Available online at www.sciencedirect.com
SCIENCE DIRECT®
European Journal of Operational Research 148 (2003) 621–643
EUROPEAN JOURNAL OF OPERATIONAL RESEARCH
www.elsevier.com/locate/dsw

O.R. Applications

An evolutionary heuristic for the index tracking problem

J.E. Beasley *, N. Meade, T.-J. Chang

The Management School, Imperial College, 53 Prince's Gate, Exhibition Road, London SW7 2AZ, UK

Received 14 November 2000; accepted 19 April 2002

Abstract

Index tracking is a popular form of passive fund management. The index tracking problem is the problem of reproducing the performance of a stock market index, but without purchasing all of the stocks that make up the index. Our formulation of the problem explicitly includes transaction costs (associated with buying or selling stocks) and a limit on the total transaction cost that can be incurred. Our formulation also includes a constraint limiting the number of stocks that can be purchased.

An evolutionary heuristic (population heuristic) is presented for the solution of the index tracking problem. Reduction tests are also presented. Computational results are presented for five data sets drawn from major world markets. These data sets are made publicly available for use by other workers.

Keywords: Index tracking; Passive fund management; Evolutionary heuristic

1. Introduction

Fund management involves the investment of many millions of dollars (or their equivalent) in the equities (stocks, shares) of companies quoted on the world's stock markets. The funds for investment are typically provided by pension contributions, insurance premiums and savings. The objective for fund managers is to provide a combination of capital growth and income over the medium to long-term. The basic strategies adopted by fund managers can be broadly classified as

(a) *Active management*, where the fund managers have a high degree of flexibility and attempt to "pick winners", stocks whose values are going to outperform (over time) other stocks. The assumption underlying this strategy is that human fund mangers can, through their expertise and judgement, add value though choosing high performing stocks and/or by the timing of their buy/sell decisions.

(b) *Passive management*, where the fund managers have much less flexibility and their role is to conform to a closely defined set of criteria. Common criteria are that the fund should achieve approximately the same return as a specified market index (such as the S&P 500 in New York, or the FTSE 100 in London) through investment in an appropriately selected set of stocks that are present in the index.

* Corresponding author. Tel.: +44-207-594-9171; fax: +44-207-823-7685.

E-mail address: j.beasley@ic.ac.uk (J.E. Beasley).

0377-2217/03/$ - see front matter
doi:10.1016/S0377-2217(02)00425-3

Index tracking describes the process of attempting to track (reproduce) the performance of an index. A passively managed fund whose objective is to reproduce the return on an index is known as an *index fund*, or *tracker fund*. Market indices that can be tracked are not only available for national stock markets, but are also available to reflect regional stock market performances (e.g. S&P Europe 350) and even global performance (e.g. FTSE All-World Index). Sutcliffe [51] details a number of the indices that are available.

The above strategies are pure strategies, but mixed strategies, whereby a portion of the fund is invested passively and the remainder actively managed, are also possible. Active and passive fund management strategies have their respective strengths and weaknesses:

(a) Active management has high fixed costs (associated with payments to the management team) and the frequent trading involved in stock picking means high transaction costs. If all goes well these costs will be offset by the returns obtained.
(b) Passive management has lower fixed costs and lower transaction costs, but has the disadvantage that if the market (as represented by the index) falls, so inevitably will the return obtained from an index fund.

In active management an investor is exposed to both market and company risk, whilst in passive management an investor is only exposed to market risk.

In recent years, both in the USA and in Europe, passive management (and in particular index tracking) has been receiving a much higher profile for two reasons:

(1) Historical empirical analysis has revealed that:
    (a) Whilst the best of the actively managed funds outperform the market in any particular year, over the long-term the majority of such funds do not (e.g. in the UK in 1998 only one quarter of actively managed funds outperformed their comparative index over a five year period [14]).
    (b) An actively managed fund which outperforms the market one year may fail to do so in subsequent years (e.g. in the UK many funds that performed well in 1992 had fallen to bottom quartile positions by 1998).
(2) As stock markets (and so their indices) have historically risen in the long-term it has become clear that reasonable returns can be obtained without incurring the extra risks associated with active management.

For more concerning these points as to the relative merits of active and passive fund management see [4,11,15,22,25,33,36,49]. It is estimated that in total over 10^{12} dollars are passively invested in index funds in the USA alone [49].

Consider now a passively managed fund that intends to track a single given index. This fund could purchase all of the stocks that make up the index (in appropriate quantities) and so perfectly reproduce the index. This approach is known as full (or complete) replication. Whilst simple, both conceptually and computationally, full replication has a number of disadvantages:

(1) Certain stocks that are in the index may be held (proportionally) in very small quantities. For example if the number of distinct stocks included in the index is large this inevitably happens. This may be administratively inconvenient and, because there may be a restricted market for such stocks, they can be expensive to buy.
(2) When the underlying composition of the index is revised the holdings of all stocks will typically need to be changed to reflect their new weightings in the index. Index revisions can occur for a number of reasons. For example, one company may grow sufficiently large to merit inclusion in the index, and this may involve excluding another company from the index. Mergers are another reason why indices are revised. In recent years a common reason for a stock to drop out of an index has been that its price has fallen so low that, in market capitalisation terms, another stock merits inclusion in the index. The composition of the S&P 500, for example, underwent 60 revisions in 2000 alone [50].

(3) Transaction costs associated with the purchase and sale of stocks need to be considered. Full replication, involving as it does a simple mathematical formula to compute the holding of each stock, provides no means of limiting transaction costs so as to prevent them becoming excessive.

For these reasons many passively managed funds hold fewer stocks than are included in the index they are tracking [17,29,41,46].

One way for investors to engage in index tracking is to purchase an exchange traded fund (ETF)—this trades like a stock but its value at any time essentially reflects the underlying index value. The advantage of ETF's are that they are continually traded and hence can be bought and sold minute by minute (for relatively small sums). They hence enable investors to buy into the index when it is low (and sell when it is high) without incurring company risk.

ETF's are currently very popular in the USA. For example on the American Stock Exchange (ASE), during 2000, average ETF daily volume nearly doubled for the sixth consecutive year. In non-US markets ETF's are very new, only being introduced in the UK in 2000 for example. At the time of writing the two largest ETF's on the ASE, comprising between them 67% of the net assets invested in ETF's, are the Standard and Poor's Depositary Receipts (SPDR's, ASE trading symbol (ticker) SPY) based on the S&P 500 and the Nasdaq 100 Trust Shares (ticker QQQ) based on the Nasdaq 100. Both of these ETF's use full replication of the indices they are tracking. Not all ETF's use full replication however. For example, at the time of writing, the ETF with ticker FFF designed to track the Fortune 500 contains 441 stocks and the ETF with ticker VTI designed to track the Wilshire 5000 (approximately 7000 stocks) contains 3436 stocks.

In the light of the problems with full replication discussed above a natural decision problem is how can we (imperfectly) reproduce index performance with fewer stocks whilst limiting total transaction cost. This is the *index tracking problem* addressed in this paper. We refer to the stocks that we choose to hold to track an index as a *tracking portfolio*. We believe that the contribution of this paper is

(1) to encompass, within a single formulation and solution procedure, the problem of creating (from cash) a tracking portfolio as well as the problem of revising (rebalancing) an existing tracking portfolio,
(2) the development of an effective evolutionary heuristic for the solution of the index tracking problem,
(3) to make the test data sets we use publicly available so that future workers will be able to assess the performance of their algorithms for index tracking against our heuristic using exactly the same data.

The structure of this paper is as follows: Section 2 presents a literature survey of the index tracking problem and Section 3 presents our formulation of the problem. Section 4 presents our evolutionary heuristic for the problem and Section 5 presents a number of problem reduction tests. Section 6 presents computational results followed by a conclusions section.

2. Literature survey

The majority of the work relating to index tracking presented in the literature does not consider transaction costs and only considers the problem of creating an initial tracking portfolio from cash. Such work deals with the problem of limiting the number of stocks in the tracking portfolio by first limiting the stocks that can be considered for selection to an approved list of stocks (e.g. designed statistically to cover all of the major market sectors included in the index) and then allowing the number of stocks from this approved list which can be in the tracking portfolio to be unrestricted.

By contrast the formulation we present in this paper does explicitly consider transaction costs, does consider the revision of an existing tracking portfolio (not simply its creation) and does explicitly limit the number of stocks which can appear in the tracking portfolio. In order to structure our literature review we consider factor models, Markowitz models and other work separately below.

2.1. Factor models

Factor models [55] relate the return on a stock to one or more underlying economic factors. A single factor model, for example, might proceed by performing a linear regression for the return on each stock in terms of the return on the index, i.e. an equation of the form return on stock $i =$ constant $+ \beta_i$ (return on the index) is established for each stock i. A variance minimisation model with a constraint requiring the tracking portfolio beta to equal one (since the index has a regression coefficient of one when regressed against itself) is then constructed and solved to find a tracking portfolio.

Rudd [46] gave a simple construction heuristic for index tracking as well as presenting a factor model. He proposed that transaction costs be incorporated into the objective function of the factor model via a weighting parameter to discourage incurring such costs with respect to an initial tracking portfolio. His approach corresponds to the case where transaction costs are not limited and are separate from the cash to be invested in the tracking portfolio. Results were reported for tracking portfolios related to the S&P 500. Haugen and Baker [26] extended the above single factor approach to a multi-factor model. Results were presented for tracking a consumer price index using stocks. Larsen and Resnick [31] used Rudd's approach to construct tracking portfolios to investigate how well a number of indices of their own devising could be tracked, and the effect of timing the decision as to when to revise the tracking portfolio.

2.2. Markowitz models

Markowitz models simply take the standard mean–variance model established for portfolio optimisation by Markowitz [19,21,34,35,47] and apply it to index tracking. Hodges [27] used a Markowitz model and compared the tradeoff curve relating variance to return in excess of the index with the tradeoff curve for the standard Markowitz portfolio optimisation model. Roll [45] discussed tradeoff curves (as in Hodges [27]) and also combined the Markowitz and factor models by adding a constraint relating to the beta of a tracking portfolio. Franks [23] used a Markowitz model in a simulation over 50 years with five stocks. Rohweder [44] presented a Markowitz model which includes in the objective function a term relating to transaction cost. Wang [54] presented a Markowitz model which includes in the objective function terms relating to tracking more than one index, as well as a term relating to transaction cost.

