Accepted Manuscript

Tracking and Outperforming Large Stock-Market Indices

M. Gnägi, O. Strub

PII: S0305-0483(18)30264-0
DOI: https://doi.org/10.1016/j.omega.2018.11.008
Reference: OME 1999

To appear in: *Omega*

Received date: 8 March 2018
Revised date: 28 September 2018
Accepted date: 9 November 2018

Please cite this article as: M. Gnägi, O. Strub, Tracking and Outperforming Large Stock-Market Indices, *Omega* (2018), doi: https://doi.org/10.1016/j.omega.2018.11.008

Highlights

- We study the problem of tracking and outperforming large stock-market indices.
- We compare linear and quadratic objective functions used in the literature.
- We consider various real-life constraints that are relevant in practice.
- We propose novel MIP formulations and novel matheuristics.
- We find that the tracking error variance, a quadratic function, should be optimized.

Tracking and Outperforming Large Stock-Market Indices

M. Gnägi[1,*], O. Strub[1,*]

[1] *Department of Business Administration, University of Bern, Schützenmattstrasse 14, 3012 Bern, Switzerland*

Abstract

Enhanced index-tracking funds aim to achieve a small target excess return over a given financial benchmark index with minimum additional risk relative to this index, i.e., a minimum tracking error. These funds are attractive to investors, especially when the index is large and thus well diversified. We consider the problem of determining a portfolio for an enhanced index-tracking fund that is benchmarked against a large stock-market index subject to real-life constraints that may be imposed by investors, stock exchanges, or investment guidelines. In the literature, various solution approaches have been proposed to enhanced index tracking that are based on different linear and quadratic tracking-error functions. However, it remains an open question which tracking-error function should be minimized to determine good enhanced index-tracking portfolios. Moreover, the existing approaches may neglect real-life constraints such as the minimum trading values imposed by stock exchanges or may not devise good feasible portfolios within a reasonable computational time when the index is large. To overcome these shortcomings, we propose novel mixed-integer linear and quadratic programming formulations and novel matheuristics. To address the open question, we minimize different tracking-error functions by applying the proposed matheuristics and exact solution approaches based on the proposed mixed-integer programming formulations in a computational experiment using a set of problem instances based on large stock-market indices with up to more than 9,000 constituents. The results of our study suggest that minimizing the so-called tracking error variance, which is a quadratic function, is preferable to minimizing other tracking-error functions.

Keywords: Enhanced Index Tracking, Mixed-Integer Programming, Matheuristics

1. Introduction

A stock-market index reflects the overall development of the stocks that constitute that index. Examples of such indices include the Standard & Poor's 500 index, the EURO STOXX 50 index, and the Thomson Reuters Global index, which reflect the development of national, regional, and global stock markets, respectively. Stock-market indices serve as benchmarks for evaluating the performance of professional managers of both active and passive investment funds. A passive fund, also known as an index-tracking fund, aims to replicate the return of an index, whereas an active fund aims to achieve an excess return over its benchmark index.

*Corresponding author
Email addresses: mario.gnaegi@pqm.unibe.ch (M. Gnägi), oliver.strub@pqm.unibe.ch (O. Strub)

Passive funds tend to be less risky and incur lower management costs than active funds (cf. [2]). However, active funds have a higher potential return. Recently, a new type of investment fund has emerged, so-called enhanced index-tracking funds, which are based on the idea of combining the advantages of both active and passive funds by aiming at a small target excess return with minimum additional risk relative to the index, i.e., a minimum tracking error (cf. [8]). Note that we regard index-tracking funds as a special type of enhanced index-tracking funds with a target excess return of zero. Enhanced index-tracking funds are attractive to investors, especially when such a fund is benchmarked against an index that has a large number of constituents and thus is well diversified.

We consider the enhanced index-tracking problem (EITP) faced by the portfolio manager of an enhanced index-tracking fund that is benchmarked against a large stock-market index. In the EITP, the portfolio manager is given the current composition of the index and the current composition of the portfolio, which can consist of stocks from the index and cash. The portfolio manager can receive cash deposits and cash withdrawal requests. The available investment budget consists of the net cash flow from deposits and withdrawals plus the value of the current portfolio. Furthermore, the portfolio manager is given the following data from the past, i.e., the in-sample period: the values of the index, the prices of the stocks that currently constitute the index, and the interest rates on cash. The portfolio manager needs to decide how to revise (rebalance) the current portfolio such that the rebalanced portfolio will exhibit a small tracking error and achieve a given target excess return in the future, i.e., the out-of-sample period. Because future outcomes are not known in advance, the portfolio manager aims to minimize the expected tracking error subject to a constraint that prescribes some minimum expected excess return. When rebalancing the portfolio, the manager must consider a budget constraint that ensures that the investment in the stocks plus the total transaction costs spent for rebalancing do not exceed the investment budget. Furthermore, the portfolio manager must also consider various real-life constraints that may be imposed by investment guidelines, the investors, or stock exchanges. Specifically, we consider the following real-life constraints, which are common both in the literature and in practice (cf., e.g., [8, 13, 30]). The number of stocks included in the portfolio, i.e., the portfolio cardinality, must not exceed a given upper bound because investing in all constituents of a large index would be impractical due to the consequent prohibitive management costs. The trading value of each traded stock and the weight of each stock in the portfolio must be within given ranges. The total proportional and fixed transaction costs spent for rebalancing must not exceed a given fraction of the investment budget. Finally, the short selling of stocks is prohibited, and it is assumed that fractional units of stocks can be traded. Note that the EITP also includes the construction of a new portfolio as a special case when the portfolio before rebalancing consists only of cash.

In the literature, various mathematical programming formulations have been proposed for the problem of determining an enhanced index-tracking portfolio. These formulations differ with respect to the real-life constraints considered, the way the expected tracking error is attempted to be minimized, and whether and how the expected excess return is integrated. With respect to the real-life constraints, some authors have determined enhanced index-tracking portfolios without considering real-life constraints (cf., e.g., [23]), whereas others have considered all real-life constraints as defined in the EITP (cf., e.g., [30]). With respect to the expected tracking error, the earliest studies attempted to minimize the tracking error variance (TEV), which is a quadratic function of the covariances between the returns of the stocks, the weights of the stocks in the portfolio, and the weights of the stocks in the index (cf., e.g., [23]). Minimizing the TEV corresponds to minimizing the estimated variance of the return differences between the portfolio and the index in the out-of-sample period (cf. [22]). By contrast, later studies attempted to minimize the expected tracking error by using as the objective function a dissimilarity function that captured the deviation between the historical developments of the portfolio and the index. One of the most widely used dissimilarity functions is the mean-absolute deviation (MAD) between the historical values of the portfolio and the index (cf., e.g., [8, 13, 16]). In the most recent study, the goal of minimizing the TEV was revisited (cf. [22]). With respect to the expected excess return, some studies have focused on the problem of determining the portfolio for an index-tracking fund without considering the expected excess return (cf., e.g., [30]). In other studies, the expected excess return has been considered by using a bi-objective approach with the maximization of the expected excess return as a second competing objective (cf., e.g., [8]), by introducing into the objective function a second term that captures the expected excess return (cf., e.g., [2]), or by introducing a constraint that prescribes a minimum expected excess return (cf., e.g., [23]). From an optimization point of view, these various means of integrating the expected excess return are very similar because all functions used for the expected excess return are linear. Various exact approaches, such as mixed-integer programming, and metaheuristic approaches, such as population-based heuristics or local-search heuristics, have all been proposed as solution approaches for the problem of determining an enhanced index-tracking portfolio.

We have identified four gaps in the literature on enhanced index tracking. Gap 1: it remains an open question whether it is preferable in terms of the out-of-sample tracking error to use the TEV as the objective function, which, together with the real-life constraints, constitutes a cardinality-constrained quadratic optimization problem that is known to be very challenging to solve (cf. [3, 33]), or whether it is preferable to use a dissimilarity function such as the MAD, which can be formulated as a linear objective function and thus is less challenging to optimize. Gap 2: the EITP as defined above has not been previously considered because the problems studied in the literature may neglect the minimum expected excess return or some of the real-life constraints. Hence, the EITP as defined above has not been formulated as a mathematical program. Moreover, the existing mathematical programming formulations for problems related to the EITP that consider transaction costs allow the implicit holding of cash because the budget constraint is modeled as an inequality or because the modeled transaction costs correspond to merely an upper bound on the true transaction costs. Consequently, these cash holdings are not considered in the formulation of the expected tracking error and the expected excess return. Gap 3: the existing solution approaches for the related problems studied in the literature may not be appropriate for the EITP when the TEV is used as the objective function. The existing exact approaches would require the solution of a series of quadratic programming relaxations, which may become computationally very expensive when large indices are considered. The existing metaheuristic approaches would require adaptation to the real-life constraints of the EITP, which may reduce their effectiveness because they are tailored for other specific problems that are less constrained. Gap 4: there are no available instances of the EITP based on large stock-market indices; the existing instances of related problems either are based on small indices or do not provide information about the index composition.

The main contribution of this paper is to address the open question corresponding to gap 1 by providing novel theoretical arguments and novel experimental results. The theoretical arguments indicate that minimizing the TEV instead of a dissimilarity function may lead to superior out-of-sample tracking errors, especially when the index is large, because dissimilarity functions may not exploit the known index composition. To be able to provide experimental results, we first had to address the gaps 2 to 4. To address gap 2, we present a novel mixed-integer quadratic programming (MIQP) formulation and a novel mixed-integer linear programming (MILP) formulation of the EITP. In the MIQP formulation, we use the TEV as the objective function. In the MILP formulation, we use the MAD as the objective function. The novelties in these formulations are a formulation of the considered real-life constraints in which cash holdings are explicitly considered and insights that allow to remove redundant variables and constraints. To address gap 3, we present a construction matheuristic and two improvement matheuristics based on the proposed MIQP formulation that are able to determine good feasible portfolios, i.e., portfolios that satisfy all considered constraints, within a reasonable computational time for the EITP when the TEV is used as the objective function, especially for instances based on large indices. The construction heuristic, which can be used to find an initial feasible portfolio quickly, is based on a novel idea of linearizing the TEV by using the identity matrix as a simplified covariance matrix and by considering absolute instead of squared deviations in the terms of the resulting function. The first improvement heuristic is based on the concept of local branching, which has been successfully applied to various combinatorial optimization problems (cf. [9]). In local branching, starting from the initial feasible solution, the solution space to be searched is iteratively defined with an upper bound on the number of binary variables whose values flip. The novelty of this improvement heuristic is that we consider a subset of promising stocks that differs in each iteration to reduce the required computational time. The second improvement heuristic is based on the concept of iterated greedy heuristics (cf., e.g., [26]). In iterated greedy heuristics, a current feasible solution is iteratively deconstructed and subsequently reconstructed in a greedy manner to form a new feasible solution. The novelties of this improvement heuristic are that we also consider a different subset of promising stocks in each iteration and that, in contrast to existing iterated greedy heuristics (cf., e.g., [31]), we apply mixed-integer quadratic programming for the reconstruction. The proposed matheuristics are particularly suitable for the EITP because they are simple to implement and because they combine the flexibility of mathematical programming to easily incorporate complex constraints such as the considered real-life constraints with the ability of heuristics to find good feasible solutions quickly. Hence, the proposed matheuristics exhibit the properties of accuracy, speed, simplicity, and flexibility, which are the four essential attributes of good heuristics according to Cordeau et al. [7]. Finally, to address gap 4, we generated a set of novel instances of the EITP based on nine large regional and global real-world stock-market indices maintained by Thomson Reuters. The largest of these indices has more than 9,000 constituents. In a computational experiment based on these instances, we tested two heuristic solution approaches that are based on the two proposed improvement matheuristics initialized with the proposed construction matheuristic and two exact solution approaches that are based on the MIQP and the MILP formulation along with a commercial mixed-integer programming solver. This computational experiment yielded the following three main findings. 1) An exact solution approach may be appropriate for the EITP when the MAD is used as the objective function, but may not be appropriate when the TEV is minimized, which indicates the potential improvements that may be achieved by applying heuristics to the EITP when the TEV is used as the objective function. 2) The proposed matheuristics are indeed able to achieve substantial improvements in terms of the TEV compared to an exact solution approach within a limited computational time. 3) Minimizing the TEV instead of the MAD leads to superior portfolios in terms of the out-of-sample tracking error.

The remainder of this paper is organized as follows. In Section 2, we review the existing solution approaches in the literature for problems that are related to the EITP. In Section 3, we present the MIQP and the MILP formulation and provide the arguments to address gap 1 theoretically. In Section 4, we present the construction matheuristic and the two improvement matheuristics. In Section 5, we report the computational results to address gap 1 experimentally. In Section 6, we offer some concluding remarks and an outlook on future research.

2. Related literature

Various papers in the literature have studied problems that are related to the EITP. Table 1 lists, for each of these papers with a ✓-symbol, whether it considers the real-life constraints of the EITP mentioned above and whether the objective is index tracking (IT) or enhanced index tracking (EIT). We categorize the papers into two groups based on whether the objective function used is non-linear or linear. In the following, we describe the proposed solution approaches of both groups.

