risk-return model, especially for low-risk strategies. For example, in the original Markowitz model (Markowitz 1959), diversification is obtained through variance minimization. On the other hand, another important requirement, specifically in the case of IT and of EI problems, is that of finding a portfolio that uses only a limited number of assets, see, e.g., Cesarone et al. (2013, 2014). This is usually obtained by imposing cardinality constraints that typically require the use of integer variables, thus greatly increasing the computational complexity of the model.

Figure 6a and c show the number of assets having $x_i > 0$ in the EI portfolios obtained solving model (1) with different percentages of the maximum risk level, computed as

$$\tilde{K} = (K - K_{\min})/(K_{\max} - K_{\min}).$$

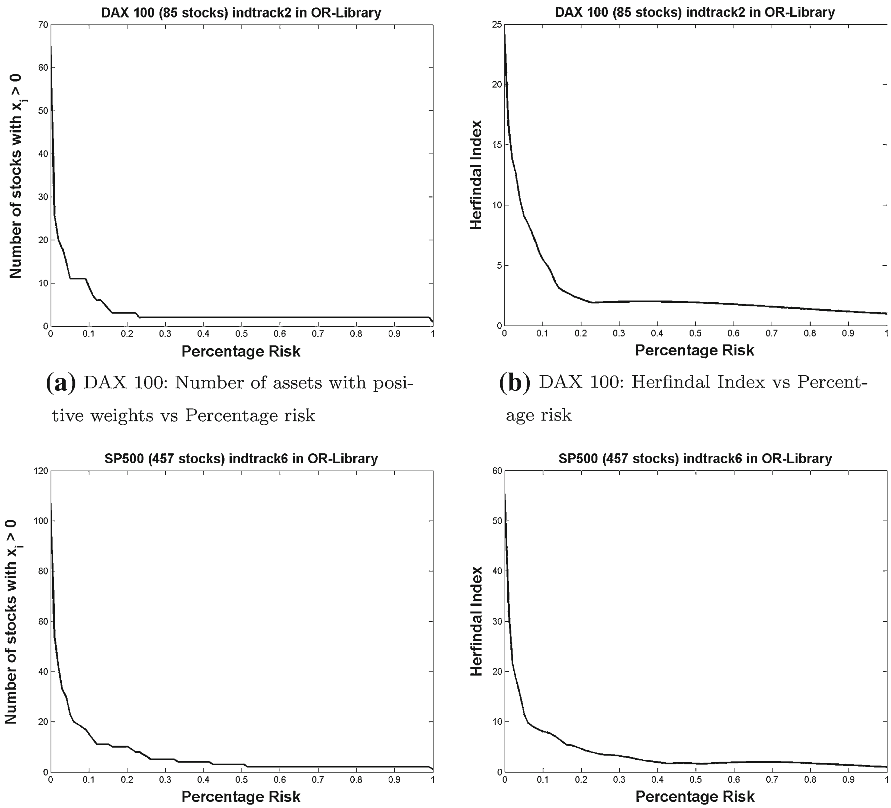
Note that the minimum risk level $K_{\min}$ and the maximum risk level $K_{\max}$ correspond to $\tilde{K}$ values of 0 and 100 %, respectively. The sample datasets considered here are DAX 100 and S&P 500. However, the results are very similar for all other instances. Figure 6b and d also report, for the same datasets, a diversification analysis obtained by computing the Herfindahl Index

Table 7 Minimum risk level $K_{\min}$ with market index as benchmark

T	Hang Seng $n = 31$ $(\times 10^{-2})$	DAX 100 $n = 85$ $(\times 10^{-2})$	FTSE 100 $n = 89$ $(\times 10^{-2})$	S&P 100 $n = 98$ $(\times 10^{-2})$	Nikkei $n = 225$ $(\times 10^{-2})$	S&P 500 $n = 457$ $(\times 10^{-2})$	Russell 2000 $n = 1,318$ $(\times 10^{-2})$	Russell 3000 $n = 2,151$ $(\times 10^{-2})$
10	−0.933	−1.089	−1.822	−1.153	−1.951	−3.644	−7.546	−9.091
30	−0.238	−0.440	−0.549	−0.452	−0.741	−0.846	−2.245	−2.105
50	−0.037	−0.135	−0.266	−0.258	−0.412	−0.725	−1.781	−1.715
70	0.090	−0.059	−0.186	−0.180	−0.245	−0.473	−1.347	−1.155
90	0.098	−0.029	−0.124	−0.115	−0.192	−0.413	−1.183	−0.919
110	0.219	−0.017	−0.086	−0.054	−0.153	−0.365	−1.001	−0.702
130	0.240	−0.003	−0.044	−0.035	−0.114	−0.307	−0.884	−0.613
150	0.278	0.013	−0.017	−0.009	−0.095	−0.261	−0.825	−0.550
170	0.280	0.022	0.013	0.006	−0.074	−0.224	−0.771	−0.511
190	0.284	0.029	0.025	0.024	−0.064	−0.171	−0.699	−0.449
210	0.311	0.034	0.038	0.033	−0.050	−0.150	−0.640	−0.424
230	0.311	0.041	0.045	0.042	−0.041	−0.128	−0.526	−0.396
250	0.313	1.886	0.061	0.085	−0.037	−0.100	−0.468	−0.371
270	0.321	1.905	0.080	0.093	−0.018	−0.087	−0.435	−0.350
290	0.322	2.015	0.119	0.104	−0.003	−0.067	−0.397	−0.327