2.3. Other papers

Toy and Zurack [53] discussed tracking portfolios for the Euro-Pac index (containing 1650 stocks). Their approach is to first construct tracking portfolios for individual country indices and then to combine these together. These individual country tracking portfolios were constructed using commercial "optimization software".

Meade and Salkin [37] defined tracking error via a mathematical approximation which enabled the problem to be solved via quadratic programming. They also considered the effect of further constraining the tracking portfolio to have characteristics similar to the index. Using data from the Tokyo stock market they found that there was no benefit to be gained from this. Meade and Salkin [38] built on this work and considered index tracking under the assumption that the stock/index returns follow an autoregressive conditional heteroscedastic process. Using this assumption they solved the problem via quadratic programming. They presented results using data drawn from the London stock market showing the effect of periodically readjusting the tracking portfolio using their approach. Although they consider transaction costs at each readjustment these are not explicitly limited, nor included in their formulation. Adcock and Meade [1] considered the problem of rebalancing a tracking portfolio over time, where transaction costs are incurred at each rebalancing. In their formulation these costs are included (via a weighting factor) in the objective function, but are not explicitly limited. Computational results were presented for a tracking portfolio of 200 stocks.

Worzel et al. [56] presented an approach based upon enumerating potential scenarios when tracking a mortgage index. Their formulation, which only looks at returns over a single period, is designed to maximise expected return, whilst limiting (over all scenarios) underperformance with respect to the index. Their formulation includes transaction costs, but does not explicitly constrain them. Results were presented based on real mortgage data. A similar approach is used by Consiglio and Zenios [18] in tracking a bond index. In order to track a mortgage index in a multi-period framework Zenios et al. [57] presented an approach based upon stochastic programming with the objective being to maximise the expected utility of terminal wealth.

Connor and Leland [17] considered the cash management problem that arises when the tracking portfolio is subjected to random inflows and outflows of cash. In such cases (to avoid transaction costs) it may be preferable to maintain some money in cash that is not invested in the tracking portfolio. They assume that the tracking portfolio exactly replicates the index. Their approach corresponds to the case where transaction costs are not limited and are separate from the cash to be invested in the tracking portfolio. They consider transaction costs defined as a fixed proportion of the monies invested/recouped. Buckley and Korn [13] extended their analysis to include fixed transaction costs. Tabata and Takeda [52] considered the problem of choosing a tracking portfolio containing a given number of stocks so as to minimise the expected value of the squared difference between the return on the tracking portfolio and the return on a benchmark portfolio. They presented a heuristic for the problem and gave results for one example involving 15 stocks.

Alexander [2] considered the problem of choosing a tracking portfolio whose constituents are most highly cointegrated with the index they are tracking. She gives an example of this approach applied to the MSCI EAFE index. Browne [12] adopted a stochastic control framework and considered problems related to outperforming a benchmark index in continuous time and in the absence of transaction costs. Dorfleitner [20] discussed the issue of rounding with respect to the number of units of each stock held that arises in the context of full replication and gave computational results for the German DAX index. Keim [30] considered a fund which has been designed to track an index of small-capitalisation stocks. Such stocks are associated with illiquid markets (lack of buyers/sellers) and consequently high transaction costs. This fund was designed using investment and trading strategy rules. Rudolf et al. [48] discussed four different linear definitions of tracking error that all make use of the absolute difference between the index and portfolio return. They presented an example relating to minimising the tracking error when tracking the MSCI world stock market index using six national stock market indices.

Bamberg and Wagner [6] considered the problem of estimating, via regression, a suitable tracking portfolio containing a fixed subset of index stocks. Computational results were presented for one problem involving 20 stocks. Ammann and Zimmermann [3] considered the tracking errors that would result were an active manager to deviate from index composition, as measured by the percentage devoted to each of five asset classes, via choice of the tracking portfolio with the lowest correlation with the index. Gilli and Këllezi [24] presented an algorithm for index tracking with transaction costs based on threshold accepting. Computational results were presented for a number of data sets involving up to 528 stocks.

3. Formulation

In this section we first outline our basic approach to index tracking and indicate the role that transaction costs and dividends play. We then formulate the index tracking problem.

3.1. Basic approach

Suppose that we observe over time $0, 1, 2, \ldots, T$ the value of N stocks, as well as the value of the index we are tracking and that our investment horizon (i.e. how far into the future we are looking) is L. The value adopted for L will depend upon the frequency with which the composition of the underlying index is revised. Since many indices are revised frequently nowadays horizons greater than six months can be unrealistic. Further suppose that we are interested in deciding the best set of K stocks to hold (where $K < N$), as well as their appropriate quantities. In index tracking we want to answer the question:

> What will be the best set of K stocks to hold (as well as their appropriate quantities) so as to best track the index over the time period $[T, T + L]$?

Our basic approach in index tracking is a historical look-back approach. To ask the historical question:

> What would have been the best set of K stocks to have held (as well as their appropriate quantities) so as to have best tracked the index over the time period $[0, T]$?

The assumption here is that the past is a guide to the future and that by answering this historical question we have an answer to our real question, namely what should we be holding into the future.

The numeric value adopted for T (i.e. how far back into the past we look in order to decide our future tracking portfolio) is determined by

(1) The selection process implies that the returns on the N stocks are multivariate random variables with a constant mean and a constant covariance matrix. Ideally, in order that the estimated covariance matrix has full rank, we need $T > N$ (although for indices with large N this ideal is not achievable). Other practical considerations, such as the age of stocks, mean that T rarely exceeds five years.
(2) Returns on equities are typically characterised as having constant mean and time dependent variance (heteroscedasticity). The effects of heteroscedasticity diminish as frequency of observation decreases, i.e. it is very significant in hourly or daily data, but not in monthly data. Weekly data has been used in the computational results presented below as a compromise between these two considerations.

3.2. Transaction costs and dividends

Our formulation of the index tracking problem includes transaction costs (associated with buying or selling stocks) and an explicit limit on the total transaction cost that can be incurred. Transaction costs appear for two basic reasons:

(a) At time T we may have a cash inflow (new cash to be invested in the fund by buying stock) or cash outflow (cash to be produced by the selling of stock). Buying or selling stock is not free and transaction costs are incurred.
(b) We may already be using some tracking portfolio containing stocks X_i $(i = 1, \dots, N)$. In the light of the information contained in the observed behaviour of these stocks in $[0, T]$ it may be appropriate to change this tracking portfolio. Again this will involve buying/selling stock and transaction costs will be incurred.

Transaction costs must be limited in order to prevent the value of the fund diminishing unduly due to buying/selling stock.

The stocks in the tracking portfolio generate dividend payments. To track an index of total return (as is possible with an appropriately constructed dividend reinvested S&P 500 index for example), it would be necessary to reinvest dividend payments on receipt, leading to high transaction costs. Typically when tracking a price portfolio any dividend payments are accumulated in an interest earning account until the next rebalancing of the portfolio. It is worth noting here that dividend yield has recently declined in importance. Companies are tending to repurchase their own stock, rather than issuing dividends, offering the shareholder return in the form of (potential) capital gains rather than income. In 1998, for example, the dividend yield on the S&P 500 reached a record low of 1.34%. To put this figure in context, even if the index tracked offered 5% per annum above the risk free interest rate and the dividend yield was 5%, over a six month rebalancing period the discrepancy due to non-reinvestment of dividends is only 0.06% of the portfolio value.

3.3. Notation

In order to formulate the index tracking problem we need to define our notation. Let:

- N be the total number of distinct stocks (companies) in which we can invest
- K be the desired number of distinct stocks in the tracking portfolio
- ε_i be the minimum proportion of the tracking portfolio that must be held in stock i $(i = 1, \ldots, N)$ if any of stock i is held. In practice ε_i represents a "minimum holding level" for stock i
- δ_i be the maximum proportion of the tracking portfolio that can be held in stock i $(i = 1, \ldots, N)$ if any of stock i is held. We must have $0 \leqslant \varepsilon_i \leqslant \delta_i \leqslant 1$ and the role of δ_i is to limit the exposure of the tracking portfolio to stock i
- X_i be the number of units of stock i $(i = 1, \ldots, N)$ in the current tracking portfolio
- T be such that we have observed historical values for stocks and the index over the time period $0, 1, 2, \ldots, T$. The time T represents a decision point, a time at which we may switch from our current tracking portfolio $[X_i]$ to a new tracking portfolio
- V_{it} be the value of one unit of stock i $(i = 1, \ldots, N)$ at time t $(t = 0, \ldots, T)$
- I_t be the value of the index at time t $(t = 0, \ldots, T)$
- R_t be the single period continuous time return given by the index, i.e. $R_t = \log_e(I_t/I_{t-1})$ $t = 1, \ldots, T$
- C_{cash} be the cash change in the tracking portfolio at time T ($C_{\text{cash}} > 0$ represents new cash to be invested in the tracking portfolio, $C_{\text{cash}} < 0$ represents cash to be taken out of the tracking portfolio)
- C be the total value ($\geqslant 0$) of the current tracking portfolio $[X_i]$ at time T plus the cash change C_{cash}, i.e. $C = \sum_{i=1}^{N} V_{iT} X_i + C_{\text{cash}}$
- $F_i(\zeta, \Theta, t)$ be a transaction cost function that gives the transaction cost ($\geqslant 0$) incurred for stock i $(i = 1, \ldots, N)$ in moving from holding ζ units of the stock to holding Θ units of the stock at time t. Here $\zeta > \Theta$ corresponds to selling $\zeta - \Theta$ units of stock, $\zeta < \Theta$ corresponds to buying $\Theta - \zeta$ units of stock, and we have $F_i(\zeta, \Theta, t) = 0$ if $\zeta = \Theta$. In the heuristic presented below there is essentially no limitation on the form (linear or non-linear) that this function can take
- γ be the limit on the proportion of C that can be consumed by transaction cost (where $0 \leqslant \gamma \leqslant 1$)

Then our decision variables are

- x_i the number of units of stock i $(i = 1, \ldots, N)$ that we choose to hold in the new tracking portfolio
- z_i 1 if any of stock i $(i = 1, \ldots, N)$ is held in the new tracking portfolio, 0 otherwise

Without significant loss of generality (since the sums of money involved are large) we allow $[x_i]$ to take fractional values. It is helpful when formulating the problem below to introduce:

- C_{trans} the total transaction cost involved in moving from the current tracking portfolio $[X_i]$ to the new tracking portfolio $[x_i]$ at time T
- r_t the single period continuous time return given by the new tracking portfolio $[x_i]$ at time t $(t = 1, \ldots, T)$

These are related to the variables given previously via:

$$C_{\text{trans}} = \sum_{i=1}^{N} F_i(X_i, x_i, T), \tag{1}$$

$$r_t = \log_e \left[\left[\sum_{i=1}^{N} V_{it} x_i \right] \bigg/ \left[\sum_{i=1}^{N} V_{it-1} x_i \right] \right], \quad t = 1, \ldots, T. \tag{2}$$

Eq. (1) defines C_{trans} to be the sum of the individual transaction costs for each stock at time T and Eq. (2) defines r_t to be the return on the new tracking portfolio (since the total value of that portfolio at time t is $\sum_{i=1}^{N} V_{it} x_i$).

