European Journal of Operational Research 196 (2008) 384–399

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

O.R. Applications

Mixed-integer programming approaches for index tracking and enhanced indexation

N.A. Canakgoz, J.E. Beasley *

CARISMA, Mathematical Sciences, Brunel University, Uxbridge UB8 3PH, United Kingdom

ARTICLE INFO

Article history:
Received 27 April 2007
Accepted 7 March 2008
Available online 15 March 2008

Keywords:
Index tracking
Enhanced indexation
Passive fund management

ABSTRACT

We consider the problem of reproducing the performance of a stock market index, but without purchasing all of the stocks that make up the index, index tracking. We also consider the problem of out-performing the index, enhanced indexation. We present mixed-integer linear programming formulations of these problems. Our formulations include transaction costs, a constraint limiting the number of stocks that can be purchased and a limit on the total transaction cost that can be incurred. As our formulations of these problems are mixed-integer linear programs we can use a standard solver (Cplex). Numeric results are presented for eight data sets drawn from major markets. The largest of these data sets involves over 2000 stocks.

1. Introduction

Index tracking is a popular form of passive fund (portfolio) management where the concern is to reproduce (track) the performance of a stock market index. The simplest way to track an index is full replication, where all of the stocks that make up the index are purchased in the same proportions as in the index. However full replication has a number of disadvantages, in particular certain stocks may be held in very small quantities; when the composition of the index is revised the holdings of all stocks will need to be changed and transaction costs associated with the subsequent sale and purchase of stocks may be high. Moreover in large indices (indices containing many stocks) the stocks associated with smaller companies may be relatively illiquid, with consequent comparatively high transaction costs. Because of these disadvantages many passively managed funds, especially those that are tracking large indices, hold fewer stocks than are included in the index they are tracking.

Enhanced indexation (sometimes referred to as enhanced index tracking) aims to reproduce the performance of a stock market index, but to generate excess return (return over and above the return achieved by the index). One phrase often encountered with regard to enhanced indexation is "adding alpha". Other phrases using the word "alpha" also exist, e.g. "index plus alpha" and "alpha tilt". All of these relate to a regression based view of enhanced indexation, where a regression of the returns achieved by an enhanced indexation portfolio, against the returns achieved by the index, has a higher regression intercept (alpha) than other portfolios. Note here that for a portfolio which perfectly tracks the index if we regress the returns from the portfolio against the returns from the index we would have a regression intercept (alpha) of zero, and a regression slope (beta) of one.

In this paper we present related mixed-integer linear programming formulations for index tracking and enhanced indexation. Our formulation for enhanced indexation is a simple extension of our formulation for index tracking. In our approach we include transaction costs, a constraint limiting the number of stocks that can be purchased and a limit on the total transaction cost that can be incurred. For convenience we use the single phrase "tracking portfolio" below to signify both the portfolio of stocks that we use to track an index, and the enhanced indexation portfolio that we use to out-perform an index, which of these two being referred to being deduced from the context.

The structure of this paper is as follows. In Section 2 we present a literature survey relating to index tracking and enhanced indexation. In Section 3 we present our formulation of the index tracking problem and extend it to deal with enhanced indexation. In Section 4 we present computational results and in Section 5 we conclude the paper.

2. Literature survey

Many earlier papers relating to index tracking were discussed previously in 2003 in Beasley et al. [5]. For this reason only literature additional to that presented there is discussed below. In order to structure our literature review we consider index tracking and enhanced indexation separately.

* Corresponding author. Tel.: +44 1895 266219; fax: +44 1895 269723.
E-mail addresses: nilgun.canakgoz@brunel.ac.uk (N.A. Canakgoz), john.beasley@brunel.ac.uk (J.E. Beasley).

0377-2217/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.ejor.2008.03.015

2.1. Index tracking

Gilli and Kellezi [14], drawing on a working paper later to appear as Beasley et al. [5] in 2003, presented a threshold accepting algorithm for index tracking. In local search algorithms of this type new solutions are accepted if they are no worse (by a specified threshold) than the best solution found so far. They move between solutions by choosing a stock to sell and a stock to buy. Computational results were presented for problems involving up to 528 stocks. The largest problem solved required 22 seconds on a Pentium III (800 MHz).

Jansen and van Dijk [16], considered the problem of tracking error minimisation when the number of stocks in the tracking portfolio is limited. In their approach they minimise a weighted objective function comprising both (continuous) tracking error and the (discrete) number of stocks in the tracking portfolio. The discrete component of this objective function is approximated by a continuous function. Once the set of stocks has been decided they optimise stock weights using a standard quadratic programming approach. They presented computational results for their approach for a number of indices involving up to 250 stocks, however no detailed computational times were given.

Derigs and Nickel [9], presented a simulated annealing based metaheuristic. In their approach stock returns and covariances are derived from a linear multi-factor model, where the factors are based on macro-economic variables. They presented a case study based around an investment trust tracking the German DAX30 index (albeit with additional restrictions, which meant that perfect tracking was not possible). Their investment universe, some 202 stocks, was taken from the DAX30 and STOXX200. Limited computational results were presented. For one problem instance they refer to their algorithm taking 118 seconds on a Pentium III (1000 MHz). They also refer to their approach taking minutes when dealing with an investment universe of 500 stocks used to track the MSCI World Developed Market index.

Okay and Akman [21], used constraint aggregation. They based their formulation on that given in Beasley et al. [5] and after applying constraint aggregation they have a mixed-integer nonlinear programming problem to be solved. Using a data set of 31 stocks from the Hang Seng they concluded that the solutions given by their original formulation and their aggregated formulation are approximately the same. For two test problem instances the total CPU time was 487 seconds for their original formulation, but only 117 seconds for their aggregated formulation.

Focardi and Fabozzi [12], argued for the use of clustering as a methodology for building index tracking portfolios. They proposed using Euclidean distances between stock price series as a basis for hierarchical clustering. Once clusters of stocks have been formed they proposed selecting one (or more) stocks from each cluster to include in the tracking portfolio. They note that their approach can also be used for enhanced indexation. No computational results were presented.

Fang and Wang [11], regarded the index tracking problem as a bi-objective programming problem. One objective relates to mean absolute downside deviation from index return, the other objective to excess return. They proposed a fuzzy decision theory approach which leads to a mathematical model that can be solved by linear programming. A numeric example was reported relating to selecting from a universe of 30 stocks so as to track the Shanghai 180 index, but no computational times were given.

Gaivoronoski et al. [13], discussed various approaches to quantifying the difference between the performance of a tracking portfolio and the performance of the index. They also discussed approaches to rebalancing a tracking portfolio. Their heuristic approach to stock selection is based on first solving an unrestricted problem, and then ranking stocks in terms of their appearance in the unrestricted portfolio. Computational results were presented for 65 stocks from the Oslo stock exchange, however no computational times were given.

Li et al. [18], regarded index tracking as being related to minimising a return weighted squared difference between stock weights in the index and stock weights in the tracking portfolio. They used Stein rule estimation to adjust returns. Results were presented for the Shenzhen 40 index, however no computational times were given.

Oh et al. [20], defined a priority function for each stock comprising a weighted sum of trading volume, market capitalisation and beta (the regression slope when the returns from the stock are regressed against the returns from the index). A simple heuristic using these stock priority functions was applied to decide the stocks in the tracking portfolio. A genetic algorithm was then used to decide the weights associated with each stock in the tracking portfolio. Computational results were given for the Korean KOPSI 200 index, however no computational times were presented.

Coleman et al. [6], considered the problem of tracking error minimisation when the number of assets in the tracking portfolio is limited. In their approach they minimise a weighted objective function comprising both (continuous) tracking error and the (discrete) number of assets in the tracking portfolio (c.f. Jansen and van Dijk [16]). The discrete component of this objective function is progressively approximated by a continuous function. They presented computational results for their approach using the publicly available data of Beasley et al. [5], however no computational times were given.

Colwell et al. [7], regarded the index tracking problem as a hedging problem in continuous time. They adopted a dynamic approach, local risk minimisation, to choose a tracking portfolio which is optimal among all continuous trading strategies that replicate a given index. Their formulation of index tracking does not include transaction costs and does not restrict short selling. Computational results were presented for portfolios formed from nine of the 10 largest stocks in the S&P 500.

Corielli and Marcellino [8], presented an index tracking approach based on assuming that stock prices are driven by a factor model. In their approach the index and its tracking portfolio share the same factor structure. They adopt a simple construction heuristic to add stocks to their tracking portfolio. This works by first ordering factors, then for each factor in turn the stocks that best correlate with it are added to the tracking portfolio. Computational results were presented for their approach as applied to a reconstructed EuroStoxx50 index, but no computation times were given. Here by reconstructed index we mean that the actual index values were replaced by index values computed on the basis of the total set of stocks available for selection.

Yao et al. [24], considered the problem of tracking a financial benchmark with a portfolio containing only a very few assets. In their approach the set of assets contained in the tracking portfolio must be known. They formulated the tracking problem as a stochastic linear quadratic control problem and used semidefinite programming to solve it. Computational results related to tracking the Hang Seng index with portfolios containing just four or five stocks were presented (short selling allowed, no transaction costs), but no computation times were given.

Yu et al. [25], presented a Markowitz model for index tracking where their approach assumes that index tracking relates to constraining the probability that the return from the tracking portfolio falls below index return (downside risk); or to higher order moments of downside risk. They assume that stock returns are jointly normally distributed and that short selling is allowed. They presented a small numeric example using stocks from the Hang Seng index.

Stoyan and Kwon [22], presented an approach to index tracking based on a two-stage stochastic program. Their objective is an unweighted sum of three elements: the absolute difference between tracking portfolio value and index value; the absolute value of deviations from desired investment in specified sectors; number of shares traded. Computational results related to selecting from a universe of 1150 stocks to track a Canadian S&P/TSX index were presented.

2.2. Enhanced indexation

Enhanced indexation is a relatively unconsidered area in the scientific literature. All of the work considered below dates from 2005 or later. Approaches developed for enhanced indexation can typically (with only minor modifications) also be applied to index tracking. As such the algorithms presented in the papers discussed below can also be used for index tracking.

Alexander and Dimitriu [3], applied a cointegration based strategy. In their approach they construct two index series by adding/subtracting from the original index values a constant excess return, alpha. Using their cointegration approach (see [1]) they construct portfolios to track these two index series. They seek to earn excess return by going long on the alpha plus tracking portfolio and shorting the alpha minus tracking portfolio. They used a very simple approach to decide the stocks to include in the tracking portfolios based on ranking stocks by price. They presented computational results for their approach applied to a reconstructed Dow Jones index. No computational times were given.

Dose and Cincotti [10], drawing on Beasley et al. [5], cluster stocks based on a distance measure between two stocks defined using the difference between their stock prices histories (amongst other approaches). This clustering is used to decide which particular stocks should be in the tracking portfolio, given a prespecified number of stocks that are to be included in the tracking portfolio. To decide the investment in each stock they use a weighting parameter lambda (as in Beasley et al. [5]) to tradeoff tracking the index and excess return. They applied their approach to a reconstructed S&P 500 index. Computational results were presented, but no computation times were given.

Konno and Hatagi [17], used a mean absolute deviation objective. In their approach this objective is the difference between index values in each time period (scaled up by a factor alpha, and normalised by the index value at the end of the time period) and the tracking portfolio value (normalised by the tracking portfolio value at the end of the time period). They include transaction cost and their formulation involves minimisation of a separable concave function with linear constraints. They performed a problem reduction test to eliminate variables. Computational results were presented for 225 stocks from the Nikkei 225. Computation times (on a 1.67 GHz pc) varied between 4 and 566 seconds.

