COMPARING METHODOLOGIES FOR INDEX TRACKING

ACOSTA-GONZÁLEZ, EDUARDO(*)
e-mail: eacosta@dmc.ulpgc.es

ARMAS-HERRERA REINALDO(*)
e-mail:reinaldo.armas102@alu.ulpgc.es

FERNÁNDEZ-RODRÍGUEZ, FERNANDO(*)
e-mail: ffernandez@dmc.ulpgc.es

(*) MÉTODOS CUANTITATIVOS EN ECONOMÍA Y GESTIÓN
UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA

Abstract

In a seminal paper Alexander and Dimitriu (2005) proposed a new methology for Index Tracking by means of cointegration. As the quality of the tracking portfolio depends highly on the stock selection procedure, we propose picking the stocks using a model selection technique based on optimizing the cointegration level of the tracking portfolio and the benchmark index instead of selecting, as in previous papers the assets by ad hoc decisions. To illustrate an empirical application of these techniques we use daily closing prices in the Dow Jones Industrial Average index from 01-01-1990 to 31-12-2003 previously used by other authors. We have compared our results with a more traditional procedure based on correlation and again our results reveal superiority. The empirical illustration not only has been focused on the tracking portfolio itself, but has also been extended to tracking the index with an added profitability of 5%, 10%, 15% or 20%. In both frameworks, our methodology improves the other methodologies in terms of TE and mean annual return of tracking portfolio.

1. Introduction

The basic strategies adopted by fund managers can broadly be classified as active and passive. Active strategies rely on the belief that skilful investors can outperform the market practicing activities such as picking winners, or choosing adequate market timing for their buy/sell decisions. In the case of passive management, there is less flexibility and the typical strategy attempts to reproduce, as closely as possible, the performance of a theoretical index representing the market, I_t, where $t = 1, 2, \ldots, T$, being T, the sample size. The process replicating an index is called Index Tracking (IT) or index replication and the portfolio, which follows the index, is called the tracking portfolio (TP). Passive strategies rely on the belief that the market is efficient and it is therefore impossible to consistently beat the aggregate market return. Passive investment strategies have become very popular because they cost less than active strategies. In this sense, Eton, Gruber and Blake (1996) have demonstrated through historical empirical analysis, over the long term, the majority of actively managed funds do not outperform the market.

There are two interconnected problems associated with IT: Asset selection and asset allocation. Asset selection involves selecting a subset of assets out of all the available assets, whereas asset allocation involves investing a proportion of the total available capital in each one of the assets selected with the aim of getting the performance of the index. In order to address these problems, two ways are normally employed. On one hand, full replication. This consists of purchasing all of the stocks that make up the index proportional to their share in it. Thus, both asset selection and asset allocation are resolved. Full replication has a number of disadvantages such as certain stocks in the index may be held in very small quantities, high transaction costs and difficulties in rebalancing the portfolio when the weights in the tracked index change. On the other hand, in order to avoid these disadvantages, or at least mitigate their effects, partial replication can also be done selecting a smaller subset of assets. In this case, some optimal procedures should be carried out to deal with asset selection and asset allocation.

Besides, a difference, which appears with the full replication, partial replication, a result of its own idiosyncrasies, produces tracking errors. A much used variable for capturing these errors is the denominated Tracking Error (TE), which is defined in the following way [Roll (1992)]:

$$TE_t = R_{TP,t} - R_{I,t} \tag{1}$$

where $R_{TP,t}$ and $R_{I,t}$ represent the TP return and the index return in the time t, respectively. A good measure that permits us to gauge the goodness of fit is the mean square error of the TE, which is usually called the *Goodness of Tracking* (GT)

$$GT = \frac{\sum_{i=1}^{T} TE_t^2}{T} \tag{2}$$

It is possible to split this measure into two parts. If to the second term we add and subtract the square of the TE, we arrive at

$$GT = variance\left(TE_t\right) + \left(\bar{R}_{TP} - \bar{R}_I\right)^2 \tag{3}$$

Where $\bar{R}_{TP}$ and $\bar{R}_I$ represent the mean return of the TP and the index I_t, respectively. The first element of the GT, the variance of the TE, gauges the lack of synchronization of the TP with the index I_t. This is a measure of the lack of capacity of TP to adapt itself, temporarily, to the behavior pattern followed by I_t. The lower the variance of the TE, the higher the synchronization between the TP and the index I_t. In the extreme case of variance equals to zero, the synchronization would be perfect, and the difference each time between the I_t and TP returns would generate a constant value.

Nevertheless, this constant must not be zero, so that even existing synchronization, the difference of profitability between the I_t and TP may not coincide, and so, the tracking could be called into question. The measure of this tracking error following I_t is obtained in the second term added in GT, which captures the square of the skew of profits. The perfect pursuit is only possible in the case where as well as the variance of the TE as the difference of profitability between I_t and TP are equal to zero.

Different approaches to the partial replication problem have been proposed in literature such as Toy and Zurack (1989), Franks (1992), Roll (1992), Clarke, Krase and Statman (1994), Connor and Leland (1995), Larsen and Renick (1998) and Rohweder (1998), among others. More recently, Dose and Cincotti (2005) employed the clustering methodology developed by Beasley et al (2003) in order to group the assets using a measure of distance between them that is based on their historical prices. This grouping procedure is used to decide the assets in the TP. Konno and Hatagi (2005) minimize the absolute difference between the index and the TP including transaction costs and solving an optimization problem with linear restrictions. Wu, Chou, Yang, and Ong (2007) face the problem from the point of view of goal programming where the goals are the rate of returns and the TE. Using a series of the Taiwan Stock Exchange, they conclude that the desired return depends on the index return. Canakgoz and Bearsley (2008) employ mixed integer lineal programming in order to get an excess return with respect to the index in a TP problem. Using data from eight different markets they select an asset portfolio, limiting the number of assets to buy and the number of transactions to carry out. Finally, Li, Sun, and Bao (2011) employ a multi-objective optimization where the goals are maximizing the index outperformance and minimizing the TE. Using data from five different markets they apply Enhanced Index Tracking, a strategy that will be described in this section, which we be carried out using an artificial immune algorithm. After all these contributions, it seems to us that two of the most important improvements to the IT problem which have recently been proposed are the use of the evolutionary heuristic algorithms and the use of cointegration techniques. Let us consider each one separately.

