A Hybrid Genetic Algorithm to track the Dutch AEX-index

Bachelor Thesis

Informatics & Economics
Faculty of Economics
Erasmus University Rotterdam

Roland Jeurissen
E-mail: rjeurissen@gmail.com

Supervisor: dr.ir. Jan van den Berg

6th October 2005

Abstract

Assuming the market is efficient, an obvious portfolio management strategy is passive where the challenge is to track a certain benchmark, such that equal returns and risks are achieved. In this paper, we investigate an approach for tracking the Dutch AEX-index where an optimal tracking portfolio (consisting of a weighted subset of stocks) is determined. The optimal portfolio is found using a hybrid genetic algorithm where the fitness function of each chromosome (possible subset of stocks) equals the minimal tracking error achievable. This minimal tracking error is determined by solving the corresponding quadratic programming problem. Finally, we compare the out-of-sample performance of the portfolio constructed by the hybrid genetic algorithm with several other portfolios.

Keywords: Index Tracking; Passive Management; Portfolio Selection; Quadratic Programming; Genetic Algorithms; Heuristic Approaches; Hybrid Methods

Acknowledgements

First of all, I would like to thank my supervisor Jan van den Berg for his supervision and the many useful suggestions he made regarding this thesis. Furthermore he provided me some exceptional insight with regard to the techniques used in this project. I am also grateful to my co-students Dennis Bron and Alexander Sprengers who inspired me to go deeply into the field of Index Tracking. Finally, I wish to thank my girlfriend Monique van Hoof who encouraged and supported me during the writing of this thesis.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 General background

1.1.1 Portfolio Management

A fund manager is faced with the problem to create a portfolio from an immense set of assets offered in the market that will perform in line with his objective. A common objective is to provide a capital growth over the medium to long-term while respecting certain constraints. Once a portfolio is created it has to be controlled and re-optimized over time due to the dynamic of asset markets. As discussed by Beasley et al. [2003], approaches to portfolio management can be divided into two broad categories:

- *Active management* relies on the belief that skillful investors are able to outperform the aggregate market return. An active manager tries to pick stocks that will outperform other stocks and attempts other active activities such as market timing. (optimal timing their buy/sell decisions) These opportunities are possible because active management assumes that financial markets are not fully efficient.

- *Passive management* on the other hand is adopted by investors who believe that financial markets are efficient. Consequently, passive managers believe that it is impossible to consistently beat the market. Their main activity is endeavoring to achieve the same returns and risk of a certain benchmark. Attempting to reproduce the performance of an index, such as the S & P 500, is often referred as *index tracking*. A passively managed fund whose objective is index tracking is called an *index fund*, or *tracker fund*.

A mixed strategy of the two is also possible; a portion of the money is invested active and a portion passive.

1.1.2 Active versus passive management

The question of active versus passive management is a major issue to pension funds and other large participants. Both strategies have their advantages and disadvantages:

Active management advantages

- Seasoned managers can profit from inefficiencies of the market because of expert analysis and experience.
- Possibility of superior returns
- Managers can make changes if they believe the market will fall

Active management disadvantages

- High transaction and management costs
- Because of the decision making process which securities to include, many mistakes are possible
- Exposed to market and company risk

Passive management advantages (compared to active management)

- Low transaction and management costs
- Much easier decision making process
- Only exposed to market risk

Passive management disadvantages

- Lack of control, managers can not take action if they think markets will fall
- Returns are determined by the performance of the benchmark

The debate about active versus passive management will always continue. However, Sharpe [1991] asserts that on average active managers can not beat passive strategies. Active investing is a zero sum game where some investors will win and some will lose relative to the return of the market or a market segment. Consequently, after costs, the return on the average actively managed dollar will be less than the return on the average passively managed dollar. Empirical researches by Malkiel [1995], Sorenson et al. [1998] and Frino and Gallagher [2001] have shown that passive strategies are found to outperform active strategies on average. So it is not a surprise that passive strategies are increasingly popular. Frino and Gallagher [2001] estimates that in total $1,5 trillion is passively invested in index funds in the USA alone. Or to quote Sorenson et al. [1998]:

> “Equity index funds have become as popular to investors as electronic cash machines have to consumers.”

1.1.3 Index Tracking

As discussed before, index tracking is a method of passive portfolio management. It attempts to match the performance of a certain benchmark. Matching the performance of an index can be performed two ways.