The first group of problems consists of those that involve the optimization of a non-linear objective function. In some papers, only indices with a small number of constituents are considered, such that exact approaches are applicable (cf. [10, 14, 24]). In other papers, the real-life constraints are neglected, which allows closed-form solutions to be devised (cf. [15, 23]). In the remaining papers, metaheuristics such as evolutionary algorithms (cf. [1, 6, 17, 21, 27, 28, 29]) or local-search heuristics (cf. [18, 22, 32]) are proposed. The majority of the papers in this first group neglect most of the real-life constraints of the EITP. An exception is the paper by Beasley et al. [2], in which the goal is to optimize the trade-off between a non-linear dissimilarity function and the expected excess return subject to a cardinality constraint, minimum and maximum weights for the stocks included in the portfolio, and a budget for proportional transaction costs. An evolutionary algorithm is presented that uses cross-over and mutation operators to combine and modify, respectively, individuals that represent feasible and infeasible solutions. The presented algorithm

Table 1: Problems related to the EITP considered in the literature.

	Paper	Real-life constraints				Objective	
		Cardinality	Min./max. weights	Transaction costs	Min./max. trades	IT	EIT
Non-linear objective function	[15]					✓	✓
	[23]					✓	✓
	[24]		✓			✓	✓
	[16]			✓		✓	
	[6]			✓		✓	
	[10]	✓		✓		✓	
	[32]	✓				✓	
	[18]	✓				✓	
	[21]	✓				✓	
	[22]	✓				✓	
	[27]	✓				✓	✓
	[28]	✓				✓	
	[1]	✓	✓			✓	✓
	[14]	✓	✓			✓	
	[29]	✓	✓			✓	
	[17]	✓	✓			✓	
	[2]	✓	✓	✓		✓	
Linear objective function	[25]					✓	
	[4]	✓	✓			✓	✓
	[11]	✓	✓			✓	✓
	[13]	✓	✓	✓		✓	✓
	[8]	✓	✓	✓		✓	✓
	[30]	✓	✓	✓	✓	✓	

includes a customized procedure for determining portfolio weights, a repair operator, and a penalty term in the objective function to handle infeasible solutions.

The second group of problems consists of those that involve the optimization of a linear objective function. For these problems, exact approaches such as linear programming and MILP approaches are able to devise good feasible solutions within a reasonable computational time, even when real-life constraints and large indices are considered (cf. [4, 8, 11, 13, 25, 30]). Among all these problems, those studied in the following papers are most similar to the EITP in terms of the real-life constraints considered. Strub and Baumann [30] introduce a MILP formulation for determining the portfolio for an index-tracking fund in which a linear dissimilarity function is minimized subject to all real-life constraints of the EITP. Guastaroba and Speranza [13] minimize the MAD between the historical values of the portfolio and the index, which is modeled as a linear dissimilarity function, subject to a budget for fixed and proportional transaction costs, minimum and maximum portfolio weights, and a cardinality constraint. They also present a heuristic called Kernel Search, which is a matheuristic that can easily handle various real-life constraints. In this heuristic, the information from the solution to the linear programming relaxation is exploited to construct different sub-problems that can be solved quickly. They also show that their heuristic can be applied for enhanced index tracking by tracking an artificial index that represents the index return plus the target excess return. Filippi et al. [8] aim to maximize a linear excess-return function and minimize the same linear dissimilarity function subject to the same real-life constraints as those of Guastaroba and Speranza [13]. They modify the Kernel Search heuristic such that it can be applied to the considered problem. In the MILP formulation presented by Strub and Baumann [30], implicit cash holdings can occur because the budget constraint is modeled as an inequality, which is necessary because the total transaction costs spent for rebalancing plus the value of the portfolio may not exactly match the investment budget. In the MILP formulations proposed by Guastaroba and Speranza [13] and Filippi et al. [8], implicit cash holdings can occur because the modeled transaction costs correspond merely to an upper bound on the true transaction costs. A drawback of these implicit cash holdings is that they are not considered in the calculation of the historical portfolio values and thus are also ignored in the dissimilarity and excess return functions.

The existing solution approaches presented in the literature may not be appropriate for the EITP when the TEV is used as the objective function. The existing exact approaches and the Kernel Search heuristic would first require the solution of the continuous relaxation of the MIQP formulation of the EITP, which is a quadratic program that becomes computationally very expensive to solve when large indices are considered. The existing metaheuristics would require adaptation to the real-life constraints of the EITP, which may reduce their effectiveness because they are tailored for other specific problems that do not include all of the real-life constraints of the EITP. A further drawback of metaheuristics is that they may investigate many infeasible solutions and thus be ineffective.

3. Mixed-integer linear and quadratic programming formulations

In this section, we present the novel MIQP formulation and the novel MILP formulation of the EITP. In Subsection 3.1, we first present the objective functions and the constraint on the expected excess return that are used in the two mixed-integer programming (MIP) formulations. In Subsection 3.2, we present new arguments that using the TEV instead of a dissimilarity function as the objective function may lead to superior portfolios in terms of the out-of-sample tracking error. In Subsection 3.3, we introduce the formulation of the real-life constraints. In Subsection 3.4, we provide insights that allow to remove redundant variables and constraints from the formulation of the real-life constraints, and we present the complete MIP formulations without the removed variables and constraints.

Table 2 shows the nomenclature used in the MIP formulations. The set of available assets consists of the set of index constituents $U = \{1, \ldots, n\}$ and an asset $n + 1$ that represents the explicitly modeled cash holdings. Note that in Table 2, the decision variables are defined only for a set of considered stocks I, with I being a subset of the set of index constituents U and a superset of the set I_s that contains the stocks that must always be included in the portfolio after rebalancing, i.e., $I_s \subseteq I \subseteq U$. Thereby, the set I_s must always contain the stocks included in the portfolio before rebalancing that cannot be sold off completely due to the minimum and maximum trading values. Then, let T be the point in time at which the EITP must be solved, P_{it} be the prices of the stocks $i \in U$ at the point in time $t \in \{1, \ldots, T\}$, and $P_{n+1,t}$ be the values of the asset that represents cash calculated as $P_{n+1,t} = P_{n+1,T} \exp(-\sum_{s=t+1}^{T} i_s)$ for $t \in \{1, \ldots, T-1\}$, with $P_{n+1,T} = 100$ and i_t for $t \in \{2, \ldots, T\}$ corresponding to the continuously compounded interest rate on the cash holdings. Furthermore, let Y_i be the number of units of the assets $i \in U \cup \{n+1\}$ in the portfolio before rebalancing, κ be the net cash flow from deposits and withdrawals, and $C = \kappa + \sum_{i \in U \cup \{n+1\}} Y_i P_{iT}$ be the investment budget. Then, the stocks that cannot be sold off completely are those that have a value in the portfolio before rebalancing of $P_{iT}Y_i$ that is greater than the maximum trading value of $\eta_i C$ or greater than zero but smaller than the minimum trading value of $\zeta_i C$, i.e., $I_s \supseteq \{i \in U : P_{iT}Y_i > \eta_i C \vee 0 < P_{iT}Y_i < \zeta_i C\}$. We define the MIP formulations in this general form based on the sets I and I_s because this simplifies the notation for the MIP formulations without the removed redundant variables and constraints presented in Subsection 3.4 and because we can then use the MIQP formulation with only minor modifications for the heuristic solution approaches presented in Section 4.

3.1. Objective functions and the constraint on the expected excess return

The two competing objectives in enhanced index tracking are the minimization of the expected tracking error and the maximization of the expected excess return. In this subsection, we present the functions used to model these objectives in the proposed MIP formulations. In the MIQP and the MILP formulation, we use the TEV and the MAD, respectively, for the expected tracking error. In both formulations, we use the function presented by Roll [23] for the expected excess return. We adjust all functions to account for the set of considered stocks I and the explicitly modeled cash holdings.

We define $X_i \geq 0$ to be the main decision variables that correspond to the number of units of the assets $i \in I \cup \{n+1\}$ in the portfolio after rebalancing. Then, the TEV, which is used in the MIQP formulation, is a function of the covariances σ_{ij} between the returns of assets $i \in U \cup \{n+1\}$ and $j \in U \cup \{n+1\}$, the weights $\frac{P_{iT}X_i}{C}$ of the assets $i \in U \cup \{n+1\}$ in the portfolio, and the weights w_i^I of the assets $i \in U \cup \{n+1\}$ in the index, with $w_{n+1}^I = 0$. Any stock that is not included in I will have a portfolio weight of zero. Thus, the following function represents the TEV:

$$\sum_{i,j \in I \cup \{n+1\}} \sigma_{ij} \left(\frac{P_{iT}X_i}{C} - w_i^I \right) \left(\frac{P_{jT}X_j}{C} - w_j^I \right) - 2 \sum_{i \in I \cup \{n+1\}} \sum_{j \in U \setminus I} \sigma_{ij} \left(\frac{P_{iT}X_i}{C} w_j^I - w_i^I w_j^I \right) + \sum_{i,j \in U \setminus I} \sigma_{ij} w_i^I w_j^I \tag{1}$$

Based on the expected returns $\overline{r}_i$ of the assets $i \in U \cup \{n+1\}$, the expected excess return is calculated as the difference between the expected return of the portfolio and the expected return of the index:

$$\sum_{i \in I \cup \{n+1\}} \frac{P_{iT}X_i}{C} \overline{r}_i - \sum_{i \in U \cup \{n+1\}} w_i^I \overline{r}_i \tag{2}$$

In the MIQP formulation, we minimize the TEV subject to a constraint that prescribes a minimum

Table 2: Nomenclature for the MIP formulations.

Sets and parameters:	
T	Point in time at which the EITP must be solved (today)
n	Number of stocks in the index
U	Set of index constituents ($U = \{1, \ldots, n\}$)
I	Set of considered stocks ($I \subseteq U$)
I_s	Set of stocks that must be included in the portfolio after rebalancing ($I_s \subseteq I$)
k	Maximum portfolio cardinality
κ	Net cash flow from deposits and withdrawals
i_t	Continuously compounded interest rate on cash for the period starting at $t-1$ and ending at t, $t \in \{2, \ldots, T\}$
I_t/P_{it}	Historical value/price of index/asset $i \in U \cup \{n+1\}$ at $t \in \{1, \ldots, T\}$
Y_i	Number of units of asset $i \in U \cup \{n+1\}$ in the portfolio before rebalancing
C	Investment budget
ζ_i/η_i	Minimum/maximum trading value of stock $i \in U$ if traded, expressed as a percentage of C
ε_i/δ_i	Minimum/maximum weight of stock $i \in U$ if included in the portfolio after rebalancing
c_i^f	Fixed transaction cost for trading stock $i \in U$
c_i^b/c_i^s	Proportional transaction cost for buying/selling stock $i \in U$ as a percentage of the trading value
γ	Maximum total transaction costs, expressed as a percentage of C
w_i^I	Weight of asset $i \in U \cup \{n+1\}$ in the index, with $w_{n+1}^I = 0$
$\overline{r}_i$	Expected return of asset $i \in U \cup \{n+1\}$
α	Prescribed minimum expected excess return
σ_{ij}	Covariance between the discrete returns of asset $i \in U \cup \{n+1\}$ and asset $j \in U \cup \{n+1\}$
Continuous non-negative decision variables:	
X_i	Number of units of asset $i \in I \cup \{n+1\}$ in the portfolio after rebalancing
G_i	Total transaction costs associated with stock $i \in I$
v_i^b/v_i^s	Value bought/sold of stock $i \in I$
u_t/d_t	Absolute upside/downside deviation between the values of the portfolio and the index at $t \in \{1, \ldots, T\}$
Binary decision variables:	
z_i	$= 1$, if $X_i > 0$; $= 0$, otherwise ($i \in I$)
z_i^b	$= 1$, if $X_i > Y_i$; $= 0$, otherwise ($i \in I$)
z_i^s	$= 1$, if $X_i < Y_i$; $= 0$, otherwise ($i \in I$)

expected excess return of α, as follows:

$$\begin{cases} \text{Min. } (1) & (3) \\ \text{s.t.} \quad \sum_{i \in I \cup \{n+1\}} \frac{P_{iT} X_i}{C} \overline{r}_i - \sum_{i \in U \cup \{n+1\}} w_i^I \overline{r}_i \geq \alpha & (4) \end{cases}$$

In the MILP formulation, the dissimilarity function captures the MAD over all in-sample time points $t \in \{1, \ldots, T\}$ between the values of the index I_t, scaled to the investment budget C at time point T, and the values of the portfolio $\sum_{i \in I \cup \{n+1\}} P_{it} X_i$. With the introduction of the non-negative decision variables u_t and d_t for $t \in \{1, \ldots, T\}$, the MAD can be minimized subject to the constraint on the expected excess return as follows:

$$\begin{cases} \text{Min. } \frac{1}{T} \sum_{t \in \{1,\ldots,T\}} (u_t + d_t) & (5) \\ \text{s.t.} \quad u_t - d_t = \sum_{i \in I \cup \{n+1\}} P_{it} X_i - I_t \frac{C}{I_T} \qquad (t \in \{1, \ldots, T\}) & (6) \\ \qquad \sum_{i \in I \cup \{n+1\}} \frac{P_{iT} X_i}{C} \overline{r}_i - \sum_{i \in U \cup \{n+1\}} w_i^I \overline{r}_i \geq \alpha & (4) \end{cases}$$

3.2. TEV: a comparison with dissimilarity functions

In this subsection, we compare the minimization of the TEV with the minimization of a dissimilarity function such as the MAD in terms of the out-of-sample tracking error. For this purpose, we consider all index constituents, i.e., $I = U$, and we consider only a budget constraint that ensures that the entire investment budget is invested in the assets. We assume that the number of available assets is much larger than the number of in-sample time points, i.e., $|U \cup \{n+1\}| \gg T$, and that the matrix consisting of the in-sample prices of each stock, where each stock corresponds to a column, has full row rank. Both assumptions are usually satisfied when the index is large. We further assume that the matrix of the covariances is positive definite. This assumption is satisfied when an appropriate estimator is used for the covariances, such as that of Ledoit and Wolf [20], but may be violated when the sample covariance is used as an estimator (cf. Ledoit and Wolf [19]).