(1) During the last decade, evolutionary heuristics for selecting the assets and the introduction of realistic restrictions have frequently been employed. The application of heuristics in the context of portfolio optimization was probably suggested for the first time in the 1992 paper by Dueck and Winker. Evolutionary heuristics are employed to resolve the non-linear mixed-integer program associated with picking the assets in order to replicate the index. This is the case with Beasley, Meade, and Chang (2003) who formulate the IT problem including a limit on the total transaction costs and a constraint limiting the number of stocks in the portfolio. Oh, Kim, and Min (2005) and Oh, Kim, and Lee (2006) propose a portfolio algorithm that engages beta in its

construction process and employs a Genetic Algorithm (GA) for optimizing the individual weights of the portfolio's assets. Heuristic algorithms were also employed by Gilli and Kellezi (2002), Derigs and Nickel (2003), Okay and Akman (2003), Maringer (2008) and Zhang and Maringer (2009).

(2) Cointegration techniques for finding long-term relationships between the subset of stocks constituting the portfolio and the index has also been developed. Since the seminal work of Markowitz (1952), the traditional starting point for portfolio construction is the correlation analysis of returns. Nevertheless, recent research on stock market linkages has found common stochastic trends for a group of stock market prices through testing for cointegrating relationships. Cointegration is a powerful technique for modelling long-run dynamics which has also been employed in building IT as in Alexander and Dimitriu (2005). This approach has also been followed by Dunis and Ho (2005). Since the seminal work of Engle and Granger (1987), cointegration has become an essential tool in time series analysis, displacing the acceptance of correlation. Several stock prices are cointegrated if there is one stationary linear combination of their prices (see Alexander, 2001, as general reference in stock market cointegration). Following Alexander and Dimitriu (2005), the advantage of using cointegration for carrying out IT is double. On one hand, the price difference between the benchmark and the portfolio is, by construction, stationary; so, the TP will be tied to the benchmark index in the long run. On the other hand stock weights, based on cointegration and a long history of prices, will have more stability producing the benefits of less frequent portfolio rebalancing with respect to using correlation for carrying out IT. These benefits rest on making full use of the information in stock prices before their detrending, which permits a long-run relationship between equity prices and the market index.

In this paper we present some improvements to deal with the asset selection and the asset allocation problems with respect to previous literature when the partial replication is chosen. These improvements are based on both cointegration techniques and heuristic algorithms. In previous papers, the stock selection is the result of technical analysis, stock picking skills of a portfolio manager or ad hoc decisions. As the quality of the benchmark tracking depends highly on the stock selection procedure, we propose a model selection technique based on optimizing the cointegration level of the TP and the index. However, in this context, as the number of portfolio candidates to track the index is extremely high, a simple mechanical stock selection procedure can be very time consuming. To deal with this problem we propose a heuristic optimization method based on genetic algorithms. So, our proposal combines the cointegration principles for describing the common stochastic trends between assets with new computational techniques of evolutionary heuristics devoted to finding the assets which currently constitute the TP.

Our proposal also presents several advantages when it is applied to other related strategies with index replications. This is specifically, the Enhanced Index Tracking strategy. In short an Enhanced Index Tracking strategy attempts to over-perform the index that is tracked by a given annual percentage return (say, plus 20%). This new index is called Enhanced Index (EI_t) and can be obtained by the recursive expression

$$EI_t = EI_{t-1}\left[1+\left(R_{I,t} + PlusReturn\right)\right] \qquad t = 2,3,\ldots,T$$

where $EI_1 = I_1\left(1+PlusReturn\right)$ and $PlusReturn$ represents the plus of profitability that the TP must obtain over the original index. In this strategy the Tracking Error would be defined as

$$TE_t = R_{TP,t} - R_{EI,t}$$

Where $R_{EI,t}$ represents the EI_t profitability. In the same way, the GT would be defined by the following expression

$$GT = variance\left(TE_t\right) + \left(\overline{R}_{TP} - \overline{R}_{EI}\right)^2$$

There are two alternatives for getting these results; the first one is using derivatives, futures or options (See Li, Sun and Bao, 2011), and the second one is only using stocks. This last alternative is the one we use in this paper.

The rest of the paper is organized as follows. In Section 2 we formulate and resolve the optimisation problem in order to select the tracking portfolio by means of cointegration and heuristic algorithms. In Section 3 some empirical applications are shown. For carrying out these applications we have used the same database used by Alexander and

Dimitriu in 2005, which contains the daily data of the Dow Jones Industrial Average (DJIA) and its stocks. This permits us to compare our results with theirs. Also, we compare our methodology with a more traditional one based on correlation. Specifically, we have carried out the following compared tracking portfolio strategies: (1) Tracking the DJIA and (2) Tracking the DJIA Enhanced. Finally, Section 4 is devoted to the concluding remarks.

2. Asset selection and asset allocation in the tracking portfolio by means of cointegration and heuristic algorithms

2.1 The set-up of the problem

Let us consider an index vector Γ composed of K stocks in such a way that:

$$\Gamma = PW \qquad (1)$$

Where:

$$\Gamma = \begin{bmatrix} I_1 \\ \vdots \\ I_t \\ \vdots \\ I_T \end{bmatrix}, \quad P = \begin{bmatrix} p_{1,1} & \cdots & p_{K,1} \\ \vdots & \cdots & \vdots \\ p_{1,t} & \cdots & p_{K,t} \\ \vdots & \cdots & \vdots \\ p_{1,T} & \cdots & p_{K,T} \end{bmatrix}, \quad W = \begin{bmatrix} w_1 \\ \vdots \\ w_K \end{bmatrix}$$

And:

- I_t: Represents the index value at time t.
- p_{jt}: Represents the price of stock j at time t.
- w_j: Represents the weight of stock j in the index. For a particular TP, w_j represents the number of units of stock j that are selected to be held in it.