In defining r_t and R_t above we have implicitly assumed that successive observations of stock/index values are made at equal time intervals. If, in terms of the underlying elapsed time, there are unequal intervals between such observations (e.g. due to weekends or holidays) r_t and R_t must be modified accordingly. This is easily done by defining G_t, a constant dependent upon the elapsed time between the observations at $t-1$ and t, and multiplying both r_t and R_t by G_t for $t = 1, \ldots, T$.

3.4. Constraints

The constraints associated with the index tracking problem are

$$\sum_{i=1}^{N} z_i = K, \tag{3}$$

$$\varepsilon_i z_i \leqslant V_{iT} x_i / C \leqslant \delta_i z_i, \quad i = 1, \ldots, N, \tag{4}$$

$$C_{\text{trans}} \leqslant \gamma C, \tag{5}$$

$$\sum_{i=1}^{N} V_{iT} x_i = C - C_{\text{trans}}, \tag{6}$$

$$x_i \geqslant 0, \quad i = 1, \ldots, N, \tag{7}$$

$$z_i \in [0, 1], \quad i = 1, \ldots, N. \tag{8}$$

Eq. (3) ensures that there are precisely K stocks in the new tracking portfolio and is a cardinality constraint. Note here however that there is no requirement for the current tracking portfolio $[X_i]$ also to contain K stocks (i.e. the cardinalities of the new and the current tracking portfolios can be different). We include K explicitly in our formulation since deciding the number of stocks to hold can only be made in the light of the tradeoffs involved. This issue will be explored computationally below.

Eq. (4) fulfils two roles. Its first role is to ensure that if a stock i is not in the new tracking portfolio ($z_i = 0$) then x_i is also zero. Its second role is to ensure that if a stock i is in the new tracking portfolio ($z_i = 1$) then x_i is limited appropriately ($\varepsilon_i \leqslant V_{iT} x_i / C \leqslant \delta_i$). Eq. (5) explicitly limits the total transaction cost incurred (γC is a known constant since C is known).

Eq. (6) is a cash balance constraint. It says that the total value of the new tracking portfolio at time T ($\sum_{i=1}^{N} V_{iT} x_i$) must exactly equal the value of the current tracking portfolio at time T plus the cash change (i.e. C) minus the total transaction cost C_{trans}. Note here that this constraint implicitly assumes that, in the absence of any (or sufficient) cash, transaction costs are paid for out of the value associated with the tracking portfolio. Hence an implicit tradeoff occurs in that (within limits) we would expect that the more tracking portfolio value we are prepared to sacrifice (through trading, buying/selling) the better will be the new tracking portfolio that we obtain.

In some fund management applications transaction costs (albeit limited) are paid out of a separate account (e.g. if a standard proportion is taken from new cash inflows and maintained in a separate account purely for transaction cost purposes). In such cases Eq. (6) becomes $\sum_{i=1}^{N} V_{iT} x_i = C$ and Eq. (5) is unchanged, except that γ is defined such that γC corresponds to the amount in the separate transaction cost account that we are prepared to pay in transaction costs. Although we do not present this below our heuristic can be easily modified to deal with this situation.

Note here that Eq. (6), in conjunction with Eq. (2), implies that the return r_T on the new tracking portfolio at time T reflects both the cash change C_{cash} as well as the transaction costs C_{trans} incurred. Amending our heuristic so that these factors do not impact directly on return at time T is easily done and is not given here.

In some fund management applications there may be capital gains/losses incurred when stock is sold and consequently the tax implications of such sales need to be considered. Although we do not present this below our heuristic can be modified to deal with the situation where the total capital gains tax payable needs to be limited.

3.5. Objective

There are a variety of objectives we can use in index tracking. The two factors that are of interest to us in deciding a new tracking portfolio are *tracking error* and *excess return*. We deal with each in turn below.

3.5.1. Tracking error

In our approach tracking error is some function of the difference between r_t (the tracking portfolio return, Eq. (2)) and R_t the index return. In particular if r_t and R_t coincide for all values of t ($t = 1, \ldots, T$) then we would have perfectly reproduced (tracked) the index, i.e. have zero tracking error. Hence in our approach the tracking error E is defined by

$$E = \left[\left[\sum_{t \in S} |r_t - R_t|^{\alpha}\right]^{(1/\alpha)}\right] \Big/ T, \tag{9}$$

where $\alpha > 0$ is the power by which we penalise differences between r_t and R_t and S is a suitable set of time periods over which we compare r_t with R_t. Two cases for S can be distinguished: $S = [t | t = 1, 2, \ldots, T]$, i.e. S comprises all time periods; and $S = [t | r_t < R_t \ t = 1, 2, \ldots, T]$, i.e. S comprises only those time periods for which the tracking portfolio return is less than the index return. The second of these definitions for S corresponds to *downside risk* and represents the fact that, in practice, we may care more about time periods in which our tracking portfolio fails to match the index than about time periods in which it exceeds the index.

The case $\alpha = 2$ effectively corresponds to defining tracking error as the root mean squared error. However there may be advantages to using different values of α. Essentially in our approach we have formulated the problem of determining a new tracking portfolio $[x_i]$ which, had we held it over $[0, T]$, would have performed well. However we are really interested in performance over $[T, T + L]$. We would make two comments here:

(a) in statistical terms we are determining a tracking portfolio that would have performed well in-sample, whereas in reality we are concerned with out-of-sample performance,
(b) in pattern recognition terms (e.g. see [10,32]) we would like to define tracking error in a way that has "generalisation capability", that when suitably tuned on in-sample data performs well on out-of-sample data.

It is clear that deciding a tracking portfolio using a particular value for α in-sample does not automatically lead to a good performance with that value of α out-of-sample.

In practice, we might want to expand Eq. (9) to include a weighting factor Δ_t for each time period t so that, for example, more recent time periods get a higher weighting than other time periods. In such cases we have

$$E = \left[\left[\sum_{t \in S} \Delta_t |r_t - R_t|^{\alpha}\right]^{(1/\alpha)}\right] \Big/ T. \tag{10}$$

In the computational results reported in this paper however we used $\Delta_t = 1 \ \forall t$.

A significant number of papers presented in the literature (e.g. [13,17,23,27,31,40,44–46,54]) essentially define tracking error to be the variance of $\{(r_t - R_t) | t = 1, \ldots, T\}$. We disagree with this definition of tracking error since it implies that a tracking portfolio that constantly underperforms the index, i.e. has $r_t = R_t - M \ \forall t$ where $M > 0$ is a constant, has a tracking error of zero. This illustrates that only using variance in the definition of tracking error has the disadvantage that it ignores the role of bias $(r_t - R_t)$, which is why we prefer our definition (Eq. (10)) above.

In Eq. (10) we have defined (total) tracking error as relating to the sum of accumulated errors. If we had defined (total) tracking error as relating to the maximum error (i.e. $E = \max[(\Delta_t |r_t - R_t|^{\alpha})^{(1/\alpha)} \ |t \in S])$ the heuristic presented in this paper would still apply.

3.5.2. Excess return

Excess return is return over and above the return on the index. This is given by

$$r^* = \sum_{t=1}^{T} (r_t - R_t)/T. \tag{11}$$

In Eq. (11) R_t is a known constant $\forall t$. Hence we could drop R_t from that equation and r^* would then correspond to the average return per period achieved by the tracking portfolio. Expanding Eq. (11) using Eq. (2) and incorporating the definition of R_t we obtain

$$r^* = \left(\log_e \left[\left[\sum_{i=1}^{N} V_{iT} x_i\right] \Big/ \left[\sum_{i=1}^{N} V_{i0} x_i\right]\right] - \log_e (I_T / I_0) \right) \Big/ T \tag{12}$$

so that r^* corresponds to a comparison between the total return achieved by the tracking portfolio and the total return achieved by the index (over the period $[0, T]$).

Excess return is of interest from a competitive viewpoint. Consider the problem of marketing an index tracking fund. If all such funds were simply to track the index to the same degree then the only competitive basis for distinguishing between them would be cost (our fund charges are lower). However if we are prepared to sacrifice a degree of tracking error, to have a tracking portfolio that (historically) would have shown returns over and above the index (excess return >0), then we have another competitive basis with which to distinguish between funds (our fund tracks the index and exceeds it).

3.5.3. Generalised objective

In the light of the discussion as to tracking error and excess return above we can identify a single generalised objective for index tracking as

$$\text{minimise}\, \lambda E - (1 - \lambda) r^*, \tag{13}$$

where λ ($0 \leqslant \lambda \leqslant 1$) represents an implicit tradeoff between tracking error and excess return (note that Eqs. (10) and (11) have been defined in such a way that E and r^* are in comparable units). For example $\lambda = 1$ corresponds to minimising tracking error and $\lambda = 0$ corresponds to maximising excess return.

We will investigate the tradeoff between tracking error and excess return that occurs as λ varies computationally below.

3.6. Complete formulation

Our complete formulation of the index tracking problem is therefore minimise (13) subject to (1)–(8), (10) and (11) and is a non-linear mixed-integer program. There are number of remarks that we can make with respect to this formulation:

(a) Our formulation addresses the problem "if we were to revise the tracking portfolio now what should we revise it to?". If *does not* answer the question "should we revise the tracking portfolio now or leave it unchanged?". Rather, solving our formulation will provide fund managers with information to answer this second question better. For example, in our formulation, by varying γ and hence constructing tradeoff curves illustrating the change in the objective function (e.g. tracking error) as transaction cost increases. This is explored further computationally below.

(b) Note the role of transaction cost. It is a "one-off" cost that we pay at time T in order to change to a tracking portfolio that (historically) would have given better tracking error/excess return. As such the decision as to whether to pay this one-off cost or not depends upon our human judgement as to the timescale into the future we are considering.