Table 8 Minimum risk level $K_{\min}$ with uniform portfolio as benchmark

T	Hang Seng	DAX 100	FTSE 100	S&P 100	Nikkei	S&P 500	Russell 2000	Russell 3000
	$n = 31$ $(\times 10^{-2})$	$n = 85$ $(\times 10^{-2})$	$n = 89$ $(\times 10^{-2})$	$n = 98$ $(\times 10^{-2})$	$n = 225$ $(\times 10^{-2})$	$n = 457$ $(\times 10^{-2})$	$n = 1,318$ $(\times 10^{-2})$	$n = 2,151$ $(\times 10^{-2})$
10	−0.813	−1.190	−1.593	−1.318	−1.788	−3.678	−7.482	−8.522
30	−0.176	−0.617	−0.510	−0.482	−0.673	−0.988	−2.127	−2.273
50	−0.039	−0.251	−0.202	−0.272	−0.353	−0.753	−1.651	−1.833
70	0	−0.134	−0.109	−0.209	−0.195	−0.540	−1.219	−1.335
90	0	−0.054	−0.055	−0.110	−0.144	−0.469	−1.054	−1.128
110	0	−0.026	−0.030	−0.049	−0.110	−0.406	−0.872	−0.937
130	0	−0.011	−0.013	−0.023	−0.077	−0.314	−0.737	−0.802
150	0	0	0	−0.010	−0.051	−0.249	−0.668	−0.717
170	0	0	0	0	−0.037	−0.232	−0.644	−0.681
190	0	0	0	0	−0.027	−0.198	−0.588	−0.633
210	0	0	0	0	−0.019	−0.173	−0.526	−0.570
230	0	0	0	0	−0.013	−0.133	−0.447	−0.487
250	0	0	0	0	−0.011	−0.112	−0.405	−0.440
270	0	0	0	0	−0.009	−0.091	−0.371	−0.409
290	0	0	0	0	−0.006	−0.081	−0.335	−0.373



(a) DAX 100: Number of assets with positive weights vs Percentage risk

(b) DAX 100: Herfindal Index vs Percentage risk

(c) S&P 500: Number of assets with positive weights vs Percentage risk

(d) S&P 500: Herfindal Index vs Percentage risk

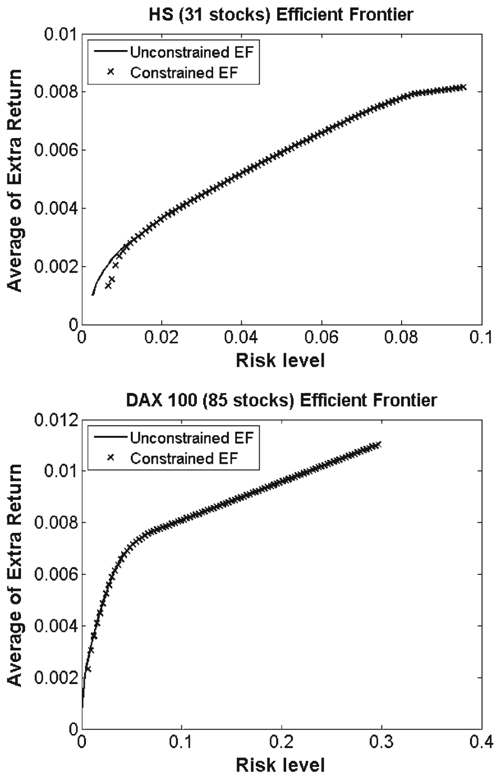
Fig. 6 Diversification analysis

$$H(x) = \left(\sum_{i=1}^{n} x_i^2 \right)^{-1},$$

which is considered a common measure of diversification, see,e.g., Adam et al. (2008). We point out that these two diversification analyses are practically equivalent.