In a full replication the vector W is predetermined by the weight that the managers of the Index use to construct it. However, when a full replication is not carried out, each element of W must be estimated subject to the restrictions set, which means that some of the elements of W remain at zero (we denote q the number of elements that remain at zero). When this happens, to hold the equation (1) it is necessary to add an error tracking term $E = [e_1,\ldots e_T]'$ as is shown in the equation (2):

$$\Gamma = PW + \mathrm{E} \qquad (2)$$

Besides, a general linear framework of restrictions is imposed in the weights with the form:

$$RW = r\,, \qquad (3)$$

Where R is a $q \times K$ matrix of known constants $\left(R = \left[r_{ij}\right]; i = 1,\ldots q;\ j = 1,\ldots K\right)$, with $q < K$, and r is a q vector of known constants. For our particular case the following restrictions must be held:

$$\sum_{j=1}^{K} r_{ij} = 1 \qquad i = 1,\ldots,q\,, \qquad (4)$$

$$\sum_{j=1}^{q} r_{ij} \leq 1 \qquad j = 1,\ldots,K \qquad (5)$$

$$r = \vec{0} \qquad (6)$$

For example, if stocks 2 and 4 are rejected as part of the portfolio, then $r_{1,2} = 1$ and $r_{2,4} = 1$, and the other r_{ij} are equal to zero. In general, once the restrictions have been established a restricted least squares estimator can be used to calculate W through the well-known expression.

$$W = W^{LS} + \left(P'P\right)^{-1} R'\left[R\left(P'P\right)^{-1} R'\right]^{-1}\left(r - RW^{LS}\right) \qquad (7)$$

Where W^{LS} is the unrestricted least square estimator of equation (2), that is $W^{LS} = \left(P'P\right)^{-1} P'\Gamma$ (see (Johnston and Dinardo, 1997)).

2.2 Optimization of the cointegration relationship

Cointegration is a statistical relationship where two time series that are both integrated of same order d can be linearly combined to produce a single time series which is integrated of order $d - b$, where $b > 0$. In its application to IT, we refer to the case where the portfolio prices and the index to be replicated are both integrated of order 1 and are combined in such a way that they produce a stationary TE or more specifically an integrated TE of order 0. In this sense, our goal is to reach a matrix R that optimizes the cointegration relationship in equation (2) between the index and the selected portfolio once the number of stocks (n) (it is easy to see that $n = K - q$) has been chosen. Following Alexander and Dimitriu we use the Augmented Dickey-Fuller (1979) test statistic (ADF) as a measure of the cointegration relationship. This test is calculated on the residuals contained in vector E [equation (2)]. That is:

$$\Delta e_t = \gamma e_{t-1} + \sum_{i=1}^{p} \alpha_i \Delta e_{t-i} + u_t \qquad (8)$$

Where Δ is the difference operator so that $\Delta e_t = e_t - e_{t-1}$. When $\gamma = 0$, it means that a cointegration relationship doesn't exist. However, when $\gamma < 0$ it does exist. On this consideration, we minimize the t-ratio based on the estimation of this coefficient, turning the cointegration relationship optimization problem into the following minimization problem:

$$\underset{R}{Min} \; f = \frac{|A|}{\sqrt{|C||B|}} \frac{1}{\hat{\sigma}_u} \qquad (9)$$

Subject to:

$$E = \Gamma - PW \quad \text{where } W \text{ is obtained through equation (7)} \qquad (10)$$

$$n = cte \qquad (11)$$

$$q = K - p \qquad (12)$$

$$\sum_{j=1}^{K} r_{ij} = 1 \qquad i = 1, \ldots, q \qquad (13)$$

$$\sum_{j=1}^{q} r_{ij} \leq 1 \qquad j = 1, \ldots, K \qquad (14)$$

Where $\hat{\sigma}_u^2$ is the variance estimation of u_t in equation (13), A is a $(p+1)\times(p+1)$ dimension matrix and B and C are two symmetric matrices of dimension $(p+1)\times(p+1)$ and $p \times p$, respectively (Appendix 1 shows the general form of matrixes A, B and C). As can be seen, the t-ratio function f depends directly on the vector E which in turn depends on matrix R .

Once the number n of assets in the replicating portfolio is decided, the optimization problem is a very complex task. For example, in order to replicate an index with K stocks using, at the most, $n \leq K$ stocks, the total amount of possible portfolio candidates is $K!/(K-n)!n!$. It becomes prohibitive for index where the number of stocks is high like the S&P 500 or the Russell 6000. So, in most methodologies for IT, the selection of the stocks in the portfolio is not generally objective and can be the result of a proprietary selection model, technical analysis or the stock-picking skills of a portfolio manager. A simple rule of thumb frequently employed is selecting the n stocks with the highest capitalization.

In order to resolve this intractable problem of stock-picking, rather than search through all possible portfolios, a heuristic optimization procedure called Genetic Algorithm (GA, from now on) is employed. A GA is a class of optimization technique, based on principles of natural evolution developed by Holland (1975) which tries to overcome problems of traditional optimization algorithms, such as an absence of continuity or differentiability in the loss function. A GA starts with a population of randomly generated solution candidates, which apply the principle of fitness to produce better approximations to optimal solution. Promising solutions, as represented by relatively better performing solutions, are selected and then bred together through a process of binary recombination. This process is called crossover and it's inspired by Mendel's natural genetics. The objective of this process is to generate successive populations solutions that are better fitted to the optimization problem than the solutions from which they were created. It is very common to introduce random mutations to avoid local optima.

3. Empirical results

In this section we illustrate the performance of our methodology on the database used by Alexander and Dimitriu (2002). The data set consists of daily closing prices over the period that goes from 01-01-1990 to 31-12-2003 for the 30 stocks in the Dow Jones Industrial Average index (DJIA) (for further information look at Alexander and Dimitriu's work).

We compare our results for TP with those of Alexander and Dimitriu which have proposed an Index Tracking methodology based on cointegration where the tracking portfolio is formed by the assets with maximum capitalization. In this case our assessment is based on comparing our results with those that Alexander and Dimitriu would have obtained if they had selected the same number of stocks we did. Also, we compare our results with those obtained when instead of using prices, we had used returns directly and, consequently, the criterion for measuring the relationship between returns is the correlation instead of cointegration.

The heuristic algorithm used has been coded and solved in Matlab 7.8. It is an adapted version of the Genetic Algorithm employed by Acosta-Gonzalez and Fernandez-Rodriguez (2007), specially designed for IT and based on the methodology given in that paper. For running this algorithm it is necessary to set three parameters. The first two parameters are the number of chromosomes (2000) and the stopping criteria (20), which is the number of repeated solutions to stop the algorithm. For a better understanding of the role of these parameters see the former reference. The third parameter is the number of stocks in the portfolio. In this sense, Alexander and Dimitriu (2005) point out that fewer than 20 stocks proved not to be sufficiently cointegrated with the market index. However, using our methodology with just five stocks we are able to get cointegration relationships as shown below in this section.