(c) At first sight it might seem surprising that our formulation includes nothing explicitly related to the performance (i.e. tracking error and excess return) of the current tracking portfolio $[X_i]$. In fact this is considered implicitly since (obviously) that performance can be obtained at zero transaction cost (when $C_{\text{cash}} = 0$). Hence our formulation implicitly takes this performance as its starting point and (effectively) answers the question "how much better than current performance can we do given our constraint upon transaction cost?".

(d) Our formulation is a strategic model in that it helps answer the question about what new tracking portfolio $[x_i]$ we should have. It is not a tactical trading model in the sense that it does not provide advice on how (and when) to trade in the market (where prices change minute by minute) to move from $[X_i]$ to $[x_i]$. However, some trading factors can be taken into account in our formulation. For example $F_i(\zeta, \Theta, t)$ can be a non-linear function designed to represent the fact that it may be more difficult to trade larger volumes than smaller volumes (owing to price changes if large buy/sell orders are placed, lack of buyers/sellers) depending upon the market for stock i. Also absolute limits on the number of units of a particular stock i that can be traded can be imposed by setting ε_i and δ_i appropriately.

(e) Nowhere in our formulation have we assumed that we have any knowledge as to how the index is constructed. In practice however we typically have such knowledge. Suppose, for example, that we know that the index value is an arithmetically weighted average, i.e. $I_t \propto \sum_{i\in\Lambda} m_i V_{it}$ where Λ is the set of index stocks and m_i is a weighting factor for stock $i \in \Lambda$. Then it would seem appropriate to choose the stocks $\{1, 2, \ldots, N\}$ we consider in our formulation such that $\Lambda \subseteq \{1, 2, \ldots, N\}$, i.e. so that the stocks we consider for inclusion in the tracking portfolio include all of the index stocks. In such cases if there are no transaction costs, $K = N$ and $\varepsilon_i = 0$, $\delta_i = 1\ \forall i$ then zero tracking error (and zero excess return) can be achieved using $x_i \propto m_i\ \forall i$ (which is in fact the basis of full replication).

(f) Any assets which give return values over time can be included in our formulation. For example ''stocks'' could be associated with the returns available from a variety of interest-bearing instruments (e.g. bonds), enabling us to investigate investing a proportion of the total fund in such assets. Also in tracking European market indices for instance (given increased economic integration within Europe) it may well be advantageous to include (say) French and Italian stocks in tracking a Germany market index. Our formulation allows such possibilities to be explored.

(g) Our formulation includes the case where we have no existing tracking portfolio (i.e. $X_i = 0\ \forall i$) and are seeking to construct a tracking portfolio from scratch using cash C_{cash}. A problem of this kind is sometimes referred to as a *creation* problem. The formulation we have presented above is sometimes referred to as a *revision*, or *rebalancing*, problem—seeking, as we do, to revise the existing tracking portfolio.

(h) Our formulation can deal with the situation which arises when the underlying composition of the index is revised. For example assume that stock values for the ''new'' stocks involved in the change are available (or can be reliably estimated). Then simply recompute the index value I_t $(t = 0, \ldots, T)$ assuming index recomposition had been carried out historically. Our formulation then enables one to determine a new tracking portfolio $[x_i]$, having regard to the transaction costs involved in changing from the current tracking portfolio $[X_i]$, that would have best tracked this assumed index. Such a tracking portfolio would seem appropriate for tracking the revised index into the future.

We would also comment here that, in recent years, index composition changes have been increasing, both in frequency of occurrence and in the number of stocks moving into and out of an index when its composition is changed. Such changes reflect the increased volatility of world markets. The heuristic presented in this paper, by enabling a tracking portfolio to be constructed out of stocks less likely to be moved out of an index (e.g. the top 250 stocks in the S&P 500 for example) means that an index tracker can achieve insulation (to a certain extent) from such changes, albeit at the cost of increased tracking error. Our heuristic enables such portfolios to be constructed and the cost of this insulation explicitly quantified.

4. A heuristic for index tracking

In this section we first briefly introduce population heuristics and then go on to outline our population heuristic for index tracking.

4.1. Population heuristics

Evolutionary algorithms are algorithms that simulate evolution in solving a specific problem. A population heuristic (henceforth abbreviated to PH) is an evolutionary algorithm that takes an initial population of individuals and applies genetic operators such as reproduction and mutation. In optimisation terms, each individual in the population is encoded into a string or *chromosome* which represents a possible *solution* to a given problem. The fitness of an individual is evaluated with respect to a given objective function. Highly fit individuals or *solutions* are given opportunities to reproduce by exchanging pieces of their genetic information, in a *crossover* procedure, with other highly fit individuals. This produces new "offspring" solutions (i.e. *children*), who share some characteristics taken from both parents. Mutation is often applied after crossover by altering some genes in the strings. The offspring can either replace the whole population (*generational* approach) or replace less fit individuals (*steady-state* approach). This evaluation–selection–reproduction cycle is repeated until a satisfactory solution is found.

The basic steps of a simple PH are shown below:

Generate an initial population
Evaluate fitness of individuals in the population
repeat
- Select individuals from the population to be parents
- Recombine (mate) parents to produce children
- Mutate the children
- Evaluate fitness of the children
- Replace some or all of the population by the children

until
- You decide to stop whereupon report the best solution encountered

More information relating to evolutionary heuristics can be found in [5,9,28,39,42,43] and information as to standard software packages for such heuristics can be found at http://evonet.dcs.napier.ac.uk/evoweb/resources/software/bbplatform.html.

4.2. Population heuristic for index tracking

Our PH for the index tracking problem is mostly easily presented if we transform our original formulation (minimise (13) subject to (1)–(8), (10) and (11)) of the problem appropriately, and this is done below. A number of elements in our PH are standard so that we, for reasons of space, have not further described them here. Readers who have not encountered PH's before are referred to the tutorial paper [9]. The key elements in our PH for index tracking are the solution representation that we adopt and the treatment of infeasibility via unfitness. We deal with each of these below in turn.

4.2.1. Transformation

Introduce variables $y_i \geqslant 0 \quad (i = 0, 1, \ldots, N)$ where:

y_0 represents the fraction of C associated with transaction cost, so that $y_0 = C_{\text{trans}}/C$

y_i $(i = 1, \ldots, N)$ represents the fraction of C associated with the total value of stock i in the new tracking portfolio at time T, so that $y_i = V_{iT}x_i/C$

Substituting into our original formulation (minimise (13) subject to (1)–(8), (10) and (11)) so as to eliminate $[x_i]$, $[r_t]$, C_{trans}, E and r^* we have that an equivalent formulation of the index tracking problem is

$$\text{minimise}\,\lambda\left[\left[\sum_{t\in S}\Delta_t\left|\log_e\left[\left[\sum_{i=1}^{N}V_{it}y_i/V_{iT}\right]\Big/\left[\sum_{i=1}^{N}V_{it-1}y_i/V_{iT}\right]\right]-R_t\right|^{\alpha}\right]^{(1/\alpha)}\right]\Big/T$$
$$-(1-\lambda)\sum_{t=1}^{T}\left[\log_e\left[\left[\sum_{i=1}^{N}V_{it}y_i/V_{iT}\right]\Big/\left[\sum_{i=1}^{N}V_{it-1}y_i/V_{iT}\right]\right]-R_t\right]\Big/T, \tag{14}$$

subject to:

$$\sum_{i=1}^{N} z_i = K, \tag{15}$$

$$\varepsilon_i z_i \leqslant y_i \leqslant \delta_i z_i, \quad i = 1, \ldots, N, \tag{16}$$

$$y_0 = \sum_{i=1}^{N} F_i(X_i, Cy_i/V_{iT}, T)/C, \tag{17}$$

$$\sum_{i=0}^{N} y_i = 1, \tag{18}$$

$$0 \leqslant y_0 \leqslant \gamma, \tag{19}$$

$$z_i \in [0, 1], \quad i = 1, \ldots, N. \tag{20}$$

Eq. (14) is derived from Eqs. (13), (11), (10) and (2); (16) from (4); (17) from (1); (18) from (6); (19) from (5).

4.2.2. Representation

In our PH the representation of a solution has two distinct parts, a set Q of K distinct stocks which are in the tracking portfolio (i.e. $z_i = 1\ \forall i \in Q$, so that Eqs. (15) and (20) are satisfied) and $K+1$ real numbers s_0 and s_i $i \in Q$ (where $0 \leqslant s_i \leqslant 1\ \forall i$). From Eq. (16) we know that the fraction $\sum_{j\in Q} \varepsilon_j$ of the total tracking portfolio (which must sum to one, Eq. (18)) is already accounted for and so we interpret s_i as relating to the share of the *free* tracking portfolio proportion $(1 - \sum_{j\in Q} \varepsilon_j)$ associated with stock $i \in Q$. We interpret s_0 as relating to the share of the free tracking portfolio proportion associated with transaction cost. These $[s_i]$ variables are translated into $[y_i]$ variables via the procedure given below:

(a) set $\varepsilon_0 = 0$, $\delta_0 = \gamma$, $Q = Q \cup [0]$, $R = \emptyset$

(b) set $y_i = \varepsilon_i + s_i\left[1 - \sum_{j\in Q} \varepsilon_j\right] \Big/ \left[\sum_{j\in Q} s_j\right]\ \forall i \in Q$

(c) while $\exists i \in Q - R$ with $y_i > \delta_i$ do:

  (1) $\forall i \in Q - R$ if $y_i > \delta_i$ then $R = R \cup [i]$

  (2) $y_i = \varepsilon_i + s_i\left[1 - \left(\sum_{j\in Q-R} \varepsilon_j + \sum_{j\in R} \delta_j\right)\right] \Big/ \left[\sum_{j\in Q-R} s_j\right]$ $\forall i \in Q - R$

  (3) $y_i = \delta_i\ \forall i \in R$

This procedure enables us to produce a set of $[y_i]$ that are guaranteed to satisfy Eqs. (16), (18) and (19).

4.2.3. Unfitness

Whilst the procedure presented above enables us to guarantee that Eqs. (16), (18) and (19) are satisfied it is not true that Eq. (17) will necessarily be satisfied after applying that procedure (indeed in general it will not). Hence a convenient measure of solution infeasibility (unfitness) is given by

$$\left| y_0 - \sum_{i=1}^{N} F_i(X_i, Cy_i/V_{iT}, T)/C \right|. \tag{21}$$

Any solution which has unfitness zero will be feasible (i.e. will satisfy Eqs. (15)–(20)). Hence associated with every solution (individual) in our PH are two numbers, fitness and unfitness. Here fitness is (as is usual) the objective function (Eq. (14)) value and unfitness, the measure of solution infeasibility, is as given by Eq. (21).

