2

Constrained Index Tracking under Loss Aversion Using Differential Evolution

Dietmar Maringer

Centre for Computational Finance and Economic Agents (CCFEA), University of Essex, UK.
dmaring@essex.ac.uk

Summary. Index tracking is concerned with forming a portfolio that mimics a benchmark index as closely as possible. Traditionally, this implies that the returns between the index and the portfolio should differ as little as possible. However, investors might happily accept positive deviations (ie, returns higher than the index's) while being particularly concerned with negative deviations. In this chapter, we model these preferences by introducing loss aversion to the index tracking problem and analyze the financial implications based on a computational study for the US stock market. In order to cope with this demanding optimization problem, we use Differential Evolution and investigate some calibration issues.

2.1 Introduction

Over recent years, passive portfolio management strategies have seen a remarkable renaissance. Assuming that the market cannot be beaten on the long run (in particular after transaction costs), these strategies aim to mimic a given market (or sector) index by investing either into a replication of the benchmark, or by selecting a portfolio which exhibits a behavior as similar to the benchmark's as possible. The market share of products such as exchange traded funds (ETFs) has increased significantly, and it is argued that passive portfolio management is becoming predominant. According to (5), almost half the capital in the Tokyo Stock exchange is subject to passive trading strategies, and (2) report that assets benchmarked against the S&P 500 exceed US$1 trillion.

In contrast, active portfolio management tries to generate excess returns by picking stocks which are expected to outperform the market and avoiding assets that are expected to underperform. Both approaches have their advantages and disadvantages: active strategies rely heavily on superior predictions while passive strategies require few assumptions about future price movements. Passive strategies will also copy the benchmark's poor behavior (in particular when segments of the market drag down the overall performance) while active strategies can react more flexibly in bear markets; etc. If investments are benchmarked against the index, a fund that aims to replicate this benchmark will, by definition, have a lower likelihood to severely

D. Maringer: *Constrained Index Tracking under Loss Aversion Using Differential Evolution*, Studies in Computational Intelligence (SCI) **100**, 7–24 (2008)
www.springerlink.com

fall below this benchmark. It is often argued that it is the actively managed funds' poor performance that makes passive funds all the more attractive (see, e.g., (2)). An alternative, more fundamental argument against active funds is the basis for works including (15) or (3), stating that all investments ought to be efficient *ex ante*, while deviations *ex post* should be transitory. The rational behind this argument is, simply speaking, that if all market participants have the same access to the market and to information, then the market index should be some sort of average of the individual investments. By definition and in the absence of transaction costs, over- and underperformers should then balance. In particular, no participant should be able to produce results that are persistently on just one side of the average. Outperforming the index could therefore be just a matter of good luck rather than skill. Constantly remaining in this position (in particular when transaction costs are incurred) becomes an increasingly rare experience the longer a period of time is considered, as empirical studies confirm.[1]

Arguably the most common approach in passive portfolio management is index tracking. In this strategy, investors select portfolios that mimic the behavior of an index representing the market or a market segment as closely as possible. To find the optimal combination, a distance measure between tracking portfolio and benchmark index is defined which is to be minimized. This so-called Tracking Error (TE) is typically defined as the mean squared deviation between the returns. Akin to volatility, this measure is more sensitive to bigger deviations, but it is oblivious to the sign of the deviation; hence, it does not distinguish between under- or over-performance of the tracking portfolio. Other things equal, investors appear to be mainly concerned with losses rather than any deviation, positive and negative, from their expected outcome. While some authors find that this might even lead to preference functions contradicting the traditional utility analysis for rational risk aversion (see (4) and (18)) the only assumption about investors' preferences made in this contribution is that they want to copy the market and that negative deviations are more "hurtful" than positive ones. Translated to index tracking, investors will try to avoid falling below the benchmark, hence they might want to particularly avoid negative deviations.

This chapter investigates how different levels of loss aversion affect the choice in asset selection for index tracking under realistic constraints. These constraints include an integer constraint on the number of assets, and that initial weights of included assets must fall within certain limits to avoid dominating assets as well as excessive fragmentation and data-fitting. As a consequence, the solution space becomes discrete, exhibits frictions and has multiple optima. Standard optimization techniques cannot deal with these problems satisfactorily and therefore tend to simplify the optimization problem. Heuristic methods, on the other hand, can deal with

[1] As (1, p. 133) put it in their opening statement: "There is overwhelming evidence that, post expenses, mutual fund managers on average underperform a combination of passive portfolios of similar risk. [. . .] Of the few studies that find that managers or a subset of managers with a common objective (such as growth) outperform passive portfolios, most, if not all, would reach opposite conclusions when survivorship bias and/or correct adjustment for risk are taken into account."

these situations (see, e.g., (8)). In particular, it has been found in (10) and (9) that Differential Evolution is well suited to solve this type of optimization problem.

The rest of this chapter is organized as follows. After formalizing the decision problem in section 2.2, section 2.3 will present how it can be solved numerically using Differential Evolution. Section 2.4 presents the results of a computational study, and section 2.5 concludes.

2.2 The Asset Selection Problem under Loss Aversion

2.2.1 Passive Asset Management and Preferences

The simplest way of tracking a given index would be to replicate it by investing into the assets included in it with their exact corresponding weights. This approach, however, is hampered by several practical limitations. For one, when the index is large and has a high number of different assets, monitoring a correspondingly large tracking portfolio becomes not only cumbersome but also rather costly. Furthermore, if the index is computed in a way that the required number of stocks is not constant over time (e.g., when the relative weights are constant), a perfect replication would require frequent portfolio revision which to some extent defeats the purpose of passive portfolio management. Finally, practical or institutional reasons can impede a perfect replication: If the investor's budget is limited and only whole-numbered quantities or lots of assets can be bought, or if there are upper and/or lower limits on the weight an individual asset can have within the portfolio, deviations of the benchmark's structure can become inevitable. Finally, this one-to-one replication strategy requires that the assets are available in the first place (which might be a problem not only for illiquid assets, but also if these assets are not easily accessible, e.g., in commodity indices). Hence, a trade-off between the revision frequency and the magnitude of the tracking error has to be accepted, and in many cases using a subset of the index's component will be preferred to a full replication.