3.1 A comparative analysis of selection methods for TP

We compare our methodology of minimizing the ADF statistic by a GA (ADF-GA, henceforth) with both the naive approach of selecting a subset of assets based on price ranking used by Alexander and Dimitriu (2005) (maximum Capitalization, henceforth) and by the classical procedure of maximizing the correlation coefficient between the tracking portfolio and the Index (maximum Correlation, henceforth) also using a GA for selecting the stocks. Table 1 and Table 2 (a,b,c) show the results of this study. Specifically, Table 1 shows the results when the task is tracking the DJIA; while Table 2 (a,b,c) shows the results when the task is tracking the enhanced DJIA. A result immediately highlights, the ADF-GA is the only method able to achieve the goal of return over all the period, as can be seen in the row “Excess return of TP over DJIA” in Table 1 and Table 2(a,b,c). The results are consistently good, especially when these are compared with those obtained by the other methods. Table 1 shows that only choosing by maximum capitalization gets closer to ADF-GA´s results when 10 or 20 assets are included in the portfolio. However, when five assets are included in the portfolio, choosing by maximum capitalization is unable to reach its goal with a single deviation of -4.13% against ADF-GA´s results which reach a single deviation of -0.05%. Table 2(a,b,c) shows the same results: A very high performance of ADF-GA and a very poor one of choosing by maximum capitalization or correlation. Let see some examples in Table 2(a): when the goal is to get a Plus Return of 20% over DJIA return and five assets are included in the portfolio, ADF-GA reaches an excess return over DJIA of 20.61%, very close to the goal. However, choosing by maximum capitalization or correlation reaches an excess return over DJIA of -7.09% and -2.82%, respectively, far away from the target. More or less this is what happen in the rest of the results of Table 2 (a,b,c), which reflected clearly the superiority of ADF-GA in terms of returns achieved.

However, when we focus on the “Goodness of Tracking”, the results are not as good for the ADF-GA procedure which obtains the worse result. As we saw in Section 1 the goodness of tracking is composed of two parts: The lack of synchronization and the returns bias squared. Table 1 and Table 2(a,b,c) show that the goodness of tracking for ADF-GA procedure is concentrated on the lack of synchronization, attaining in all cases yields at a rate of 100%. However, the returns bias squared remains at zero which is in consonance with the results shown in the above paragraph. Also, ADF-GA produces a substantial reduction of the variable “mean annual TE”, which is a key variable for IT. This reduction is not achieved when choosing by maximum capitalization or correlation. It means that ADF-GA is able to compensate positive and negative deviations in the long-run relationships between DJIA return and the tracking portfolio return up to getting an almost zero mean. The results of the rows “Rate of success” in Table 1 and Table 2(a,b,c) reinforce this conclusion.

Another interesting aspect that shows Tables 1 and 2 (a, b, c) is that ADF-GA produces portfolios with higher risk, as it is shown in the row “TP Volatility” of these tables. However, taking into account that the returns obtained by ADF-GA are far larger, this risk is properly compensated. This can be seen in the row “Information ratio”. In all cases this ratio plays in favour of ADF-GA, in the sense that for each unit of risk ADF-GA gets greater return from either of the other two methods.

Finally, observing the critical values of Engle Granger test for cointegration [MacKinnon's (1991)] and comparing it with the in sample ADF statistic we can conclude that both ways of choosing by maximum capitalization or correlation select portfolios which do reject the alternative hypothesis of cointegration at the 5% of significance level. This means that having several highly correlated assets is not trivial in getting a portfolio properly cointegrated with the index. In contrast, the portfolios that we have got by ADF-GA are fully cointegrated with the index.

[INSERT TABLE 1 HERE]

[INSERT TABLE 2a HERE]

[INSERT TABLE 2b HERE]

[INSERT TABLE 2c HERE]

4. Conclusions

The process replicating an index is called Index Tracking or index replication and the portfolio which follows the index is called the tracking portfolio. It is passive strategies which rely on the belief that the market is efficient and it is therefore impossible to consistently beat the aggregate market return. Index Tracking investment strategies have become very popular because they cost less than active strategies. In this work we provide some significant improvements for index tracking by means of cointegration.

Cointegration is a powerful technique for modelling long-run dynamics which has also been employed in building index tracking as in Alexander and Dimitriu (2005). This approach has also been followed by Dunis and Ho (2005). Several stock prices are cointegrated if there is one stationary linear combination of their prices. The advantage of using cointegration for carrying out index tracking is double. On one hand, the price difference between the benchmark and the portfolio is, by construction, stationary; so, the tracking portfolio will be tied to the benchmark index in the long run. On the other hand stock weights, based on cointegration and a long history of prices, will have more stability producing the benefits of less frequent portfolio rebalancing with respect to using correlation for implementing Index Tracking. These benefits rest on making full use of the information on stock prices before their detrending, which permits a long-run relationship between equity prices and the market index.

In this paper we present some improvements to deal with the asset selection and the asset allocation problems with respect to previous literature when the partial replication is chosen. These improvements are based on both cointegration techniques and heuristic algorithms. In previous papers, the stock selection is the result of technical analysis, stock picking skills of a portfolio manager, assets with more capitalization in the benchmark index, or ad hoc decisions. As the quality of the benchmark tracking depends highly on the stock selection procedure, we propose a model selection technique based on optimizing the cointegration level of the tracking portfolio and the index. However, in this context, as the number of portfolio candidates to track the index is extremely high, a simple mechanical stock selection procedure can be very time consuming. To deal with this problem we propose a heuristic optimization method based on a genetic algorithm. It permits us to obtain tracking portfolios with very few stocks (five in our empirical illustration) that are able to follow the benchmark index properly.

To illustrate an empirical application of these methodologies, and looking for comparison with previous works, we use daily closing prices over the period going from 01-01-1990 to 31-12-2003 for the 30 stocks, in the Dow Jones Industrial Average index. Our methodology is able to improve the results from the Alexander's methodology and the correlation ones in terms of TE and mean annual return of TP. The only disadvantage of our methodology is that the correlation between the index and the TP is lower. When we compare our methodology with the alternative ones in an Enhanced Index Tracking framework our results are very similar to the Index Tracking results.