Consider now placing a newly generated child in the population (steady-state approach). Fig. 1 shows the situation conceptually. In that figure we have plotted a newly generated child (which will have fitness and unfitness values). However, members of the population will also have fitness and unfitness values and these too are shown in Fig. 1. We need to choose one member of the population to kill (replace by the child).

Drawing, as in Fig. 1, horizontal and vertical lines through the child will divide the population into four groups (labelled G1 to G4 in Fig. 1). Recall now that we wish to obtain solutions that have unfitness zero (i.e. are feasible) and which minimise fitness, colloquially we wish to obtain solutions that are as low down the y-axis in Fig. 1 as possible. Hence:

(a) Any members of the population that lie in group G1 are worse than the child with respect to both measures (fitness and unfitness) and so choosing to kill a member from this group would be ideal.

(b) It may be however that there are no members of the population in G1, where should we look next? Plainly we would look in group G4 last, since members of this group are better than the child with respect to both measures. Whilst there are differences relating to whether we look in G2 or G3 first (e.g. see [16]) here we will look first in G2, then if that is empty in G3, and finally if that is empty in G4.

If we do have to look in G4 (because G1, G2 and G3 are empty) we will be replacing a member of the population by a child that is worse with respect to both measures. Our reasoning here is that unless this is done the PH may fail to make progress.

Once the group in which to look has been decided as above a logical approach is to kill the member of that group with the worst unfitness (i.e. most infeasible), ties broken by worst fitness.

Our unfitness measure (Eq. (21)) essentially captures the difference (discrepancy) between the (fractional) transaction cost ($y_0 = C_{\text{trans}}/C$) given directly by the representation and the (fractional) transaction cost given indirectly (in terms of the change in stockholding, $\sum_{i=1}^{N} F_i(X_i, Cy_i/V_{iT}, T)/C$).

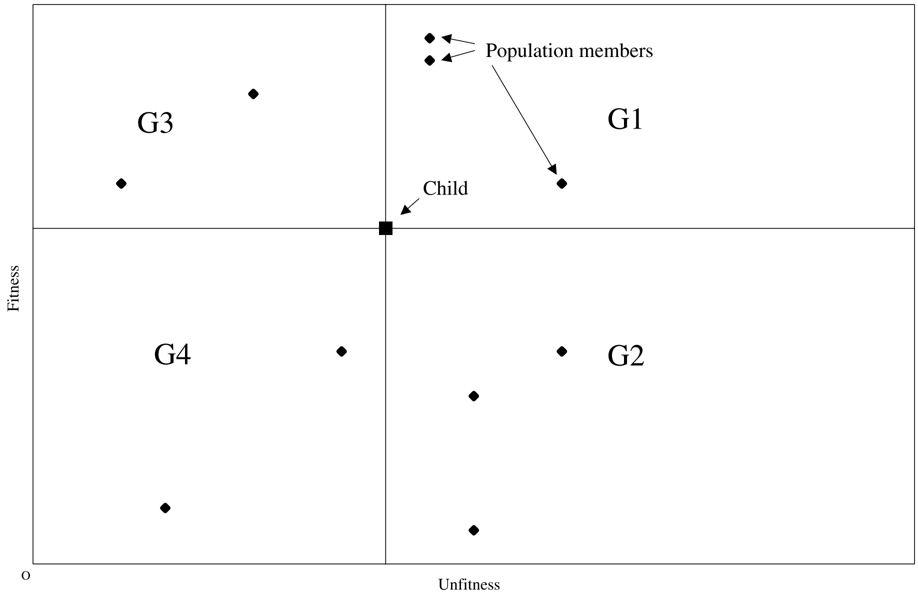

Fig. 1. Population replacement.

One point to note here is that, in terms of the underlying practical problem being solved, very small unfitness values, whilst mathematically infeasible, are to all intents and purposes feasible from a practical viewpoint. Hence, in the computational results reported later, we regard a solution as feasible if its unfitness is $\leqslant 0.01\gamma$. For example if $C = 10^6$ ($ say) and $\gamma = 0.0025$ (i.e. a transaction cost limit of 0.25%) then this corresponds to a transaction cost discrepancy of $25 in a portfolio worth one million dollars when we are willing to spend up to $2500 in transaction costs.

4.2.4. Complete PH

We used a population size of 100 with the initial population being randomly generated (except for one member of the initial population which corresponded to the current tracking portfolio $[X_i]$). Parents were chosen by binary tournament selection which works by forming two pools of individuals, each consisting of two individuals drawn from the population randomly. The individuals with the best fitness, one taken from each of the two tournament pools, are chosen to be parents. We do not use unfitness in parent selection as our previous experience with PH's (see http://mscmga.ms.ic.ac.uk/jeb/jeb.html) has indicated that this is not worthwhile provided population replacement is carried out as indicated above.

Children in our PH are generated by uniform crossover in which two parents have a single child. If a stock i is present in both parents it is present in the child (with an associated value s_i randomly chosen from one or other parent). If a stock i is present in just one parent it has probability 0.5 of being present in the child. Children are also mutated, changing by $\pm 0.5\%$ the value associated with a randomly selected stock.

If a child, after crossover and mutation, did not contain exactly K stocks then a heuristic procedure was used to add or remove stocks until the child contained exactly K stocks. Evaluating the fitness and unfitness values for the child, and selecting a member of the population to kill (replace by the child), were carried out as described above.

5. Reduction tests

There are a number of tests that we can carry out to reduce the size of the search space and, hopefully, enable the PH to be more effective. In this section we outline those tests that we found computationally effective.

Consider the current portfolio $[X_i]$. Given the transaction cost constraint (transaction cost must be $\leqslant \gamma C$, Eq. (5)) it may be that we can identify certain stocks that cannot enter/leave the tracking portfolio and/or can update the proportion limits ε_i and δ_i on the proportion of the tracking portfolio associated with stock i. Without significant loss of generality assume that the transaction cost function $F_i(\zeta, \Theta, t)$, the transaction cost ($\geqslant 0$) incurred for stock i in moving from holding ζ units of the stock to holding Θ units of the stock at time t, is a non-decreasing function about ζ, i.e. that $F_i(\zeta, \Theta, t) \geqslant F_i(\zeta, \phi, t)$ if $\Theta \geqslant \phi \geqslant \zeta$; and $F_i(\zeta, \Theta, t) \geqslant F_i(\zeta, \phi, t)$ if $\Theta \leqslant \phi \leqslant \zeta$. These conditions simply say that buying/selling more stock cannot lead to a decrease in transaction cost. Let:

N_{out} be the stocks not in the current tracking portfolio $[X_i]$, i.e. $N_{\text{out}} = [i | X_i = 0 \; i = 1, \ldots, N]$

N_{in} be the stocks in the current tracking portfolio $[X_i]$, i.e. $N_{\text{in}} = [i | X_i > 0 \; i = 1, \ldots, N]$

W be the stocks that must appear (at some stockholding) in the new tracking portfolio $[x_i]$, where initially $W = \emptyset$

Then our reduction tests (for the case when $[X_i]$ contains exactly K stocks) are as follows:

(a) For $i \in N_{\text{out}}$ if:

$$F_i(0, \varepsilon_i C/V_{iT}, T) + \min[F_j(X_j, 0, T) | j \in N_{\text{in}} - W] > \gamma C \tag{22}$$

then we can delete stock i from the problem since it can never appear in the tracking portfolio. In Eq. (22) the first term represents the cost of bringing stock i into the tracking portfolio at its minimum stockholding whilst the second term represents the minimum cost of removing some stock j ($j \in N_{\text{in}} - W$) from the tracking portfolio (to maintain exactly K stocks in the tracking portfolio).

There is one technical issue here, namely that strictly the left-hand side of Eq. (22) does not completely capture the transaction cost incurred if we were to introduce stock i (since we might need to adjust the proportions associated with the stocks in the tracking portfolio to ensure feasibility). However the assumptions we have made about the nature of the transaction cost function ensure that the left-hand side of Eq. (22) is a lower bound on the cost that would be incurred in such a situation and so the reduction is valid. Similar considerations apply below.

(b) For $i \in N_{\text{in}} - W$ if:

$$F_i(X_i, 0, T) + \min[F_j(0, \varepsilon_j C/V_{jT}, T) | j \in N_{\text{out}}] > \gamma C \tag{23}$$

then stock i must always be in the tracking portfolio (at some stockholding) and so we can set $W = W \cup [i]$. In Eq. (23) the first term represents the cost of removing stock i from the tracking portfolio whilst the second term represents the minimum cost of adding some stock j ($j \in N_{\text{out}}$) to the tracking portfolio.

(c) For $i \in N_{\text{out}}$ find the minimum value of $\tau \geqslant \varepsilon_i$ such that

$$F_i(0, \tau C/V_{iT}, T) + \min[F_j(X_j, 0, T) | j \in N_{\text{in}} - W] > \gamma C \tag{24}$$

and set $\delta_i = \min[\tau, \delta_i]$. In Eq. (24) the first term represents the cost of bringing stock i into the tracking portfolio at a proportion level τ whilst the second term represents the minimum cost of removing some stock j ($j \in N_{\text{in}} - W$) from the tracking portfolio. Given the nature of our assumptions about the transaction cost function the problem of determining τ can, in the worst case, be done using a simple line search procedure.

Irrespective of whether $[X_i]$ contains exactly K stocks or not we also have:

(a) For $i \in W$ find the maximum value of $\tau \leqslant V_{iT} X_i / C$ such that

$$F_i(X_i, \tau C/V_{iT}, T) > \gamma C \tag{25}$$

and set $\varepsilon_i = \max[\tau, \varepsilon_i]$.

(b) For $i \in N_{\text{in}}$ find the minimum value of $\tau \geqslant V_{iT} X_i / C$ such that

$$F_i(X_i, \tau C/V_{iT}, T) > \gamma C \tag{26}$$

and set $\delta_i = \min[\tau, \delta_i]$.

In Eqs. (25) and (26) the transaction cost term is associated with reducing/increasing the current holding of stock i. These reduction tests can be applied repeatedly until no further reduction can be achieved. Modifying our heuristic to deal with stocks deleted from the problem, as well as stocks $i \in W$ which must always be present in the tracking portfolio, is straightforward.

6. Computational results

In this section we present computational results for our heuristic for index tracking.