Wu et al. [23], presented a goal programming approach. In their formulation, which has two goals, one goal relates to the desired rate of return, the other goal to a desired tracking error (defined in a nonlinear, but additive, fashion). Implicit in the title of their paper and their discussion appears to be the assumption that the desired rate of return is connected (somehow) to index return. Computational results for portfolios relating to the Taiwanese market were presented, however no computational times were given.

In the discussion above we have interpreted enhanced indexation to mean that inherent in the approach is some user defined parameter which is used in an attempt to influence the return, over and above the return from the index, obtained. In Alexander and Dimitriu [3] this is their factor alpha; in Dose and Cincotti [10] their weighting parameter lambda; in Konno and Hatagi [17] their factor alpha; in Wu et al. [23] their desired rate of return. There are papers in the literature relating to approaches to generating excess returns, but they contain no explicit parameter in an attempt to influence such returns. These include:

- Alexander and Dimitriu [2], 2005, who for a cointegration based index tracking strategy (see [1]) found evidence of excess (abnormal) returns for a number of indices.
- Meade and Beasley [19], 2007, who, drawing on Beasley et al. [5], proposed a Sortino ratio objective to decide a tracking portfolio. Their paper is focused on investigating momentum, that performance observed in the recent past will continue into the near future. They find evidence of significant momentum profits (including reasonable transaction costs) for a number of indices.

In summary there is no single accepted underlying mathematical model neither for index tracking, nor for enhanced indexation. Authors typically solve only one or two problems, often not giving detailed computational results/times and use just data of their own, making it impossible to perform a systematic data driven comparison of different approaches. Overall the literature with respect to both index tracking and enhanced indexation is "fragmented", different models and different data result in papers that stand in isolation, but with little connecting them in a mathematical/data sense.

3. Problem formulation

In this paper we adopt the regression based view of index tracking and enhanced indexation. This view has the advantage that it enables us to develop linear formulations, more technically our formulations for both index tracking and enhanced indexation are mixed-integer linear programs. Algorithmically, standard packages for the numeric solution of such programs are well developed and very sophisticated. This contrasts with the approach for index tracking presented in Beasley et al. [5] which relied on a specially developed evolutionary heuristic.

In this section we first give our notation and present the constraints and objective for the index tracking problem from the regression viewpoint. Here, for convenience, we adopt the same notation (and indeed where applicable the same constraints) as in [5]. We then discuss how our formulation can be extended for enhanced indexation.

3.1. Notation

We observe over time $0, 1, 2, \ldots, T$ the value of N stocks, as well as the index we are tracking. We are interested in deciding the best set of K stocks to hold (where $K < N$), as well as their appropriate quantities. Let:

ε_i	be the minimum proportion of the tracking portfolio (henceforth TP) that must be held in stock i if any of the stock is held
δ_i	be the maximum proportion of the TP that can be held in stock i if any of the stock is held
X_i	be the number of units of stock i in the current TP
V_{it}	be the value (price) of one unit of stock i at time t
I_t	be the value of the index at time t
R_t	be the single period continuous time return for the index at time t, i.e. $R_t = \ln(I_t/I_{t-1})$
r_{it}	be the single period continuous time return for stock i at time t, i.e. $r_{it} = \ln(V_{it}/V_{it-1})$
C	be the total value ($\geqslant 0$) of the current TP $[X_i]$ at time $T, \sum_{i=1}^{n} X_i V_{iT}$, plus cash change (either new cash to be invested or cash to be taken out)
f_i^s	be the fractional cost of selling one unit of stock i at time T
f_i^b	be the fractional cost of buying one unit of stock i at time T
γ	be the limit ($0 \leqslant \gamma \leqslant 1$) on the proportion of C that can be consumed by transaction cost

Then our decision variables are:

x_i the number of units ($\geqslant 0$) of stock i that we choose to hold in the new TP
G_i the transaction cost ($\geqslant 0$) incurred in selling/buying stock i
z_i =1 if any stock of i is held in the new TP, = 0 otherwise

Without significant loss of generality (since the sums of money involved are large) we allow $[x_i]$ to take fractional values.

With respect to deciding the universe of N stocks to consider for use in tracking the index it is usual to include in that universe stocks that are included in the index that is being tracked (index composition typically being known). However, from the mathematical viewpoint, this is not strictly necessary and we can include any stocks that we wish in our universe of N stocks. In fact, more broadly, we can include in our universe any asset at all provided that it can be traded and has a price history $[V_{it}]$.

3.2. Index tracking constraints

The constraints associated with the index tracking problem are:

$$\sum_{i=1}^{N} z_i = K, \tag{1}$$
$$\varepsilon_i z_i \leqslant x_i V_{iT}/C \leqslant \delta_i z_i \quad i = 1, \ldots, N, \tag{2}$$
$$G_i \geqslant f_i^s (X_i - x_i) V_{iT} \quad i = 1, \ldots, N, \tag{3}$$
$$G_i \geqslant f_i^b (x_i - X_i) V_{iT} \quad i = 1, \ldots, N, \tag{4}$$
$$\sum_{i=1}^{N} G_i \leqslant \gamma C, \tag{5}$$
$$\sum_{i=1}^{N} x_i V_{iT} = C - \sum_{i=1}^{N} G_i, \tag{6}$$
$$x_i, G_i \geqslant 0 \quad i = 1, \ldots, N, \tag{7}$$
$$z_i \in [0, 1] \quad i = 1, \ldots, N. \tag{8}$$

Eq. (1) ensures that there are exactly K stocks in the new TP. Eq. (2) ensures that if a stock i is not in the new TP ($z_i = 0$) then x_i is also zero; it also ensures that if the stock is chosen to be in the new TP ($z_i = 1$) then the amount of the stock held satisfies the proportion limits defined. Eqs. (3) and (4) define the transaction cost and Eq. (5) limits the total transaction cost incurred. Eq. (6) is a balance constraint such that the total value of the new TP at time T equals the value of the current TP at time T plus the cash change (i.e. C) minus the total transaction cost. On a computational note benefit can be gained by strengthening the linear programming relaxation of the above constraints. This can be done by setting δ_i=min [δ_i, 1 – the sum of the $(K-1)$ smallest $\varepsilon_j,\ j = 1, \ldots, N,\ j \neq i$].

There is one technical subtlety here, namely that the equations defining G_i above do not constrain G_i to be equal to the correct transaction cost for stock i. Rather G_i is bounded below by the correct transaction cost (as the inequalities in Eqs. (3) and (4) indicate), and hence G_i is an over-estimate of the transaction cost for stock i. In order to address this issue we, as will become apparent below, adopt in our procedure a stage relating to minimising total transaction cost $\left(\sum_{i=1}^{N} G_i\right)$ in an attempt to ensure that the numeric values we obtain for G_i do correctly correspond to the transaction cost incurred in selling/buying stock i.

3.3. Index tracking objective

Adopting the regression viewpoint for index tracking we have that ideally our TP would be chosen such that when we perform a regression of TP returns against index returns we would find that the regression line has an intercept (alpha) of zero, and a slope (beta) of one. However achieving this ideal is not straightforward. There are a number of complications to achieving this ideal within the context of a linear (or linearisable) formulation, as will become apparent below.

The single period continuous time return for the TP (in period t), $\ln\left(\sum_{i=1}^{N} x_i V_{it} / \sum_{i=1}^{N} x_i V_{it-1}\right)$, is a nonlinear function of the decision variables. In order to linearise (in an approximate fashion) this return we shall assume that it can be expressed as a linear weighted sum of individual stock returns, where the weights, summing to one, reflect the proportion invested in each stock at time t. The assumption that portfolio return can be expressed as a linear weighted sum of individual stock returns is a common assumption in finance. Hence we have an equation of the form: return on the TP at time $t = \sum_{i=1}^{N} W_{it} r_{it}$ where $W_{it} = x_i V_{it} / \sum_{j=1}^{N} x_j V_{jt}$ is the weight associated with the investment in stock i at time t and $\sum_{i=1}^{N} W_{it} = 1\ \forall t$.

Now the W_{it} are nonlinear expressions involving our decision variables x_i, as well as the value $\sum_{j=1}^{N} x_j V_{jt}$ of the TP at time t. In order to proceed we shall approximate W_{it} by a constant term which is independent of time, namely replace W_{it} by w_i where $w_i = x_i V_{iT} / \sum_{j=1}^{N} x_j V_{jT}$ represents the proportion invested in stock i at time T. Hence we have an equation of the form: return on the TP at time $t = \sum_{i=1}^{N} w_i r_{it}$.

Our expression for $w_i\left(x_i V_{iT} / \sum_{j=1}^{N} x_j V_{jT}\right)$ is nonlinear so to linearise it we first use Eq. (6) to replace $\sum_{j=1}^{N} x_j V_{jT}$ by $C - \sum_{j=1}^{N} G_j$. This is also a function of our variables, but we know from Eq. (5) that $\sum_{j=1}^{N} G_j$ is bounded above by γC. Hence we approximate w_i using the linear expression:

$$w_i = x_i V_{iT}/(C - \gamma C) \quad i = 1, \ldots, N. \tag{9}$$

Finally therefore we have a linear expression (approximation) for the returns on the TP as $\left[\sum_{i=1}^{N} w_i r_{it}, t = 1, \ldots, T\right]$. Using the log-sum inequality it is possible to prove (assuming we spend γC in transaction cost) that the average return from this approximation $\left[\sum_{t=1}^{T}\sum_{i=1}^{N} w_i r_{it}/T\right]$ over-estimates average tracking portfolio return $\left[\sum_{t=1}^{T} \ln(\sum_{i=1}^{N} x_i V_{it} / \sum_{i=1}^{N} x_i V_{it-1})/T\right]$.

If we regress these TP returns against the index returns $[R_t, t = 1, \ldots, T]$ then we have from standard regression theory that the ordinary least-squares estimates, $\hat{\alpha}$ and $\hat{\beta}$, for the intercept and slope of the regression line are given by:

$$\hat{\alpha} = \sum_{i=1}^{N} w_i \hat{\alpha}_i, \tag{10}$$

$$\hat{\beta} = \sum_{i=1}^{N} w_i \hat{\beta}_i, \tag{11}$$

where $\hat{\alpha}_i$ and $\hat{\beta}_i$ are the ordinary least-squares regression intercept and slope when we regress the returns from stock $i, [r_{it}, t = 1, \ldots, T]$ against the index returns $[R_t, t = 1, \ldots, T]$.

Ideally we would like, for index tracking, to choose K stocks and their associated quantities $[x_i, i = 1, \ldots, N]$ such that we achieve $\hat{\alpha} = 0$ and $\hat{\beta} = 1$. For real life data this may not, however, be achievable. There are a number of possible ways to proceed:

- Adopt a single-stage approach using a simple weighted objective:

$$\text{minimise } \lambda_\alpha |\hat{\alpha} - 0| + \lambda_\beta |\hat{\beta} - 1|, \tag{12}$$

where $\lambda_\alpha, \lambda_\beta \geqslant 0$ are our user defined weighting values for achieving our desired intercept of zero and our desired slope of one. We can impose if we wish that $\lambda_\alpha + \lambda_\beta = 1$, but this is essentially just a scaling of our weighted objective above and is not strictly necessary here.

- Adopt a two-stage approach, where the primary objective is to achieve the desired intercept of zero and the secondary objective is to achieve the desired slope of one, i.e.

$$\text{first minimise } |\hat{\alpha} - 0| \text{ and then minimise } |\hat{\beta} - 1|. \tag{13}$$

This approach is equivalent to having $\lambda_\alpha \gg \lambda_\beta$ in Eq. (12), but avoids the numeric difficulties that occur computationally when one weight is much larger than the other.