Deviating from the index's structure might be reasonable if the investor does not want to copy all the market movements: in line with the usual non-satiation assumption of investor preferences, he might be willing to accept a (slightly) less well diversified portfolio with a higher risk if the average return is higher. On a similar note, a symmetric measure for the tracking error will not necessarily capture the investor's actual preferences: while the returns of the tracking portfolio should not fall below the index's, the investor will probably not object if the portfolio outperforms the benchmark. Putting more emphasis on losses lessens the impact of positive deviations when evaluating the TE. Explicitly aiming at outperforming the benchmark blends the index tracking approach with active strategies where, typically, predictions about individual assets' future returns are required. In this contribution, however, the original paradigm of passive strategies is maintained where no specific predictions are required and the only assumption about returns is that they follow some but (more or less) stable distribution without further specification.

2.2.2 Risk Aversion and Loss Aversion

Traditionally, the measure for the tracking error is the (root of the) mean squared difference between portfolio and index returns. This measure does not capture the investor's preferences for positive deviation, nor does it distinguish bullish (i.e., increasing) and bearish (i.e., decreasing) markets, and minimizing the tracking error will not maximize the investor's utility.

Findings in behavioral finance suggest that an individual's utility is not only affected by the future level of wealth, but also whether this level is above or below a certain threshold which is usually the current level of wealth. The same level of future wealth provides a higher utility if it represents an increase in wealth rather than a decrease in wealth. The marginal utility will be lower (higher) if this given level of wealth depending on it represents a profit (loss). In other words, decision makers exhibit loss aversion by reacting more sensitively to losses than suggested by the traditional assumptions on risk aversion.[2] A simple way to model this behavior is to introduce a measure of loss aversion, λ, and transform the actual terminal levels of wealth, w_T, into "perceived" wealth, $\widetilde{w}_T$, where losses are amplified while profits are kept unchanged:

$$\widetilde{w}_T = \begin{cases} w_0 \cdot \exp(r) & r \geq 0 \\ w_0 \cdot \exp(r \cdot \lambda) & r < 0 \end{cases} \tag{2.1}$$

$$= w_0 \cdot \exp\left(r \cdot (1 + (\lambda - 1)\Im_{r<0})\right) \tag{2.2}$$

where r is the log return, $r = \ln(w_T/w_0)$ and $\Im_{r<0}$ is a binary indicator for losses. If $\lambda = 1$, the decision maker has no extra attitude to losses beyond the (still intact) risk aversion, while $\lambda > 1$ models additional loss aversion. Cases where $\lambda < 1$ indicate loss seeking behavior which contradicts the other assumptions and can therefore be neglected.

One could find several ways to translate this into an index tracking framework. However, an investor following a passive strategy can be assumed to accept losses if this also reflects the development in the benchmark index. It therefore appears more plausible to regard losses in terms of opportunity costs, i.e., if the tracking portfolio's returns fall below the index's. This has the advantage that it also covers situations where the tracking portfolio loses not as much as the benchmark which, *ceteris paribus*, can be assumed favorable, while being outperformed in a bullish market is not favorable. Hence, in this contribution the decision maker is concerned with perceived deviations of the tracking portfolio's returns, r_P, from the index's, r_I,

[2] Kahneman and Tversky (4) found in their experiments that, as a consequence of loss aversion, individuals are reluctant to realize losses and therefore tend to stick to assets that have generated losses, making their behavior that of an irrational risk seeker (prospect theory). Since this chapter is not concerned with repeated investment decisions and, more importantly, assumes that decision makers comply to the usual assumptions of rationality and risk aversion, their model will not be applied in this contribution. For a critical assessment of prospect theory, see, e.g., (6) and (7).

$$\widetilde{\Delta r} = \begin{cases} r_P - r_I & r_P \geq r_I \\ (r_P - r_I) \cdot \lambda & r_P < r_I \end{cases} \tag{2.3}$$

$$= (r_P - r_I) \cdot (1 + (\lambda - 1)\Im_{r_P < r_I}) \tag{2.4}$$

where $\Im_{r_P<r_I}$ indicates outperformance (and hence opportunity costs). The parameter λ can be interpreted akin to loss aversion for levels of wealth: If $\lambda > 1$, then losses are amplified, and their effects are bigger than under risk aversion alone. The more λ exceeds one, the more harmful losses are perceived and the more keen the decision maker will be to avoid, in particular, these deviations. As a consequence, asymmetric preferences are introduced and reinforced. Furthermore, high loss aversion reduces the (relative) contribution of positive deviations to the TE.

2.2.3 The Constrained Index Tracking Problem under Loss Aversion

Under real life conditions, index tracking is hampered by several practical limitations. Following (10), it can be assumed that an investor has a certain initial budget B_0 and that only a non-negative and integer quantity of n_i of stock i can be bought. If this stock has an initial price of $S_{0,i}$, then the initial fraction invested in this stock is $x_{0,i} = (n_i \cdot S_{0,i})/B_0$. As prices will change over time while quantities are kept constant, the fractions of asset i will also change over time. In the presence of lower and upper limits on the initial weights for included assets, x^ℓ and x^u, respectively, and the absence of short selling, the quantities n_i must either fall within certain bandwidths or be equal to zero. By assumption, these limits must be kept only at time $t = 0$. If price changes over the holding period lead to violations of these limits, portfolio revisions are not required. Note that introducing lower and upper limits also introduces implicit cardinality constraints: if no asset must exceed an initial weight of x^u, then at least $k^{\min} = \lceil 1/x^u \rceil$ must be included, also the lower weight limit allows for at most $k^{\max} = \lfloor 1/x^\ell \rfloor$ positive weights.