6.1. Test data sets and computational considerations

To test our heuristic we constructed five test data sets by considering the stocks involved in five different capital market indices drawn from around the world. Specifically we considered the Hang Seng (Hong Kong), DAX 100 (Germany), FTSE 100 (UK), S&P 100 (USA) and Nikkei 225 (Japan). We used DATASTREAM to obtain weekly price data from March 1992 to September 1997 for the stocks in these indices. Stocks with missing values were dropped. We had 291 values for each stock from which to calculate (weekly) returns and the size of our five test problems ranged from $N = 31$ (Hang Seng) to $N = 225$ (Nikkei).

All of the test problems solved in this paper, but with the identity of each stock disguised, are publicly available from OR-Library [7,8]. To obtain them, email the message *indtrackinfo* to o.rlibrary@ic.ac.uk or see http://mscmga.ms.ic.ac.uk/jeb/orlib/indtrackinfo.html. This means that future workers will be able to assess the performance of their algorithms for index tracking against our heuristic using *exactly* the same data.

All of the computational results presented below are for our heuristic as coded in FORTRAN and run on a Silicon Graphics O2 workstation (R10000 chip, 225 MHz, 128MB main memory). Unless otherwise stated we:

(a) used an initial tracking portfolio composed of the first K stocks in equal proportions, i.e. $X_i = (C/K)/V_{i0} \quad i = 1, \ldots, K; \quad X_i = 0 \quad \forall i > K$ with $C = 10^6$, $C_{\text{cash}} = 0$ and $K = 10$, i.e. a tracking portfolio containing ten stocks
(b) 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
(c) used $F_i(\zeta, \Theta, t) = 0.01|\zeta - \Theta|V_{it}$, i.e. transaction cost was one percent of the value of the stocks bought/sold
(d) used $\alpha = 2$ and $S = [1, 2, \ldots, T]$ (see Eq. (10)), so tracking error effectively corresponds to root mean squared error
(e) used $\lambda = 1$ (see Eq. (13)) so we were minimising tracking error
(f) stopped our heuristic after $1500N$ children had been generated
(g) replicated (repeated) each heuristic run five times with the same set of parameters and took the best solution found.

6.2. Artificial index

One of the difficulties in evaluating the quality of our heuristic is that, except for problems of trivial size, we cannot solve the problem optimally. Hence we have no benchmark optimal solutions against which to compare our heuristic solutions. In order to overcome this and provide data on which we can test the quality of our heuristic we artificially adjusted our problem instances by setting $I_t = \sum_{i=N-K+1}^{N} V_{it}$, so that the index value is composed of the last K stocks (all with an equal weighting); and solved the problem in the situation corresponding to zero transaction cost (i.e. $F_i(\zeta, \Theta, t) = 0 \ \forall i, \zeta, \Theta, t$) with $\varepsilon_i = 0$, $\delta_i = 1 \ \forall i$ and an arbitrary value of γ. In such cases we automatically know that it is possible to track the index with a tracking error of zero. The tracking error returned by our heuristic therefore will provide some insight into the quality of our heuristic. Note here that we have defined the index value, and the initial tracking portfolio $[X_i]$, such that they have no stocks in common.

Table 1 shows the results obtained (with $T = 290$). In that table we show, for each of our five data sets the tracking error, both in absolute (Eq. (10)) terms and in percentage terms 100 (tracking error)/(standard deviation of

Table 1
Artificial index results

Index	Number of stocks (N)	Tracking error		Time (minute)
		Value	Percentage	
Hang Seng	31	2.560×10^{-8}	0.000	1.7
DAX	85	6.405×10^{-5}	0.258	5.0
FTSE	89	1.911×10^{-4}	1.041	5.3
S&P	98	1.528×10^{-4}	0.847	5.8
Nikkei	225	2.294×10^{-4}	0.762	15.7
Average				6.7

$\{R_t | t = 0, \ldots, 290\}$); and the total time taken (in minutes). It is clear from Table 1 that our heuristic produces good quality results, with the average tracking error being just over one-half of one percent of the standard deviation in index return.

6.3. In-sample versus out-of-sample tracking errors

Recall from our discussion previously that we are choosing a tracking portfolio that performs well in-sample, when in reality we would really like a tracking portfolio that performs well out-of-sample. In order to investigate this issue we took the time period $[0, 145]$ for each of our five data sets and, for a range of values of γ, used our heuristic to decide a tracking portfolio. We then measured the tracking error associated with this tracking portfolio out-of-sample (in $[145, 290]$). Table 2 shows the results obtained. In that table we show (for each of our five data sets and each of the values of γ considered) the tracking errors achieved in-sample and out-of-sample, expressed both in absolute terms and as percentages (cf. Table 1), and the total time taken (in minutes).

In the absence of a known optimal solution and the absence of previous heuristic solutions in the literature, the performance of our heuristic cannot be measured in absolute terms at this time (although we would note here that since we have made our data sets publicly available future workers will be able to compare their results against ours). Rather the performance of our heuristic must be judged in terms of its consistency with reasonable expectations. These expectations are that:

(a) *Increasing the transaction cost limit γ reduces tracking error*. Logically as our constraint relating to transaction cost (Eq. (5)) is an inequality if we were to solve the index tracking problem optimally we would find that as γ increases tracking error decreases (or remains constant). From Table 2 it can be seen that, in all cases, the in-sample tracking error decreases monotonically as γ increases. The out-of-sample tracking error also decreases as γ increases (with three exceptions for the S&P and the Nikkei out of the 20 increases in γ shown in Table 2). Hence our heuristic results do conform to this expectation.
(b) *Reducing in-sample tracking error reduces out-of-sample tracking error*. Utilising Table 2 the Pearson product moment correlation coefficient between the in-sample and out-of-sample percentage tracking errors shown over all data sets is $+0.72$. This is highly significant statistically (probability of 0.00006 at a two-sided level) and indicative of the fact that as in-sample tracking error reduces so too does out-of-sample tracking error.

Table 3 shows, for each value of γ, the geometric mean in-sample and out-of-sample errors (when expressed as proportions and averaged across the five data sets). It can be seen from Table 3 that both in-sample and out-of-sample percentage error decrease monotonically as the transaction cost limit γ increases. This performance accords with both of our expectations above.

In a further attempt to gain insight into the performance of our heuristic we have also used our heuristic to calculate the "best" in-sample tracking error over the period $[145, 290]$ given our initial tracking portfolio. This provides a further measure by which we can judge the out-of-sample tracking

Table 2
In-sample versus out-of-sample tracking errors

Index	Number of stocks (N)	Transaction cost limit γ	Tracking error				Ratio of out-of-sample tracking error to best tracking error	Time (minute)
			In-sample		Out-of-sample			
			Value	Percentage	Value	Percentage		
Hang Seng	31	0	1.028×10^{-3}	3.104	1.267×10^{-3}	3.827	1.00	4.1
		0.0025	8.728×10^{-4}	2.636	1.066×10^{-3}	3.221	1.15	
		0.0050	6.836×10^{-4}	2.065	8.714×10^{-4}	2.632	1.23	
		0.0075	5.245×10^{-4}	1.584	6.407×10^{-4}	1.935	1.35	
		0.0100	4.591×10^{-4}	1.387	5.705×10^{-4}	1.723	1.76	
DAX	85	0	1.173×10^{-3}	5.777	2.049×10^{-3}	10.091	1.00	11.8
		0.0025	9.575×10^{-4}	4.716	1.776×10^{-3}	8.750	1.14	
		0.0050	7.668×10^{-4}	3.777	1.613×10^{-3}	7.945	1.27	
		0.0075	5.937×10^{-4}	2.925	1.394×10^{-3}	6.867	1.49	
		0.0100	4.514×10^{-4}	2.223	1.146×10^{-3}	5.643	1.47	
FTSE	89	0	1.021×10^{-3}	5.890	9.584×10^{-4}	5.527	1.00	12.8
		0.0025	8.379×10^{-4}	4.832	9.223×10^{-4}	5.319	1.16	
		0.0050	6.605×10^{-4}	3.809	8.153×10^{-4}	4.702	1.26	
		0.0075	5.795×10^{-4}	3.342	7.705×10^{-4}	4.443	1.42	
		0.0100	5.186×10^{-4}	2.990	6.464×10^{-4}	3.728	1.37	
S&P	98	0	1.038×10^{-3}	6.843	1.032×10^{-3}	6.805	1.00	14.0
		0.0025	8.110×10^{-4}	5.345	7.047×10^{-4}	4.645	1.00	
		0.0050	6.388×10^{-4}	4.210	6.452×10^{-4}	4.252	1.13	
		0.0075	5.050×10^{-4}	3.328	6.653×10^{-4}	4.385	1.31	
		0.0100	4.821×10^{-4}	3.177	6.451×10^{-4}	4.252	1.37	
Nikkei	225	0	7.803×10^{-4}	2.746	8.213×10^{-4}	2.890	1.00	33.1
		0.0025	6.011×10^{-4}	2.115	9.949×10^{-4}	3.501	1.45	
		0.0050	5.530×10^{-4}	1.946	7.049×10^{-4}	2.480	1.20	
		0.0075	5.034×10^{-4}	1.771	8.427×10^{-4}	2.965	1.58	
		0.0100	4.587×10^{-4}	1.614	6.537×10^{-4}	2.300	1.36	
Average								15.2

Table 3
Variation in average error with transaction cost limit

Transaction cost limit γ	Average percentage tracking error	
	In-sample	Out-of-sample
0	4.86	5.80
0.0025	3.92	5.07
0.0050	3.16	4.38
0.0075	2.59	4.11
0.0100	2.28	3.52

errors shown in Table 2 since it indicates how much better we could have done had we known in advance what the out-of-sample data was (i.e. if we had possessed a perfect prediction of the future). Computationally this was done by fixing the initial tracking portfolio, setting $I_{145-t} = I_{145+t}$, $t = 1, \ldots, 145$ and $V_{i,145-t} = V_{i,145+t}$ $\forall i$, $t = 1, \ldots, 145$ and applying our heuristic but for 50, instead of 5, replications. The ratio of our out-of-sample tracking error to this ‘‘best’’ tracking error is shown in Table 2. Excluding $\gamma = 0$ our out-of-sample tracking error is on average approximately one-third greater than the ‘‘best’’ tracking error. Note that the computation times given in Table 2 exclude the times taken to calculate this ‘‘best’’ tracking error.