- Adopt a two-stage approach, where the primary objective is to achieve the desired slope of one and the secondary objective is to achieve the desired intercept of zero, i.e.

$$\text{first minimise } |\hat{\beta} - 1| \text{ and then minimise } |\hat{\alpha} - 0|. \tag{14}$$

This approach is equivalent to having $\lambda_\beta \gg \lambda_\alpha$ in Eq. (12).

We have investigated all of the above approaches. For reasons of space we will not report results for all of these approaches here, rather we will report results for the approach that seems, to us, to give the best results. This is to adopt a two-stage approach, where the primary objective is to achieve the desired intercept of zero and the secondary objective is to achieve the desired slope of one, i.e.

$$\text{first minimise } |\hat{\alpha} - 0| \text{ and then minimise } |\hat{\beta} - 1|. \tag{15}$$

3.4. Index tracking formulation

Although the modulus objectives in Eq. (15) above are nonlinear they can be linearised in a standard way. Introduce variables D and E where

$$D \geqslant \hat{\alpha}, \tag{16}$$

$$D \geqslant -\hat{\alpha}, \tag{17}$$

$$E \geqslant \hat{\beta} - 1, \tag{18}$$

$$E \geqslant -(\hat{\beta} - 1), \tag{19}$$

$$D, E \geqslant 0. \tag{20}$$

Then our full mixed-integer linear programming formulation for solving the index tracking problem in the first-stage (primary objective to achieve the desired intercept of zero) is

$$\text{minimise} \quad D, \tag{21}$$

subject to Eqs. (1)–(11) and (16)–(20).

This formulation (neglecting any algebraic substitution to eliminate variables/constraints) has approximately $3N$ continuous variables, N zero-one variables and approximately $4N$ constraints. For $N = 2000$, for example, this equates to 6000 continuous variables, 2000 zero-one variables and 8000 constraints. In modern mathematical programming terms this is not a large mixed-integer linear program.

Algorithmically our formulation can be tackled using a standard solver such as Cplex [15], to find the optimal solution to our *approximation* of the index tracking problem – where the approximation here relates to the approximation of tracking portfolio return by a weighted sum of individual stock returns and to the approximation of the time dependent weights by constant weights. Note here that our approximation only relates to the objective function. The constraints of the problem (which we require to be satisfied for feasibility) are not approximated, but are represented exactly.

In the second-stage, where the emphasis is on achieving the desired slope of one, then when we solve we constrain the intercept $\hat{\alpha}$ so that it retains the value that it achieved at the first-stage. Formally let α^{opt} be the numeric value for $\hat{\alpha}$ when our formulation minimise Eq. (21) subject to Eqs. (1)–(11) and (16)–(20) is solved. Then in the second-stage we:

$$\text{minimise} \quad E, \tag{22}$$

subject to (1)–(11) and (16)–(20) and

$$\hat{\alpha} = \alpha^{\text{opt}}. \tag{23}$$

3.5. Three-stage approach

Computational experience (reported below) indicated that benefit could be gained by adopting a three-stage approach. The third-stage relates to minimising transaction cost subject to the regression coefficients $\hat{\alpha}$ and $\hat{\beta}$ retaining the values that they achieved at the first two-stages. Formally let β^{opt} be the numeric value for $\hat{\beta}$ when our formulation minimise Eq. (22) subject to Eqs. (1)–(11) and (16)–(20), (23), is solved at the second-stage. Then in the third-stage we:

$$\text{minimise} \quad \sum_{i=1}^{N} G_i, \tag{24}$$

subject to (1)–(11) and (16)–(20) and

$$\hat{\alpha} = \alpha^{\text{opt}}, \tag{25}$$

$$\hat{\beta} = \beta^{\text{opt}}. \tag{26}$$

This third-stage can be regarded as making use of any flexibility in the problem, achieving precisely the same values for $\hat{\alpha}$ and $\hat{\beta}$ as were achieved at the first and second-stage respectively, but minimising the transaction cost involved in so doing. On a numeric issue solvers such as Cplex [15] only solve to a limited number of decimal places and so in practice the equality seen in Eqs. (23) and (25) above must be replaced by $(\alpha^{\text{opt}} - \epsilon) \leqslant \hat{\alpha} \leqslant (\alpha^{\text{opt}} + \epsilon)$ and that in (26) must be replaced by $(\beta^{\text{opt}} - \epsilon) \leqslant \hat{\beta} \leqslant (\beta^{\text{opt}} + \epsilon)$ where $\epsilon > 0$ is a small constant value.

We referred above to the issue that in the formulation we have adopted G_i is an over-estimate of the transaction cost for stock i. When we minimise total transaction cost $(\sum_{i=1}^{N} G_i)$ in this third-stage the four constraints in our formulation that involve G_i are:

$$G_i \geqslant f_i^s (X_i - x_i) V_{iT} \quad i = 1, \ldots, N, \tag{27}$$

$$G_i \geqslant f_i^b (x_i - X_i) V_{iT} \quad i = 1, \ldots, N, \tag{28}$$

$$\sum_{i=1}^{N} G_i \leqslant \gamma C, \tag{29}$$

$$\sum_{i=1}^{N} x_i V_{iT} = C - \sum_{i=1}^{N} G_i. \tag{30}$$

Minimising $\sum_{i=1}^{N} G_i$ implies we wish to make each G_i value as small as possible (consistent with the constraints). If the equality relationship, Eq. (30), was not present then in achieving this minimisation we would have equality achieved in one (or other) of Eqs. (27) and (28), and hence G_i would correctly and precisely represent the transaction cost for stock i.

However, because in our formulation the equality relationship, equation (30), is present it is possible that when we minimise $\sum_{i=1}^{N} G_i$ we will not achieve equality in either of the two inequalities, Eqs. (27) and (28). Such a case corresponds to a situation where the transaction cost incurred in practice in moving from our old TP $[X_i, i = 1, \ldots, N]$ to our new TP $[x_i, i = 1, \ldots, N]$ will be lower than the transaction cost given by the solution to our formulation.

We might expect however that the likelihood of this situation occurring would be low (and indeed our numeric experience reported below supports this view). Moreover as we know of no way within a linear formulation to correctly represent the transaction cost, we believe that it is acceptable to use our three-stage approach whilst recognising this deficiency.

3.6. Enhanced indexation

In regression based enhanced indexation there are a number of approaches we might follow. For example we might wish to achieve (if possible) a regression slope of one, as it would be if we were to perfectly track the index, and then maximise the value of the regression intercept that can be achieved, a positive regression intercept being interpreted as excess return (return over and above the return achieved by the index). But other possibilities also exist, as discussed below:

- Adopt a two-stage approach, where the primary objective is to achieve (if possible) the desired slope of one, and the secondary objective is to maximise the intercept, i.e.

$$\text{first minimise } |\hat{\beta} - 1| \text{ and then maximise } \hat{\alpha}. \tag{31}$$

- Adopt a two-stage approach, where the primary objective is to maximise the intercept and the secondary objective is to achieve the desired slope of one, i.e.

$$\text{first maximise } \hat{\alpha} \text{ and then minimise } |\hat{\beta} - 1|. \tag{32}$$

- Adopt a one-stage approach, where the objective is to achieve the desired slope of one, subject to the intercept being constrained to be a desired value, i.e.

$$\text{minimise } |\hat{\beta} - 1| \text{ subject to } \hat{\alpha} = \alpha^*, \tag{33}$$

where α^* is the desired excess return (return over and above index return).

In our discussion of enhanced indexation above we referred to approaches which have no explicit parameter to influence excess return. Of the three approaches listed above the first two fall into this category, the third does include a explicit parameter α^* that influences excess return. Note here that these three approaches do not exhaust the range of possibilities, for example there are approaches based on maximising $\hat{\beta}$ and approaches based on introducing a desired regression slope β^*.

For space reasons we will not report results for all of the three approaches given above here, rather we will report results for an approach that seems, to us, to give good quality results. This is to adopt a one-stage approach, where the objective is to achieve the desired slope of one, subject to the intercept being constrained to be a desired value (α^*). Formally therefore our one-stage approach to enhanced indexation is:

$$\text{minimise} \quad E, \tag{34}$$

subject to (1)–(11) and (16)–(20) and

$$\hat{\alpha} = \alpha^*. \tag{35}$$

As for index tracking we can extend this by one-stage, here to a two-stage approach where, as before, the final-stage relates to minimising transaction cost subject to the regression coefficients $\hat{\alpha}$ and $\hat{\beta}$ retaining the values that they achieved previously. Let β^{opt} be the numeric value for $\hat{\beta}$ when our formulation minimise Eq. (34) subject to Eqs. (1)–(11) and (16)–(20), (35) is solved. Then the second-stage is minimise Eq. (24) subject to Eqs. (1)–(11) and (16)–(20), (35) and $\hat{\beta} = \beta^{\text{opt}}$.

3.7. Extensions

Above we have formulations of both index tracking and enhanced indexation as mixed-integer linear programming problems. Standard packages (such as Cplex [15]) for the computational solution of such problems are well developed and highly sophisticated. Although, for reasons of space, we will not enlarge on it here in any mathematical detail it is very easy to incorporate into our formulations additional constraints that may, depending upon the practical situation being considered, be useful. These include:

- Fixed transaction costs, where we incur a fixed cost associated with any change (from X_i to $x_i \neq X_i$) in our holding of stock i.
- Sector constraints, where we want to constrain between lower and upper limits the total investment in stocks identified as being in the same sector. These are sometimes also known as class constraints.
- Lot size constraints, these are of two types:
  - where the amount (number of units) x_i of any stock i that is held must be an integer multiplier of a given constant,
  - where the amount (number of units) traded, $|X_i - x_i|$, of any stock i must be an integer multiplier of a given constant.

All of these extensions to the basic index tracking/enhanced indexation problem can be formulated in a linear manner (albeit possibly involving additional integer variables). Hence using a standard solution package (such as Cplex [15]) to find the optimal solution to the problem remains a valid approach.

4. Computational results

4.1. Data sets and computational considerations

We used the same data sets as in [5] but added three more data sets for (large) capital market indices. More specifically, we considered Hang Seng (Hong Kong), DAX 100 (Germany), FTSE 100 (UK), S&P 100 (USA), Nikkei 225 (Japan) (as in [5]), but added the S&P 500 (USA), Russell 2000 (USA) and Russell 3000 (USA). Stock price data was obtained from DATASTREAM, stocks with missing values were dropped and for each index we had 290 (weekly) stock returns. Table 1 summarises our data sets, which are publicly available from OR-Library [4], http://people.brunel.ac.uk/~mastjjb/jeb/info.html.

The computational results presented below are for our approach as coded in C++ and solved using Cplex (version 9) [15] on a pc (Intel Pentium 4, 3.00 GHz, 3,619,568 KB RAM). Unless otherwise stated we:

- Used an initial tracking portfolio composed of the first K stocks in equal proportions, i.e. $X_i = (C/K)/V_{i0},\ i = 1,\ldots,K;\ X_i = 0\ \forall i > K$ with $C = 10^6$.
- Used $\varepsilon_i = 0.01$ and $\delta_i = 1\ \forall i$, i.e. a lower proportion limit for any stock that appears in the tracking portfolio of one percent.
- Used $f_i^s = f_i^b = 0.01$, i.e. transaction cost was one percent of the value of the stocks sold/bought.