The value of the portfolio at this time is $P_t = \sum_{i=1}^{N} n_i \cdot S_{t,i}$ where $S_{t,i}$ denotes the price of stock i at time t. In this contribution, tracking an index I with a portfolio P requires that over a given holding period their daily returns $r_{t,P}$ and $r_{t,I}$, respectively, are as similar as possible. In addition, negative deviations are perceived less favorable than positive ones. Hence, the investor wants to minimize the perceived differences where, depending on the level of loss aversion, there is different additional concern with losses. A popular measure for the tracking error, TE, is the root of the mean squared deviations of returns.[3] This measure is adopted and applied to perceived deviations, $\widetilde{\Delta r}$. The optimization problem for the constrained index tracking portfolio under loss aversion can therefore be summarised as follows:

$$\min TE = \sqrt{\frac{1}{T} \sum_{t=1}^{T} \widetilde{\Delta r_t}^2} \tag{2.5a}$$

[3] See, e.g., (11) or (14); this definition is also widely used in the industry. Alternatively, authors including (13) define the tracking error as the difference between returns.

subject to

$$\widetilde{\Delta r_t} = (r_{t,P} - r_{t,I})(1 + (\lambda - 1)\Im_{r_{t,P} < r_{t,I}}) \tag{2.5b}$$

$$r_{t,P} = \ln(P_t / P_{t-1}), \qquad r_{t,I} = \ln(I_t / I_{t-1}) \tag{2.5c}$$

$$\Im_{r_{t,P} < r_{t,I}} = \begin{cases} 1 & \text{if } r_{t,P} < r_{t,I} \\ 0 & \text{if } r_{t,P} \geq r_{t,I} \end{cases} \tag{2.5d}$$

$$P_t = \sum_{j=1}^{N} n_i \cdot S_{t,i} \tag{2.5e}$$

$$n_i \in \mathbb{N}_0^+ \tag{2.5f}$$

$$n_i : \begin{cases} x^\ell \leq \frac{n_i \cdot S_{0,i}}{B_0} \leq x^u & \text{if asset } i \text{ is included} \\ 0 & \text{otherwise} \end{cases} \tag{2.5g}$$

$$n_i \cdot S_{0,i} = B_0 \tag{2.5h}$$

The solution space for this problem is non-convex, and the integer constraint on the number of assets makes it a discrete problem, traditional numerical methods can therefore not be employed. Heuristic methods, on the other hand, are more flexible with respect to the shape of the solution space and the constraints that have to be met, as will be shown in the following section.

2.3 Differential Evolution for Portfolio Selection

2.3.1 The Principle and Application to Asset Selection Problems

Differential Evolution (DE) is a population based optimization heuristic for continuous search spaces, suggested by Storn and Price (16, 17); a comprehensive presentation can be found in (12). The basic idea is to generate new solutions by linearly combining three distinct current solutions plus crossing over with a fourth; a tournament principle is then employed to decide over replacement. In the course of iterations, the population should evolve and converge towards the (global) optimum.

More specifically, the typical DE implementation is structured as follows. The algorithm starts by generating P random initial solutions where P is the population size. DE is designed for continuous problems, and solutions are represented as vectors $\mathbf{v}_p, p = 1, \ldots, P$, containing the values of the decision variables, which, for the Index Tracking problem at hand, will be the asset weights. $v_p[i]$ therefore represents the weight of asset i in p's solution. Each of the subsequent iterations comprises of the following steps. First, for each of current solution p, one new solution $\widetilde{\mathbf{v}}_p$ is generated. This is done by randomly picking three further distinct current members of the population, $c_1 \neq c_2 \neq c_3$, and linearly combining their corresponding solution vectors. To do so, the first solutions is chosen as the base vector, $\mathbf{v}_{c_1}$ to which the weighted difference of the other two solutions, $F \cdot (\mathbf{v}_{c_2} - \mathbf{v}_{c_3})$, is added: $\mathbf{v}_c := \mathbf{v}_{c_1} + F \cdot (\mathbf{v}_{c_2} - \mathbf{v}_{c_3})$. Next, this combined solution is crossed over with the original candidate solution, p, $\widetilde{\mathbf{v}}_p = \text{crossover}(\mathbf{v}_p, \mathbf{v}_c, \pi)$, where π is the cross over probability that element i comes from parent p. The i-th element of the new solution is therefore computed as follows:

$$\widetilde{v}_p[i] := \begin{cases} v_p[i] & \text{with probability } \pi \\ v_{c_1} + F \cdot (v_{c_2}[i] - v_{c_3}[i]) & \text{otherwise} \end{cases} \tag{2.6}$$

Graphically speaking, the difference vector "moves" the base solution within the solution space. The larger the difference in the i-th element, the larger the move in this dimension, while no (or small) differences in other elements preserve the current position in those dimensions. The latter is the case in particular when the population has converged and "flocks" around a certain point in the solution space. In order to avoid premature convergence to local optima, it is common practise to add noise ("jitter"). In this case, new solutions are generated according to

$$\widetilde{v}_p[i] := \begin{cases} v_p[i] & \text{with probability } \pi \\ v_{c_1} + (F + z_1[i]) \cdot (v_{c_2}[i] - v_{c_3}[i] + z_2[i]) & \text{otherwise} \end{cases} \tag{2.6*}$$

where the vectors $\mathbf{z}_j, j = 1,2$ are vectors with either zero (with probability π_j) or normally distributed values with expected value zero and some predefined standard deviation σ_j. Further extensions can contain a second difference vector where another two current solutions are picked randomly or where the distance from the best solution so far is introduced.

Once a new solution $\widetilde{\mathbf{v}}_p$ has been generated for each current solution $\mathbf{v}_p$, a tournament is run where $\widetilde{\mathbf{v}}_p$ replaces $\mathbf{v}_p$ if it has a better fitness value. The updated population of candidate solutions then enters the next iteration, and the process is repeated until some halting criterion is met, e.g., when the population has converged or a given number of iterations has been passed.