6.4. Tradeoff between tracking error and excess return

Fig. 2 shows, for the Hang Seng data set, tradeoff curves relating excess return to tracking error. These curves were plotted under the as-

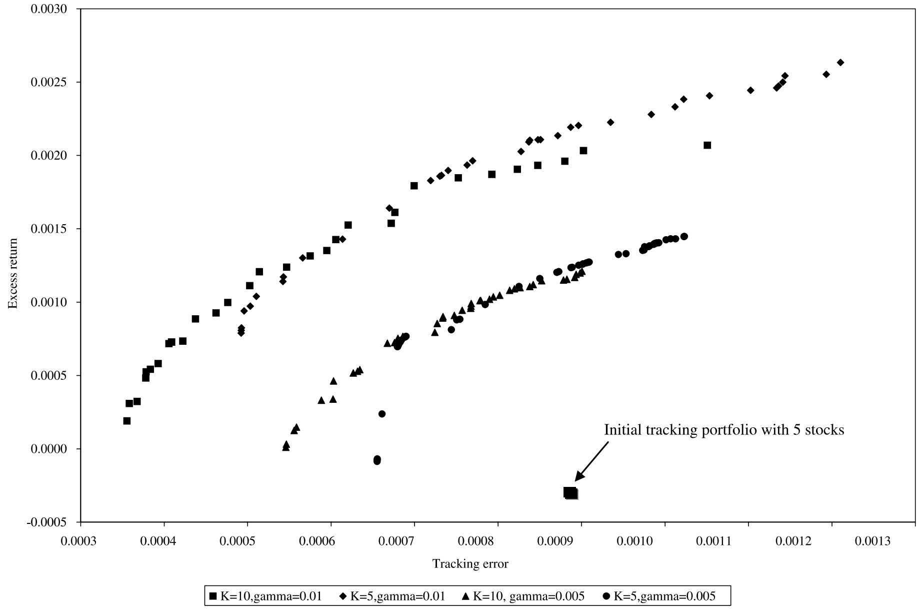

Fig. 2. Tradeoff curves—Hang Seng.

sumption that we were changing from an initial tracking portfolio containing five stocks to new tracking portfolios with $K = 5$, 10 and $\gamma = 0.005$, 0.01 (arbitrarily chosen). We used values of λ (Eq. (13)) equal to $0, 0.01, 0.02, \ldots, 1$ with $T = 290$ and these curves show the in-sample [0, 290] tracking error and excess return associated with the tracking portfolios found by our heuristic. Producing this figure required 10.9 hours on our Silicon Graphics workstation.

From Fig. 2 it can be seen that, for $\gamma = 0.005$, the $K = 10$ and $K = 5$ curves are close together at high excess returns, i.e. there are only small differences in tracking errors between tracking portfolios containing 10 stocks and tracking portfolios containing 5 stocks. By contrast, for $\gamma = 0.01$, the $K = 10$ and $K = 5$ curves are not close together at high excess returns. The tradeoff curves shown in Fig. 2, composed of tracking portfolios in which transaction cost is explicitly constrained, provide fund managers with information that, prior to the work presented in this paper, was simply not available in a systematic fashion.

In order to provide some insight into the composition of the tracking portfolios shown in Fig. 2, Fig. 3 illustrates how these portfolios change along the tradeoff curves. Specifically Fig. 3 plots the stocks involved in every tracking portfolio shown in Fig. 2. Each small square in Fig. 3 represents a stock (where stocks that never appear in any tradeoff curve tracking portfolio for any K, γ pair are not represented). The same pattern appearing in successive columns in Fig. 3 means that the tracking portfolios had the same composition in terms of the underlying stocks (although the amount of each stock held may have been different). In Fig. 3 the average difference (in terms of stocks held) between successive tracking portfolios as we move up the tradeoff curves shown in Fig. 2 is 0.28 for $K = 5$, $\gamma = 0.005$; 0.44 for $K = 5, \gamma = 0.01$; 1.44 for $K = 10$, $\gamma = 0.005$ and 1.73 for $K = 10$, $\gamma = 0.01$. This implies that for $K = 10$, $\gamma = 0.01$ (for example) on average approximately two of the ten stocks in the tracking portfolios are different between successive points on the tradeoff curve.

6.5. Tracking errors associated with systematic revision

In order to investigate the performance of our heuristic over time (as we continually and systematically revise our tracking portfolio) we took the

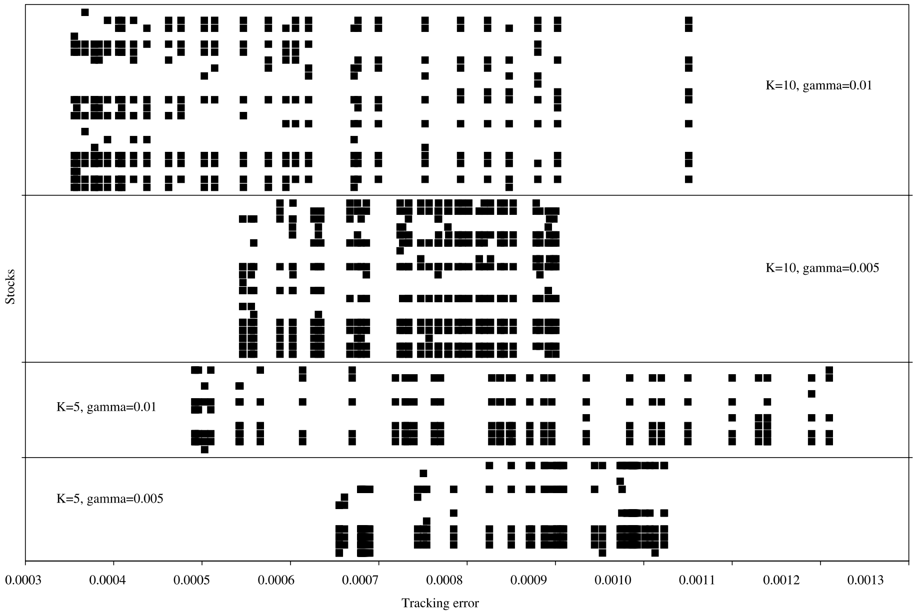

Fig. 3. Hang Seng tracking portfolios structure.

time period [0, 290] and chose within it, seven decision points ($T = 150, 170, 190, 210, 230, 250, 270$). Essentially these decision points correspond to revising our tracking portfolio 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 (using our heuristic) a decision as to an appropriate new tracking portfolio to have, using all the information contained in $[0, T]$. Formally:

(a) set $T = 150$
(b) use the heuristic to decide the tracking portfolio $[x_i]$ that would have been best over $[0, T]$ within the constraint relating to the transaction cost limit γ
(c) set $[X_i] = [x_i]$ (replace the current tracking portfolio by the new tracking portfolio)
(d) set $T = T + 20$ and if $T \leqslant 270$ go to (b)
(e) calculate the total tracking error associated with the tracking portfolios held over [0, 290] (as we know, for each time period, the tracking portfolio that applies in that time period this is easily done).

Table 4 shows the results obtained. In that table we show (for each of our five data sets and each of the values of γ considered) the tracking errors achieved, expressed both in absolute terms and as percentages (cf. Table 1) and the total time taken (in minutes). It is clear from Table 4 that, in nearly all cases, as the transaction cost limit γ increases, the tracking error decreases.

6.6. Other test data sets

We have also tested our heuristic using a number of other data sets drawn from the S&P 500 index (which contains 500 stocks), the UK FTSE All Share index (approximately 800 stocks), the Russell 2000 index (2000 stocks), the Russell 3000 index (3000 stocks), and the Wilshire 5000 index (approximately 7000 stocks). Unfortunately commercial restrictions preclude us from making this data publicly available so we have decided not to include detailed results for that work within this paper. Results for all of these data sets can be seen at http://mscmga.ms.ic.ac.uk/jeb/track.html.

As an indication of our results however Fig. 4 shows graphically the performance of our heuristic for the Wilshire 5000 index with $K = 250$. Since we lacked price data for all stocks in the Wilshire 5000 we limited ourselves to choosing 250 stocks from

Table 4
Tracking errors associated with systematic revision

Index	Number of stocks (N)	Transaction cost limit γ	Tracking error		Time (minute)
			Value	Percentage	
Hang Seng	31	0	8.157×10^{-4}	2.464	39.0
		0.0025	6.481×10^{-4}	1.958	
		0.0050	5.820×10^{-4}	1.758	
		0.0075	5.717×10^{-4}	1.727	
		0.0100	5.625×10^{-4}	1.699	
DAX	85	0	1.180×10^{-3}	5.814	98.0
		0.0025	8.832×10^{-4}	4.350	
		0.0050	7.656×10^{-4}	3.771	
		0.0075	7.284×10^{-4}	3.588	
		0.0100	7.187×10^{-4}	3.540	
FTSE	89	0	7.003×10^{-4}	4.039	111.5
		0.0025	6.214×10^{-4}	3.583	
		0.0050	6.004×10^{-4}	3.462	
		0.0075	5.900×10^{-4}	3.403	
		0.0100	5.750×10^{-4}	3.316	
S&P	98	0	7.321×10^{-4}	4.825	112.3
		0.0025	6.028×10^{-4}	3.973	
		0.0050	5.939×10^{-4}	3.914	
		0.0075	5.903×10^{-4}	3.891	
		0.0100	5.806×10^{-4}	3.826	
Nikkei	225	0	5.664×10^{-4}	1.993	285.4
		0.0025	5.374×10^{-4}	1.891	
		0.0050	5.660×10^{-4}	1.992	
		0.0075	5.362×10^{-4}	1.887	
		0.0100	5.183×10^{-4}	1.824	
Average					129.2

the Russell 3000 in order to track the Wilshire 5000 (note here that the Wilshire 5000 does include all stocks in the Russell 3000). Fig. 4 has been produced by taking weekly price data from January 1997 to September 1999 and assuming that in September 1998 we had applied our heuristic and bought (at zero transaction cost) the tracking portfolio containing 250 stocks it recommended where the desired objective in-sample was to minimise tracking error. Fig. 4 shows the performance of our tracking portfolio, and of the index, over the one year out-of-sample period from September 1998 onward.

To decide the tracking portfolio shown in Fig. 4 using our heuristic required 12.6 hours on our Silicon Graphics workstation. On a practical note it is clear that tracking the Wilshire 5000 index by holding just 250 stocks is a much more desirable proposition than tracking that index using full replication, which would require holding approximately 7000 stocks.

7. Conclusions

In this paper we considered the index tracking problem and presented a population heuristic for its solution. Computational results were presented for our heuristic for five data sets drawn from major world markets. These data sets have been made publicly available for use by other workers.