When the TEV is to be minimized with a positive-definite matrix of covariances, the only solution with zero TEV is the portfolio that has the same composition as the index. By contrast, when a dissimilarity function is used, i.e., when the known index composition is ignored, infinitely many different portfolios can exist that achieve a dissimilarity of zero with respect to the index. To see this, note that finding a portfolio with zero dissimilarity is equivalent to solving a system of T linear equations with $n + 1$ unknowns, where these T equations state that the portfolio value at each time point $t \in \{1, \ldots, T\}$ must match the scaled index value at that time point. Note that the equation for time point T also ensures that the budget constraint is satisfied because of the scaling of the index values. Under the assumption of a full row rank matrix of stock prices, infinitely many solutions to this linear system exist, which means that infinitely many portfolios with zero dissimilarity exist.

Based on the arguments above, one drawback of minimizing a dissimilarity function is that, in contrast to the case of minimizing the TEV, many different portfolios can exist that each have an objective function value of zero but a composition that strongly differs from that of the index. These portfolios may have very high out-of-sample tracking errors. Hence, for our computational experiment reported in Section 5, we expect that the compositions of the portfolios obtained when minimizing the MAD will differ more strongly from the composition of the index than the compositions of the portfolios obtained when minimizing the TEV. Consequently, we also expect that, over all considered problem instances, the average and the worst-case tracking error for the out-of-sample period will be worse when the MAD is minimized instead of the TEV.

3.3. Real-life constraints

Next, we model the real-life constraints. The constraints expressed in (7) assign at least the absolute value bought or sold of each stock $i \in I$ to the non-negative decision variable v_i^b or v_i^s, respectively. These decision variables are used to model the transaction costs and the minimum and maximum trading values.

$$v_i^b - v_i^s = P_{iT}(X_i - Y_i) \qquad (i \in I) \tag{7}$$

The purpose of constraints (8) and (9) is twofold. First, the binary variables z_i^b and z_i^s are assigned a value of one if the variables v_i^b and v_i^s, respectively, take a positive value and a value of zero otherwise. Second, the constraints prescribe minimum and maximum values of $\zeta_i C$ and $\eta_i C$, respectively, for v_i^b and v_i^s.

$$\zeta_i C z_i^b \leq v_i^b \leq \eta_i C z_i^b \qquad (i \in I) \tag{8}$$

$$\zeta_i C z_i^s \leq v_i^s \leq \eta_i C z_i^s \qquad (i \in I) \tag{9}$$

The constraints defined in (10) ensure that for each stock $i \in I$, at most one of the binary variables z_i^b and z_i^s can be set to one.

$$z_i^b + z_i^s \leq 1 \qquad (i \in I) \tag{10}$$

Together, constraints (8), (9), and (10) ensure that for each stock $i \in I$, either v_i^b or v_i^s must be set to zero. Because it is not possible for both variables v_i^b and v_i^s to take positive values simultaneously for a given stock $i \in I$, the constraints defined in (7) assign the actual values bought or sold of each stock $i \in I$ to the variables v_i^b or v_i^s, respectively. These actual values are necessary to model the minimum and maximum trading values $\zeta_i C$ and $\eta_i C$ using constraints (8) and (9).

Let c_i^f, c_i^b, and c_i^s be the parameters that determine the fixed transaction costs for trading the stocks $i \in U$, the proportional transaction costs for buying units of the stocks $i \in U$, and the proportional transaction costs for selling units of the stocks $i \in U$, respectively. Then, based on the variables v_i^b, v_i^s, z_i^b, and z_i^s, the transaction costs G_i for each stock $i \in I$ are calculated using the constraints defined in (11). Note that the variables G_i take values equal to the actual transaction costs associated with each stock $i \in I$, because we ensure that the variables v_i^b and v_i^s take the actual values bought and sold of each stock, and that at most one of the binary variables z_i^b and z_i^s can be set to one if stock $i \in I$ is traded, whereas both variables z_i^b and z_i^s are set to zero otherwise.

$$G_i = c_i^b v_i^b + c_i^s v_i^s + c_i^f (z_i^b + z_i^s) \qquad (i \in I) \tag{11}$$

The budget constraint (12) states that the available investment budget C must be either held in cash, invested in the stocks that constitute the index, or spent for transaction costs. Note the possibility that some stocks were included in the portfolio before rebalancing but are not included in the set of considered stocks I. Hence, the shares of these stocks must be sold, incurring total transaction costs of $\sum_{i \in U \setminus I : Y_i > 0} \left(c_i^s Y_i P_{iT} + c_i^f \right)$. Since the variables G_i take values equal to the actual transaction costs associated with each stock $i \in I$, constraint (12) ensures that the variable X_{n+1} corresponds exactly to the part of the investment budget that is not invested in stocks or spent for transaction costs. Hence, we can explicitly account for these cash holdings when formulating the TEV, the MAD, and the expected excess return.

$$\sum_{i \in I \cup \{n+1\}} P_{iT} X_i + \sum_{i \in I} G_i + \sum_{i \in U \setminus I : Y_i > 0} \left(c_i^s Y_i P_{iT} + c_i^f \right) = C \tag{12}$$

Constraint (13) prescribes a budget of γC for the total transaction costs.

$$\sum_{i \in I} G_i + \sum_{i \in U \setminus I : Y_i > 0} \left(c_i^s Y_i P_{iT} + c_i^f \right) \leq \gamma C \tag{13}$$

The constraints (14) ensure that each binary variable z_i takes a value of one if stock $i \in I$ is included in the portfolio after rebalancing and a value of zero otherwise. Furthermore, these constraints define minimum and maximum values of ε_i and δ_i, respectively, for the weight of each stock $i \in I$ in the portfolio.

$$\varepsilon_i z_i \leq \frac{P_{iT} X_i}{C} \leq \delta_i z_i \qquad (i \in I) \tag{14}$$

Based on the binary variables z_i, the cardinality constraint (15) is formulated as follows.

$$\sum_{i \in I} z_i \leq k \tag{15}$$

The domains of the decision variables are specified by (16) and (17).

$$X_i \geq 0 \qquad (i \in I \cup \{n+1\}) \tag{16}$$

$$z_i^b, z_i^s, z_i \in \{0,1\};\ v_i^b, v_i^s, G_i \geq 0 \qquad (i \in I) \tag{17}$$

3.4. Removing redundant variables and constraints from the mixed-integer programming formulations

From the formulation of the real-life constraints presented in Subsection 3.3, we can remove some redundant variables and constraints based on the following three insights. First, we note that some stocks that are included in the portfolio before rebalancing must always be included in the portfolio after rebalancing due to the specified minimum and maximum trading values. As mentioned above, these stocks are included in the set $I_s \supseteq \{i \in U : P_{iT} Y_i > \eta_i C \vee 0 < P_{iT} Y_i < \zeta_i C\}$. Second, Filippi et al. [8] note that if a stock is not included in the portfolio before rebalancing and is traded, then this stock will always be included in the portfolio after rebalancing. Third, Strub and Baumann [30] note that stocks that are not included in the portfolio before rebalancing cannot be sold because short selling is not allowed. We combine all three insights to obtain the following restrictions on stocks based on their values in the portfolio before rebalancing. For each stock i that is not included in the portfolio before rebalancing, i.e., $Y_i = 0$, selling stock i is not possible, and trading stock i means that it will be included in the portfolio after rebalancing. For each stock i that has a value in the portfolio before rebalancing that is positive but smaller than the minimum trading value, i.e., $0 < P_{iT}Y_i < \zeta_i C$, selling stock i is not possible, and thus, stock i must be included in the portfolio after rebalancing. For each stock i that has a value in the portfolio before rebalancing that is larger than the maximum trading value, i.e., $\eta_i C < P_{iT}Y_i$, selling all units of stock i is not possible, and thus, stock i must be included in the portfolio after rebalancing.

Based on the restrictions above, we can eliminate certain variables and constraints. For each stock i that cannot be sold, the binary variable z_i^s and the continuous variable v_i^s must both be zero and thus can be removed. Additionally, the continuous variable v_i^b can be replaced with $P_{iT}(X_i - Y_i)$. Furthermore, for each stock i that is not included in the portfolio before rebalancing, i.e., $Y_i = 0$, we can replace the binary variable z_i^b with the binary variable z_i because selling is not possible and buying stock i means that it will be included in the portfolio after rebalancing. For each stock i that must be included in the portfolio after rebalancing, we can set the binary variable z_i equal to one.

The novel MIQP formulation (M-Q) and the novel MILP formulation (M-L) below include the real-life constraints (without redundant variables and constraints) along with the constraint on the expected excess return and their corresponding objective functions.

$$
\text{(M-Q)}\begin{cases}
\text{Min.} & (1) & & \\
\text{s.t.} & (4), (12), (13) & & \\
& v_i^b - v_i^s = P_{iT}(X_i - Y_i) & (i \in I : P_{iT}Y_i \geq \zeta_i C) & (18)\\
& \zeta_i C \leq X_i P_{iT} \leq \eta_i C & (i \in I_s : Y_i = 0) & (19)\\
& \zeta_i C z_i \leq X_i P_{iT} \leq \eta_i C z_i & (i \in I \setminus I_s : Y_i = 0) & (20)\\
& \zeta_i C z_i^b \leq (X_i - Y_i) P_{iT} \leq \eta_i C z_i^b & (i \in I : 0 < P_{iT}Y_i < \zeta_i C) & (21)\\
& \zeta_i C z_i^b \leq v_i^b \leq \eta_i C z_i^b & (i \in I : P_{iT}Y_i \geq \zeta_i C) & (22)\\
& \zeta_i C z_i^s \leq v_i^s \leq \eta_i C z_i^s & (i \in I : P_{iT}Y_i \geq \zeta_i C) & (23)\\
& z_i^b + z_i^s \leq 1 & (i \in I : P_{iT}Y_i \geq \zeta_i C) & (24)\\
& G_i = c_i^b P_{iT} X_i + c_i^f & (i \in I_s : Y_i = 0) & (25)\\
& G_i = c_i^b P_{iT} X_i + c_i^f z_i & (i \in I \setminus I_s : Y_i = 0) & (26)\\
& G_i = c_i^b P_{iT}(X_i - Y_i) + c_i^f z_i^b & (i \in I : 0 < P_{iT}Y_i < \zeta_i C) & (27)\\
& G_i = c_i^b v_i^b + c_i^s v_i^s + c_i^f (z_i^b + z_i^s) & (i \in I : P_{iT}Y_i \geq \zeta_i C) & (28)\\
& \varepsilon_i \leq \dfrac{P_{iT}X_i}{C} \leq \delta_i & (i \in I_s) & (29)\\
& \varepsilon_i z_i \leq \dfrac{P_{iT}X_i}{C} \leq \delta_i z_i & (i \in I \setminus I_s) & (30)\\
& \displaystyle\sum_{i \in I \setminus I_s} z_i \leq k - |I_s| & & (31)\\
& X_i \geq 0 & (i \in I \cup \{n+1\}) & (32)\\
& G_i \geq 0 & (i \in I) & (33)\\
& z_i \in \{0, 1\} & (i \in I \setminus I_s) & (34)\\
& v_i^b, v_i^s \geq 0,\ z_i^s \in \{0, 1\} & (i \in I : P_{iT}Y_i \geq \zeta_i C) & (35)\\
& z_i^b \in \{0, 1\} & (i \in I : Y_i > 0) & (36)
\end{cases}
$$

$$
\text{(M-L)}\begin{cases}
\text{Min.} & (5) & & \\
\text{s.t.} & (6), (4), (12), (13), (18), (19), (20), (21), (22), (23), (24), (25), & & \\
& (26), (27), (28), (29), (30), (31), (32), (33), (34), (35), (36) & & \\
& u_t, d_t \geq 0 & (t \in \{1, \ldots, T\}) & (37)
\end{cases}
$$

4. Heuristic solution approaches

In this section, we present the matheuristics for the EITP when the TEV is used as the objective function. In Subsection 4.1, we present the construction heuristic for determining an initial feasible portfolio. In

Table 3: Additional nomenclature for the heuristic solution approaches.

Parameters:	
q	Parameter that defines the number of considered stocks
Δ	Number of stocks that can be added and removed
$\overline{\nu}$	Maximum number of iterations without improvement
d	Maximum number of stocks to be removed
Continuous non-negative decision variables:	
ξ_i	Absolute deviation between the portfolio weight and index weight of asset $i \in I \cup \{n+1\}$

Subsections 4.2 and 4.3, we present the two improvement heuristics based on local branching and the concept of iterated greedy heuristics, respectively, for improving a given initial feasible portfolio. Table 3 defines the additional nomenclature used in formulating these matheuristics.