The first is *full (or complete) replication*, a method in which all the shares making up an index are held in their respective market weights. This method perfectly reproduces the index, but has some important disadvantages. It incurs high transaction costs; imagine buying all the stocks of the Wilshire 5000 (composed of more than 6000 companies). Moreover, when stock weights comprising the index change (as a result of a merger, stock split etc.) high transaction costs will occur again. Unfortunately, composition changes of indices happens a lot. The process of adjusting a tracking portfolio due to changes of an index is referred as *rebalancing*.

For these reasons many index fund managers hold only a subset of stocks from the index. This method is called *partial replication* and involves lower transaction and management costs.

However, this strategy introduces *tracking error*, the measure of the deviation of the chosen portfolio from the index.(Shapcott [1992]) Obviously, index fund managers try to minimize this tracking error. Consequently the index tracking problem exists of minimizing tracking error and trying to minimize transaction costs. Fore more information concerning the challenge facing index managers see Frino and Gallagher [2001].

1.2 Goal

In this thesis we try to reproduce the performance of the Dutch AEX-index with only a subset of shares comprising the index (partial replication). Tracking error is minimized in order to accomplish this objective. The goal of the research described in this thesis is to provide insight with regard to the sophisticated methods used to minimize tracking error and to provide insight how a tracking portfolio performs.

1.3 Methodology

A genetic algorithm is used to search the solution space efficiently for possible subsets of stocks in order to create a tracking portfolio. For each subset quadratic programming will be used to determine the weights of the chosen stocks by minimizing tracking error. This in sample tracking error is based on historical calculated returns and covariances. After the optimal portfolio is determined, out-of-sample performance of the tracking portfolio compared to the index is evaluated.

1.4 Structure

The structure of this thesis is as follows. Firstly existing literature in the field of index tracking is discussed in chapter 2. Next the research approach with the research problem, theoretical model and the development of the genetic algorithm will be further explained in chapter 3. The experimental setup and results of the tracking portfolio will be reported in chapter 4. In chapter 5 conclusions and thoughts about further research will be discussed. References and figures can be found at the end.

Chapter 2

The Existing Literature of Index Tracking

2.1 Literature Survey

In spite of the great performance of index tracking recent years, researches about the topic are relatively scarce compared to active management. Whereas the objectives of index tracking are simple and well known, problems arise when trying to put this simple objectives into an optimal solvable model. In order to structure this literature review we first consider empirical insights on index tracking. After that, we consider different techniques to find optimal tracking portfolios including factor models, Markowitz models, quadratic programming and heuristic approaches.

2.1.1 Empirical Insights on Index Tracking

As discussed before, the main objective of index tracking is to reproduce the performance of an index. Performance deviations of the tracking portfolio, i.e. tracking error, arise through several factors. The main factors identified by Frino and Gallagher [2001] are transaction costs (the amount of transaction costs is determined by the liquidity of the underlying index and the efficiency of the market), the way dividends are treated by the index, the volatility of the index, timing effects, front-running by arbitrageurs and index composition changes. Frino and Gallagher [2001] also noticed seasonality in tracking error. Tracking error is significantly higher in January due to "January effect" and during months when stocks go ex-dividend.

Liquid indices such as the S & P 500 tend to have low transaction costs compared to non-liquid ones, but investors have to pay a substantial price premium for stocks in these markets as identified by Petajisto [2004]. When a new stock is added to the index, this generally goes up, this brings extra costs for index managers because they have to buy the stock at the new price.

Larsen-Jr. and Resnick [1998] have shown that tracking portfolios for value weighted indices (based on market capitalization) have less tracking error and lower standard deviations of tracking error than they have for equal weighted indices.

2.1.2 Traditional Methods

There are several ways described in literature to formulate a model that tries to reproduce the performance of an index using only a subset of stocks. Meade and Beasley [2004] and Larsen-Jr. and Resnick [1998] used a single factor model in order to minimize tracking error given a subset of shares. Factor models contribute security returns on one ore more underlying sources. In the case of a single factor, this is usually the market return.

Meade and Salkin [1990] and Meade and Salkin [1989] discuss some assumptions related to the tracking error in order to solve the problem by using quadratic programming. They also considered the effect of industry stratification within a tracking portfolio. That is, shares are selected in order to achieve the same sector representation as the index. However, this stratification strategy did not boost performance in their research. On the other hand, Larsen-Jr. and Resnick [1998] found out that stratification does works, especially for high capitalization indices.