Finally we would comment that index tracking is an important problem that, in our view, has received insufficient attention in the literature. We

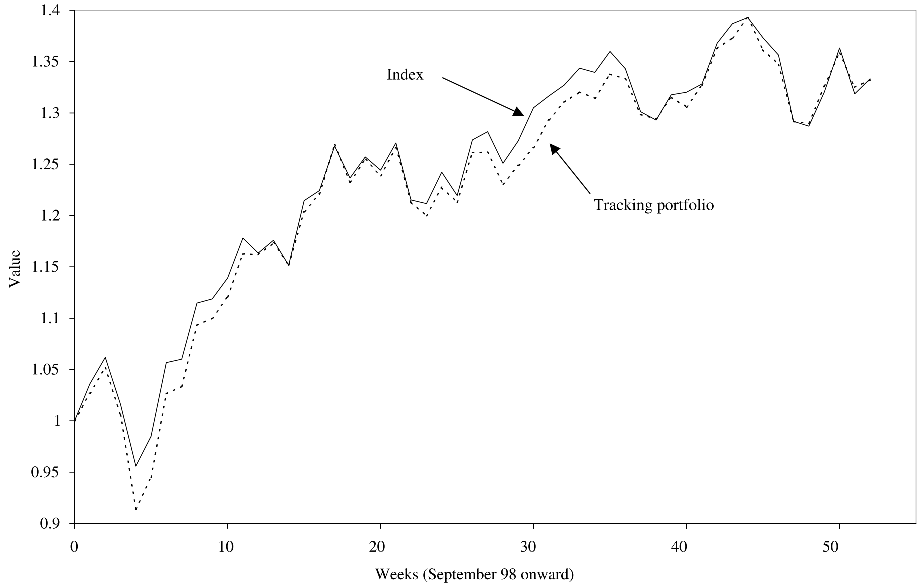

Fig. 4. Out-of-sample tracking performance for the Wilshire 5000 index with $K = 250$.

hope that this paper will help to highlight the problem and encourage others to work on it.

References

[1] C.J. Adcock, N. Meade, A simple algorithm to incorporate transaction costs in quadratic optimisation, European Journal of Operational Research 79 (1994) 85–94.

[2] 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.

[3] M. Ammann, H. Zimmermann, Tracking error and tactical asset allocation, Financial Analysts Journal 57 (2) (2001) 32–43.

[4] C. Andrews, D. Ford, K. Mallinson, The design of index funds and alternative methods of replication, The Investment Analyst 82 (1986) 16–23.

[5] T. Bäck, D.B. Fogel, Z. Michalewicz (Eds.), Handbook of Evolutionary Computation, Oxford University Press, Oxford, 1997.

[6] G. Bamberg, N. Wagner, Equity index replication with standard and robust regression estimators, OR Spektrum 22 (2000) 525–543.

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

[8] J.E. Beasley, Obtaining test problems via Internet, Journal of Global Optimization 8 (1996) 429–433.

[9] J.E. Beasley, Population heuristics, in: P.M. Pardalos, M.G.C. Resende (Eds.), Handbook of Applied Optimization, Oxford University Press, Oxford, 2002, pp. 138–157.

[10] J.L. Blue, G.T. Candela, P.J. Grother, R. Chellappa, C.L. Wilson, Evaluation of pattern classifiers for fingerprint and OCR applications, Pattern Recognition 27 (4) (1994) 485–501.

[11] J.C. Bogle, Selecting equity mutual funds, The Journal of Portfolio Management 18 (2) (1992) 94–100.

[12] S. Browne, Beating a moving target: Optimal portfolio strategies for outperforming a stochastic benchmark, Finance and Stochastics 3 (1999) 275–294.

[13] I.R.C. Buckley, R. Korn, Optimal index tracking under transaction costs and impulse control, International Journal of Theoretical and Applied Finance 1 (3) (1998) 315–330.

[14] CAPS Investment Information Services, 1999. Available from <http://www.caps-iis.com/site/pressdetails.asp?press_id = 38>.

[15] A. Chan, C.R. Chen, How well do asset allocation mutual fund managers allocate assets, The Journal of Portfolio Management 18 (3) (1992) 81–91.

[16] P.C. Chu, J.E. Beasley, Constraint handling in genetic algorithms: The set partitioning problem, Journal of Heuristics 4 (1998) 323–357.

[17] G. Connor, H. Leland, Cash management for index tracking, Financial Analysts Journal 51 (6) (1995) 75–80.

[18] A. Consiglio, S.A. Zenios, Integrated simulation and optimization models for tracking international fixed income indices, Mathematical Programming 89 (2001) 311–339.

[19] H. Dahl, A. Meeraus, S.A. Zenios, Some financial optimization models. I. Risk management, in: S.A. Zenios (Ed.), Financial Optimization, Cambridge University Press, Cambridge, 1993, pp. 3–36.
[20] G. Dorfleitner, A note on the exact replication of a stock index with a multiplier rounding method, OR Spektrum 21 (1999) 493–502.
[21] E.J. Elton, M.J. Gruber, Modern Portfolio Theory and Investment Analysis, Wiley, New York, 1995.
[22] E.J. Elton, M.J. Gruber, C.R. Blake, The persistence of risk-adjusted mutual fund performance, Journal of Business 69 (2) (1996) 133–157.
[23] E.C. Franks, Targeting excess-of-benchmark returns, The Journal of Portfolio Management 18 (4) (1992) 6–12.
[24] M. Gilli, E. Këllezi, Threshold accepting for index tracking, Working paper available from the first author at Department of Econometrics, University of Geneva, 1211 Geneva 4, Switzerland, 2001.
[25] M.J. Gruber, Another puzzle: The growth in actively managed mutual funds, Journal of Finance 51 (3) (1996) 783–810.
[26] R.A. Haugen, N.L. Baker, Dedicated stock portfolios, The Journal of Portfolio Management 16 (4) (1990) 17–22.
[27] S.D. Hodges, Problems in the application of portfolio selection models, Omega 4 (6) (1976) 699–709.
[28] J.H. Holland, Adaptation in Natural and Artificial Systems: An Introductory Analysis With Applications to Biology, Control, and Artificial Intelligence, University of Michigan Press, Ann Arbor, MI, 1975.
[29] Investment Week News, 1998. Available from <http://www.invweek.co.uk/invweek/news/98dec/7insid4.htm>.
[30] D.B. Keim, An analysis of mutual fund design: The case of investing in small-cap stocks, Journal of Financial Economics 51 (1999) 173–194.
[31] G.A. Larsen Jr., B.G. Resnick, Empirical insights on indexing, The Journal of Portfolio Management 25 (1) (1998) 51–60.
[32] C. Liu, H. Wechsler, Evolutionary pursuit and its application to face recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (6) (2000) 570–582.
[33] B.G. Malkiel, A. Radisich, The growth of index funds and the pricing of equity securities, The Journal of Portfolio Management 27 (2) (2001) 9–21.
[34] H. Markowitz, Portfolio selection, Journal of Finance 7 (1952) 77–91.
[35] H.M. Markowitz, Portfolio Selection: Efficient Diversification of Investments, Wiley, New York, 1959.
[36] S.J. Masters, The problem with emerging markets indexes, The Journal of Portfolio Management 24 (2) (1998) 93–100.
[37] N. Meade, G.R. Salkin, Index funds—construction and performance measurement, Journal of the Operational Research Society 40 (1989) 871–879.
[38] N. Meade, G.R. Salkin, Developing and maintaining an equity index fund, Journal of the Operational Research Society 41 (1990) 599–607.
[39] M. Mitchell, An Introduction to Genetic Algorithms, MIT Press, Cambridge, MA, 1996.
[40] P.F. Pope, P.K. Yadav, Discovering errors in tracking error, The Journal of Portfolio Management 20 (2) (1994) 27–32.
[41] S. Reed, On the trail of the tracker funds, What Investment (May 1999) 8–12.
[42] C.R. Reeves (Ed.), Modern Heuristic Techniques for Combinatorial Problems, Blackwell Scientific Publications, Oxford, 1993.
[43] C.R. Reeves, Genetic algorithms for the operations researcher, INFORMS Journal on Computing 9 (1997) 231–250.
[44] H.C. Rohweder, Implementing stock selection ideas: Does tracking error optimization do any good? The Journal of Portfolio Management 24 (3) (1998) 49–59.
[45] R. Roll, A mean/variance analysis of tracking error, The Journal of Portfolio Management 18 (4) (1992) 13–22.
[46] A. Rudd, Optimal selection of passive portfolios, Financial Management (Spring 1980) 57–66.
[47] A. Rudd, B. Rosenberg, Realistic portfolio optimization, in: E.J. Elton, M.J. Gruber (Eds.), Portfolio Theory, 25 Years After, TIMS Studies in the Management Sciences 11, North-Holland, 1979, pp. 21–46.
[48] M. Rudolf, H.-J. Wolter, H. Zimmermann, A linear model for tracking error minimization, Journal of Banking & Finance 23 (1999) 85–103.
[49] E.H. Sorenson, K.L. Miller, V. Samak, Allocating between active and passive management, Financial Analysts Journal 54 (5) (1998) 18–31.
[50] Standard & Poor, 2001. Available from <http://www.spglobal.com/indexmain500_changes00.html>.
[51] C.M.S. Sutcliffe, Stock Index Futures: Theories and International Evidence, second ed., International Thompson Business Press, 1997.
[52] Y. Tabata, E. Takeda, Bicriteria optimization problem of designing an index fund, Journal of the Operational Research Society 46 (1995) 1023–1032.
[53] W.M. Toy, M.A. Zurack, Tracking the Euro-Pac index, The Journal of Portfolio Management 15 (2) (1989) 55–58.
[54] M.Y. Wang, Multiple-benchmark and multiple-portfolio optimization, Financial Analysts Journal 55 (1) (1999) 63–72.
[55] P. Wilmott, Derivatives: The Theory and Practice of Financial Engineering, Wiley, New York, 1998, pp. 540–542.
[56] K.J. Worzel, C. Vassiadou-Zeniou, S.A. Zenios, Integrated simulation and optimization models for tracking indices of fixed-income securities, Operations Research 42 (1994) 223–233.
[57] S.A. Zenios, M.R. Holmer, R. McKendall, C. Vassiadou-Zeniou, Dynamic models for fixed-income portfolio management under uncertainty, Journal of Economic Dynamics and Control 22 (1998) 1517–1541.