An important aspect in the implementation of any heuristic is constraint satisfaction. For the given asset selection problem, the asset weights have to be chosen such that (ideally) they add up to one and must fall within certain ranges; furthermore the integer constraint on number of assets makes it a discrete problem. In the suggested implementation, these constraints are met by introducing an interpretation or mapping function that converts a candidate solution $\mathbf{v}$ into a valid solution of asset weights $\mathbf{x}$. This function comprises of the following steps. Due to undesirable properties as well as minimum weight constraints, it might be favorable not to include all of the available assets in the tracking portfolio. Hence, it must first be decided which assets shall be assigned positive weights. This is done by finding the elements i where the corresponding element in the solution vector is positive, $v[i] > 0$. If this applies to fewer than $k^{\min} = \lceil 1/x^u \rceil$ assets, the budget could not be spent without violating the upper weight limit; if it applies to more than $k^{\max} = \lfloor 1/x^\ell \rfloor$ then the lower weight limit would be exceeded. In these cases, the elements of $\mathbf{v}$ with the $k^{\min}$ and $k^{\max}$ largest values, respectively, are picked. Next, the included assets are assigned the minimum weight while all the other weights are set equal to zero. Finally, the weights of the included assets are increased proportional to their values in $\mathbf{v}$ and the

Algorithm 1: Pseudocode for tracking error (TE) minimization with Differential Evolution

```
randomly initialize population of vectors v_p, p = 1...P;
repeat
    %% generate new solutions ṽ_p;
    for current solutions v_p, p = 1...P do
        randomly pick c1 ≠ c2 ≠ c3 ≠ p;
        for all elements i in the solution vector do
            with probability π1 : z1[i] ← N(0,σ1) else z1[i] ← 0 ;
            with probability π2 : z2[i] ← N(0,σ2) else z2[i] ← 0 ;
            randomly pick u[i] ~ U(0,1);
            if u[i] < π then
                ṽ_p[i] ← v_p[i];
            else
                ṽ_p[i] ← v_c1[i] + (F + z1[i]) · (v_c2[i] − v_c3[i] + z2[i]);

    %% select new population;
    for current solutions v_p, p = 1...P do
        if TE(ṽ_p) < TE(v_p) then
            v_p ← ṽ_p;

until halting criterion met ;
```

Fig. 2.1. Differential evolution algorithm

weights add up to one. If for some asset(s) the upper weight limit is violated then the excess weight is redistributed proportionally to the remaining included assets until this constraint is also met. The actual number of stocks is determined according to $n_i \leftarrow \langle x_i V_0 / S_{0,i} \rangle$ where $\langle \cdot \rangle$ is the rounding operator.[4] This approach allows the algorithm to operate in the continuous space, but encourages convergence to solutions with suitable discrete counterparts.

2.3.2 Calibrating the Heuristic

Considerations and Experimental Setting

Unlike traditional deterministic methods, heuristics use stochastic ingredients in their search process. As a consequence, they are non-deterministic and independent restarts can lead to different reported solutions. It is therefore common practise to solve a problem repeatedly in independent runs and use the best of these results for

[4] A more strict version would round up (down) if this simple rounding operator violated the lower (upper) weight limit. Given the granularities $S_{0,i}/V_0$ for the chosen assets and initial budget, however, these violations are rather small. Preliminary experiments showed that this more sophisticated method is computationally more expensive, but has only negligible effects on the quality of results.

the subsequent analysis. To investigate the properties of the heuristic, however, the reported results have to be looked at in more detail. Due to their stochastic ingredients, the results reported by a heuristic must be considered random with a certain distribution, and the quality of a heuristic method can be described by the statistical properties of these reported results. These statistical properties can then be used to derive convergence proofs to show that a heuristics is capable of identifying the optimum with a given probability. These properties can also be used to find technical parameters that favor the heuristic's convergence to the global optimum by searching for parameters where the optimum itself or a sufficiently close solution is found with a given probability.

Ideally, the heuristic ought to find the global optimum in each run, resulting in a degenerated distribution with just one realization. In practical applications, however, the reported results can and will differ. Though the resulting distribution is (theoretically) truncated at the global optimum, it might well be that the empirical distribution of reported results from a series of restarts is not: when the global optimum is never found then this bound is never reached.

If a heuristic is to be evaluated for several test problems, it might be beneficial to "standardize" the reported results in a way that they become comparable. As will be presented in Section 2.4, the main computational study distinguishes six different time windows and three different levels of loss aversion. For each of these 18 cases, the optimal tracking error will differ. To evaluate the heuristic, a total of 28 075 independent restarts were performed with different combinations of parameters, resulting in approximately 1550 reported solutions per case c. From these reported solutions $TE_{r,c}$, the best was chosen, $TE_c^* = \min_r TE_{r,c}$, and the relative deviations of reported from optimal solutions, $D_{r,c} = TE_{r,c}/TE_c^* - 1$, computed. While the distributions of the $D_{r,c}$'s might still differ between cases, they are more similar than that of the $TE_{r,c}$'s; moreover, by definition they are all truncated at the same value, namely 0.

An indicator for the reliability of the implementation is how often this limit is actually reached. More recently, sometimes statistics such as the mean and standard deviation of the analyzed indicator are reported; this makes sense only when the assumption of a Gaussian distribution is reasonable (and the truncation at the global optimum is considered in the estimation process); otherwise they are not very illuminating. Usually statistically more stable and more meaningful is information on how likely a certain deviation is to be reached or exceeded. In the lack of a suitable parametric distribution, the latter can be measured by the quantiles of the empirical distribution of the $D_{r,c}$. These quantiles represent which values are exceeded with the respective frequency; they therefore indicate which values are also likely to be exceeded with a certain probability in future restarts. These quantiles can also be used to compare the results of different experimental settings: in superior settings, exceedances beyond a given deviation should happen less frequently, and for a given confidence level, the maximum exceedance should be smaller. Finally, these quantiles allow some indication on how many restarts are advisable.