We observe that the EI portfolios obtained with our model are very diversified for small risk values, while the number of assets included in the optimal portfolios becomes rapidly small for slightly larger risk values. This avoids the use of complicating cardinality constraints for constructing the efficient frontiers. However, our model easily allows the introduction of additional real-world constrains. Indeed, the constrained version of model (1) with cardinality constraints and buy-in thresholds can be reformulated as a mixed integer linear program by adding n binary variables y_i:

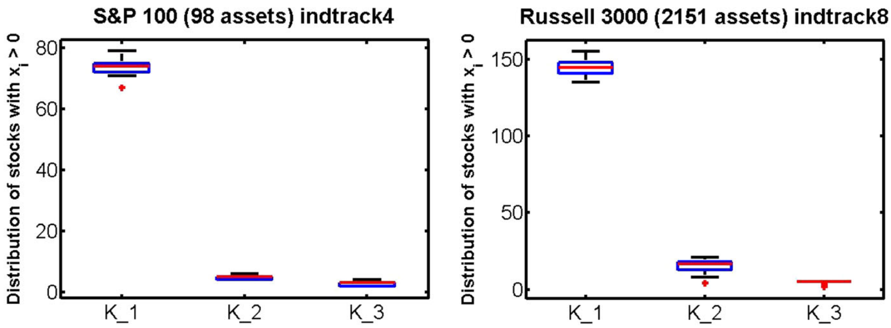
Fig. 7 Examples of cardinality constrained efficient frontiers



$$\begin{array}{lll}\max\limits_{x,y} & \frac{1}{T}\sum\limits_{t=1}^{T}\delta_t(x) & \\ s.t. & -\delta_t(x) \leq K & t=1,\ldots,T \\ & \sum\limits_{i=1}^{n} x_i = 1 & \\ & \ell_i y_i \leq x_i \leq u_i y_i & i=1,\ldots,n \\ & \sum\limits_{i=1}^{n} y_i \leq m & \\ & x_i \geq 0 & i=1,\ldots,n \\ & y_i \in \{0,1\} & i=1,\ldots,n \end{array} \quad (4)$$

For moderate sizes of n, this problem can be solved to optimality by general purpose mixed integer linear programming (MILP) solvers like CPLEX.

Nevertheless, for larger problems, specialized and possibly approximate methods are required. As an example, in Fig. 7, we report the efficient frontiers for the Hang Seng (31 assets) and DAX 100 (85 assets) datasets for 100 equally spaced values of K in $[K_{\min}, K_{\max}]$. This is done both in the unconstrained case and in the cardinality constrained case with at most ten assets ($m = 10$). As expected from previous



(a) S&P 100 **(b)** Russell 3000

Fig. 8 Distributions of the number of selected assets (RTW approach)

diversification analysis, the two efficient frontiers coincide for all but the smallest risk levels.

Models (1) and (4) have been coded in MATLAB 7.11.0 and executed on a PC with Intel Core i3 CPU M330 2.13 G Hz with 4 Gb RAM under MS Windows 7, using the exact solver CPLEX 11.0, which is called from MATLAB with the TOMLAB/CPLEX toolbox (Holmstrom et al. 2007). When solving model (4) for computing the efficient frontiers, computational times are 45 secs for Hang Seng and 958 secs for DAX 100 (see Fig. 7). When solving model (1), on the other hand, the corresponding computations require less than 1 s.

Finally, we analyze the number of selected assets in the RTW approach on different in-sample windows, and again for three risk levels K1, K2, K3. We observe that the number of selected assets is fairly stable in all in-sample windows, as reported in Fig. 8a and in Fig. 8b that show the box plots of the distribution of those number of assets for the datasets S&P 100 and Russel 3000. This analysis confirms that imposing cardinality constraints in our model is necessary only for the smallest risk levels.

4 Conclusions

We proposed a new and efficiently solvable risk-return approach to the Enhanced Indexation problem. In spite of its simplicity, our model is able to find portfolios that exhibit out-of-sample performances that seem comparable or even superior, to those reported in previous works on the same problem. We chose to avoid cluttering the presentation of our model with complicating real-world constraints and also to highlight some theoretical connections between a No Arbitrage condition and the existence of a portfolio outperforming the index. However, the linearity of our model easily allows for the addition of further constraints coming from real-world practice, such as the cardinality constraints and buy-in thresholds mentioned in Sect. 3, or the turn-over or UCITS constraints described in Scozzari et al. (2013). Also, to overcome possible bias due to some inaccuracies contained in several financial datasets, a data cleaning step would be profitable, as described in Bruni (2005). A detailed analysis of the effect of such extensions on our model is left for future research.