4.2. Index tracking

4.2.1. In-sample and out-of-sample results

We want to chose a TP that performs well in-sample, but we also hope that the chosen TP will perform well out-of-sample. In order to investigate this issue we took the time period [0, 145] for each of our eight data sets and, for a range of values for the transaction cost limit γ, used the approach given above to decide

Table 1
Data sets

Index	Number of stocks, N	Number of selected stocks, K
Hang Seng	31	10
DAX 100	85	10
FTSE 100	89	10
S&P 100	98	10
Nikkei 225	225	10
S&P 500	457	40
Russell 2000	1318	90
Russell 3000	2151	70

a TP. We then calculated the alpha and beta values associated with this TP out-of sample (in [145, 290]).

In Table 2 we report the alpha and beta values and the time taken (in seconds), for each data set and each of the values of γ, when the two-stage and three-stage approaches are adopted. In-sample values for alpha and beta given in Table 2 (as in the other in-sample table results in this paper, unless otherwise stated) relate to the values $\hat{\alpha}$ and $\hat{\beta}$ as defined in Eqs. (10) and (11). These values are appropriate in-sample as these are the values involved in our in-sample objective (here in the first two-stages). Out-of-sample values for alpha and beta given in Table 2 (as in the other out-of-sample table results in this paper, unless otherwise stated) are the ordinary least-squares regression estimates as calculated by regressing the returns from the chosen TP $[x_i, i = 1, \ldots, N]$ against the returns from the index in the out-of-sample period.

It is clear from Table 2 that, in all cases, in-sample alpha values approach zero, and beta values approach one, as the transaction cost limit γ increases. In-sample averages of alpha and beta, 0.00013 and 0.98668 respectively, show that our two-stage approach provides values of alpha and beta close to zero and one respectively. By adopting a three-stage approach, instead of a two-stage, approach we have an increase in the out-of-sample value for alpha (albeit remaining very close to zero), whilst the out-of-sample value for beta comes closer to one, although the change in beta is not high.

For reasons of space we have not shown the transaction cost in Table 2, but examination of the detailed results revealed that (on average) the transaction cost at the third-stage was 39.08% less than the transaction cost at the second-stage. Reducing transaction cost by approximately 39%, whilst retaining the same $\hat{\alpha}$ and $\hat{\beta}$ values as were achieved at stages one and two respectively, and with only a slight effect on (average) out-of-sample results, seems beneficial. Hence we believe that overall the three-stage approach provides better results and consequently we will use just the three-stage approach in the following sections concerned with index tracking.

It can be seen from Table 2 that computation times are very low, on average approximately 3 seconds for the three-stage approach, with no case taking more than 35 seconds. With respect to the issue of the G_i correctly representing the transaction cost for stock i we found that in all 31 cases in Table 2 the values of G_i in the third-stage solution did correctly represent the transaction cost for stock i.

The result shown as infeasible in Table 2 corresponds to a case where it is not possible to change from the current TP $[X_i, i = 1, \ldots, N]$ to any new TP $[x_i, i = 1, \ldots, N]$ whilst satisfying the constraints of the problem (e.g. because of the limit on transaction cost).

4.2.2. Standard errors, skewness and kurtosis

The results presented in Table 2 give insight into the in-sample and out-of-sample behaviour of the tracking portfolios chosen. To provide further insight, and also to give insight into higher order moments, we present in Table 3 further details as to these in-sample and out-of-sample portfolios. In this table we have taken

Table 2
In-sample and out-of-sample results

Index (N, K)	Transaction cost limit γ	In-sample		Out-of-sample					
		α	β	Two-stage approach			Three-stage approach		
				α	β	Time	α	β	Time
Hang Seng	0.0025	0.00076	0.94513	−0.00106	0.98759	0.1	−0.00106	0.98759	0.1
(31,10)	0.005	0.00000	1.00000	−0.00131	1.02845	0.1	−0.00070	1.00952	0.1
	0.0075	0.00000	1.00000	−0.00200	1.06124	0.0	−0.00072	1.00900	0.1
	0.01	0.00000	1.00000	−0.00083	0.96187	0.0	−0.00075	1.00849	0.1
DAX 100	0.0025	0.00053	0.68501	0.00176	0.87469	0.1	0.00176	0.87469	0.2
(85,10)	0.005	0.00000	0.82413	0.00160	0.98815	0.1	0.00160	0.98815	0.2
	0.0075	0.00000	0.96548	−0.00023	0.94477	0.1	−0.00023	0.94477	0.1
	0.01	0.00000	1.00000	0.00029	0.79576	0.0	−0.00018	0.95194	0.1
FTSE 100	0.0025	0.00077	1.00937	0.00173	0.76300	0.1	0.00173	0.76300	0.2
(89,10)	0.005	0.00000	1.00000	0.00214	0.74130	0.1	0.00133	0.73679	0.1
	0.0075	0.00000	1.00000	0.00093	0.80163	0.1	0.00130	0.73567	0.1
	0.01	0.00000	1.00000	0.00106	0.79921	0.0	0.00127	0.73455	0.1
S&P 100	0.0025	0.00000	1.00000	−0.00032	0.93979	0.1	−0.00032	0.93962	0.2
(98,10)	0.005	0.00000	1.00000	−0.00056	0.89706	0.0	−0.00034	0.93836	0.1
	0.0075	0.00000	1.00000	−0.00007	0.83406	0.0	−0.00035	0.93711	0.1
	0.01	0.00000	1.00000	−0.00056	0.86672	0.1	−0.00037	0.93585	0.1
Nikkei 225	0.0025	0.00000	1.00000	−0.00015	0.96541	0.2	−0.00027	0.94549	0.8
(225,10)	0.005	0.00000	1.00000	0.00021	0.97332	0.2	−0.00028	0.94822	0.5
	0.0075	0.00000	1.00000	0.00045	0.94481	0.1	−0.00027	0.94780	0.4
	0.01	0.00000	1.00000	0.00129	0.94209	0.1	−0.00025	0.95055	0.5
S&P 500	0.0025	0.00209	1.15789	0.00227	1.20867	0.6	0.00227	1.20867	0.9
(457,40)	0.005	0.00000	1.00000	0.00164	1.20305	0.4	0.00259	1.00861	1.0
	0.0075	0.00000	1.00000	0.00238	1.22159	0.2	0.00258	1.00563	0.9
	0.01	0.00000	1.00000	0.00162	1.21086	0.2	0.00273	1.00796	0.6
Russell 2000	0.0025	Infeasible							
(1318,90)	0.005	0.00000	1.00000	0.00177	1.17542	1.5	0.00198	1.23064	5.1
	0.0075	0.00000	1.00000	0.00181	1.04470	4.3	0.00204	1.35907	7.8
	0.01	0.00000	1.00000	0.00102	1.09754	3.6	0.00199	1.36683	5.5
Russell 3000	0.0025	0.00000	1.00000	0.00363	1.11200	14.1	0.00347	1.10492	31.8
(2151,70)	0.005	0.00000	1.00000	0.00310	1.21382	4.2	0.00349	1.12587	15.1
	0.0075	0.00000	1.00000	0.00380	1.04089	6.9	0.00345	1.10701	13.2
	0.01	0.00000	1.00000	0.00298	1.14642	1.9	0.00346	1.10233	9.6
Average		0.00013	0.98668	0.00098	0.99309	1.3	0.00106	0.99725	3.1

Table 3
Standard errors, skewness and kurtosis

Index (N, K)	Transaction cost limit γ	In-sample				Out-of-sample			
		S.E α	S.E β	Skewness	Kurtosis	S.E α	S.E β	Skewness	Kurtosis
Hang Seng (31,10)	0.0025	0.00101	0.02726	−0.25650	0.56543	0.00098	0.03430	−0.03847	2.06707
	0.005	0.00118	0.03167	−0.19837	1.10975	0.00104	0.03620	0.19210	1.88728
	0.0075	0.00117	0.03130	−0.20442	1.06934	0.00102	0.03577	0.18407	1.89372
	0.01	0.00115	0.03096	−0.21014	1.02941	0.00101	0.03535	0.17610	1.89976
DAX 100 (85,10)	0.0025	0.00135	0.06873	−0.16248	0.51703	0.00181	0.08617	−0.48711	1.31625
	0.005	0.00136	0.06895	−0.18230	0.42990	0.00187	0.08903	−1.02678	3.24812
	0.0075	0.00079	0.03992	0.01382	0.77380	0.00112	0.05313	−0.42370	0.53573
	0.01	0.00080	0.04060	0.04441	0.63505	0.00110	0.05246	−0.39514	0.55708
FTSE 100 (89,10)	0.0025	0.00123	0.06400	1.37908	7.82412	0.00115	0.07384	0.15226	1.05780
	0.005	0.00112	0.05840	1.21327	5.82561	0.00107	0.06863	0.08063	0.90705
	0.0075	0.00112	0.05856	1.19688	5.69004	0.00106	0.06838	0.07111	0.88662
	0.01	0.00113	0.05878	1.17942	5.54961	0.00106	0.06817	0.06155	0.86443
S&P 100 (98,10)	0.0025	0.00116	0.09447	−0.15836	−0.21386	0.00114	0.06309	−0.07828	−0.17897
	0.005	0.00115	0.09413	−0.15465	−0.21269	0.00113	0.06266	−0.08117	−0.17624
	0.0075	0.00115	0.09380	−0.15083	−0.21112	0.00112	0.06225	−0.08410	−0.17394
	0.01	0.00115	0.09349	−0.14690	−0.20915	0.00111	0.06185	−0.08708	−0.17208
Nikkei 225 (225,10)	0.0025	0.00085	0.02910	1.04409	2.60121	0.00080	0.02929	−0.14706	1.53522
	0.005	0.00085	0.02903	1.03794	2.58595	0.00080	0.02925	−0.15752	1.54778
	0.0075	0.00084	0.02883	1.02307	2.53922	0.00080	0.02908	−0.15808	1.55287
	0.01	0.00084	0.02890	0.99822	2.47475	0.00079	0.02894	−0.17080	1.55080
S&P 500 (457,40)	0.0025	0.00162	0.06908	−0.33207	0.45628	0.00136	0.05109	−0.03004	0.70311
	0.005	0.00187	0.07971	−0.62725	0.71212	0.00139	0.05120	−0.00950	0.69801
	0.0075	0.00188	0.07993	−0.63055	0.72965	0.00139	0.05199	−0.00727	0.68480
	0.01	0.00188	0.08016	−0.63380	0.74791	0.00139	0.05199	−0.00502	0.67142
Russell 2000 (1318,90)	0.0025	Infeasible							
	0.005	0.00206	0.07770	−0.85896	2.79166	0.00157	0.05038	−0.32992	0.16723
	0.0075	0.00205	0.07712	−0.89203	2.95057	0.00175	0.05619	−0.23624	0.98563
	0.01	0.00197	0.07434	−0.69569	1.65767	0.00181	0.05801	−0.18346	1.00283
Russell 3000 (2151,70)	0.0025	0.00133	0.05666	−0.70962	1.96168	0.00197	0.07217	−0.19805	0.12551
	0.005	0.00135	0.05748	−0.61936	1.68167	0.00205	0.07488	−0.17255	0.13406
	0.0075	0.00135	0.05736	−0.61982	1.67064	0.00200	0.07305	−0.16657	0.13069
	0.01	0.00135	0.05726	−0.62242	1.67585	0.00199	0.07271	−0.17102	0.13260
Average		0.00129	0.05928	0.00205	1.85191	0.00131	0.05585	−0.12668	0.90459
Significance (%)				–	0.00			0.62	0.00

the TP as decided by our three-stage approach and examined its return distributions (in-sample and out-of-sample) and:

- Computed the standard errors for the regression coefficients (alpha and beta) when TP returns are regressed against index returns.
- Computed the skewness and kurtosis associated with these distributions (using Excel).