The Trade-off Between Computational Time, Population Size and Number of Generations

One of DE's advantages is the low number of technical parameters it requires. Furthermore, it is often found that DE requires less tuning than comparable methods and that standard values often yield reasonable stable results already. Since the solution space for the given problem is multi-modal with frictions and interpretation function adds further complexity, experiments for different parameter settings appear reasonable.

The crucial ingredient to any heuristic search method is the number of function evaluations (FEs) during the optimization process. If the algorithm stops not when a convergence criterion is met but when an exogenously given number of FEs has been reached, then for population based methods, a decision has to be made whether these FEs should rather be spent on small populations with many generations or a larger, more diverse population with less iterations to converge. The number of conceded FEs for this implementation are randomly set to 10 000, 25 000, 50 000, 100 000 or 250 000. Next, alternatives for the population size P are 50, 75, 100 and 150. Depending on the combination of these parameter values, the number of generations ranges from 67 (FEs = 10 000, P = 150) to 5 000 (FEs = 250 000, P = 50).

Table 2.1 lists the respective 1%, 10% and 50% quantiles for the different combinations. Other things equal, the figures suggest that increasing the number of function evaluations is highly beneficial: if the FEs are doubled, the deviation often reduces by half or more, in some cases even by more, in magnitude. When there are at least 100 000 FEs, then one in ten restarts is likely to find a solution with a Tracking Error at most 1% above the optimal one's. With 250 000 FEs, every other run can be expected to find a solution which deviates by at most 5% from the global one, and one in 100 restarts is likely to end in the global optimum.

Table 2.1. 1% (10%, 50%) Quantiles for $D_{r,c}$ for different numbers of function evaluations (FEs) and population size (P)

	population size, P			
FEs	50	75	100	150
10 000	0.0334 (0.1189, 0.3083)	0.0613 (0.1483, 0.3446)	0.0941 (0.1738, 0.4153)	0.1751 (0.2728, 0.5264)
25 000	0.0126 (0.0335, 0.1912)	0.0075 (0.0484, 0.1764)	0.0118 (0.0638, 0.1855)	0.031 (0.0604, 0.2335)
50 000	0.0023 (0.0133, 0.1474)	0.0036 (0.0182, 0.128)	0.0015 (0.0223, 0.1254)	0.0033 (0.0278, 0.1357)
100 000	0.0003 (0.0087, 0.1139)	0.0005 (0.0072, 0.1011)	0.0005 (0.0098, 0.1067)	0.0009 (0.0123, 0.1011)
250 000	0 (0.0027, 0.0537)	0.0001 (0.0019, 0.0287)	0 (0.0011, 0.017)	0 (0.0024, 0.0205)

These figures also shed light on another relevant question: other things equal, should a large number of function evaluations be used on one (or a few) run(s) with many iterations or on more frequent restarts with fewer iterations per run? If a total of 1 million FEs can be used for 10 restarts with 100 000 FEs each or 100 restarts with 10 000 FEs each, then 10 restarts of the short runs are equally expensive as one long run. Hence, the 1% quantile of the short runs can be compared to the 10% quantile of the long ones. For all of the tested combinations, having longer runs yields better results than having more restarts with more reported results to choose from.

When looking at the different population sizes, the effect is less consistent. In general smaller populations can converge faster as they have less diversity within them. This might be an advantage when the number of generations is small and converging to a local optimum might be better than not converging at all. With more generations available, however, slower convergence reduces the chances of getting stuck in local optima, and it is the larger populations that benefit. Not surprisingly, it is the latter that also see the strongest improvements in reported results, and one can expect that all the quantiles will be superior to those of smaller population sizes if FEs were increased further. It is particularly noteworthy that increasing the population while keeping the number of generations constant is beneficial. Since in these experiments, the number of generations equals FEs divided by population size, doubling both FEs and population size, e.g., leaves the number of generations untouched. Comparing the quantiles for the different settings favors larger population sizes – at least for the tested ranges. This suggests that larger (and hence more diverse) populations have an advantage over smaller ones when conceded the same time to evolve in terms of generations. A general rule of thumb suggests the population size should be about three times the number of variables; with 64 (65) asset weights to optimize, this would correspond to $P \approx 200$. The empirical results from this study suggest that this would require a substantially higher number of FEs than the ones investigated. From a practical point of view, population sizes of 75 and 100 appear to work sufficiently reliably (provided a sufficiently high number of FEs).

Using Additional Noise in the Optimization Process

When generating new solutions, the scaling factor F and the cross over probability π play an important role (see equations (2.6) and (2.6*)). In most implementations, they are both often chosen in the range between 0.5 and 0.9. In this study, either parameter can take the values 0.25, 0.5 or 0.75. In addition, the cross over probability can be zero, implying that the new solution is equal to the linear combination of three current solutions, but inherits no element of the solution against which it is compared in the subsequent tournament. Table 2.2 summarizes the quantiles of deviations from the global optimum under different parameter constellations. Introducing crossover is beneficial as it increases the probability of finding good solutions. This is particularly true for when small populations and low numbers of FEs are allowed (top half of the table). It is noteworthy, however, that low values for F have a positive effect. In combination with high values for π, this implies that the base vector is not dramatically changed by adding the weighted difference vector, and that the new candidate solution is mainly inheriting properties of its future opponent in the tournament, crossed over with the base vector. The figures also confirm that the results are more sensitive to F than to π.