References

Adam A, Houkari M, Laurent J (2008) Spectral risk measure and portfolio selection. J Bank Financ 32:1870–1882
Alexander C, Dimitriu A (2005a) Indexing and statistical arbitrage. J Portf Manag 31:50–63
Alexander C, Dimitriu A (2005b) Indexing, cointegration and equity market regimes. Int J Financ Econ 10:213–231
Beasley JE (1990) OR-library: distributing test problems by electronic mail. J Oper Res Soc 41:1069–1072
Beasley JE, Meade N, Chang TJ (2003) An evolutionary heuristic for the index tracking problem. Eur J Oper Res 148:621–643
Benson HP (1998) An outer approximation algorithm for generating all efficient extreme points in the outcome set of a multiple objective linear programming problem. J Glob Optim 13:1–24
Bruni R (2005) Error correction for massive data sets. Optim Method Softw 20:295–314
Bruni R, Cesarone F, Scozzari A, Tardella F (2012) A new stochastic dominance approach to enhanced index tracking problems. Econ Bull 32:3460–3470
Canakgoz NA, Beasley JE (2008) Mixed-integer programming approaches for index tracking and enhanced indexation. Eur J Oper Res 196:384–399
Cesarone F, Scozzari A, Tardella F (2013) A new method for mean-variance portfolio optimization with cardinality constraints. Ann Oper Res 205:213–234
Cesarone F, Scozzari A, Tardella F (2014) Linear vs. quadratic portfolio selection models with hard real-world constraints. Comput Manag Sci (in press). doi:10.1007/s10287-014-0210-1
Dose C, Cincotti S (2005) Clustering of financial time series with application to index and enhanced index tracking portfolio. Physica A 355:145–151
Duffie D (2001) Dynamic asset pricing theory, 3rd edn. Princeton University Press, Princeton NJ. ISBN 069109022
Fábián CI, Mitra G, Roman D, Zverovich V (2011) An enhanced model for portfolio choice with ssd criteria: a constructive approach. Quant Financ 11:1525–1534
Geyer A, Hanke M, Weissensteiner A (2010) No-arbitrage conditions, scenario trees, and multi-asset financial optimization. Eur J Oper Res 206:609–613
Guastaroba G, Speranza MG (2012) Kernel search: an application to the index tracking problem. Eur J Oper Res 217:54–68
Hamel AH, Löhne A, Rudloff B (2014) Benson type algorithms for linear vector optimization and applications. J Glob Optim 59:811–836
Holmstrom K, Goran AO, Edvall MM (2007) Users guide for TOMLAB
Klaassen P (1998) Financial asset-pricing theory and stochastic programming models for asset/liability management: a synthesis. Manag Sci 44:31–48
Konno H, Hatagi T (2005) Index-plus-alpha tracking under concave transaction cost. J Ind Manag Optim 1:87–98
Koshizuka T, Konno H, Yamamoto R (2009) Index-plus-alpha tracking subject to correlation constraint. Int J Optim 1:215–224
Li Q, Sun L, Bao L (2011) Enhanced index tracking based on multi-objective immune algorithm. Expert Syst Appl 38:6101–6106
Löhne A (2011) Vector Optimization with Infimum and Supremum. Springer-Verlag, Berlin. ISBN 978-3-642-18350-8
Markowitz HM (1959) Portfolio selection: efficient diversification of investments. Wiley, New York
Meade N, Beasley JE (2011) Detection of momentum effects using an index out-performance strategy. Quant Financ 11:313–326
Murty KG (1983) Linear programming. Wiley, New York
Prisman EZ (1986) Valuation of risky assets in arbitrage free economies with frictions. J Financ 41:545–557
Rachev S, Biglova A, Ortobelli S, Stoyanov S (2004) Different approaches to risk estimation in portfolio theory. J Portf Manag 31:103–112
Roman D, Darby-Dowman K, Mitra G (2006) Portfolio construction based on stochastic dominance and target return distributions. Math Program 108:541–569
Roman D, Mitra G, Zverovich V (2013) Enhanced indexation based on second-order stochastic dominance. Eur J Oper Res 228:273–281
Ruszczynski A, Vanderbei RJ (2003) Frontiers of stochastically nondominated portfolios. Econometrica 71:1287–1297

Scozzari A, Tardella F, Paterlini S, Krink T (2013) Exact and heuristic approaches for the index tracking problem with UCITS constraints. Ann Oper Res 205:235–250

Sharpe WF (1966) Mutual fund performance. J Bus 39:119–138

Sharpe WF (1994) The sharpe ratio. J Portf Manag 21:49–58

Sortino FA, Satchell S (2001) Managing downside risk in financial markets. Batterworth-Heinemann Finance, Oxford

Thomaidis NS (2012) A soft computing approach to enhanced indexation. In: Natural computing in computational finance, Springer, pp 61–77

Wu LC, Chou SC, Yang CC, Ong CS (2007) Enhanced index investing based on goal programming. J Portf Manag 33:49–56