Appendix 1

$$A=\begin{bmatrix}
\sum_{t=1}^{T-p} e_{t+(p-1)}\Delta e_{t+p} & \sum_{t=1}^{T-p} e_{t+(p-1)}\Delta e_{t+(p-1)} & \sum_{t=1}^{T-p} e_{t+(p-1)}\Delta e_{t+(p-2)} & \cdots & \sum_{t=1}^{T-p} e_{t+(p-1)}\Delta e_{t} \\
\sum_{t=1}^{T-p} \Delta e_{t+(p-1)}\Delta e_{t+p} & \sum_{t=1}^{T-p} \Delta e^2_{t+(p-1)} & \sum_{t=1}^{T-p} \Delta e_{t+(p-1)}\Delta e_{t+(p-2)} & \cdots & \sum_{t=1}^{T-p} \Delta e_{t+(p-1)}\Delta e_{t} \\
\sum_{t=1}^{T-p} \Delta e_{t+(p-2)}\Delta e_{t+p} & \sum_{t=1}^{T-p} \Delta e_{t+(p-1)}\Delta e_{t+(p-2)} & \sum_{t=1}^{T-p} \Delta e^2_{t+(p-2)} & \cdots & \sum_{t=1}^{T-p} \Delta e_{t+(p-2)}\Delta e_{t} \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
\sum_{t=1}^{T-p} \Delta e_{t}\Delta e_{t+p} & \sum_{t=1}^{T-p} \Delta e_{t+(p-1)}\Delta e_{t} & \sum_{t=1}^{T-p} \Delta e_{t+(p-2)}\Delta e_{t} & \cdots & \sum_{t=1}^{T-p} \Delta e^2_{t}
\end{bmatrix}$$

$$B=\begin{bmatrix}
\sum_{t=1}^{T-p} \Delta e^2_{t+(p-1)} & \sum_{t=1}^{T-p} \Delta e_{t+(p-1)}\Delta e_{t+(p-2)} & \cdots & \sum_{t=1}^{T-p} \Delta e_{t+(p-1)}\Delta e_{t} \\
 & \sum_{t=1}^{T-p} \Delta e^2_{t+(p-2)} & \cdots & \sum_{t=1}^{T-p} \Delta e_{t+(p-2)}\Delta e_{t} \\
 & & & \vdots \\
 & & & \sum_{t=1}^{T-p} \Delta e^2_{t}
\end{bmatrix}$$

$$C=\begin{bmatrix}
\sum_{t=1}^{T-p} e^2_{t+(p-1)} & \sum_{t=1}^{T-p} e_{t+(p-1)}\Delta e_{t+(p-1)} & \sum_{t=1}^{T-p} e_{t+(p-1)}\Delta e_{t+(p-2)} & \cdots & \sum_{t=1}^{T-p} e_{t+(p-1)}\Delta e_{t} \\
 & \sum_{t=1}^{T-p} \Delta e^2_{t+(p-1)} & \sum_{t=1}^{T-p} \Delta e_{t+(p-1)}\Delta e_{t+(p-2)} & \cdots & \sum_{t=1}^{T-p} \Delta e_{t+(p-1)}\Delta e_{t} \\
 & & \sum_{t=1}^{T-p} \Delta e^2_{t+(p-2)} & \cdots & \sum_{t=1}^{T-p} \Delta e_{t+(p-2)}\Delta e_{t} \\
 & & & & \vdots \\
 & & & & \sum_{t=1}^{T-p} \Delta e^2_{t}
\end{bmatrix}$$

References

Acosta-González, E., and Fernández-Rodríguez, F., (2007). Model selection via genetic algorithms illustrated with cross country growth data. *Emprical Economics,* (33), 313-337.

Alexander, C. (2001). *Market Models.* Wiley.

Alexander, C. (1999). Optimal hedging using cointegration. *Philosophical Transactions of the Royal Society Series A* (357), 2039-2058.

Alexander, C., and Dimitriu, A. (2002). The cointegration alpha: Enhanced Index Tracking and Long – Short Equity Market Neutral Strategies. ISMA Centre, University of Reading. Working Paper.

Alexander, C., and Dimitriu, A. (2005). Indexing and statistical arbitrage: Tracking error or cointegration? *Journal of Portfolio Management* (31), 50-63.

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

Canakgoz, N., and Bearsley, J. (2008). Mixed-integer programming approaches for index tracking and enhanced indexation. *European Journal of Operational Research* (196), 384–399.

Canakgoz, N., and Bearsley, J. (2008). Mixed-integer programming approaches for index tracking and enhanced indexation. *European Journal of Operational Research* (196), 384–399.

Clarke, R. C., Krase, S., and Statman, M. (1994). Tracking errors, regret, and tactical asset allocation. *The journal of Portfolio management* (20), 16-24.

Colemnan, T., Li, Y., and Henniger, J. (2006). Minimizing tracking error while restricting the number of assets. *Journal of Risk* (8), 33-56.

Colwell, D., El-Hassan, N., and Kwon, K. (2007). Hedging diffusion processes by local risk minimization with applications to index tracking. *Journal of Economic Dynamics and Control* (31), 2135-2151.

Connor, G., and Leland, H. (1995). Cash management for index tracking. *Financial Analysts Journal , 6* (51), 75-80.

Corielli, F., and Marcellino, M. (2006). Factor based index tracking. *Journal of Banking and Finance* (30), 2215–2233.

Derigs, U., and Nickel, N. (2003). Meta-heuristic based decision support for portfolio optimization with a case study on tracking error minimization in passive portfolio management. *OR Spectrum* (25), 345–378.

Dose, C., and Cincotti, S. (2005). Clustering of financial time series with application to indexand enhanced index tracking portfolio. *Physica A:Statistical Mechanics and its Applications , 1* (355), 154-151.

Dueck, G, and Winker, P. (1992). New concepts and algorithms for portfolio choice. *Applied Stochastic Models and Data Analysis,* 8, 195-178.

Dunis, C., and Ho, R. (2005). Cointegration portfolios of European equities for index tracking and market neutral strategies. *Journal of Asset Management , 6* (1), 33-52.