Table 2.2. 1% (10%, 50%) Quantiles for $D_{r,c}$ for different values of the scaling factor F and cross over probability π for all reported results (top half) and results where FEs $\geq$ 50000 and $P \geq 75$ (bottom half)

	cross over probability π			
F	0	0.25	0.5	0.75
0.25	0.1153 (0.3831, 0.8095)	0.0088 (0.0331, 0.1439)	0.0006 (0.0098, 0.0588)	0.0001 (0.0041, 0.0286)
0.5	0.0093 (0.0918, 0.5506)	0.0002 (0.0053, 0.04)	0.0002 (0.0043, 0.0464)	0.0003 (0.0054, 0.0543)
0.75	0.0224 (0.1149, 0.6762)	0.0172 (0.0655, 0.2003)	0.018 (0.0477, 0.1475)	0.0131 (0.0325, 0.1053)
0.25	0.0828 (0.2648, 0.6864)	0.0044 (0.0193, 0.08)	0.0002 (0.0049, 0.0203)	0 (0.0009, 0.0103)
0.5	0.0053 (0.0427, 0.3913)	0.0001 (0.0027, 0.0178)	0.0003 (0.0028, 0.0178)	0.0004 (0.0043, 0.0215)
0.75	0.0207 (0.0777, 0.3913)	0.0132 (0.0448, 0.1331)	0.0155 (0.0361, 0.0895)	0.0099 (0.0258, 0.0565)

Tables 2.3 and 2.4 summarize the quantiles for the different setting for the noise terms when equation (2.6*) is used to generate new solutions. Note that $\pi_1 = \pi_2 = 0$ reduces this model to the basic version without noise. Differences in the quantiles for different values of σ_i when $\pi_i = 0$ are due to sampling errors and give some indication about the Monte Carlo error for the estimated quantiles. Noting this, the results suggest that adding some noise to the scaling factor F ($\pi_1 > 0$) improves the distribution of reported results; the actual magnitude of the noise, however, seems less important. Adding noise to the difference vector (π_2), on the other hand, has hardly any noticeable effect unless small populations and low numbers of FEs might also be used.

Table 2.3. 1% (10%, 50%) Quantiles for $D_{r,c}$ for different probabilities that the noise term $z_1[i]$ is non-zero (π_1) and its standard deviation in that case (σ_1) for all reported results (top half) and results where FEs $\geq$ 50000 and $P \geq 75$ (bottom half)

	standard deviation for the noise term, σ_1		
π_1	0.05	0.25	0.5
0	0.0012 (0.0205, 0.2316)	0.0011 (0.0198, 0.2561)	0.001 (0.019, 0.2442)
0.025	0.0009 (0.0179, 0.2512)	0.0007 (0.0158, 0.2)	0.001 (0.014, 0.159)
0.1	0.0011 (0.0177, 0.2245)	0.0007 (0.0167, 0.1673)	0.0005 (0.0131, 0.1307)
0	0.0002 (0.0086, 0.0985)	0.0004 (0.0083, 0.1058)	0.0004 (0.0083, 0.0937)
0.025	0.0002 (0.0077, 0.1049)	0.0004 (0.0073, 0.0782)	0.0004 (0.0071, 0.0497)
0.1	0.0004 (0.0089, 0.094)	0.0003 (0.0088, 0.0711)	0.0003 (0.0067, 0.0427)

Comparing all of the results so far, one can confirm the often purported claim that Differential Evolution is usually stable with respect to the chosen parameters. Once

Table 2.4. 1% (10%, 50%) Quantiles for $D_{r,c}$ for different probabilities that the noise term $z_2[i]$ is non-zero (π_1) and its standard deviation in that case (σ_2) for all reported results (top half) and results where FEs $\geq$ 50 000 and $P \geq 75$ (bottom half)

	standard deviation for the noise term, σ_2		
π_2	0.005	0.025	0.05
0	0.0007 (0.0179, 0.2163)	0.0012 (0.0182, 0.1992)	0.0009 (0.0165, 0.1938)
0.025	0.0006 (0.018, 0.2067)	0.0009 (0.0143, 0.1912)	0.0012 (0.0146, 0.1824)
0.1	0.0008 (0.0192, 0.1983)	0.001 (0.0176, 0.1991)	0.0007 (0.0156, 0.2003)
0	0.0004 (0.0089, 0.0846)	0.0004 (0.0089, 0.0769)	0.0004 (0.0078, 0.0722)
0.025	0.0002 (0.0085, 0.0831)	0.0003 (0.0081, 0.0661)	0.0005 (0.0072, 0.0685)
0.1	0.0003 (0.0089, 0.0766)	0.0003 (0.0075, 0.0776)	0.0003 (0.0061, 0.0706)

a sufficiently large number of function evaluations is chosen, together with a reasonable number of restarts and a populations not too small in size, the algorithm is hardly affected by the values of the remaining parameters as long as they fall within certain (but generally broad) bandwidths.

2.4 Computational Study

2.4.1 The Data

For the computational study, the Dow Jones Industrial Average (DJIA64) is to be tracked by using a subset of the stocks included in it. Adjusted daily prices for 65 stocks[5] were downloaded from `finance.yahoo.com` for the period March 2000 to November 2006, leading to a total of 1648 days with observations. Nine missing data are replaced by the averages of the prices of the adjacent days, and one stock has to be excluded for all windows preceding 2004 due to missing data.

For the financial analysis, the in sample periods consist of 500 observations each, representing about two years; the out of sample tests are performed on the subsequent 250 trading days (i.e., the subsequent year). The initial budget is set to 100 000, and the weight limits are $x^{\ell} = 0.01$ and $x^h = 0.5$.