4.1. Construction heuristic

Constructing a feasible portfolio is not straightforward. It is possible that no feasible portfolio exists, e.g., when the prescribed minimum excess return is set too high or when a current portfolio must be rebalanced so heavily to ensure the prescribed minimum and maximum weights that the prescribed budget for transaction costs is too low. Even if a feasible portfolio exists, when selecting the stocks that should be included in the portfolio by applying, e.g., a random or greedy algorithm, it might not be possible to find weights for these selected stocks such that the portfolio is feasible with respect to all constraints.

Therefore, we propose a MILP-based construction heuristic that is able to find a feasible portfolio quickly if one exists, and can also prove the nonexistence of a feasible portfolio. To simplify the search for a feasible portfolio for the MIQP formulation (M-Q), we use the identity matrix as a simplified covariance matrix. Thus, the objective function (1) reduces to the following terms:

$$\sum_{i \in I \cup \{n+1\}} \left(\frac{P_{iT} X_i}{C} - w_i^I \right)^2 + \sum_{i \in U \setminus I} (w_i^I)^2 \tag{38}$$

Furthermore, we consider the sum of the absolute deviations (instead of the squared deviations) between the weights of the assets in the portfolio and the weights of the assets in the index, and we ignore the second sum because it is constant. The resulting objective function can be optimized subject to the constraints of the EITP using the following MILP formulation:

$$
(\text{M-C}) \begin{cases}
\text{Min.} \quad \sum_{i \in I \cup \{n+1\}} \xi_i & & (39) \\
\text{s.t.} \quad \xi_i \geq \frac{P_{iT} X_i}{C} - w_i^I & (i \in I \cup \{n+1\}) & (40) \\
\qquad \xi_i \geq w_i^I - \frac{P_{iT} X_i}{C} & (i \in I \cup \{n+1\}) & (41) \\
\qquad (4), (12), (13), (18), (19), (20), (21), (22), (23), (24), (25), & & \\
\qquad (26), (27), (28), (29), (30), (31), (32), (33), (34), (35), (36) & & \\
\qquad \xi_i \geq 0 & (i \in I \cup \{n+1\}) & (42)
\end{cases}
$$

To further simplify the search for a feasible portfolio, we consider only a limited set of promising stocks I. If no feasible portfolio can be found based on a given set I, we increase the cardinality of the set I. This preselection is crucial for finding good feasible portfolios for the MILP formulation (M-C). In general, the set I should contain the stocks with the highest weights in the index to allow a small objective function value to be achieved. Moreover, the set I should contain the stocks that are in the current portfolio, because not including these stocks in the set I would mean that we were required to sell all units of these stocks, which would incur high transaction costs.

Algorithm 1 describes the construction heuristic. First, we initialize ν, and we include in the set I all stocks that are held in the portfolio before rebalancing and all stocks that are in the set I_s. Then, we gradually expand the set I by including the k stocks that have the highest weights in the index and are not yet included in the set I. Thereafter, based on the expanded set I, the MILP formulation (M-C) is solved. This process is repeated until a feasible portfolio is found or until $I = U$. If no feasible portfolio is found with $I = U$, this proves that no feasible portfolio exists.

4.2. Local branching heuristic

Local branching refers to a local-search framework for MIP formulations that is based on so-called local-branching cuts. Given a feasible solution, these local-branching cuts iteratively define the neighborhood to be searched by placing an upper bound on the number of binary variables whose values can be flipped, either from one to zero or from zero to one. Based on this framework, we extend the MIQP formulation (M-Q) by incorporating constraints (43) and (44). Starting from the best feasible portfolio found so far, which includes the stocks in the set I^*, these constraints restrict the search space to all feasible portfolios that can be reached by adding at most Δ stocks to the portfolio and by removing at most Δ stocks from the portfolio. When removing stocks, we ensure that no stocks from the set I_s are removed because these stocks must be included in the portfolio after rebalancing. This results in the MIQP formulation (M-LBH) shown below.

Algorithm 1 Construction heuristic

procedure CONSTRUCTIONHEURISTIC()
  $\nu \leftarrow 0$; $I \leftarrow \{i \in U : Y_i > 0\} \cup I_s$;
  while true **do**
    while $I \neq U$ **and** $|I| < |\{i \in U : Y_i > 0\} \cup I_s| + k(1+\nu)$ **do**
      $I \leftarrow I \cup \{a\}$, where $a \in \text{argmax}_{i \in U \setminus I}\, w_i^I$;
    end while
    Solve (M-C);
    if feasible portfolio found **then**
      return set of stocks included in the feasible portfolio;
    else if $I = U$ **then**
      return no feasible portfolio exists;
    end if
    $\nu \leftarrow \nu + 1$;
  end while
end procedure

$$
\text{(M-LBH)}\begin{cases}
\text{Min.} & (1) \\
\text{s.t.} & (4), (12), (13), (18), (19), (20), (21), (22), (23), (24), (25), \\
& (26), (27), (28), (29), (30), (31), (32), (33), (34), (35), (36) \\
& \displaystyle\sum_{i \in I^* \setminus I_s} (1 - z_i) \leq \Delta \qquad (43) \\
& \displaystyle\sum_{i \in I \setminus I^*} z_i \leq \Delta \qquad (44)
\end{cases}
$$

The local branching framework requires the solution of a series of quadratic programs, which is computationally expensive for large indices when all available stocks are considered in each iteration, i.e., when $I = U$, even when the search space is restricted by local-branching cuts. Therefore, we propose a novel approach in which the search space is further restricted by considering only a limited set of promising stocks I. To prevent the exclusion of high-quality solutions from the search space due to a poor preselection of the stocks to be included in the set I, we use a randomly selected set I in each iteration. Because we consider stocks with higher index weights to be more promising for obtaining low objective function values, we define the probability that a stock will be included in the set I in each iteration to be proportional to its weight in the index.

Algorithm 2 describes the local branching heuristic. First, we initialize ν and Δ. Then, we include in the set I the stocks from the set I^*, which contains the stocks that are included in the best feasible portfolio found so far. Subsequently, we iteratively expand the set I until it contains $k + q$ stocks. In a given iteration of this process, each stock $i \in U \setminus I$ has a probability $\frac{w_i^I}{\sum_{j \in U \setminus I} w_j^I}$ of being selected for inclusion in the set I. Thereafter, the MIQP formulation (M-LBH) is solved. If a better feasible portfolio is found, we update the set I^* to contain the selected stocks in this new best feasible portfolio, and we reset the number of iterations elapsed without finding a better feasible portfolio ν to zero; otherwise, we increase ν by one. If ν reaches $\overline{\nu}$, i.e., the maximum number of iterations without a better feasible portfolio having been found, we increase Δ by one to enlarge the search space. As soon as a new best feasible portfolio is found, we reset Δ to one. This process is repeated until a given termination criterion is satisfied. Finally, the best feasible portfolio found so far is returned.

Algorithm 2 Local branching heuristic

procedure LocalBranchingHeuristic(I^*, q, $\overline{\nu}$)
  $\nu \leftarrow 0$; $\Delta \leftarrow 1$;
  while termination criterion not satisfied **do**
    $I \leftarrow I^*$;
    while $I \neq U$ **and** $|I| < k + q$ **do**
      $a \leftarrow$ select stock from set $U \setminus I$ with probability $\frac{w_i^I}{\sum_{j \in U \setminus I} w_j^I}$ of the selection of stock $i \in U \setminus I$;
      $I \leftarrow I \cup \{a\}$;
    end while
    Solve (M-LBH) to obtain a feasible portfolio by adding at most Δ stocks and by removing at most Δ stocks;
    if new best feasible portfolio found **then**
      $I^* \leftarrow$ set of selected stocks in the new best feasible portfolio;
      $\nu \leftarrow 0$; $\Delta \leftarrow 1$;
    else
      $\nu \leftarrow \nu + 1$;
      if $\nu = \overline{\nu}$ **then**
        $\nu \leftarrow 0$; $\Delta \leftarrow \Delta + 1$;
      end if
    end if
  end while
  return best feasible portfolio found;
end procedure

4.3. Iterated greedy heuristic

In an iterated greedy heuristic, two phases are performed repeatedly: deconstruction and reconstruction. During the deconstruction phase, we remove several randomly selected stocks from the current best feasible portfolio. During the subsequent reconstruction phase, we add stocks back into the deconstructed portfolio in a greedy manner to obtain a new feasible portfolio.

In contrast to existing iterated greedy heuristics, we restrict the search space by considering only a limited set of promising stocks I, as in the local branching heuristic, and we repeatedly solve an MIQP formulation during the reconstruction phase to add the myopic best stock to the deconstructed portfolio, which allows all constraints in the MIQP formulation (M-Q) to be easily considered. Specifically, we solve the

MIQP formulation (M-IGH) below that corresponds to the MIQP formulation (M-Q) without the cardinality constraint (31), but with the additional constraint (45). This additional constraint prescribes that at most one stock from the set I that is not included in the set I_s can be added to the portfolio. During the execution of the iterated greedy heuristic, we modify the set I_s such that it contains the stocks that must be included in the reconstructed portfolio, i.e., the stocks that are included in the current best feasible portfolio and were not removed during the most recent deconstruction phase.

$$
\text{(M-IGH)} \begin{cases} \text{Min.} & (1) \\ \text{s.t.} & (4), (12), (13), (18), (19), (20), (21), (22), (23), (24), (25), \\ & (26), (27), (28), (29), (30), (32), (33), (34), (35), (36) \\ & \sum_{i \in I \setminus I_s} z_i \leq 1 \end{cases} \tag{45}
$$

Algorithm 3 describes the iterated greedy heuristic. First, we store the stocks that must be included in the portfolio after rebalancing in the set I'_s because the algorithm modifies the set I_s during its execution. Then, we enter the main loop, which consists of the deconstruction, reconstruction, and acceptance phases. Before beginning the deconstruction phase, we include in the set I the stocks that are included in the best feasible portfolio found so far. Then, during the deconstruction phase, we first remove p randomly selected stocks from the set I. When removing stocks, we must ensure that no stocks from the set I'_s are removed because these stocks must be included in the portfolio after rebalancing. After having removed p stocks from the set I, we define the set of stocks that must be included in the reconstructed portfolio, i.e., the set I_s, as the set of stocks currently in the set I. Then, the reconstruction phase begins. During the reconstruction phase, we expand the set I by adding stocks based on probabilities that depend on the weights of the stocks in the index, as is done in the local branching heuristic. As soon as the set I consists of $k + q$ stocks, we iteratively solve (M-IGH) to add to the portfolio at most one new stock in each iteration from the set $I \setminus I_s$. If a feasible portfolio with one new stock from the set $I \setminus I_s$ can be found, then the newly selected stock is added to the set I_s. This process is repeated until either k stocks are included in the portfolio, no new stock is added to the portfolio, or it is found that no feasible portfolio for (M-IGH) exists. After the reconstruction phase, we check whether a new best feasible portfolio has been found. If this is the case, we update the set I^* to contain the selected stocks in the new best feasible portfolio. The deconstruction, reconstruction, and acceptance phases are repeated until a specified termination criterion is met. Finally, we reset the set I_s and return the best feasible portfolio found so far.

5. Computational results

In this section, we address the open question whether it is preferable in terms of the out-of-sample tracking error to use the TEV or the MAD as the objective function by providing computational results. For this purpose, we test the performance of two exact solution approaches and two heuristic solution approaches for

Algorithm 3 Iterated greedy heuristic

procedure ITERATEDGREEDYHEURISTIC(I^*, q, d)
  $I'_s \leftarrow I_s$;
  while termination criterion not satisfied **do**
    $I \leftarrow I^*$;
    $p \leftarrow$ random integer from set $\{1, \ldots, d\}$; ▷ Deconstruction
    for $i \leftarrow 1$ **to** p **do**
      Remove a randomly selected element from I that is not included in I'_s;
    end for
    $I_s \leftarrow I$;
    while $I \neq U$ **and** $|I| < k + q$ **do** ▷ Reconstruction
      $a \leftarrow$ select stock from set $U \setminus I$ with probability $\frac{w_i^I}{\sum_{j \in U \setminus I} w_j^I}$ of the selection of stock $i \in U \setminus I$;
      $I \leftarrow I \cup \{a\}$;
    end while
    while $|I_s| < k$ **do**
      Solve (M-IGH) to add at most one new stock to the portfolio;
      if a feasible portfolio with one new stock has been found **then**
        $I_s \leftarrow$ set of selected stocks in the feasible portfolio;
      else
        break;
      end if
    end while
    if new best feasible portfolio found **then** ▷ Acceptance
      $I^* \leftarrow$ set of selected stocks in the new best feasible portfolio;
    end if
  end while
  $I_s \leftarrow I'_s$
  return best feasible portfolio found;
end procedure

three scenarios that differ in terms of the composition of the portfolio before rebalancing. The two exact solution approaches are based on the MIQP formulation (M-Q) and the MILP formulation (M-L); we refer to these MIP approaches as M-Q and M-L, respectively. In these MIP approaches, we consider the entire set of index constituents, i.e., $I = U$. For the two heuristic solution approaches, the construction heuristic (cf. Algorithm 1) is used to determine an initial feasible portfolio, and the two improvement heuristics (cf. Algorithms 2 and 3) are used to improve this initial feasible portfolio; we refer to these approaches as LBH and IGH, respectively.