Tabata and Takeda [1995] considered the traditional asset allocation of the Markowitz type (Markowitz [1952],Markowitz [1959]) and used an efficient approach to find a local optimal solution. They also pointed out that the problem is a bi-criteria optimization problem: firstly the zero-one integer problem deciding which stocks to include in the portfolio must be solved and afterwards an algorithm solving the weights of the chosen stocks by minimizing tracking error must be applied. This problem has no tractable analytic solution because of its combinatorial explosion and belongs to the class of NP complete problems.

Consider the Dutch AEX-index, which is generally composed of 25 shares. If we want to track this index with a portfolio consisting of 10 stocks, we can build

$$\frac{25!}{10! * 15!} = 3.268.760 \text{ portfolios.} \tag{2.1}$$

In order to minimize tracking error, we have to solve the optimal weights for the stocks for each of these portfolios. This might be still feasible with nowadays computers, but what happens if we want to track the S & P 500 with 200 stocks?

$$\frac{500!}{200! * 300!} = 5,0549498499355322213139907819098 * 10^{144} \text{ portfolios.} \tag{2.2}$$

As we can see, the solution space is very large because of its combinatorial explosion. That is why Tabata and Takeda [1995] proposed a heuristic approach for this problem.

2.1.3 Heuristic Optimization Techniques

Stochastic and deterministic search procedures - in order to find a subset of shares out of a large universal set that tracks an index well - have been successfully implemented in different ways.

Threshold Accepting

Gilli and Kllezi [2001] used a *Threshold Accepting Algorithm*.

> "The Threshold Accepting Algorithm is a refined local search procedure which escapes local minima by accepting solutions which are not worse by more than a given threshold. The algorithm is deterministic as it does not depend on some probability. The number of steps where we explore the neighborhood for improving the solution is fixed. The threshold is decreased successively and reaches the value of zero after a given number of steps."

The constructed portfolios performed rather well for different data sets. In their objective function they also considered transaction costs and they performed various experiments with different constraints.

Simulated Annealing

Another heuristic search approach is *simulated annealing*. The underlying principle arises from an analogy from metallurgy: reach a minimum energy state upon cooling a substance, but not too quickly in order to avoid reaching an undesirable final state.(Corana et al. [1987]) This procedure is implemented for the tracking error problem by Derichs and Nickel [2003]. Besides the creation of a successful tracking portfolio, they also considered the revision - including the transaction costs involved - of an existing tracking portfolio over time.

Hierarchical Clustering

Dose and Cincotti [2005] formed a tracking portfolio based on *complete-link hierarchical clustering*. They cluster homogeneous groups based on similarity of returns and then take one stock from each cluster to form a subset. After the selection they perform weight optimization of the subset. An interesting method, but not as good as the previous methods.

Genetic Algorithm

Last but not least is the use of *genetic algorithms* in order to search the solution space for a subset of shares that performs well in tracking an index. When this subset is determined, optimization techniques such as quadratic programming are used to solve the optimal proportions to hold the shares. Shapcott [1992] is to our knowledge the first who successfully employed this strategy. He tracked the FTSE 100 index (UK) effectively with a portfolio consisting of 20 stocks. He also proved the superiority of the algorithm compared to random search algorithms.

Eddelbütter [1996] used a similar strategy, but he tracked the German Xetra DAX 30 index. The research included also a section that the genetic algorithm was able to converge to the global minimum tracking error portfolio.

Beasley et al. [2003] extended these approaches by considering transaction costs and the revision of an existing tracking portfolio.(not just its creation) They tested the approach on the S & P 500 (USA), the FTSE 100 (UK), Hang Seng (Hong Kong), Xetra DAX 100 (Germany) and the Nikkei 225 (Japan). The results are quite impressing. They also conducted reduction tests in order to reduce the size of the search space and hence enabling the algorithm to be more effective. Because of the completeness of their research, this is the leading paper for index tracking using genetic algorithms.

Inspired by the outstanding results of the genetic algorithm, we try to track the Dutch AEX-index in this paper. The next paper describes our research approach.