2.4.2 Financial Results

The returns of an ideal tracking portfolio should show no deviations from the index's returns. Given the real world constraints, however, which have been considered in this contribution, this is not achievable, yet the decision maker will aim to come as close to this ideal as possible. Traditionally, this means to minimize the root of the

[5] Note that the composition of the DJIA64 has changed during the observed period.

mean squared deviations between portfolio and index returns. Under loss aversion, investors are more sensitive towards losses. This means that the reaction to losses appears exaggerated when compared to a traditional risk aversion setting and can manifest itself in two ways: the investor will try reduce losses either in magnitude or in frequency, and the investor will request even bigger profits on the positive side to restore an acceptable balance between (expected) return and risk. In either case, the investor will develop a stronger preference for positive skewness, for which he even might accept a (slight) increase in volatility.[6]

For the index tracking problem for the given DJIA data set, loss aversion ($\lambda > 1$) shows predominately in the changed magnitude in returns when compared optimal solutions for investors who are loss neutral ($\lambda = 1$). While the frequency of losses remains more or less unchanged, the mean of the deviations between the tracking portfolio's and the index's returns increases because the mean return increases on days where the tracking portfolio outperforms the index ($r_D > 0$) while it decreases otherwise. As a consequence, the skewness of these deviations increases (as predicted for loss averse investors) (see table 2.5). The standard deviation of r_D remains virtually unchanged, while their kurtosis increases in two years, decreases in another two years and remains constant in the remaining two years.

The financial results presented so far are, strictly speaking, in sample results: the optimized assets weights would have been achievable only under perfect foresight. A more realistic approach is to use a history of returns to optimize the weights and then form a portfolio with exactly these weights. Table 2.6 reports the statistics for the differences between the returns of tracking portfolio and index when decision makers invest for one year, do not readjust their portfolios and chose their asset weights such that it would have been the optimal choice for the preceding two year period. To some extent, these out of sample results reflect the in sample findings: the frequency of days with portfolio returns lower than the index's is hardly effected by the level of loss aversion, nor is the kurtosis. At the same time, with increasing loss aversion, the skewness tends to be higher, and the same is true for the Sharpe ratio. It is noteworthy, however, that the out of sample deviations do not show the high levels of kurtosis that could be observed for the in sample results in later years. The reasons for this are twofold: For one, out of sample results are – by definition – less prone to data fitting. Secondly and more important, the similarity of in and out of sample results for an optimization problem like the one considered here is also an indicator of the stability of the underlying assets' returns. Hence, it is not surprising that the actual tracking errors are bigger (in particular when the composition of the index changes and the decision maker does not adjust the tracking portfolio).

2.5 Conclusion

This chapter investigates the index tracking problem under realistic constraints where, in addition, decision makers can have different levels of loss aversion. Due to

[6] See also (9) for the effects in actively managed portfolios.

Table 2.5. Frequency of losses and statistics of differences in (conditioned) returns between tracking portfolio and index, $r_D = r_P - r_I$, under loss neutrality ($\lambda = 1$) and loss aversion ($\lambda > 1$)

λ	2000–02	2001–03	2002–2004	2003–05	2004–06	2005–2007
			frequency of losses, mean($\Im_{r_D<0}$)			
1.00	0.41483	0.45691	0.41884	0.44289	0.40281	0.45622
1.25	0.45892	0.45691	0.42084	0.44689	0.39479	0.46313
1.50	0.44890	0.45491	0.42285	0.43287	0.40481	0.46774
			mean(r_D)			
1.00	0.00020	0.00014	0.00013	0.00009	0.00009	0.00005
1.25	0.00021	0.00016	0.00013	0.00010	0.00010	0.00006
1.50	0.00022	0.00016	0.00014	0.00011	0.00012	0.00006
			mean($r_D \mid r_D < 0$)			
1.00	-0.00112	-0.00062	-0.00053	-0.00042	-0.00034	-0.00031
1.25	-0.00099	-0.00060	-0.00052	-0.00042	-0.00034	-0.00029
1.50	-0.00101	-0.00060	-0.00051	-0.00042	-0.00031	-0.00029
			mean($r_D \mid r_D > 0$)			
1.00	0.00114	0.00079	0.00060	0.00049	0.00038	0.00036
1.25	0.00124	0.00080	0.00060	0.00051	0.00038	0.00036
1.50	0.00123	0.00080	0.00061	0.00051	0.00040	0.00037
			standard deviation(r_D)			
1.00	0.001447	0.000896	0.000773	0.000661	0.000546	0.000527
1.25	0.001448	0.000896	0.000773	0.000662	0.000546	0.000527
1.50	0.001453	0.000899	0.000775	0.000666	0.000548	0.000529
			skewness(r_D)			
1.00	-0.05705	0.18258	-1.28895	-1.28101	-0.65755	-0.74783
1.25	0.04730	0.26553	-1.18525	-0.99069	-0.46092	-0.56830
1.50	0.10531	0.33182	-1.09341	-0.69125	-0.17956	-0.39504
			kurtosis(r_D)			
1.00	3.94	3.06	17.73	17.20	19.74	25.28
1.25	3.92	3.10	16.79	15.29	19.90	25.89
1.50	3.92	3.12	15.97	13.33	21.24	25.94
			Sharpe Ratio(r_D)			
1.00	0.139830	0.160291	0.162276	0.133908	0.162969	0.101599
1.25	0.147869	0.173980	0.169229	0.143540	0.178148	0.112160
1.50	0.154523	0.182161	0.174573	0.160756	0.211543	0.121664
			actual Tracking Error, mean(r_D^2)			
1.00	0.001460	0.000907	0.000782	0.000666	0.000552	0.000529
1.25	0.001463	0.000909	0.000783	0.000668	0.000554	0.000530
1.50	0.001469	0.000913	0.000786	0.000674	0.000560	0.000533

Table 2.6. Out of sample differences in returns between tracking portfolio and index, $r_D = r_P - r_I$, under loss neutrality ($\lambda = 1$) and loss aversion ($\lambda > 1$)