This section is organized as follows. In Subsections 5.1 and 5.2, we explain the design of our experiment and describe the novel problem instances, respectively. In Subsection 5.3, we investigate the index-tracking capabilities of the portfolios given in the three scenarios before rebalancing to provide a reference against which to assess the index-tracking capabilities of the portfolios after rebalancing. In Subsection 5.4, we show that in contrast to M-L, M-Q leads to feasible portfolios within a limited computational time that may be improved substantially in terms of the objective function value, which indicates the potential of applying LBH and IGH. In Subsection 5.5, we show that LBH and IGH are indeed able to achieve substantial improvements in terms of the objective function value compared to M-Q. In Subsection 5.6, we provide insights that minimizing the TEV may be superior to minimizing the MAD in terms of the out-of-sample tracking error. In Subsection 5.7, we offer further insights that this superior out-of-sample performance may be attributed to the different compositions of the rebalanced portfolios.

5.1. Experimental design

We use an experimental design similar to those of Guastaroba and Speranza [13] and Filippi et al. [8]. We assume that the manager of an investment fund rebalances a portfolio at the end of an in-sample period that consists of 104 weeks, i.e., $T = 104$. The portfolio is then left unchanged for the entirety of an out-of-sample period that consists of 52 weeks. We also assume that the fixed transaction cost for trading is 12 for all stocks (i.e., $c_i^f = 12$ for all $i \in U$), that the proportional transaction costs for buying and selling are each 1% of the trading value for all stocks (i.e., $c_i^b = c_i^s = 0.01$ for all $i \in U$), and that the budget available for transaction costs is 1.5% of the investment budget (i.e., $\gamma = 0.015$).

Furthermore, we define three scenarios, I, II, and III, which differ in terms of the composition of the investment fund's portfolio before rebalancing. In scenarios I and II, a portfolio of stocks already exists. In scenario III, a new portfolio must be constructed from cash. In scenarios I and II, the portfolios before rebalancing consist of the k stocks with the highest and lowest weights in the index, respectively. The weight of each stock in the portfolio before rebalancing is set such that it is proportional to the weight of that stock in the index and such that the sum of the weights of all stocks in the portfolio is equal to one. The portfolio before rebalancing in scenario I is a portfolio with a rather good index-tracking capability, whereas the portfolio before rebalancing in scenario II is a portfolio with a rather poor index-tracking capability. This claim is supported by the results presented in Subsection 5.3.

The values of the remaining parameters deviate from the values used by Guastaroba and Speranza [13] and Filippi et al. [8]. The reason for these deviations is that we are considering much larger indices, and thus, we wish to allow the portfolio to have a larger cardinality. Specifically, we use two different values for the maximum portfolio cardinality, namely, $k = 100$ and $k = 200$. Because of the larger portfolio cardinality, we must also allow the portfolio weights to take smaller values. For the minimum and maximum portfolio weights, we adopt values of 0.2% and 20%, respectively, for all stocks (i.e., $\varepsilon_i = 0.002$ and $\delta_i = 0.2$ for all $i \in U$). For the parameters that define the minimum and maximum trading values, we use the same values as for the parameters ε_i and δ_i for all stocks (i.e., $\zeta_i = 0.002$ and $\eta_i = 0.2$ for all $i \in U$). For scenarios I and II, we assume a value for each portfolio of 10,000,000 and a net change in cash of $\kappa = 0$. For scenario III, we assume a portfolio value of 0 and a κ of 10,000,000; this is the same as assuming that the portfolio before rebalancing consists of cash only and that the net change in cash is zero (i.e., $Y_{n+1} = \frac{10{,}000{,}000}{P_{n+1,T}}$, $Y_i = 0$ for all $i \in U$, and $\kappa = 0$). Hence, in all three scenarios, the investment budget C is 10,000,000.

Since the prescribed minimum expected excess return α and the interest rate i_t on cash were not used by Guastaroba and Speranza [13] and Filippi et al. [8], we define new values. Specifically, we use three α values of 0, 0.0001914, and 0.0005686, which correspond to annualized α values of 0%, 1%, and 2%, respectively, and a value of zero for the interest rate on cash at all time points (i.e., $i_t = 0$ for all $t \in \{2, \ldots, T\}$). The values of the parameters n, I_t, P_{it}, and w_i^I depend on the considered index.

We use the estimator of Ledoit and Wolf [20] to estimate the covariances σ_{ij} based on the discrete in-sample returns of the assets $i, j \in U \cup \{n+1\}$. As Ledoit and Wolf [19] note, using this estimator ensures that the matrix of the covariances is positive definite, and thus, the TEV is a convex quadratic function of the weights of the stocks in the portfolio, which allows commercial MIP solvers such as CPLEX and Gurobi to be applied. The expected returns of the assets $i \in U \cup \{n+1\}$ are estimated as follows: $\overline{r}_i = \frac{1}{T-1} \sum_{t \in \{2,\ldots,T\}} \frac{P_{it} - P_{i,t-1}}{P_{i,t-1}}$.

After preliminary experiments, we adopt the following values for the input parameters of the heuristic solution approaches for all problem instances. For LBH, we set the number of stocks considered to $k + 50$ and the maximum number of iterations without improvement to ten (i.e., $q = 50$ and $\overline{\nu} = 10$). For IGH, we adopt the same number of stocks considered as for LBH and a value of two as the maximum number of stocks to be removed from the best feasible portfolio found so far (i.e., $q = 50$ and $d = 2$).

We evaluate the in-sample performance of the tested solution approaches by using the following performance measures:

- OFV: objective function value of the best feasible portfolio found within the prescribed computational time limit. Note that the OFV is scaled by a factor of 100,000.
- MIP gap [%]: relative deviation between the OFV and the best lower bound (LB) provided by the solver within the prescribed computational time limit, calculated as: (OFV − LB)/OFV.

To evaluate the out-of-sample performance of the tested solution approaches we report the following performance measures. Thereby, we let T' be the end of the out-of-sample period, i.e., $T' = 156$:

- TE_{TEV} [%]: annualized standard deviation of the differences between the portfolio returns and the index returns during the out-of-sample period, calculated as:

$$\sqrt{\frac{52}{T'-T} \sum_{t \in \{T+1,\ldots,T'\}} \left(r_t^D - \overline{r}^D\right)^2} \tag{46}$$

where $r_t^D = \left(\frac{\sum_{i \in U \cup \{n+1\}} P_{it} X_i}{\sum_{i \in U \cup \{n+1\}} P_{i,t-1} X_i} - 1\right) - \left(\frac{I_t}{I_{t-1}} - 1\right)$ for $t \in \{T+1, \ldots, T'\}$ and $\overline{r}^D = \frac{1}{T'-T} \sum_{t \in \{T+1,\ldots,T'\}} r_t^D$. Note that the TE_{TEV} is the out-of-sample tracking-error measure that corresponds to the TEV that is optimized in-sample by LBH, IGH, and M-Q.

- TE_{MAD}: annualized average absolute difference between the portfolio values and the index values during the out-of-sample period, calculated as:

$$\frac{52}{T'-T} \sum_{t \in \{T+1,\ldots,T'\}} \left| \sum_{i \in U \cup \{n+1\}} P_{it} X_i - I_t \right| \frac{100}{C} \tag{47}$$

Note that the TE_{MAD} is the out-of-sample tracking-error measure that corresponds to the MAD that is optimized in-sample by M-L. Also note that the TE_{MAD} is scaled to an investment budget of 100.

- ER [%]: annualized difference between the cumulated portfolio return and the cumulated index return during the out-of-sample period, calculated as:

$$\left(\frac{\sum\limits_{i\in U\cup\{n+1\}} P_{iT'}X_i}{\sum\limits_{i\in U\cup\{n+1\}} P_{iT}X_i}\right)^{\frac{52}{T'-T}} - \left(\frac{I_{T'}}{I_T}\right)^{\frac{52}{T'-T}} \tag{48}$$

All calculations were performed on an HP Z820 workstation with two 3.1 GHz Intel Xeon CPUs and 128 GB of RAM. As the termination criterion, we prescribed a computational time limit of 60 seconds for the heuristic solution approaches and a fairly longer computational time limit of 300 seconds for the exact solution approaches. We implemented the two exact solution approaches and the two heuristic solution approaches in C, and we used Gurobi 7.5 as the solver. We used the default solver settings, except for the MILP formulation (M-C) that is solved in Algorithm 1, in which we stopped the Gurobi solver as soon as the MIP gap reached a value of 10% or lower.

5.2. Novel problem instances

To the best of our knowledge, there is no set of instances available in the literature for the EITP. In the existing sets of problem instances (cf., e.g., Beasley et al. [2], Canakgoz and Beasley [5], Guastaroba et al. [12], and Strub and Baumann [30]), the weights of the stocks in the index at the time of rebalancing, i.e., at the end of week T, are not provided. Furthermore, no available set of problem instances contains very large regional and global stock-market indices. The largest existing problem instance that corresponds to the Russell 3000 index consists of fewer than 2,500 US stocks. Hence, we here provide a set of novel instances of the EITP. These instances are based on real-world data from nine different stock-market indices maintained by Thomson Reuters (TR) for two different time periods. Table 4 lists the names of the indices, the number of stocks n in the indices, the considered time periods, and the applicable values for k. Combining the nine different indices, the two periods, the applicable values for k, the three considered scenarios, and the three consider α values, we obtain a set of 297 instances in total. We used DATASTREAM to download 156 weekly values of each index and 156 weekly closing prices of the constituents of each index during the corresponding time period. We consider the constituents of the index at the end of the in-sample period, i.e., at the end of week 104. As done by Beasley et al. [2], Canakgoz and Beasley [5], and Strub and Baumann [30], we disregard the constituents for which the price data for the considered 156 weeks are incomplete; thus, the number of stocks n in the index can differ between the two different time periods. We also provide the weight of each constituent in the index at the end of the in-sample period. For all indices, the sum of the original weights of the stocks with complete price data is at least 95% for both periods. The weights of these index constituents are then scaled for each index such that their sum is equal to one.

Table 4: Problem instances.

	Index	n	Time period	k
Period 1	TR Africa	168	08/2012–07/2015	100
	TR Latin America	194	08/2012–07/2015	100
	TR Europe	1,310	08/2012–07/2015	100, 200
	TR United States	1,592	08/2012–07/2015	100, 200
	TR North America	1,866	08/2012–07/2015	100, 200
	TR Global Emerging Markets	2,912	08/2012–07/2015	100, 200
	TR Asia Pacific	5,018	08/2012–07/2015	100, 200
	TR Global Developed Markets	5,965	08/2012–07/2015	100, 200
	TR Global	8,877	08/2012–07/2015	100, 200
Period 2	TR Africa	168	08/2013–07/2016	100
	TR Latin America	246	08/2013–07/2016	100, 200
	TR Europe	1,504	08/2013–07/2016	100, 200
	TR United States	2,222	08/2013–07/2016	100, 200
	TR Global Emerging Markets	2,532	08/2013–07/2016	100, 200
	TR North America	2,620	08/2013–07/2016	100, 200
	TR Asia Pacific	4,663	08/2013–07/2016	100, 200
	TR Global Developed Markets	6,896	08/2013–07/2016	100, 200
	TR Global	9,427	08/2013–07/2016	100, 200

Figure 1 shows the evolution of the value of the TR Global index over the two overlapping periods from August 2012 to July 2015 (period 1) and from August 2013 to July 2016 (period 2). In the figure, both periods are split into three equal parts consisting of 52 weeks each. The first two parts of each period correspond to the in-sample period, and the third part corresponds to the out-of-sample period. As Figure 1 shows, the Black Monday market crash, a drastic downward revision of the growth expectations for China's economy, and the UK referendum regarding the European Union all led to high market volatility during the time frame marked in red. These events did not affect the out-of-sample period of period 1, but strongly impacted the out-of-sample period of period 2. Hence, the differentiation between periods 1 and 2 enables investigation of the performance of portfolios during out-of-sample periods characterized by both low and high market volatility.

5.3. Portfolios without rebalancing: in-sample and out-of-sample performance analysis

In this subsection, we investigate the index-tracking capabilities of the portfolios given in the three scenarios before rebalancing to provide a reference against which to assess the index-tracking capabilities of the portfolios after rebalancing. For this purpose, we present in-sample and out-of-sample results for scenarios I, II, and III under the assumption that the given portfolio is not rebalanced by the investment fund's manager at the end of the in-sample period and thus remains unchanged for the out-of-sample period, i.e., $X_i = Y_i$ for all $i \in U \cup \{n+1\}$. For scenario III, this means that the portfolio after rebalancing still consists of cash only, i.e., $X_{n+1} = Y_{n+1} = \frac{C}{P_{n+1,T}}$ and $X_i = Y_i = 0$ for each stock $i \in U$. Note that these portfolios are not necessarily feasible portfolios; for example, the constraint regarding the prescribed minimum excess return

Figure 1: Considered time periods.