From the average values for standard errors at the foot of Table 3 it appears that the standard errors for alpha and beta in-sample and out-of-sample are the same. Of interest in Table 3 is whether, or not, the return distributions associated with our TPs are skewed, or exhibit abnormal kurtosis. In order to judge this we conducted the hypothesis tests H_0: skewness=0 versus H_1: skewness $\neq 0$ and H_0: kurtosis = 0 versus H_1: kurtosis $\neq 0$ (since the Excel KURT function for kurtosis returns zero for the kurtosis of the standard Normal distribution) with respect to the average values shown at the foot of Table 3. The last line in Table 3 shows (using a Student's t-distribution) the significance level (expressed as a percentage) associated with these hypothesis tests. If the significance level is more than 5% then it is not shown.

It is clear from these significance values that the return distributions, both in-sample and out-of-sample, display nonzero kurtosis. In-sample there is no statistical evidence of skewness, out-of-sample there is statistical evidence that skewness is negative.

4.2.3. Varying K

The number of stocks K in the tracking portfolio is a user defined parameter. One of the advantages of our approach (as the computation time required is small, as seen in Table 2) is that the user can easily vary K, to see the effect of changing the number of stocks that they choose to hold. To illustrate this we show for one instance from Table 2, the S&P 500 with $\gamma = 0.01$, the change in out-of-sample alpha and beta values as K changes. Fig. 1 shows the change in alpha, Fig. 2 the change in beta.

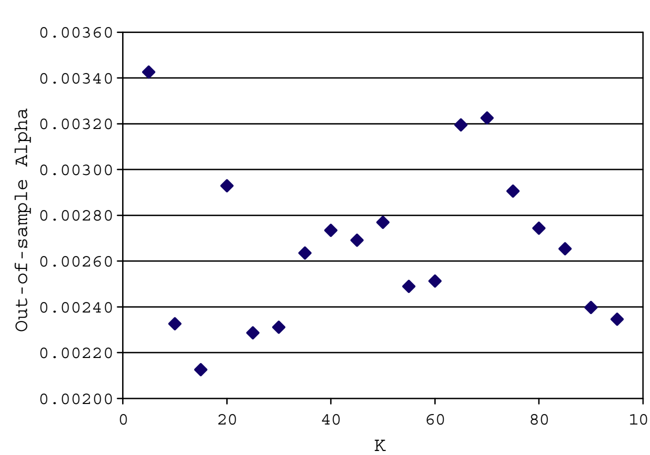

Fig. 1. S&P 500, $\gamma = 0.01$. Change in out-of-sample alpha as K varies.

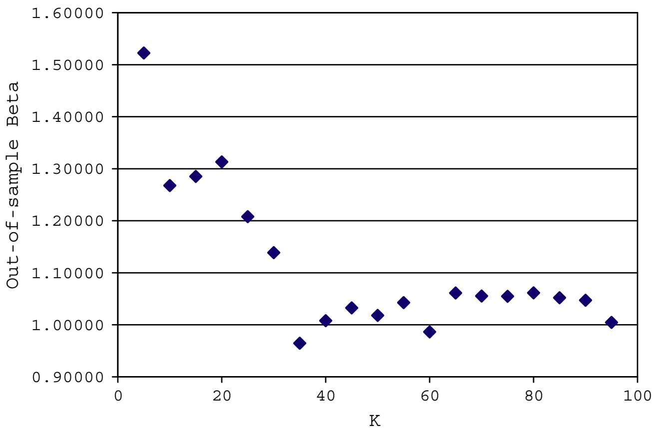

Fig. 2. S&P 500, $\gamma = 0.01$. Change in out-of-sample beta as K varies.

Graphically Fig. 1 shows no clear pattern in the out-of-sample value of alpha as K varies (the correlation coefficient between K and alpha is just 0.03, which is not statistically significant). Note that Fig. 1 does not consider values of K over 100 as the lower proportion limit of one percent (see Section 4.1) on the investment in each stock means that such cases are infeasible.

Graphically Fig. 2 shows an initial (approximately linear) decrease in the out-of-sample value of beta as K increases to approximately 35, but beta seems to remains approximately constant thereafter. The correlation coefficient between K and beta is -0.91 for $K \leqslant 35$, which is statistically significant at the 1% level. The correlation coefficient between K and beta is 0.31 for $K > 35$, which is not statistically significant.

4.2.4. Systematic revision

In order to investigate the performance of our approach over time as we systematically revise our TP, we took the time period [0, 290] and choose within it seven decision points ($T = 150, 170, 190, 210, 230, 250, 270$). These decision points correspond to revising our TP every 20 periods after an initial period has elapsed. For each of our data sets, and each value of γ that we consider, we make at time T a decision as to an appropriate new TP to have, using all the information contained in [0, T]. Formally:

(a) set $T = 150$
(b) use our three-stage approach to decide the new TP $[x_i]$
(c) set $[X_i] = [x_i]$ (replace the current TP by the new TP)
(d) set $T = T + 20$ and if $T \leqslant 270$ go to (b).

In Table 4 we show (for each of our data sets and each of the values of γ considered) the alpha and beta values achieved in the in-sample and out-of-sample periods and the total time taken (in seconds). It can be seen in Table 4 that, in all cases, in-sample alpha values approach zero, and beta values approach one, as the transaction cost limit γ increases. In this table the values given

Table 4
Systematic revision

Index (N, K)	Transaction cost limit γ	In-sample		Out-of-sample		Time
		α	β	α	β	
Hang Seng (31,10)	0.0025	0.00014	0.99270	−0.00104	0.97932	0.5
	0.005	0.00000	1.00000	−0.00150	0.97285	0.4
	0.0075	0.00000	1.00000	−0.00159	0.97260	0.4
	0.01	0.00000	1.00000	−0.00169	0.97234	0.6
DAX 100 (85,10)	0.0025	0.00017	0.89856	0.00122	0.87882	0.7
	0.005	0.00000	0.97381	0.00104	0.95320	0.9
	0.0075	0.00000	0.99353	−0.00033	0.96983	0.8
	0.01	0.00000	1.00000	−0.00080	0.95902	1.2
FTSE 100 (89,10)	0.0025	0.00012	1.00579	0.00091	0.85643	1.2
	0.005	0.00000	1.00000	0.00071	0.78958	1.2
	0.0075	0.00000	1.00000	0.00074	0.78332	1.6
	0.01	0.00000	1.00000	0.00065	0.79081	1.2
S&P 100 (98,10)	0.0025	0.00000	1.00000	−0.00063	0.93415	1.6
	0.005	0.00000	1.00000	−0.00059	0.93230	1.5
	0.0075	0.00000	1.00000	−0.00063	0.92968	1.6
	0.01	0.00000	1.00000	−0.00059	0.92697	1.1
Nikkei 225 (225,10)	0.0025	0.00000	1.00000	−0.00010	0.98941	3.2
	0.005	0.00000	1.00000	−0.00010	0.98734	3.0
	0.0075	0.00000	1.00000	−0.00011	0.98527	2.9
	0.01	0.00000	1.00000	−0.00012	0.98321	2.5
S&P 500 (457,40)	0.0025	0.00016	1.05923	0.00373	1.00193	7.4
	0.005	0.00000	1.00000	0.00358	1.04723	7.4
	0.0075	0.00000	1.00000	0.00355	1.04493	6.4
	0.01	0.00000	1.00000	0.00354	1.04365	6.6
Russell 2000 (1318,90)	0.0025	Infeasible				
	0.005	0.00000	1.00000	0.00809	0.90000	38.4
	0.0075	0.00000	1.00000	0.00799	0.90713	33.6
	0.01	0.00000	1.00000	0.00735	0.88762	27.3
Russell 3000 (2151,70)	0.0025	Infeasible				
	0.005	0.00000	1.00000	0.00960	0.89661	53.5
	0.0075	0.00000	1.00000	0.00961	0.91911	51.9
	0.01	0.00000	1.00000	0.01027	0.87808	47.1
Average		0.00002	0.99745	0.00209	0.93576	10.3
Average Std. Error		0.00011	0.05233	0.00065	0.15997	
Average Skewness		0.01530		−0.01424		
Average Kurtosis		2.15409		0.75259		

are averages over all the in-sample and out-of-sample periods considered in the course of systematic revision. It is clear from Table 4 that computation times are very reasonable, an average of approximately 10 seconds, with no case taking more than 55 seconds.

4.2.5. Comparison with Beasley et al. [5] and other previous work

The smaller data sets considered in this paper (up to the Nikkei 225 in Table 1) have been considered previously in Beasley et al. [5]. The difficulty with doing a direct comparison with that work lies in the different objectives adopted. In [5] a nonlinear objective related to minimising the squared difference between TP return and index return was adopted, whereas in this paper we adopt a linear objective based on regression. As the objective was nonlinear the work presented in [5] uses a specially developed evolutionary heuristic, so there is no guarantee of finding the globally optimum solution. Here, as we have a mixed-integer linear programming formulation, we do find the globally optimum solution. Overall we can say that:

- As our approach optimises on a different objective in-sample when we evaluate it in terms of the objective adopted in [5] it performs less well than the approach given [5], on average some 4.75% worse.
- As our approach finds the globally optimal solution in-sample it performs better than would the approach given in [5] were it to be particularised to our objective.
- It is clear from the computational times given that our approach is much faster than the approach given in [5]. Taking Table 2 and the Nikkei 225, for example, the approach given in [5] requires (on average) 6.6 minutes per γ value on a Silicon Graphics workstation, our approach requires 0.55 seconds on a 3Ghz Pentium pc.

In addition, because our approach uses a standard mixed-integer formulation it is much easier to add to it additional (practical) restrictions, such as sector and lot size constraints discussed previously, than it would be to add the same constraints to the approach given in [5].

With regard to comparing our approach to other work given previously in the literature a key difficulty is the fact that few authors use the same data sets. As such it is extremely hard to judge how, in terms of quality of results, different approaches compare. As noted above we have made the data sets we use publicly available from OR-Library [4], http://people.brunel.ac.uk/~mastjjb/jeb/info.html and we would hope that future workers will make use of these data sets to enable a clearer computational comparison to be made between different approaches for index tracking and enhanced indexation than is possible currently.

4.3. Enhanced indexation

4.3.1. In-sample and out-of-sample results

In this paper, for the data sets we consider, we are interested in achieving a desired level of alpha (excess return) in the in-sample period of [0,145]. Clearly we will also be interested in how well our chosen TP does in terms of excess return (return over and index return) in the out-of-sample period of [145,290].

Table 5 (in columns 4–8) shows in-sample and out-of sample results, for each of our data sets and each of the values of γ considered. In that table we have used values for α^* (the desired excess return level) corresponding to yearly returns of 1%, 2%, 3%, 4%, 5% and 10%. In Table 5 we show the in-sample beta value and the out-of-sample alpha and beta values. In order to provide a more easily grasped value for the reader we also give the average yearly out-of-sample (percentage) excess return, return over and above index return in the same period, (denoted here by AER), as computed directly from the TP returns. AER values are important as they reveal the potential for our approach to produce TPs that, out-of-sample, out-perform the index over time.