Dikey, D. and Fuller, W. (1979). Distribution of the Estimators for Autorregressive Time Series with a Unit Root. Journal of the American Statistical Association, 74, 427-431.

Engle, R., and Granger, C. (1987). Cointegration and error correction: Representation, estimation, and testing. *Econometrica* (55), 251-276.

Eton, E., Gruber, G., and Blake, C. (1996). Survivorship bias and mutual fund performance. *Review of Financial Studies* (9), 1097-1120.

Fang, Y., and Wang, S. (2005). A fuzzy index tracking portfolio selection model. *Lecture Notes in Computer Science 3516*, (págs. 554-561).

Focardi, S., and Fabozzi, F. (2004). A methodology for index tracking based on time series. *Quantitative Finance* (4), 417–425.

Franks, E. C. (1992). Targeting excess of benchmark returns. *The journal of Portfolio Management , 4* (18), 6-12.

Gaivoroki, A., Krylov, S., and Van der Wisjt, N. (2005). Optimal portfolio selection and dynamic benchmark tracking. *European Journal of Operational Research , 163* (1), 115-131.

Gilli, M., and Kellezi, E. (2002). The threshold accepting heuristic for index tracking. En V. T. P. Pardalos (Ed.), *Financial Engineering E-Commerce and Supply Chain* (págs. 1-18). Kluwer.

Holland, J. (1975). *Adaptation in Natural and Artificial Systems.* Ann Arbor: The University of Michigan Press.

Jansen, R., and Van Dijk, R. (2002). Optimal benchmark tracking with small portfolios. *Journal of Portfolio Management* (28), 33-39.

Johnston, J., and Dinardo, J. (1997). *Econometric methods* (Cuarta ed.). McGraw-Hill/Irwin.

Konno, H., and Hatagi, T. (2005). Index plus alpha tracking under concave transaction cost. *Journal of Industrial and Management Optimization* (1), 87-98.

Larsen, J., and Renick, B. (1998). Empirical insights on indexing. *The journal of Portfolio management , 1* (25), 51-60.

Li, J., Ma, Y., and Zeng, Y. (2005). Research on application of Stein rule estimation for index tracking problem. *Proceedings of the 2005 International Conference on Management Science and Engineering 12th (1–3)*, 434–439.

Li, Q., Sun, L., and Bao, L. (2011). Enhanced index tracking based on multi-objetive immune algorithm. *Expert Systems with Applications* (38), 6101-6106.

MacKinnon J. (1991). *Critical values for the cointegration tests*. In Engle RF, Granger CWJ (eds). Long-run Economic Relationships, pages 267–276. Oxford University Press: Oxford.

Maringer, D. 2008. Constrained Index Tracking under Loss Aversion Using Differential Evolution. In A. Brabazon, & M. Oneill (Eds.), Natural Computing in Computational Finance, Vol. 100: 7-24. New York: Springer.

Markowitz, H. (1952). *Portfolio Selection.* Journal of Finance 7(1), 77-91.

Oh, K. J., Kim, T. Y., and Lee, H. Y. (2006). Portfolio algorithm based on portfolio beta using genetic algorithm. *Expert Systems with Applications* (30), 527-534.

Oh, K. J., Kim, T. Y., and Min, S. (2005). Using genetic algorithm to support optimization for index fund management. *Expert Systems with Applications* (28), 371-379.

Okay, N., and Akman, U. (2003). Index tracking with constraint aggregation. *Applied Economics Letters* (10), 913-916.

Pearson, K. (1895), Royal Society Proceedings, 58, 241.

Rohweder, H. C. (1998). Implementing stock selection ideas: does tracking error optimization does any good? *The journal of Portfolio management , 3* (24), 49-59.

Roll, R. (1992). A mean/variance analysis of tracking error. *The Journal of Portfolio Management , 4* (18), 13-22.

Rudolf, M., Wolter, H., and Zimmermann, H. (1999). A linear model for tracking error minimization. *Journal of Banking and Finance* (23), 85-103.

Stock, J. H., and Watson, M. W. (1988). Testing for Common Trends. *Journal of the American Statistical Association* , 1097-1107.

Stoyan, A., and Kwon, B. (2007). A two-stage stochastic mixed-integer programming approach to the index tracking problem. *Working paper available from the first author at Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, Canada* .

Toy, W. M., and Zurack, M. A. (1989). Tracking the Euro-Pac index. *The Journal of Portfolio Management , 2* (15), 55-58.

Wu, L. C., Chou, S. C., Yang, C. C., and Ong, C. S. (2007). Enhanced index investing based on goal programming. *Journal of Portfolio Management , 3* (33), 49-56.

Yao, D., Zhang, S., and Zhou, X. (2006). Tracking a financial benchmark using a few assets. *Operations Research* (54), 232-246.

Yu, L., Zhang, S., and Zhou, X. (2006). A downside risk analysis based on financial index tracking models. En M. G. A.N. Shiryaev (Ed.), *Stochastic Finance* (págs. 213–236). Springer.

Zhang, J., & Maringer, D. 2009. An Application of Differential Evolution in Index Mutual Fund Replication. In M. Radek (Ed.), Mendell 2009: 303-308. Brno: Brno Univ Technology Vut Press.