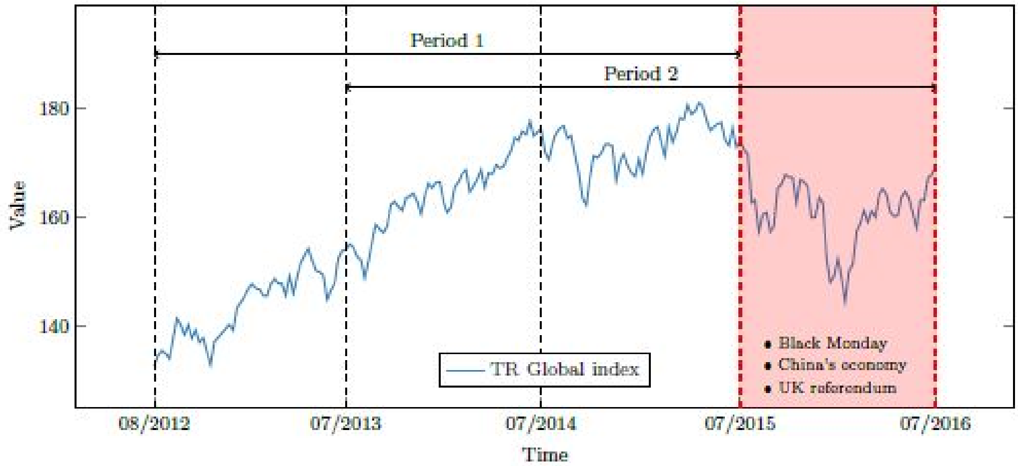


Table 5: In-sample and out-of-sample results for portfolios without rebalancing.

	Period 1				Period 2			
	# INST	OFV	TE_{TEV}	TE_{MAD}	# INST	OFV	TE_{TEV}	TE_{MAD}
Scenario I	16	0.71	2.86	37.09	17	0.75	3.39	58.83
Scenario II	16	19.20	13.13	314.42	17	15.61	13.37	245.51
Scenario III	16	124.62	13.37	244.88	17	137.00	19.49	457.20

or the constraints regarding the minimum and maximum portfolio weights might be violated.

Table 5 summarizes the in-sample and out-of-sample results for these portfolios for period 1 and period 2. Column one indicates the considered scenario and column two shows the number of considered instances for period 1. Columns three, four, and five show the average OFV, specifically the TEV according to (1), the average TE_{TEV}, and the average TE_{MAD}, respectively, for period 1. Columns six to nine report the corresponding results for period 2.

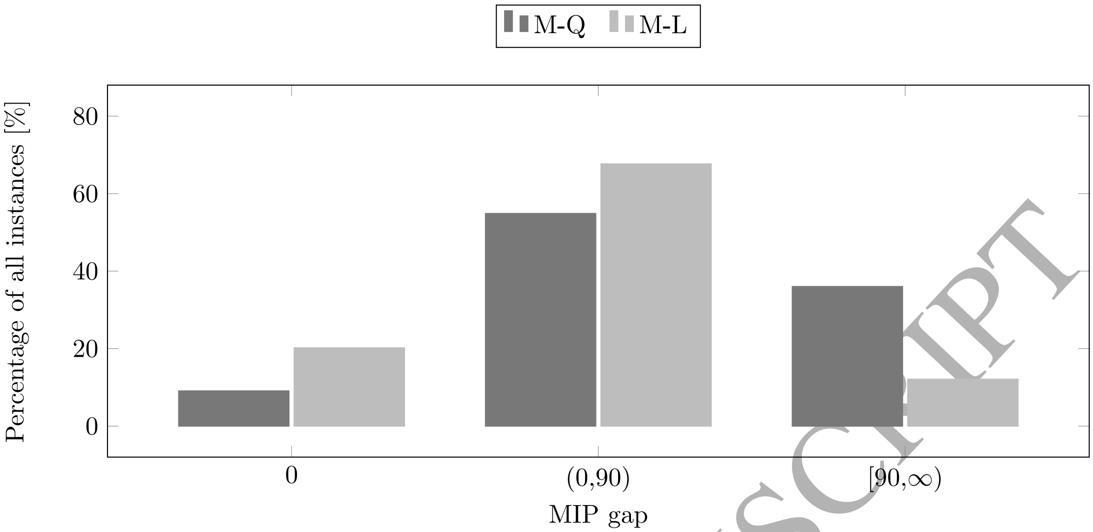
From Table 5, we can gain the following insights. The portfolios of scenario I clearly outperform those of scenarios II and III in terms of the objective function value and both out-of-sample tracking-error measures, regardless of the considered period. Therefore, portfolios that consist of stocks with high weights in the index tend to have a better index-tracking capability than portfolios that consist of stocks with low index weights.

5.4. MIP gaps: M-Q in comparison with M-L

In this subsection, we show that in contrast to M-L, M-Q leads to feasible portfolios within a limited computational time that may be improved substantially in terms of the objective function value. For this purpose, we investigate the in-sample performance in terms of the MIP gap of the portfolios obtained with M-Q and M-L within the prescribed computational time limit of 300 seconds.

Figure 2 shows the distribution of the in-sample results in terms of the MIP gap for M-Q and M-L. M-L determines provably optimal portfolios for about 20% of all problem instances, which is twice as much as

Figure 2: In-sample results for rebalanced portfolios in terms of the MIP gap.



M-Q does. Furthermore, M-L determines feasible portfolios with a MIP gap larger than 90% for only about 10% of all problem instances. In contrast, the MIP gap of the feasible portfolios determined with M-Q are larger than 90% for more than one third of all problem instances. These results suggest that an exact solution approach may be appropriate to the EITP when the MAD is used as the objective function, but may not be appropriate when the TEV is minimized. Hence, improvements in terms of the objective function value might be achieved by applying heuristics to the EITP when the TEV is used as the objective function, which is supported by the results provided in the following.

5.5. In-sample performance analysis: LBH and IGH in comparison with M-Q

In this subsection, we show that LBH and IGH achieve substantial improvements in terms of the objective function value compared to M-Q within a limited computational time. For this purpose, we present in-sample results for the portfolios of scenarios I, II, and III after rebalancing using the three considered solution approaches.

Table 6 summarizes the in-sample results in terms of the OFV, i.e., the TEV. Columns one to four indicate the considered scenario, the maximum portfolio cardinality k, the annualized prescribed minimum expected excess return α, and the number of considered instances, respectively. Columns five, seven, and nine show the average OFV for the solution approaches M-Q, LBH, and IGH, respectively; the best average OFV in each row is shown in bold. Columns six, eight, and ten show the numbers of instances for which M-Q, LBH, and IGH, respectively, are not able to find a feasible portfolio within the prescribed computational time limit. Note that we exclude all instances for which not all three considered solution approaches devise at least a feasible portfolio from the calculation of the average OFV.

From Table 6, we can gain the following insights:

Table 6: In-sample results for rebalanced portfolios in terms of the OFV.

		α p.a.	# INST	M-Q (300s) OFV	M-Q (300s) # NFP	LBH (60s) OFV	LBH (60s) # NFP	IGH (60s) OFV	IGH (60s) # NFP
Scenario I	$k = 100$	0%	18	13.22	0	0.55	0	**0.48**	0
		1%	18	13.05	0	0.57	0	**0.49**	0
		2%	18	16.33	1	0.57	0	**0.49**	0
	$k = 200$	0%	15	17.72	0	**0.24**	0	0.26	0
		1%	15	11.01	0	**0.25**	0	0.26	0
		2%	15	16.64	0	**0.27**	0	**0.27**	0
Scenario II	$k = 100$	0%	18	16.77	0	13.98	0	**13.25**	0
		1%	18	17.33	0	14.12	0	**12.71**	0
		2%	18	18.44	0	**12.31**	0	12.92	0
	$k = 200$	0%	15	15.77	1	5.10	0	**4.98**	0
		1%	15	17.07	1	4.31	0	**4.27**	0
		2%	15	14.30	1	4.24	0	**4.06**	0
Scenario III	$k = 100$	0%	18	55.73	0	**0.42**	0	0.43	0
		1%	18	55.96	0	**0.42**	0	0.43	0
		2%	18	56.21	0	**0.43**	0	0.44	0
	$k = 200$	0%	15	66.75	0	**0.17**	0	**0.17**	0
		1%	15	67.02	0	**0.17**	0	0.18	0
		2%	15	67.31	0	**0.17**	0	0.18	0

- Within the prescribed computational time limit of 60 seconds, LBH and IGH are able to find much better portfolios in terms of the objective function value than M-Q is within 300 seconds.
- LBH and IGH are able to find feasible portfolios for all considered problem instances within the prescribed computational time limit, whereas this is not the case for M-Q. This finding demonstrates that constructing a feasible portfolio for the considered problem is not straightforward.
- Compared with the portfolios before rebalancing (cf. Table 5), the portfolios found by using LBH and IGH have considerably lower objective function values.
- IGH tends to find better portfolios in terms of the objective function value than LBH does for scenarios I and II. The opposite is true for scenario III, in which the investment fund has a portfolio before rebalancing that consists only of cash. Hence, LBH should be applied to construct a portfolio from cash, and IGH should be applied to rebalance an existing stock portfolio.

5.6. Out-of-sample performance analysis

In this subsection, we provide insights that minimizing the TEV may be superior to minimizing the MAD in terms of the out-of-sample tracking error. In Subsection 5.6.1, considering only the instances that could be solved to proven optimality by M-Q and M-L, we show that the portfolios obtained with M-Q are superior to those obtained with M-L, interestingly in terms of both the TE_{TEV} and the TE_{MAD}. In Subsection 5.6.2, we show that the feasible portfolios determined with LBH and IGH are superior to the feasible portfolios determined with M-L also in terms of both the TE_{TEV} and the TE_{MAD}.

Table 7: Out-of-sample tracking errors for instances solved to proven optimality by both M-Q and M-L.

		TE_{TEV}		TE_{MAD}	
	# INST	M-Q (300s)	M-L (300s)	M-Q (300s)	M-L (300s)
Scenario I	9	**3.92**	6.54	**27.06**	71.48
Scenario II	9	**4.16**	7.99	**91.61**	105.42
Scenario III	9	**4.00**	6.50	**53.93**	65.48
Average		**4.03**	7.01	**57.54**	80.80

5.6.1. M-Q in comparison with M-L

We compare the out-of-sample performance of the portfolios of scenarios I, II, and III after rebalancing using M-Q and M-L for the instances that could be solved to proven optimality by both solution approaches within the prescribed computational time limit.

Table 7 summarizes the out-of-sample tracking errors for these portfolios. Column one indicates the considered scenario and column two shows the number of considered instances for each scenario. Columns three and four show the average TE_{TEV} for the approaches M-Q and M-L, respectively. Columns five and six report the average TE_{MAD} for the same approaches.

From Table 7, we can gain the insight that M-Q leads to lower out-of-sample tracking errors in terms of the TE_{TEV} but also in terms of the TE_{MAD}, even though the portfolios determined with M-L have the lowest possible in-sample MAD.

Figure 3 shows the out-of-sample excess returns against the out-of-sample tracking errors of the rebalanced portfolios obtained using the solution approaches M-Q and M-L for the instances that could be solved to proven optimality by both solution approaches within the prescribed computational time limit. This figure indicates that M-Q is superior to M-L with respect to both objectives in enhanced index tracking, i.e., obtaining a low tracking error and achieving a small target excess return in the out-of-sample period.

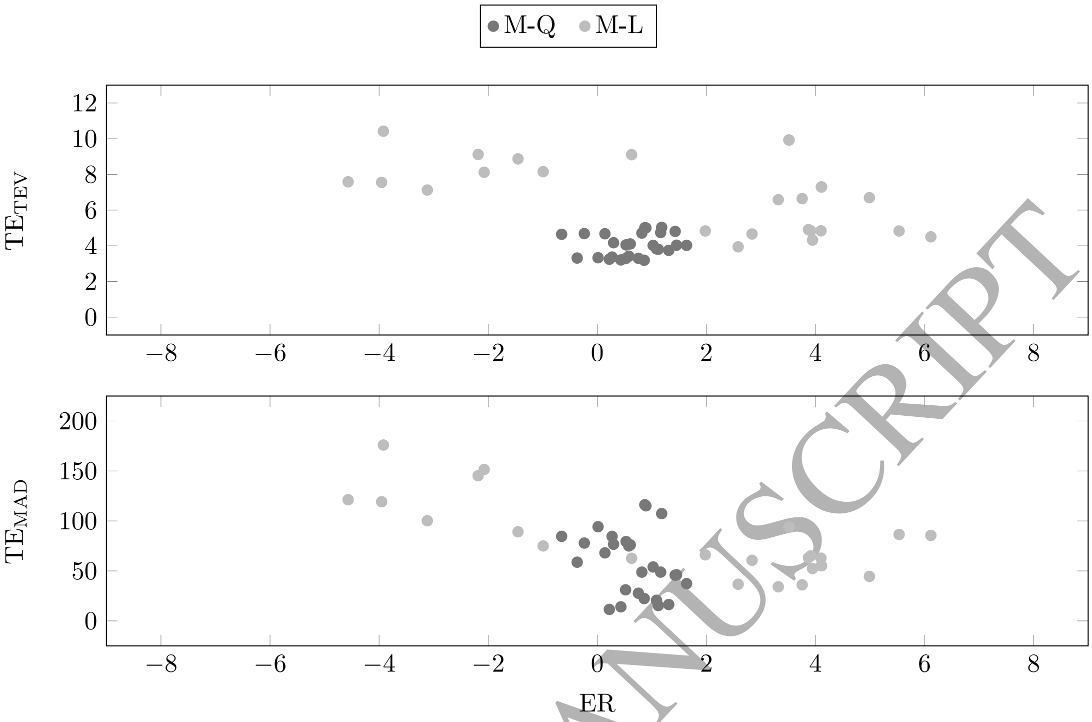
These findings suggest that minimizing the TEV leads to better enhanced index-tracking portfolios than minimizing the MAD.