In order to judge the significance of the AER values quoted we conducted the one-sided hypothesis test $H_0 : \text{AER} = 0$ versus $H_1 : \text{AER} > 0$. Using a Student's t-distribution Table 5 also shows (where appropriate) the significance level (expressed as a percentage) associated with this hypothesis test. This significance level can be regarded as the probability of achieving (by chance) the AER value seen when (in a statistical sense) the actual underlying AER value is indeed zero (as under H_0). If the significance level is more than 5% then it is not shown.

For the DAX 100, for example, the significance level is 0.00% out-of-sample, implying that we would reject H_0 and accept H_1, that the value of AER achieved (4.74%) is a clear indication that AER is indeed greater than zero. The lower the significance level the more significant our results are in terms of achieving AER values greater than zero.

It is noticeable from Table 5 (columns 4 to 8) that the large indices, namely the S&P 500, Russell 2000 and Russell 3000 all achieve AER's that are highly significant statistically. Also note here that computation times were in all cases very reasonable. As was observed for index tracking, Table 3, out-of-sample we typically have negative skewness and nonzero kurtosis.

4.3.2. Systematic revision

In order to investigate the performance of our approach with respect to enhanced indexation over time, we applied systematic revision as described in Section 4.2.4 above.

Table 5 (in columns 9 to 13) shows the in-sample beta value, the out-of-sample alpha and beta values, the average yearly excess return (AER) and the time taken (in seconds) for systematic revision. Again it is noticeable that the large indices, namely the S&P 500, Russell 2000 and Russell 3000 all achieve AER's that are highly significant statistically. Computation times are also reasonable, with an average time of approximately 40 seconds for the largest data set (the Russell 3000). The largest computation time seen (over all cases in Table 5) was no more than 70 seconds.

In our enhanced indexation two-stage approach (whether for a single period, or for systematic revision) we fix alpha and optimise so as to achieve beta as close to one as possible (after which we minimise transaction cost, whilst leaving alpha and beta unchanged). The excess returns (significant AER's) seen in Table 5, at least for the three largest indices, may arise from the approach we have adopted, or may be a consequence of the return distribution associated with the TPs chosen. Here evidence is mixed. For the single out-of-sample period results average skewness over the three largest indices is negative, −0.11474, whilst for systematic revision (over a number of out-of-sample periods) average skewness for these indices is positive, 0.11487. Kurtosis is clearly nonzero in both cases.

5. Conclusions

In this paper we have presented related mixed-integer linear programming formulations for index tracking and enhanced indexation. We proposed a three-stage solution procedure for index tracking, with the first-stage being associated with achieving a regression intercept as close to zero as possible; the second-stage being associated with achieving a regression slope as close to one as possible; and the third-stage being associated with minimising transaction cost subject to retaining the values for intercept and slope achieved at the first two-stages.

Table 5
Enhanced indexation

Index (N,K)	Desired level of return	Trans. cost limit γ	In-sample	Out-of-sample			Time	Systematic revision				
			β	α	β	AER		In-sample	Out-of-sample			Time
								β	α	β	AER	
Hang Seng (31,10)	1%	0.0025	Infeasible					Infeasible				
		0.005	1.00000	−0.00045	0.99581	−2.43	0.1	1.00000	−0.00132	0.97225	−6.58	0.3
		0.0075	1.00000	−0.00048	0.99521	−2.56	0.1	1.00000	−0.00141	0.97191	−7.06	0.3
		0.01	1.00000	−0.00050	0.99462	−2.70	0.1	1.00000	−0.00149	0.97156	−7.53	0.3
	2%	0.0025	Infeasible					Infeasible				
		0.005	1.00000	−0.00047	0.99504	−2.50	0.1	1.00000	−0.00113	0.97158	−5.51	0.4
		0.0075	1.00000	−0.00047	0.99365	−2.53	0.1	1.00000	−0.00122	0.97128	−6.03	0.3
		0.01	1.00000	−0.00046	0.99226	−2.55	0.1	1.00000	−0.00131	0.97098	−6.54	0.9
	3%	0.0025	Infeasible					Infeasible				
		0.005	1.00000	−0.00048	0.99433	−2.58	0.1	1.00000	−0.00095	0.97113	−4.53	0.4
		0.0075	1.00000	−0.00048	0.99295	−2.60	0.1	1.00000	−0.00104	0.97079	−5.02	0.4
		0.01	1.00000	−0.00048	0.99157	−2.63	0.1	1.00000	−0.00112	0.97045	−5.52	0.3
	4%	0.0025	Infeasible					Infeasible				
		0.005	1.00000	−0.00049	0.99365	−2.65	0.1	1.00000	−0.00092	0.97129	−4.35	0.3
		0.0075	1.00000	−0.00049	0.99227	−2.68	0.1	1.00000	−0.00092	0.97052	−4.37	0.3
		0.01	1.00000	−0.00049	0.99089	−2.71	0.1	1.00000	−0.00096	0.97011	−4.58	0.4
	5%	0.0025	0.98690	−0.00050	0.98714	−2.84	0.1	Infeasible				
		0.005	1.00000	−0.00050	0.99299	−2.73	0.1	1.00000	−0.00081	0.97332	−3.74	0.4
		0.0075	1.00000	−0.00050	0.99162	−2.76	0.1	1.00000	−0.00081	0.97254	−3.77	0.3
		0.01	1.00000	−0.00050	0.99024	−2.78	0.1	1.00000	−0.00081	0.97175	−3.80	0.3
	10%	0.0025	1.00000	−0.00054	0.99057	−2.98	0.1	1.00000	−0.00015	0.97849	−0.37	0.3
		0.005	1.00000	−0.00056	0.99004	−3.11	0.1	1.00000	0.00006	0.97704	0.69	0.4
		0.0075	1.00000	−0.00056	0.98863	−3.13	0.0	1.00000	0.00009	0.97578	0.84	0.5
		0.01	1.00000	−0.00056	0.98727	−3.15	0.1	1.00000	0.00002	0.97612	0.47	0.4
Average			0.99935	−0.00050	0.99204	−2.73	0.1	1.00000	−0.00085	0.97257	−4.07	0.4
Average S.E.			0.03002	0.00017	0.03941			0.02564	0.00061	0.10476		
Ave. Skewness			−0.20265	0.07072				−0.06689	−0.17736			
Ave. Kurtosis			1.18915	1.95495				1.24973	0.65664			
DAX 100 (85,10)	1%	0.0025	Infeasible					Infeasible				
		0.005	0.82274	0.00156	0.99943	8.43	0.2	0.97351	0.00096	0.97058	4.99	0.8
		0.0075	0.95929	0.00004	0.96694	−0.47	0.1	0.99266	−0.00053	0.98800	−1.72	0.8
		0.01	1.00000	−0.00004	0.97305	−0.80	0.0	1.00000	−0.00062	0.96727	−2.31	0.5
	2%	0.0025	Infeasible					Infeasible				
		0.005	0.82137	0.00151	1.01061	8.44	0.2	0.97319	0.00093	0.98708	5.12	0.8
		0.0075	0.95277	0.00026	0.98796	1.08	0.1	0.99170	−0.00057	0.99822	−1.72	1.3
		0.01	1.00000	0.00005	0.98903	0.02	0.1	1.00000	−0.00049	0.97798	−1.51	0.8
	3%	0.0025	0.68434	0.00186	0.88510	7.49	0.1	Infeasible				
		0.005	0.83450	0.00072	0.99751	3.75	0.1	0.97440	0.00044	0.99158	2.69	0.7
		0.0075	0.94582	0.00039	1.00678	2.19	0.0	0.99068	−0.00017	1.00154	0.21	0.9
		0.01	1.00000	0.00010	0.98250	0.16	0.0	1.00000	−0.00065	0.98212	−2.04	0.6
	4%	0.0025	0.68315	0.00223	0.94012	10.84	0.1	0.90310	0.00157	0.93394	8.16	0.5
		0.005	0.82869	0.00063	1.00579	3.44	0.0	0.97405	0.00072	1.01053	4.33	0.5
		0.0075	0.93895	0.00052	1.02491	3.27	0.0	0.98968	−0.00004	1.00797	1.30	0.9
		0.01	1.00000	0.00011	0.95378	−0.43	0.0	1.00000	−0.00070	0.98421	−2.13	0.7
	5%	0.0025	0.67677	0.00232	0.96132	11.88	0.1	Infeasible				
		0.005	0.82274	0.00084	1.02438	5.00	0.0	0.97320	0.00080	1.02644	5.07	0.5
		0.0075	0.93180	0.00063	1.03809	4.18	0.0	0.98867	−0.00010	1.02867	1.16	0.8
		0.01	1.00000	0.00019	0.95222	−0.05	0.1	1.00000	−0.00067	0.99430	−1.79	0.4
	10%	0.0025	0.67536	0.00199	1.00034	10.90	0.2	Infeasible				
		0.005	0.79091	0.00154	1.09645	10.60	0.1	0.96690	0.00124	1.09870	8.80	0.7
		0.0075	0.88698	0.00116	1.06247	7.62	0.1	0.98303	0.00056	1.07819	5.13	0.4