Table 1. Comparing methodologies in DJIA data

	Minimizing ADF statistic	Alexander's Capitalization ranking	Maximizing Correlation coeff.	Minimizing ADF statistic	Alexander's Capitalization ranking	Maximizing Correlation coeff.	Minimizing ADF statistics	Alexander's Capitalization ranking	Maximizing Correlation coeff.
Assets	5	5	5	10	10	10	20	20	20
Mean annual return of TP after TC	11,03%	7.69%	9.52%	11,28%	11,92%	6.34%	11,79%	11.34%	7.77%
Mean annual return of TP	12,21%	8.13%	10.06%	12,21%	12,30%	6.70%	12,45%	11.65%	7.9%
Excess return of TP over DJIA	-0,05%	-4.13%	-2.17%	-0,04%	-0.039%	-5.53%	0,19%	-0.61%	-4.34%
Goodness of Tracking (annualized)	0.0228	0.0125	0.0099	0.0106	0.0061	0.0035	0.0034	0.0019	0.00094
% Synchronization	100%	99.94%	99.98%	100%	100%	99.65%	100%	100%	99.2%
% Return bias square	0%	0.06%	0.02%	0%	0%	0.35%	0%	0%	0.8%
Average annual TE	-0.05%	-4.35%	2.310%	-0.04%	-0.03%	-5.53%	0.19%	-0.61%	-4.34%
TE Volatility	15,11%	11,16%	9,95%	10,30%	7,81%	5.95%	5,81%	4.35%	3.06%
TP Volatility	22%	19.71%	19.44%	18.88%	17.75%	17.63%	17.31%	16.24%	16.59%
Correlation TP/ DJIA index returns	0,7271	0,8241	0,8598	0.8381	0.8983	0.9419	0.9420	0.9638	0.9830
Correlation TE/DJIA index returns	-0,0043	0,0158	0,0657	-0.0236	-0.0152	0.0924	0.0414	-0.0944	0.0811
In sample ADF statistic	-7.0989	-3.5296	--	-8.8092	-4.861	--	-10.326	-6.772	--
Betas of the TP	0,9959	1.0110	1,0407	0.9849	0.9926	1.0342	1,015	0.9744	1.0154
Rate of success	49,71%	52.01%	52,07%	49,44%	50.83%	50.99%	50.29%	51.42%	53.51%
Information ratio	0.5014	0.3901	0.4897	0.5975	0.6715	0.3596	0.6811	0.6982	0.4683
Critical values of Engle Granger test(*)	-4.7295	-4.7295		-5.9509	-5.9509		-7.7045	-7.7045	

(*) 5% of significance

Table 2(a). Comparing methodologies when five assets are included in the portfolio to track the DJIA enhanced

	Minimizing ADF statistic	Alexander's Capitalization ranking	Maximizing Correlation coeff.	Minimizing ADF statistic	Alexander's Capitalization ranking	Maximizing Correlation coeff.
Plus	5%	5%	5%	10%	10%	10%
Assets	5	5	5	5	5	5
Mean annual return of TP after TC	16.31%	6.67%	8.99%	22.25%	5.85%	8.98%
Mean annual return of TP	17.30%	7.06%	9.44%	23.15%	6.23%	9.43%
Excess return of TP over DJIA	5.04%	-5.39%	-2.79%	10.89%	-6.03%	-2.8%
Goodness of Tracking (annualized)	0.0253	0.0131	0.0104	0.0328	0.0139	0.0105
% Synchronization	100%	99.68%	99.76%	100%	99.26%	99.37%
% Return bias square	0%	0.32%	0.24%	0%	0.74%	0.63%
Average annual TE	0.04%	-10.20%	-7.79%	0.90%	-16.02%	-12.79%
TE Volatility	15.9%	11.43%	10.21%	18.11%	11.73%	10.21%
TP Volatility	22.61%	20.09%	19.98%	25.24%	20.25%	19.98%
Correlation TP/ DJIA index returns	0.7108	0.8230	0.8617	0.6998	0.8156	0.8617
Correlation TE/DJIA index returns	0.0006	0.0412	0.1134	0.0887	0.0392	0.1139
In sample ADF statistic	-7.3229	-3.8693		-7.3149	-3.8693	
Betas of the TP	1.0004	1.0291	1.0719	1.0996	1.0282	1.0720
Rate of success	52.28%	53.22%	53.92%	51.19%	54.26%	55.41%
Information ratio	0.7214	0.3320	0.4499	0.8815	0.2888	0.4494
Critical values of Engle Granger test(*)	-4,7295	-4,7295		-4,7295	-4,7295	
Plus	15%	15%	15%	20%	20%	20%
Assets	5	5	5	5	5	5
Mean annual return of TP after TC	27.15%	5.25%	8.97%	31.89%	4.78%	8.96%
Mean annual return of TP	28.13%	5.64%	9.42%	32.87%	5.17%	9.41%
Excess return of TP over DJIA	15.88%	-6.62%	-2.81%	20.61%	-7.09%	-2.82%
Goodness of Tracking (annualized)	0.0587	0.0146	0.0105	0.0767	0.0152	0.0106
% Synchronization	100%	98.72%	98.8%	100%	98.06%	98.04%
% Return bias square	0%	1.28%	1.2%	0%	1.94%	1.96%
Average annual TE	0.89%	-21.61%	-17.8%	0.63%	-27.07%	-22.81%
TE Volatility	24.22%	11.99%	10.21%	27.69%	12.19%	10.21%
TP Volatility	30.53%	20.38%	19.98%	33.11%	20.50%	19.99%
Correlation TP/ DJIA index returns	0.6153	0.8091	0.8617	0.5524	0.8041	0.8617
Correlation TE/DJIA index returns	0.1127	0.0363	0.1144	0.0809	0.0351	0.1150
In sample ADF statistic	-7.2770	-3.7735		-7.1215	-3.6922	
Betas of the TP	1.1693	1.0265	1.0721	1.1386	1.0259	1.0723
Rate of success	52.73%	54.84%	56.58%	53.72%	55.70%	58.29%
Information ratio	0.8893	0.2576	0.4489	0.9632	0.2331	0.4482
Critical values of Engle Granger test(*)	-4,7295	-4,7295		-4,7295	-4,7295	

(*) 5% of significance

Table 2(b). Comparing methodologies when ten assets are included in the portfolio to track the DJIA enhanced