5.6.2. LBH and IGH in comparison with M-L

We compare the out-of-sample tracking errors of the portfolios of scenarios I, II, and III after rebalancing using LBH, IGH, and M-L. Furthermore, we provide insights on the impact of the prescribed computational time limit on the out-of-sample tracking errors of the portfolios obtained with M-L. Finally, we provide out-of-sample risk-return characteristics of the rebalanced portfolios.

Table 8 summarizes the out-of-sample results for periods 1 and 2 individually. Column one indicates the considered period. The contents of columns two to five are the same as those of columns one to four of Table 6. Columns six to eight show the average TE_{TEV} for the considered instances for the solution approaches M-L, LBH, and IGH, respectively. Columns nine to eleven present the average TE_{MAD} for the same instances and the same solution approaches. The best average TE_{TEV} and the best average TE_{MAD} in each row are shown

Figure 3: Out-of-sample risk-return characteristics for instances solved to proven optimality by M-Q and M-L.



in bold. Note that all considered solution approaches found at least a feasible solution for all instances.

From Table 8, we can gain the following insights:

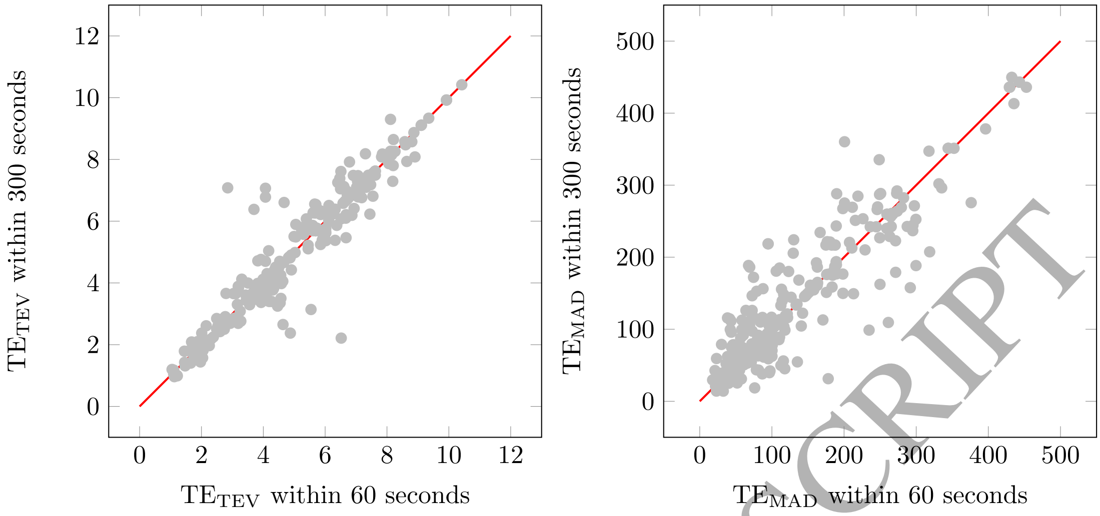
- Regardless of the period considered, the average TE_{TEV} for the portfolios obtained with LBH and IGH within 60 seconds are much lower than those for the portfolios obtained with M-L within 300 seconds. Interestingly, LBH and IGH also devise portfolios with a much lower average TE_{MAD}, even though M-L explicitly aims at minimizing the MAD during the in-sample period.
- LBH and IGH lead to considerably lower worst-case TE_{TEV} and TE_{MAD} than M-L does, regardless of the period considered. These empirical findings support the arguments presented in Subsection 3.2.
- For all considered solution approaches, the average TE_{TEV} is higher for period 2 than for period 1 because the out-of-sample period of period 2 exhibits higher market volatility than that of period 1.
- For scenario II, in which the index-tracking capability of the portfolio before rebalancing is rather poor, the average TE_{TEV} are higher than those for the other scenarios. This is consistent with the higher objective function values, i.e., the higher TEV, for scenario II (cf. Table 6).

Figure 4 shows the out-of-sample tracking errors of the rebalanced portfolios obtained with the exact solution approach M-L within 60 and 300 seconds. For the portfolios that are below the red line, the longer computational time limit leads to a performance improvement in terms of the TE_{TEV} or the TE_{MAD}. By contrast, the corresponding out-of-sample tracking-error measure deteriorates for the portfolios that are above

Table 8: Out-of-sample tracking errors for rebalanced portfolios.

					$\mathrm{TE_{TEV}}$			$\mathrm{TE_{MAD}}$		
			α p.a.	# INST	M-L (300s)	LBH (60s)	IGH (60s)	M-L (300s)	LBH (60s)	IGH (60s)
Period 1	Scenario I	$k=100$	0%	9	3.81	2.89	**2.77**	74.06	**36.99**	44.10
			1%	9	3.77	2.90	**2.79**	88.49	**38.63**	41.50
			2%	9	4.10	2.91	**2.80**	70.49	**39.98**	49.12
		$k=200$	0%	7	2.40	**2.01**	2.03	72.63	**19.87**	21.29
			1%	7	2.33	2.07	**2.06**	75.18	**21.75**	24.63
			2%	7	2.25	2.12	**2.02**	68.61	**19.60**	27.54
	Scenario II	$k=100$	0%	9	6.38	4.44	**4.05**	195.97	**123.75**	167.05
			1%	9	6.70	4.32	**4.12**	200.17	109.18	**106.90**
			2%	9	6.75	**4.17**	4.29	211.49	148.26	**128.11**
		$k=200$	0%	7	5.94	3.86	**3.82**	196.66	129.62	**126.75**
			1%	7	5.94	3.79	**3.63**	164.26	128.18	**106.23**
			2%	7	5.90	3.79	**3.77**	198.45	137.32	**131.99**
	Scenario III	$k=100$	0%	9	5.55	**2.72**	2.75	139.05	61.91	**60.22**
			1%	9	5.63	**2.67**	2.73	158.43	61.75	**54.25**
			2%	9	5.72	2.76	**2.73**	150.18	**51.97**	54.91
		$k=200$	0%	7	4.77	**1.93**	1.94	153.18	57.14	**56.20**
			1%	7	5.09	**1.92**	**1.92**	172.92	**45.32**	51.31
			2%	7	5.17	**1.91**	1.94	169.58	60.90	**58.09**
				Average	4.96	3.00	**2.94**	142.33	**72.15**	73.50
				Worst case	10.42	5.80	**5.76**	360.26	325.82	**288.16**
Period 2	Scenario I	$k=100$	0%	9	3.43	3.11	**2.81**	55.23	**47.06**	48.91
			1%	9	3.49	3.16	**2.89**	57.81	**48.44**	61.07
			2%	9	3.43	3.15	**2.98**	84.97	**51.51**	52.45
		$k=200$	0%	8	3.40	**2.30**	2.31	51.92	41.71	**38.36**
			1%	8	3.03	2.33	**2.27**	50.91	43.52	**41.97**
			2%	8	2.91	2.35	**2.31**	64.69	41.96	**40.40**
	Scenario II	$k=100$	0%	9	6.08	5.23	**4.69**	155.74	90.58	**85.23**
			1%	9	5.83	**5.11**	5.14	179.66	94.28	**80.88**
			2%	9	6.04	**4.85**	5.12	187.62	**83.59**	96.51
		$k=200$	0%	8	7.13	4.14	**4.10**	219.27	80.06	**67.09**
			1%	8	7.12	**4.18**	4.23	213.18	77.50	**67.84**
			2%	8	7.21	4.18	**4.13**	219.26	78.55	**77.00**
	Scenario III	$k=100$	0%	9	5.63	**2.79**	2.87	90.81	72.54	**64.17**
			1%	9	5.71	2.84	**2.82**	115.90	**66.57**	70.31
			2%	9	5.81	**2.83**	2.87	118.59	**66.38**	75.49
		$k=200$	0%	8	5.77	2.27	**2.21**	110.04	69.82	**65.64**
			1%	8	5.67	**2.20**	2.23	118.48	74.46	**73.53**
			2%	8	5.79	**2.25**	**2.25**	127.51	**68.33**	71.38
				Average	5.18	3.31	**3.25**	123.00	66.64	**65.76**
				Worst case	9.92	**7.19**	7.78	449.74	168.53	**168.45**

Figure 4: Out-of-sample tracking errors for rebalanced portfolios obtained with M-L within 60 and 300 seconds.



the red line. Figure 4 shows that the out-of-sample tracking errors of the rebalanced portfolios obtained within 60 seconds are similar to those of the portfolios devised within 300 seconds; for many portfolios obtained within 300 seconds the $\mathrm{TE}_{\mathrm{TEV}}$ and the $\mathrm{TE}_{\mathrm{MAD}}$ are even worse. This suggests that the results provided in Table 8 for M-L would not change considerably if a computational time limit of more than 300 seconds was imposed.

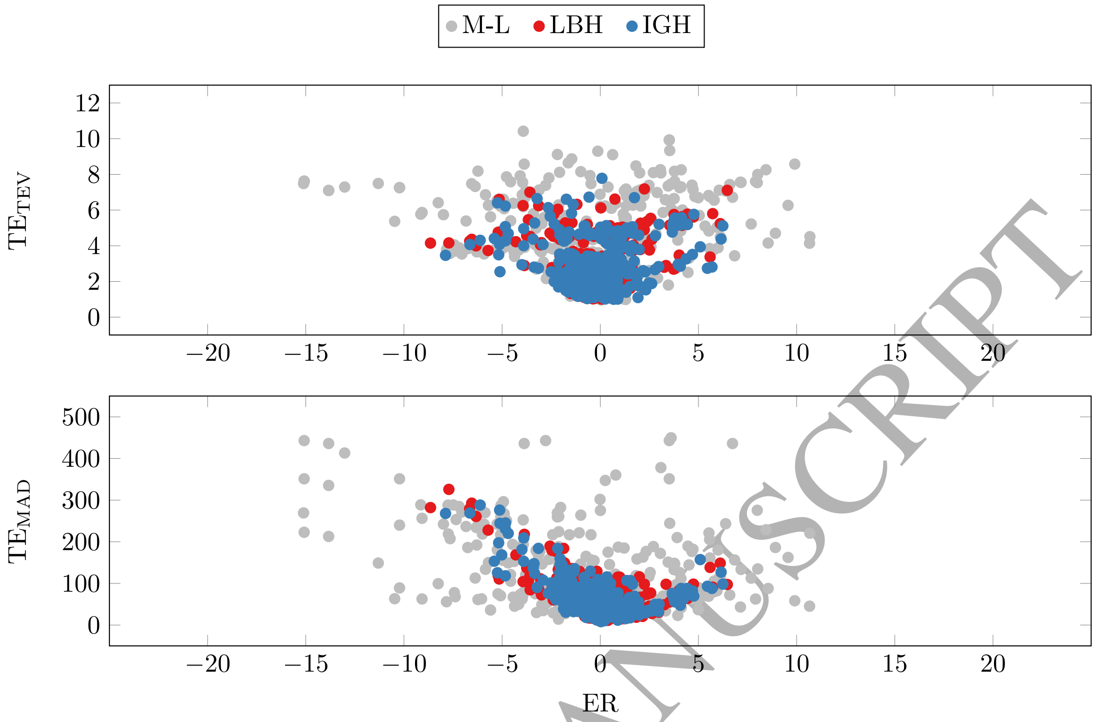
Figure 5 shows the out-of-sample excess returns against the out-of-sample tracking errors of the rebalanced portfolios obtained using the solution approaches M-L, LBH, and IGH. LBH and IGH devise portfolios with a performance that is similar to M-L in terms of the out-of-sample excess returns. Furthermore, the out-of-sample excess returns of the portfolios obtained with LBH and IGH are much less volatile. Finally, Figure 5 shows that a portfolio that achieves the prescribed minimum expected excess return during the in-sample period is not necessarily guaranteed to achieve an excess return during the out-of-sample period as there are many portfolios with a negative out-of-sample excess return.

5.7. Portfolio compositional characteristics: LBH and IGH in comparison with M-L

In this subsection, we offer further insights with respect to the compositions of the rebalanced portfolios. For this purpose, we report various compositional characteristics of portfolios that have been rebalanced using the solution approaches LBH, IGH, and M-L.

Figure 6 shows the following compositional characteristics for the portfolios of scenarios I, II, and III after rebalancing: the active share, i.e., the sum of the absolute differences between the weights of the assets in the portfolio and the weights of the assets in the index ($\frac{1}{2}\sum_{i\in U\cup\{n+1\}}\left|\frac{P_{iT}X_i}{C}-w_i^I\right|$); the portfolio cardinality, i.e., the number of different stocks that are selected after rebalancing ($|\{i \in I : X_i > 0\}|$); the transaction costs, i.e., the sum of the fixed and proportional transaction costs relative to the transaction cost budget ($\frac{1}{\gamma C}(\sum_{i\in I} G_i + \sum_{i\in U\setminus I: Y_i>0}(c_i^s Y_i P_{iT} + c_i^f))$); and the weight of the cash asset ($\frac{P_{n+1,T}X_{n+1}}{C}$). We present the

Figure 5: Out-of-sample risk-return characteristics for rebalanced portfolios.



compositional characteristics for all instances with $k = 200$ and $\alpha = 0\%$ sorted in non-decreasing order of n. The compositional characteristics for the instances with $k = 100$ and $\alpha > 0\%$ are not shown because they are similar to those presented in Figure 6. For reference, we also present the compositional characteristics in the case that no rebalancing is performed.