		0.01	0.97459	0.00106	1.04868	6.77	0.0	0.99583	0.00015	1.07814	3.73	0.7
Average			0.87413	0.00089	0.99579	4.74	0.1	0.98266	0.00015	1.00555	1.97	0.7
Average S.E.			0.04912	0.00031	0.07490			0.04356	0.00055	0.14864		
Significance (%)						0.00					1.53	
Ave. Skewness			−0.13878	−0.66023				−0.16930	−0.75875			
Ave. Kurtosis			0.82794	1.66079				0.74627	1.34523			
FTSE 100 (89,10)	1%	0.0025	Infeasible					Infeasible				
		0.005	1.00000	0.00148	0.72745	2.62	0.1	1.00000	0.00072	0.80743	−1.32	0.9
		0.0075	1.00000	0.00145	0.72632	2.42	0.1	1.00000	0.00069	0.80448	−1.52	0.8
		0.01	1.00000	0.00141	0.72518	2.22	0.0	1.00000	0.00065	0.80148	−1.74	0.9
	2%	0.0025	Infeasible					Infeasible				
		0.005	1.00000	0.00160	0.73410	3.40	0.0	1.00000	0.00073	0.81700	−1.23	1.1
		0.0075	1.00000	0.00157	0.73295	3.19	0.1	1.00000	0.00098	0.80937	0.03	0.8
		0.01	1.00000	0.00153	0.73181	2.98	0.1	1.00000	0.00066	0.81118	−1.66	1.0
	3%	0.0025	Infeasible					Infeasible				
		0.005	1.00000	0.00170	0.73907	4.02	0.0	1.00000	0.00113	0.82522	1.05	0.8
		0.0075	1.00000	0.00166	0.73791	3.81	0.1	1.00000	0.00111	0.82268	0.90	0.8
		0.01	1.00000	0.00163	0.73675	3.60	0.1	1.00000	0.00109	0.82015	0.75	0.9
	4%	0.0025	Infeasible					Infeasible				
		0.005	1.00000	0.00179	0.74401	4.62	0.0	1.00000	0.00123	0.83433	1.67	0.7
		0.0075	1.00000	0.00176	0.74284	4.42	0.1	1.00000	0.00121	0.83109	1.51	0.7
		0.01	1.00000	0.00172	0.74167	4.21	0.0	1.00000	0.00119	0.82834	1.35	0.7
	5%	0.0025	1.00000	0.00186	0.74164	4.95	0.1	1.00000	0.00135	0.84834	2.48	0.7
		0.005	1.00000	0.00186	0.74517	5.01	0.1	1.00000	0.00133	0.84556	2.32	0.7
		0.0075	1.00000	0.00185	0.74775	5.01	0.1	1.00000	0.00130	0.84281	2.15	0.7
		0.01	1.00000	0.00182	0.74656	4.80	0.1	1.00000	0.00128	0.84006	1.98	0.7
	10%	0.0025	1.00000	0.00082	0.80057	0.52	0.2	1.00000	0.00086	0.88190	0.90	1.2
		0.005	1.00000	0.00083	0.79871	0.55	0.2	1.00000	0.00112	0.86631	1.81	1.0
		0.0075	1.00000	0.00085	0.79686	0.58	0.2	1.00000	0.00089	0.87665	0.92	0.9
		0.01	1.00000	0.00086	0.79500	0.62	0.2	1.00000	0.00098	0.86510	1.17	0.9
Average			1.00000	0.00150	0.74962	3.18	0.1	1.00000	0.00103	0.83397	0.68	0.8
Average S.E.			0.05787	0.00024	0.06718			0.04963	0.00059	0.17721		
Significance (%)						0.00					2.28	
Ave. Skewness			1.34892	0.08376				1.14255	0.12064			
Ave. Kurtosis			7.23965	1.10883				6.79460	1.01342			
S&P 100 (98,10)	1%	0.0025	1.00000	−0.00028	0.94065	−3.08	0.1	1.00000	−0.00053	0.94472	−5.08	1.3
		0.005	1.00000	−0.00029	0.93943	−3.19	0.1	1.00000	−0.00055	0.94396	−5.22	1.3
		0.0075	1.00000	−0.00031	0.93822	−3.30	0.1	1.00000	−0.00080	0.94307	−6.84	1.1
		0.01	1.00000	−0.00033	0.93700	−3.41	0.1	1.00000	−0.00078	0.93915	−6.85	1.4
	2%	0.0025	1.00000	−0.00018	0.93921	−2.61	0.1	1.00000	−0.00036	0.95318	−3.95	1.2
		0.005	1.00000	−0.00019	0.93805	−2.71	0.3	1.00000	−0.00017	0.94136	−3.32	1.2
		0.0075	1.00000	−0.00021	0.93688	−2.82	0.1	1.00000	−0.00012	0.93678	−3.14	1.2
		0.01	1.00000	−0.00022	0.93572	−2.93	0.1	1.00000	−0.00038	0.94824	−4.23	1.1
	3%	0.0025	1.00000	−0.00008	0.93775	−2.13	0.1	1.00000	0.00003	0.94495	−2.02	1.4
		0.005	1.00000	−0.00009	0.93660	−2.24	0.1	1.00000	0.00002	0.94311	−2.12	1.1
		0.0075	1.00000	−0.00011	0.93546	−2.35	0.2	1.00000	0.00001	0.94128	−2.22	1.2
		0.01	1.00000	−0.00012	0.93432	−2.46	0.1	1.00000	−0.00003	0.94023	−2.45	1.0
	4%	0.0025	1.00000	0.00006	0.94649	−1.20	0.1	1.00000	−0.00027	0.96676	−3.05	1.0
		0.005	1.00000	0.00005	0.94621	−1.27	0.2	1.00000	−0.00030	0.96627	−3.18	1.1
		0.0075	1.00000	0.00003	0.94593	−1.34	0.2	1.00000	−0.00031	0.96458	−3.27	1.0
		0.01	1.00000	0.00002	0.94564	−1.42	0.1	1.00000	−0.00030	0.96150	−3.30	0.8
	5%	0.0025	1.00000	0.00017	0.94777	−0.60	0.4	1.00000	−0.00009	0.97189	−1.79	1.0
		0.005	1.00000	0.00016	0.94831	−0.63	0.2	1.00000	−0.00010	0.97160	−1.87	1.1
		0.0075	1.00000	0.00015	0.94885	−0.67	0.2	1.00000	−0.00008	0.99262	−1.38	1.1
		0.01	1.00000	0.00014	0.94938	−0.70	0.1	1.00000	0.00009	0.98111	−0.83	1.3
	10%	0.0025	1.00000	0.00059	0.95947	1.94	0.2	1.00000	0.00061	0.97413	1.55	0.8
		0.005	1.00000	0.00059	0.97104	2.30	0.2	1.00000	0.00059	0.97263	1.41	0.8

(continued on next page)

Table 5 (*continued*)

Index (N, K)	Desired level of return	Trans. cost limit γ	In-sample	Out-of-sample			Time	Systematic revision				
			β	α	β	AER		In-sample	Out-of-sample			Time
								β	α	β	AER	
		0.0075	1.00000	0.00058	0.96986	2.17	0.2	1.00000	0.00057	0.97113	1.27	0.9
		0.01	1.00000	0.00056	0.96901	2.08	0.1	1.00000	0.00054	0.96962	1.13	0.8
Average			1.00000	0.00003	0.94572	−1.36	0.2	1.00000	−0.00011	0.95766	−2.53	1.1
Average S.E.			0.08576	0.00032	0.05919			0.07428	0.00106	0.20068		
Ave. Skewness			−0.16845	−0.00421				−0.15756	−0.00036			
Ave. Kurtosis			−0.36387	0.05956				−0.14454	−0.03175			
Nikkei 225 (225,10)	1%	0.0025	1.00000	−0.00011	0.97682	−0.51	0.8	1.00000	−0.00056	1.00758	−2.25	2.9
		0.005	1.00000	−0.00011	0.97536	−0.50	0.5	1.00000	−0.00051	1.01158	−2.08	2.7
		0.0075	1.00000	−0.00010	0.97389	−0.49	0.4	1.00000	−0.00015	0.97895	−0.13	1.9
		0.01	1.00000	−0.00010	0.97243	−0.48	0.5	1.00000	−0.00010	0.97891	0.12	1.9
	2%	0.0025	1.00000	−0.00027	1.01934	−1.45	0.3	1.00000	−0.00039	1.00076	−1.38	2.1
		0.005	1.00000	−0.00027	1.01788	−1.43	0.7	1.00000	−0.00038	1.00041	−1.35	2.5
		0.0075	1.00000	−0.00006	0.97310	−0.26	0.4	1.00000	−0.00001	0.97434	0.59	2.4
		0.01	1.00000	−0.00006	0.97163	−0.25	0.4	1.00000	−0.00001	0.97300	0.61	2.1
	3%	0.0025	1.00000	−0.00022	1.01814	−1.18	0.3	0.99996	−0.00006	0.96879	0.73	1.8
		0.005	1.00000	−0.00023	1.01748	−1.20	0.3	1.00000	−0.00006	0.96774	0.72	2.2
		0.0075	1.00000	−0.00002	0.97233	−0.03	0.3	1.00000	−0.00006	0.96630	0.76	1.7
		0.01	1.00000	−0.00001	0.97087	−0.02	0.2	1.00000	−0.00005	0.96472	0.81	1.9
	4%	0.0025	0.98797	0.00003	0.98102	0.20	0.1	0.98896	0.00061	0.94835	4.13	2.7
		0.005	1.00000	0.00002	0.98995	0.14	0.2	1.00000	0.00001	0.96608	1.21	1.9
		0.0075	1.00000	0.00002	0.98756	0.15	0.2	1.00000	0.00013	0.96437	1.67	1.5
		0.01	1.00000	0.00003	0.98517	0.16	0.1	1.00000	0.00014	0.96256	1.74	1.5
	5%	0.0025	0.92858	0.00045	0.91387	2.52	0.1	Infeasible				
		0.005	1.00000	0.00007	0.99320	0.35	0.2	1.00000	0.00027	0.96617	2.59	2.1
		0.0075	1.00000	0.00007	0.99080	0.36	0.3	1.00000	0.00029	0.96543	2.68	1.5
		0.01	1.00000	0.00007	0.98840	0.38	0.2	1.00000	0.00031	0.96151	2.81	1.8
	10%	0.0025	Infeasible					Infeasible				
		0.005	Infeasible					Infeasible				
		0.0075	1.00000	0.00141	0.94626	7.71	0.2	1.00000	0.00185	0.89046	11.25	1.5
		0.01	1.00000	0.00143	0.94379	7.81	0.1	1.00000	0.00192	0.88820	11.60	1.5
Average			0.99621	0.00009	0.98088	0.54	0.3	0.99947	0.00015	0.96696	1.75	2.0
Average S.E.			0.03139	0.00001	0.03147			0.02735	0.00047	0.10359		
Significance (%)						-					1.85	
Ave. Skewness			0.81712	−0.20054				0.67118	−0.15191			
Ave. Kurtosis			1.80609	2.01130				1.72537	0.40130			
S&P 500 (457,40)	1%	0.0025	Infeasible					Infeasible				
		0.005	1.00000	0.00267	1.01072	14.78	0.8	1.00000	0.00348	1.05607	19.06	4.6
		0.0075	1.00000	0.00266	1.00775	14.76	0.5	1.00000	0.00375	1.06448	20.50	4.9
		0.01	1.00000	0.00265	1.00477	14.75	0.8	1.00000	0.00373	1.06212	20.48	5.7
	2%	0.0025	Infeasible					Infeasible				
		0.005	1.00000	0.00259	1.00747	14.32	0.6	1.00000	0.00361	1.07397	19.59	4.8
		0.0075	1.00000	0.00258	1.00447	14.31	0.8	1.00000	0.00360	1.07170	19.63	5.0
		0.01	1.00000	0.00257	1.00146	14.30	0.9	1.00000	0.00358	1.06979	19.61	4.5
	3%	0.0025	Infeasible					Infeasible				
		0.005	1.00000	0.00250	1.00410	13.87	0.7	1.00000	0.00356	1.07617	19.61	3.9
		0.0075	1.00000	0.00250	1.00107	13.86	0.8	1.00000	0.00355	1.07427	19.59	3.3
		0.01	1.00000	0.00249	0.99803	13.86	0.7	1.00000	0.00353	1.07237	19.57	4.2
	4%	0.0025	Infeasible					Infeasible				
		0.005	1.00000	0.00219	1.05692	11.53	0.7	1.00000	0.00313	1.13930	17.76	4.1
		0.0075	1.00000	0.00219	1.05085	11.62	0.7	1.00000	0.00314	1.13508	17.82	5.6