	Minimizing ADF statistic	Alexander's Capitalization ranking	Maximizing Correlation coeff.	Minimizing ADF statistic	Alexander's Capitalization ranking	Maximizing Correlation coeff.
Plus	5%	5%	5%	10%	10%	10%
Assets	10	10	10	10	10	10
Mean annual return of TP after TC	16.19%	10.72%	6.35%	21.01%	9.75%	6.22%
Mean annual return of TP	16.88%	11.07%	6.69%	21.68%	10.12	6.59%
Excess return of TP over DJIA	4.62%	-1.18	-5.54%	9.43%	-2.14%	-5.64%
Goodness of Tracking (annualized)	0.0106	0.0056	0.0036	0.0156	0.0062	0.0036
% Synchronization	100%	99.72%	98.75%	100%	99.04%	97.3%
% Return bias square	0%	0.28%	1.25%	0%	0.96%	2.7%
Average annual TE	-0.37%	-6.18%	-10.54%	-0.57%	-12.13%	-15.64%
TE Volatility	10.32%	7.46%	5.93%	12.47%	7.83%	5.94%
TP Volatility	19.18%	17.97%	17.66%	21.37%	18.33%	17.63%
Correlation TP/ DJIA index returns	0.8429	0.91	0.9424	0.8144	0.9047	0.9421
Correlation TE/DJIA index returns	0.01	0.0390	0.0980	0.1076	0.0669	0.0935
In sample ADF statistic	--9.0599	-5.2120		-8.7305	-5.1958	
Betas of the TP	1.0062	1.0179	1.0360	1.0832	1.0322	1.0342
Rate of success	49.71%	53.63%	54.01%	52.32%	55.11%	56.94%
Information ratio	0.8441	0.5965	0.3595	0.9832	0.5319	0.3528
Critical values of Engle Granger test(*)	-5.9509	-5.9509		-5.9509	-5.9509	
Plus	15%	15%	15%	20%	20%	20%
Assets	10	10	10	10	10	10
Mean annual return of TP after TC	27.02%	8.98%	6.30%	31.56%	8.58%	7.36%
Mean annual return of TP	27.88%	9.38%	6.67%	32.49%	9.00%	7.70%
Excess return of TP over DJIA	15.62%	-2.28%	-5.57%	20.23%	-3.29	-4.53%
Goodness of Tracking (annualized)	0.0368	0.0073	0.0037	0.0682	0.0084	0.0036
% Synchronization	100%	98.24%	95.42%	100%	97.41%	93.41%
% Return bias square	0%	1.76%	4.58%	0%	2.59%	6.59%
Average annual TE	0.63%	-17.87%	-10.54%	0.25%	-23.24%	-24.52%
TE Volatility	19.18%	8.46%	5.94%	26.11%	9.03%	5.84%
TP Volatility	26.42%	18.73%	17.65%	32.26%	19.06%	17.62%
Correlation TP/ DJIA index returns	0.6931	0.8928	0.9422	0.5954	0.8816	0.9441
Correlation TE/DJIA index returns	0.1178	0.0785	0.0954	0.1209	0.0833	0.1008
In sample ADF statistic	-8.2766	-4.9554		-7.8747	-4.7361	
Betas of the TP	1.14	1.0407	1.0353	1.1957	1.0460	1.0358
Rate of success	52.64%	57.14%	58.92%	53.85%	57.86%	61.26%
Information ratio	1.0227	0.4794	0.3569	0.9783	0.4501	0.4177
Critical values of Engle Granger test(*)	-5.9509	-5.9509		-5.9509	-5.9509	

For additional information see foot, Table 1

(*) 5% of significance

Table 2(c). Comparing methodologies when twenty assets are included in the portfolio to track the DJIA enhanced

	Minimizing ADF statistic	Alexander's Capitalization ranking	Maximizing Correlation coeff.	Minimizing ADF statistic	Alexander's Capitalization ranking	Maximizing Correlation coeff.
Plus	5%	5%	5%	10%	10%	10%
Assets	20	20	20	20	20	20
Mean annual return of TP after TC	16.65%	10.77%	7.76%	20.86%	10.40%	7.75%
Mean annual return of TP	17.19%	11.06%	7.89%	21.5%	10.74%	7.87%
Excess return of TP over DJIA	4.94%	-1.20	-4.35%	9.32%	-1.56%	-4.36%
Goodness of Tracking (annualized)	0.0055	0.0018	0.00095	0.0138	0.0028	0.00099
% Synchronization	100%	99.16%	96.36%	100%	98.11%	91.75%
% Return bias square	0%	0.84%	3.64%	0%	1.89%	8.25%
Average annual TE	-0.061%	-6.20	-9.34%	-0.68%	-11.51%	-14.35%
TE Volatility	7.41%	4.27%	3.04%	11.73%	5.26%	3.03%
TP Volatility	18.02%	16.68%	16.6%	21.14%	17.32%	16.61%
Correlation TP/ DJIA index returns	0.9117	0.9667	0.9832	0.8354	0.9531	0.9834
Correlation TE/DJIA index returns	0.0497	0.0156	0.0860	0.1364	0.0848	0.0910
In sample ADF statistic	-10.0943	-6.8701		-9.0363	-6.5119	
Betas of the TP	1.0227	1.0040	1.0161	1.0992	1.0274	1.0167
Rate of success	51.6%	54.08%	59.37%	52.95%	56.20%	64.77%
Information ratio	0.9240	0.6456	0.4674	0.9869	0.6004	0.4665
Critical values of Engle Granger test(*)	-7.7045	-7.7045		-7.7045	-7.7045	
Plus	15%	15%	15%	20%	20%	20%
Assets	20	20	20	20	20	20
Mean annual return of TP after TC	25.24%	10.17%	7.73%	31.21%	10.29%	7.81%
Mean annual return of TP	26.03%	10.56%	7.86%	32.09%	10.72%	7.94%
Excess return of TP over DJIA	13.77%	-1.70%	-4.37%	19.83%	-1.54%	-4.3%
Goodness of Tracking (annualized)	0.0307	0.0040	0.0011	0.0595	0.0051	0.0011
% Synchronization	100%	97.19%	85.84%	100%	96.37%	79.27%
% Return bias square	0%	2.81%	14.16%	0%	3.63%	20.73%
Average annual TE	-1.22%	-16.69%	-19.36%	-0.15%	-21.52%	-24.28%
TE Volatility	17.52%	6.22%	3.01%	24.38%	7.02%	3%
TP Volatility	25.50%	17.86%	16.61%	31.06%	18.29%	0.1663
Correlation TP/ DJIA index returns	0.7342	0.9382	0.9836	0.6295	0.9247	0.9837
Correlation TE/DJIA index returns	0.1519	0.1125	0.096	0.1435	0.1224	0.1032
In sample ADF statistic	-8.2764	-6.030		-7.8439	-5.7075	
Betas of the TP	1.1656	1.0429	1.0174	1.217	1.0526	1.0185
Rate of success	53.04%	58.95%	69.64%	53.27%	59.85%	74.73%
Information ratio	0.9898	0.5694	0.4653	1.0048	0.5626	0.4696
Critical values of Engle Granger test(*)	-7.7045	-7.7045		-7.7045	-7.7045	

(*) 5% of significance