λ	2003	2004	2005	2006	2007
			frequency of losses, mean($\Im_{r_D<0}$)		
1.00	0.542169	0.461847	0.409639	0.502008	0.451087
1.25	0.534137	0.461847	0.413655	0.481928	0.451087
1.50	0.526104	0.461847	0.417671	0.477912	0.467391
			mean(r_D)		
1.00	-0.000023	0.000778	0.000112	0.000040	0.000129
1.25	-0.000027	0.000769	0.000110	0.000048	0.000127
1.50	-0.000019	0.000766	0.000107	0.000057	0.000128
			std(r_D)		
1.00	0.002271	0.005331	0.000851	0.001048	0.001103
1.25	0.002272	0.005269	0.000839	0.001040	0.001069
1.50	0.002233	0.005258	0.000830	0.001035	0.001057
			skewness(r_D)		
1.00	0.076169	0.482286	-0.410592	0.179581	0.273688
1.25	0.134023	0.481608	-0.404146	0.142729	0.265465
1.50	0.161411	0.490486	-0.364425	0.109294	0.110819
			kurtosis(r_D)		
1.00	4.18	3.74	6.89	7.35	3.03
1.25	4.28	3.74	6.84	7.66	3.02
1.50	4.25	3.77	6.69	8.07	3.02
			Sharpe Ratio(r_D)		
1.00	-0.010261	0.145876	0.131729	0.038443	0.116941
1.25	-0.012062	0.145945	0.130525	0.046456	0.118676
1.50	-0.008420	0.145654	0.128627	0.055289	0.120888
			actual Tracking Error, mean(r_D^2)		
1.00	0.002266	0.005377	0.000857	0.001047	0.001107
1.25	0.002267	0.005315	0.000845	0.001039	0.001074
1.50	0.002228	0.005303	0.000835	0.001034	0.001061

the nature of the constraints, the solution space exhibits frictions and is non-convex and discrete; traditional optimization methods based on first order conditions are therefore not appropriate. Heuristic optimization methods such as Differential Evolution (DE), on the other hand, can deal with such demanding solution spaces. It was discussed how constraint satisfaction can be dealt, and experiments were performed to test different variants of DE and find values for the required technical parameters. The main findings of these experiments shows that (in line with the literature) DE is considerably stable with respect to its parameters. It was also found that for this problem the population size needs not to be substantially higher than the number of decision variables and that large populations are favorable merely with large number of generations.

Meanwhile, there exist numerous variants of DE. A common extension is to add noise which has an effect similar to mutation in other evolutionary methods and which should help to reduce the likelihood of getting stuck in local optima. For the given problem, the experiments suggested that this is not necessary when the other parameters are chosen appropriately. Further variants such as using several difference vectors or enforcing the elitist were not considered here but shall be investigated in future studies.

From a financial point of view, the main result is that the presence of loss aversion has an effect on the investment choice, albeit a small one. As predicted, decision makers will accept slightly bigger deviations from the benchmark if this allows for a higher positive (or a less negative) skewness. Typically, these are matched with slightly higher mean returns because both negative deviations are lowered while positive ones are increased. At the same time, the frequency of falling below the benchmark, however, is hardly affected. Out of sample, these effects mostly persist; however, with neither asset returns nor the index's composition being as stable as assumed by the optimization model, imprecisions are inevitable, and a closer analysis had to be omitted. Extensions to the index tracking model could account for these aspects and include stability measures; furthermore opportunities to readjust the portfolio and the inclusion of transaction costs are of great practical interest, yet have to be left to future research.

Acknowledgements

Valuable comments by the editors, two anonymous referees, and seminar participants at the Universities of Graz and Essex are gratefully acknowledged.

References

[1] Elton EJ, Gruber MJ, Blake CR (1996) The persistence of risk-adjusted mutual fund performance. The Journal of Business 69(2):133–157

[2] Frino A, Gallagher DR, Oetomo TN (2005) The index tracking strategies of passive and enhanced index equity funds. Australian Journal of Management 30(1):23–56

[3] Jensen M (1968) The performance of mutual funds in the period 1945–1964. The Journal of Finance 23(2):389–416

[4] Kahneman D, Tversky A (1979) Prospect theory: An analysis of decision under risk. Econometrica 47:263–291

[5] Konno H, Hatagi T (2005) Index-plus-alpha tracking under concave transaction cost. Journal of industrial and management optimization 1(1):87–98

[6] Levy M, Levy H (2002) Prospect theory: Much ado about nothing? Management Science 48(10):1334–1349

[7] List JA (2004) Neoclassical theory: Evidence from the market place. Econometrica 72(2):615–625

[8] Maringer D (2005) Portfolio Management with Heuristic Optimization. Springer-Verlag

[9] Maringer D (forthcoming) Computational Methods in Financial Engineering, Springer, chap Risk Preferences and Loss Aversion in Portfolio Optimization
[10] Maringer D, Oyewumi O (2007) Index tracking with constrained portfolios. Intelligent Systems in Accounting, Finance and Management 15:57–71
[11] Pope P, Yadav PK (1994) Discovering error in tracking error. Journal of Portfolio Management 20:27–32
[12] Price K, Storn R, Lampinen J (2005) Differential Evolution: A Practical Approach to Global Optimization. Springer
[13] Roll R (1992) A mean/variance analysis of tracking error. Journal of Portfolio Management 18(4):12–22
[14] Rudolf M, Wolter H, Zimmerman H (1999) A linear model for tracking error minimization. Journal of Banking and Finance 23:85–103
[15] Sharpe WF (1966) Mutual fund performance. Journal of Business 39:119–138
[16] Storn R, Price K (1995) Differential evolution – a simple and efficient adaptive scheme for global optimization over continuous spaces. Technical report, International Computer Science Institute, Berkeley
[17] Storn R, Price K (1997) Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. Journal of Global Optimization 11(4):341–359
[18] Thaler R (1980) Toward a positive theory of consumer choice. Journal of Economic Behaviour and Organization 1:39–60

http://www.springer.com/978-3-540-77476-1

Natural Computing in Computational Finance
Brabazon, A.; O'Neill, M. (Eds.)
2008, X, 303 p., Hardcover
ISBN: 978-3-540-77476-1