From Figure 6, we can gain the following insights:

- LBH and IGH lead to portfolios with lower active shares and higher cardinalities than M-L does, regardless of the considered scenario. Thus, the compositions of the portfolios that have been rebalanced using LBH and IGH are more similar to the composition of the index. This finding is consistent with the arguments presented in Subsection 3.2 and offers a possible explanation for the superior performance of LBH and IGH in terms of both out-of-sample tracking-error measures, the $\mathrm{TE}_{\mathrm{TEV}}$ and the $\mathrm{TE}_{\mathrm{MAD}}$, reported in Table 8.

- In scenario I, LBH and IGH incur lower transaction costs than M-L does because the portfolio before rebalancing is already invested in the k stocks with the highest index weights. In scenarios II and III, the transaction costs for LBH and IGH are higher than those for M-L. To rebalance the portfolios with a poor index-tracking capability of scenario II, LBH and IGH completely exhaust the transaction cost budget for all problem instances.

- For scenarios I and III, LBH and IGH lead to portfolios in which the weight of the cash asset is almost zero for all instances. By contrast, M-L leads to portfolios that hold a substantial amount of cash after rebalancing for scenario III. For scenario II, in which the index-tracking capacity of the portfolio before rebalancing is rather poor, LBH and IGH also lead to portfolios that contain a considerable proportion of cash. This might be because the transaction cost budget is not sufficiently large to sell all currently held stocks with low index weights and exchange them for stocks with high index weights. In this case, it is more beneficial in terms of the objective function value to sell the stocks with low index weights and maintain the revenue in cash.

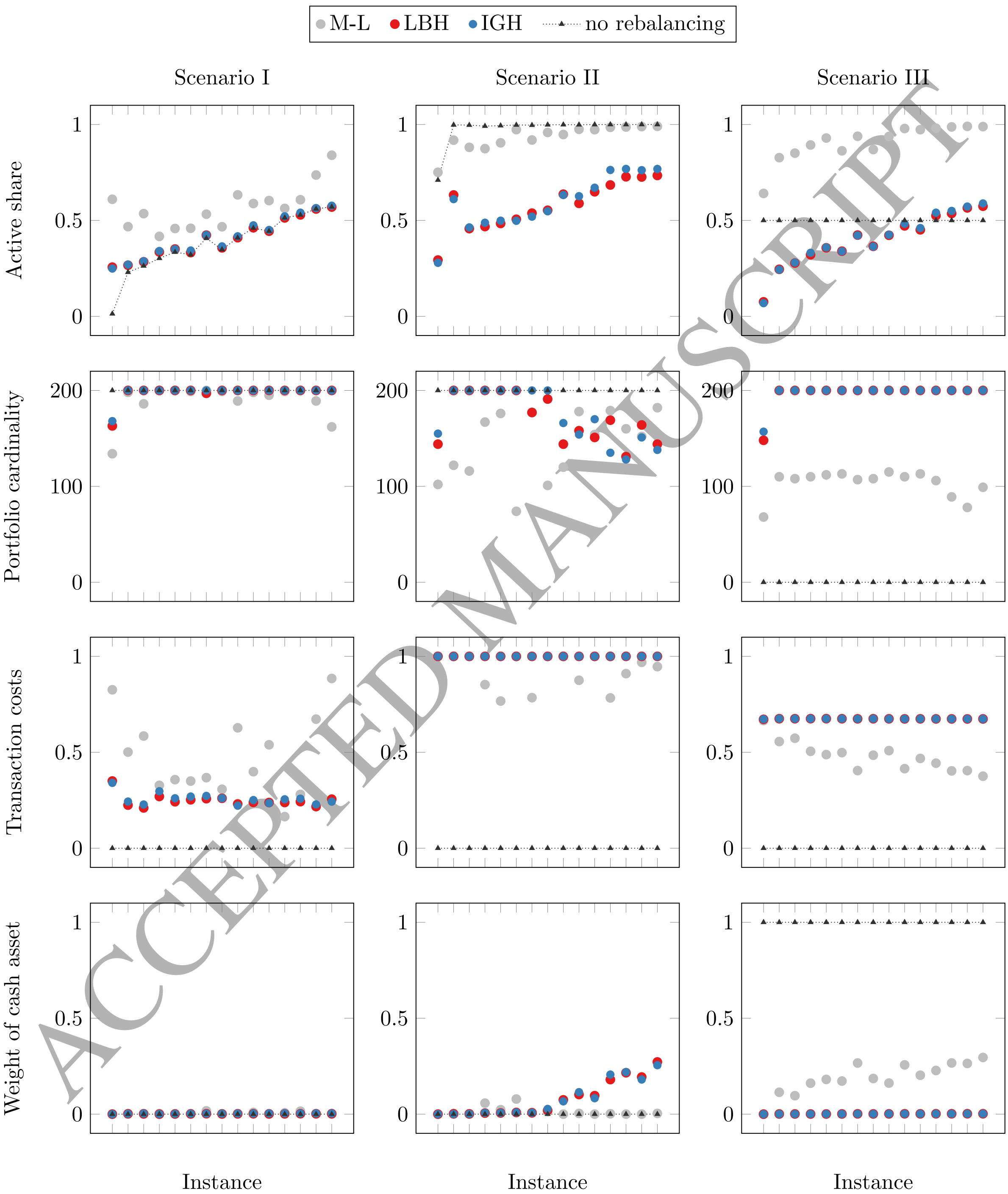
From the results provided in this section, the six main findings are as follows. 1) In contrast to M-L, M-Q leads to feasible portfolios within a limited computational time that may be improved substantially in terms of the objective function value, which indicates the possible improvements that might be achieved by applying LBH and IGH. 2) LBH and IGH are indeed able to determine considerably better feasible portfolios in terms of the objective function value within 60 seconds than M-Q is within 300 seconds. 3) LBH should be used to construct a portfolio from cash. 4) IGH should be used to rebalance an existing stock portfolio. 5) In terms of both out-of-sample tracking-error measures, the TE_{TEV} and the TE_{MAD}, considering only the instances that could be solved to proven optimality by both M-Q and M-L, the portfolios obtained with M-Q are superior to those obtained with M-L. 6) In terms of both out-of-sample tracking-error measures, the TE_{TEV} and the TE_{MAD}, the feasible portfolios determined by running LBH and IGH for 60 seconds are superior to the feasible portfolios determined with M-L within 300 seconds. These findings suggest that minimizing the TEV is superior to minimizing a dissimilarity function such as the MAD in terms of the out-of-sample tracking error.

6. Conclusions

In this paper, we consider the problem of determining the portfolio for an enhanced index-tracking fund. For this problem, we propose novel mixed-integer linear and quadratic programming formulations and novel matheuristics to minimize the tracking error variance or the mean-absolute deviation between the historical values of the portfolio and the index. The results of a computational experiment using the proposed matheuristics and exact solution approaches based on the mixed-integer linear and quadratic programming formulations suggest that it is superior in terms of the out-of-sample tracking error to minimize the tracking error variance instead of the mean-absolute deviation.

In future research, it would be interesting to investigate whether the out-of-sample tracking error achieved here could be further improved by using more sophisticated estimators for the covariance matrix. Periodic portfolio rebalancing could be applied to further improve the out-of-sample performance of the presented solution approaches, especially in market environments with high volatility. Moreover, a promising direction for future research is to combine the proposed matheuristics based on the findings regarding their individual strengths to further improve the out-of-sample tracking error.

Figure 6: Compositional characteristics of rebalanced portfolios.



References

References

[1] Andriosopoulos, K., Nomikos, N., 2014. Performance replication of the Spot Energy Index with optimal equity portfolio selection: Evidence from the UK, US and Brazilian markets. European Journal of Operational Research 234 (2), 571–582.

[2] Beasley, J. E., Meade, N., Chang, T.-J., 2003. An evolutionary heuristic for the index tracking problem. European Journal of Operational Research 148 (3), 621–643.

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

[4] Bruni, R., Cesarone, F., Scozzari, A., Tardella, F., 2015. A linear risk-return model for enhanced indexation in portfolio optimization. OR Spectrum 37 (3), 735–759.

[5] Canakgoz, N. A., Beasley, J. E., 2008. Mixed-integer programming approaches for index tracking and enhanced indexation. European Journal of Operational Research 196 (1), 384–399.

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

[7] Cordeau, J.-F., Gendreau, M., Laporte, G., Potvin, J.-Y., Semet, F., 2002. A guide to vehicle routing heuristics. Journal of the Operational Research society 53 (5), 512–522.

[8] Filippi, C., Guastaroba, G., Speranza, M., 2016. A heuristic framework for the bi-objective enhanced index tracking problem. Omega 65, 122–137.

[9] Fischetti, M., Lodi, A., 2003. Local branching. Mathematical Programming 98 (1–3), 23–47.

[10] Gaivoronski, A. A., Krylov, S., Van der Wijst, N., 2005. Optimal portfolio selection and dynamic benchmark tracking. European Journal of Operational Research 163 (1), 115–131.

[11] Guastaroba, G., Mansini, R., Ogryczak, W., Speranza, M. G., 2016. Linear programming models based on Omega ratio for the enhanced index tracking problem. European Journal of Operational Research 251 (3), 938–956.

[12] Guastaroba, G., Mansini, R., Speranza, M. G., 2009. On the effectiveness of scenario generation techniques in single-period portfolio optimization. European Journal of Operational Research 192 (2), 500–511.

[13] Guastaroba, G., Speranza, M. G., 2012. Kernel Search: An application to the index tracking problem. European Journal of Operational Research 217 (1), 54–68.

[14] Jansen, R., Van Dijk, R., 2002. Optimal benchmark tracking with small portfolios. The Journal of Portfolio Management 28 (2), 33–39.

[15] Jorion, P., 2003. Portfolio optimization with tracking-error constraints. Financial Analysts Journal 59 (5), 70–82.

[16] Konno, H., Wijayanayake, A., 2001. Minimal cost index tracking under nonlinear transaction costs and minimal transaction unit constraints. International Journal of Theoretical and Applied Finance 4 (6), 939–957.

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

[18] Kwiatkowski, J. W., 1992. Algorithms for index tracking. IMA Journal of Management Mathematics 4 (3), 279–299.

[19] Ledoit, O., Wolf, M., 2004. Honey, I shrunk the sample covariance matrix. The Journal of Portfolio Management 30 (4), 110–119.

[20] Ledoit, O., Wolf, M., 2004. A well-conditioned estimator for large-dimensional covariance matrices. Journal of Multivariate Analysis 88 (2), 365–411.

[21] Maringer, D., Oyewumi, O., 2007. Index tracking with constrained portfolios. Intelligent Systems in Accounting, Finance and Management 15 (1-2), 57–71.

[22] Mutunge, P., Haugland, D., 2018. Minimizing the tracking error of cardinality constrained portfolios. Computers & Operations Research 90, 33–41.

[23] Roll, R., 1992. A mean/variance analysis of tracking error. The Journal of Portfolio Management 18 (4), 13–22.

[24] Rudd, A., 1980. Optimal selection of passive portfolios. Financial Management, 57–66.

[25] Rudolf, M., Wolter, H.-J., Zimmermann, H., 1999. A linear model for tracking error minimization. Journal of Banking & Finance 23 (1), 85–103.

[26] Ruiz, R., Stützle, T., 2007. A simple and effective iterated greedy algorithm for the permutation flowshop scheduling problem. European Journal of Operational Research 177, 2033–2049.

[27] Sant'Anna, L. R., Filomena, T. P., Caldeira, J. F., 2017. Index tracking and enhanced indexing using cointegration and correlation with endogenous portfolio selection. The Quarterly Review of Economics and Finance 65, 146–157.

[28] Sant'Anna, L. R., Filomena, T. P., Guedes, P. C., Borenstein, D., 2017. Index tracking with controlled number of assets using a hybrid heuristic combining genetic algorithm and non-linear programming. Annals of Operations Research 258 (2), 849–867.

[29] Scozzari, A., Tardella, F., Paterlini, S., Krink, T., 2013. Exact and heuristic approaches for the index tracking problem with UCITS constraints. Annals of Operations Research 205 (1), 235–250.

[30] Strub, O., Baumann, P., 2018. Optimal construction and rebalancing of index-tracking portfolios. European Journal of Operational Research 264 (1), 370–387.

[31] Strub, O., Trautmann, N., 2016. An iterated greedy heuristic for the 1/N portfolio tracking problem. In: Vitoriano, B., Parlier, G., de Werra, D. (Eds.), Proceedings of the 5th International Conference on Operations Research and Enterprise Systems. Rome, pp. 424–431.

[32] Takeda, A., Niranjan, M., Gotoh, J., Kawahara, Y., 2013. Simultaneous pursuit of out-of-sample performance and sparsity in index tracking portfolios. Computational Management Science 10 (1), 21–49.

[33] Wu, D., Kwon, R. H., Costa, G., 2017. A constrained cluster-based approach for tracking the S&P 500 index. International Journal of Production Economics 193, 222–243.