		0.01	1.00000	0.00220	1.04479	11.71	0.7	1.00000	0.00259	1.12353	15.36	4.5
	5%	0.0025	Infeasible					Infeasible				
		0.005	1.00000	0.00219	1.03344	11.74	0.6	1.00000	0.00256	1.11800	15.15	6.3
		0.0075	1.00000	0.00219	1.02743	11.83	0.6	1.00000	0.00256	1.11365	15.20	3.9
		0.01	1.00000	0.00220	1.02142	11.92	0.8	1.00000	0.00256	1.10930	15.24	4.0
	10%	0.0025	1.18490	0.00247	1.22130	11.69	0.5	1.02159	0.00354	1.10937	18.62	5.8
		0.005	1.00000	0.00206	1.04734	10.89	0.3	1.00000	0.00250	1.11877	14.46	3.3
		0.0075	1.00000	0.00206	1.04731	10.90	0.5	1.00000	0.00251	1.11703	14.53	3.1
		0.01	1.00000	0.00206	1.04729	10.91	0.3	1.00000	0.00252	1.11529	14.61	4.0
Average			1.00973	0.00237	1.03358	12.82	0.6	1.00114	0.00316	1.09580	17.70	4.5
Average S.E.			0.07765	0.00008	0.05151			0.06124	0.00049	0.13306		
Significance (%)						0.00					0.00	
Ave. Skewness			−0.62029	0.06207				−0.15884	0.22175			
Ave. Kurtosis			1.21915	0.39217				1.29845	0.58325			
Russell 2000 (1318,90)	1%	0.0025	Infeasible					Infeasible				
		0.005	1.00000	0.00193	1.35638	12.19	5.0	1.00000	0.00535	1.07769	34.03	34.5
		0.0075	1.00000	0.00193	1.34753	12.16	2.2	1.00000	0.00721	0.93747	48.11	19.3
		0.01	1.00000	0.00194	1.33837	12.18	1.5	1.00000	0.00820	0.94107	56.99	23.0
	2%	0.0025	Infeasible					Infeasible				
		0.005	1.00000	0.00191	1.35153	12.07	3.7	1.00000	0.00752	0.93292	51.79	44.5
		0.0075	1.00000	0.00184	1.35546	11.69	5.3	1.00000	0.00433	1.07091	28.02	33.1
		0.01	1.00000	0.00183	1.33731	11.52	4.4	1.00000	0.00737	0.94712	50.73	24.5
	3%	0.0025	Infeasible					Infeasible				
		0.005	1.00000	0.00185	1.33993	11.67	1.6	1.00000	0.00747	0.97268	51.83	49.7
		0.0075	1.00000	0.00188	1.35091	11.90	2.1	1.00000	0.00467	1.07969	29.92	60.1
		0.01	1.00000	0.00179	1.33355	11.31	1.6	1.00000	0.00467	1.09747	30.07	28.2
	4%	0.0025	Infeasible					Infeasible				
		0.005	1.00000	0.00193	1.32775	12.09	4.0	1.00000	0.00365	1.09321	24.10	33.2
		0.0075	1.00000	0.00176	1.30150	10.95	3.4	1.00000	0.00617	0.95761	41.73	36.6
		0.01	1.00000	0.00194	1.33416	12.17	4.4	1.00000	0.00580	0.99296	39.36	19.7
	5%	0.0025	Infeasible					Infeasible				
		0.005	1.00000	0.00185	1.31548	11.53	4.3	1.00000	0.00287	1.05442	17.28	44.7
		0.0075	1.00000	0.00179	1.31270	11.18	3.4	1.00000	0.00529	0.95550	33.86	46.8
		0.01	1.00000	0.00186	1.33769	11.71	2.5	1.00000	0.00493	0.96637	30.93	20.2
	10%	0.0025	Infeasible					Infeasible				
		0.005	1.00000	0.00155	1.16507	9.15	4.3	1.00000	0.00225	1.05548	13.09	25.5
		0.0075	1.00000	0.00169	1.15047	9.86	2.8	1.00000	0.00408	0.98031	24.66	26.5
		0.01	1.00000	0.00169	1.16496	9.86	2.9	1.00000	0.00605	0.93773	37.88	36.9
Average			1.00000	0.00183	1.30671	11.40	3.3	1.00000	0.00544	1.00281	35.80	33.7
Average S.E.			0.03888	0.00004	0.05523			0.04002	0.00066	0.14473		
Significance (%)						0.00					0.00	
Ave. Skewness			−0.95534	−0.16490				−0.64751	0.03548			
Ave. Kurtosis			2.55635	1.02989				2.28519	0.62337			
Russell 3000 (2151,70)	1%	0.0025	1.00000	0.00365	1.12423	19.98	16.3	Infeasible				
		0.005	1.00000	0.00363	1.11693	19.90	8.7	1.00000	0.00740	1.06702	46.34	47.6
		0.0075	1.00000	0.00361	1.10989	19.83	3.8	1.00000	0.00944	0.89458	66.11	34.9
		0.01	1.00000	0.00406	1.11034	22.62	5.3	1.00000	0.00906	0.86168	62.60	28.6
	2%	0.0025	1.00000	0.00366	1.13631	19.92	4.8	Infeasible				
		0.005	1.00000	0.00357	1.09651	19.67	7.5	1.00000	0.00923	0.88927	64.43	52.5
		0.0075	1.00000	0.00357	1.09491	19.70	6.8	1.00000	0.00732	1.06398	46.28	34.9
		0.01	1.00000	0.00381	1.07706	21.30	8.0	1.00000	0.00922	0.89429	64.77	28.4
	3%	0.0025	1.00000	0.00363	1.14108	19.71	4.2	Infeasible				
		0.005	1.00000	0.00366	1.13616	19.94	3.3	1.00000	0.00672	1.04776	41.63	42.2
		0.0075	1.00000	0.00369	1.13193	20.16	5.1	1.00000	0.00837	0.90060	57.34	38.4
		0.01	1.00000	0.00364	1.12272	19.93	5.6	1.00000	0.00616	1.06325	36.95	35.8

(*continued on next page*)

Table 5 (*continued*)

Index (N, K)	Desired level of return	Trans. cost limit γ	In-sample	Out-of-sample			Time	Systematic revision				
			β	α	β	AER		In-sample	Out-of-sample			Time
								β	α	β	AER	
	4%	0.0025	1.00000	0.00359	1.17056	19.23	7.2	Infeasible				
		0.005	1.00000	0.00349	1.09745	19.14	6.9	1.00000	0.00539	1.04041	31.48	48.6
		0.0075	1.00000	0.00361	1.13378	19.63	4.0	1.00000	0.00840	0.90046	57.95	30.7
		0.01	1.00000	0.00359	1.12714	19.67	4.4	1.00000	0.00580	1.07496	34.26	35.6
	5%	0.0025	1.00000	0.00354	1.10715	19.40	9.0	Infeasible				
		0.005	1.00000	0.00354	1.10753	19.42	5.2	1.00000	0.00615	1.03633	37.57	40.3
		0.0075	1.00000	0.00355	1.10691	19.45	4.2	1.00000	0.00525	1.05618	30.22	43.4
		0.01	1.00000	0.00355	1.10714	19.48	4.3	1.00000	0.00577	1.04605	34.90	32.4
	10%	0.0025	1.00000	0.00369	1.13258	20.12	9.7	Infeasible				
		0.005	1.00000	0.00369	1.14375	20.03	5.7	1.00000	0.00541	1.06016	32.23	42.7
		0.0075	1.00000	0.00391	1.10934	21.39	5.7	1.00000	0.00537	1.08005	31.64	66.2
		0.01	1.00000	0.00395	1.11460	21.93	4.9	1.00000	0.00576	1.07502	35.01	42.1
Average			1.00000	0.00366	1.11900	20.06	6.3	1.00000	0.00701	1.00289	45.10	40.3
Average S.E.			0.05650	0.00009	0.07640			0.05689	0.00063	0.18956		
Significance (%)						0.00					0.00	
Ave. Skewness			−0.67213	−0.24140				−0.47638	0.08737			
Ave. Kurtosis			1.67539	0.15316				1.94453	0.44065			

For enhanced indexation we proposed a two-stage solution procedure with the first-stage being associated with achieving a regression slope as close to one as possible, subject to a constraint on the regression intercept; and the second-stage being associated with minimising transaction cost subject to retaining the value for slope achieved at the first-stage.

Computational results were presented for a number of data sets involving up to 2151 stocks. Problems of this size are much larger than have been considered previously by other authors in the literature. Computational times were in all cases reasonable.

References

[1] C. Alexander, Optimal hedging using cointegration, Philosophical Transactions of the Royal Society of London Series A – Mathematical Physical and Engineering Sciences 357 (1999) 2039–2058.
[2] C. Alexander, A. Dimitriu, Indexing, cointegration and equity market regimes, International Journal of Finance and Economics 10 (2005) 213–231.
[3] C. Alexander, A. Dimitriu, Indexing and statistical arbitrage: Tracking error or cointegration?, Journal of Portfolio Management 31 (2005) 50–63
[4] J.E. Beasley, OR-Library: Distributing test problems by electronic mail, Journal of the Operational Research Society 41 (1990) 1069–1072.
[5] J.E. Beasley, N. Meade, T.-J. Chang, An evolutionary heuristic for the index tracking problem, European Journal of Operational Research 148 (2003) 621–643.
[6] T.F. Coleman, Y. Li, J. Henniger, Minimizing tracking error while restricting the number of assets, Journal of Risk 8 (2006) 33–56.
[7] D. Colwell, N. El-Hassan, O.K. Kwon, Hedging diffusion processes by local risk minimization with applications to index tracking, Journal of Economic Dynamics and Control 31 (2007) 2135–2151.
[8] F. Corielli, M. Marcellino, Factor based index tracking, Journal of Banking and Finance 30 (2006) 2215–2233.
[9] U. Derigs, N.-H. Nickel, Meta-heuristic based decision support for portfolio optimisation with a case study on tracking error minimization in passive portfolio management, OR Spectrum 25 (2003) 345–378.
[10] C. Dose, S. Cincotti, Clustering of financial time series with application to index and enhanced index tracking portfolio, Physica A – Statistical Mechanics and Its Applications 355 (2005) 145–151.
[11] Y. Fang, S.-Y. Wang, A fuzzy index tracking portfolio selection model, Lecture Notes in Computer Science 3516 (2005) 554–561.
[12] S.M. Focardi, F.J. Fabozzi, A methodology for index tracking based on time-series clustering, Quantitative Finance 4 (2004) 417–425.
[13] A.A. Gaivoronoski, S. Krylov, N. van der Wijst, Optimal portfolio selection and dynamic benchmark tracking, European Journal of Operational Research 163 (2005) 115–131.
[14] M. Gilli, E. Kellezi, The threshold accepting heuristic for index tracking, in: P. Pardalos, V.K. Tsitsiringos (Eds.), Financial Engineering E-Commerce and Supply Chain Kluwer Applied Optimization Series, 2002, pp. 1–18.
[15] ILOG CPLEX. http://www.ilog.com/products/cplex/news/whatsnew.cfm#cplex90 (2005).
[16] R. Jansen, R. van Dijk, Optimal benchmark tracking with small portfolios, Journal of Portfolio Management 28 (2002) 33–39.
[17] H. Konno, T. Hatagi, Index-plus-alpha tracking under concave transaction cost, Journal of Industrial and Management Optimization 1 (2005) 87–98.
[18] J. Li, Y. Ma, Y. Zeng, Research on application of Stein rule estimation for index tracking problem, in: Proceedings of the 2005 International Conference on Management Science and Engineering, 12th (1–3), 2005, pp. 434–439.
[19] N. Meade, J.E. Beasley, Detection of momentum effects using an index out-performance strategy, Working paper (2007) available from the second author at CARISMA Mathematical Sciences Brunel University, Uxbridge UB8 3PH, United Kingdom.
[20] K.J. Oh, T.Y. Kim, S. Min, Using genetic algorithm to support portfolio optimization for index fund management, Expert Systems with Applications 28 (2005) 371–379.
[21] N. Okay, U. Akman, Index tracking with constraint aggregation, Applied Economics Letters 10 (2003) 913–916.
[22] A.S.J. Stoyan, B.R.H. Kwon, A two-stage stochastic mixed-integer programming approach to the index tracking problem, Working paper (2007) available from the first author at Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, Canada.
[23] L.C. Wu, S.C. Chou, C.C. Yang, C.S. Ong, Enhanced index investing based on goal programming, Journal of Portfolio Management 33 (3) (2007) 49–56.
[24] D.D. Yao, S. Zhang, X.Y. Zhou, Tracking a financial benchmark using a few assets, Operations Research 54 (2006) 232–246.
[25] L. Yu, S. Zhang, X.Y. Zhou, A downside risk analysis based on financial index tracking models, in: A.N. Shiryaev, M.R. Grossinho, P.E. Oliveira, M.L. Esquivel (Eds.), Stochastic Finance, Springer, 2006, pp. 213–236.