Chapter 3

Research Approach

3.1 Research Problem

Index tracking involves building an investment portfolio designed to reproduce the performance of a particular benchmark index. Because of cost reduction, only a subset of stocks is included in our replicating portfolio. The objective function is to minimize the tracking error of this subset. Following Grinold and Kahn [1999] we define tracking error as the standard deviation of the difference in returns between our portfolio and the benchmark. Calculated as

$$TE_{1,p} = stdev(R_{Pt} - R_{Pb}) = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}((R_{Pt} - R_{Pb}) - (\overline{R}_{Pt} - \overline{R}_{Pb}))^2} \quad (3.1)$$

The symbols are defined as follows,

- $TE_{1,p}$ = Tracking error of the portfolio compared to the benchmark
- R_{Pt} = Return of the portfolio in period t
- R_{Bt} = Return of the benchmark in period t

An alternative measure mentioned by Roll [1992] is to define tracking error as the absolute difference in returns of the index portfolio and the benchmark. Calculated as

$$TE_{2,p} = \frac{\sum_{t=1}^{T} |R_{Pt} - R_{Pb}|}{T} \quad (3.2)$$

We will use $TE_{1,p}$ because when the tracking portfolio consistently outperforms the index, this measure will result in zero tracking error. $TE_{1,p}$ is an ex-post measure because we need R_{Pt} and R_{Bt}. However, we want to predict tracking error in order to minimize tracking error of the portfolio to be created. The expected tracking error following Eddelbütter [1996] is calculated as,

$$E(TE_p) = \sqrt{(h_p^T - h_b^T)V(h_p - h_b)} \quad (3.3)$$

The symbols are defined as follows,

- h_p = the vector defining the weights of the stocks of our portfolio (subset)
- h_b = the vector defining the weights of the stocks of the index (benchmark)
- V = the symmetric matrix of sample covariance's between members of the index

Quadratic programming involves minimization of a quadratic function subject to linear constraints. We can formulate our problem as a quadratic programming problem as

$$\begin{aligned} \text{Minimize } & \sqrt{x^T V x} \\ \text{Subject to } & s \cdot x = 0 \\ \text{with } & -h_b \leq x \leq (-h_b + 1) \end{aligned} \tag{3.4}$$

Where x = the vector of active weights (that is, $h_p - h_b$) and s is a vector containing only ones. Because the weights of the index (h_b) and our portfolio (h_p) sum to 1, follows the constraint that the sum of the active weights (x) must be 0.

3.2 Approach

A hybrid genetic algorithm is used to search the solution space efficiently for possible subsets of stocks and quadratic programming will be used to determine the optimal weights of the stocks in order to minimize expected tracking error. In order to minimize expected tracking error (3.4) we need the benchmarks weights (h_b) and the covariance matrix (V).

3.2.1 Benchmark weights

There are two major ways of constructing an index:

- Equally weighted index (every company has equal weight)
- Market capitalization weighted index (Each of the company is weighted relative to their market capitalization)

An example of the former is the Dow Jones Industrial Average. The latter is the most common one, an example is the S & P 500. [1]

3.2.2 Covariance matrix

The *covariance matrix* defines the relationship between all of the companies in the index. This is important information because our tracking portfolio performances optimal when it has roughly the same similarities and differences between the stocks. In this section we consider three risk models to estimate this important matrix.

[1] In this thesis we track the Dutch AEX-index. This is a real time market capitalization based index.

Single Index Model

The *Single Index Model* supposes that systematic risk can be captured by a single factor; the market. Returns are calculated as

$$r_n = \beta_n \cdot r_m + \theta_n \tag{3.5}$$

Where r_n is stock *n*'s return, θ_n is stock *n*'s residual return,β_n is stock *n*'s beta and r_m is the return of the market. As discussed by Grinold and Kahn [1999], the single index model assumes that the residual returns θ_n are uncorrelated, and hence

$$cov(r_n, r_m) = \beta_n \cdot \beta_m \cdot \sigma_m^2 \tag{3.6}$$

Where σ_m^2 is the variance of the market.

Multi-factor Model

The *multi-factor model* is a more advanced version of the single index model and is based on the notion that returns can be explained by a collection of common factors plus an idiosyncratic element that belongs to that particular stock. Following Grinold and Kahn [1999], returns are calculated as

$$r_n(t) = \sum_k \chi_{n,k}(t) \cdot b_k(t) + u_n(t) \tag{3.7}$$

The symbols are defined as follows,

- $r_n(t)$ = return on stock n above the risk-free return from time t to time t + 1
- $\chi_{n,k}(t)$ = exposure of asset n to factor k at time t
- $b_k(t)$ = factor return to factor k from time t to time t + 1
- $u_n(t)$ = stock $n's$ idiosyncratic return to the stock from time t to time t + 1

If we assume that the idiosyncratic returns are not correlated with the factor returns and are not correlated with each other, the covariance matrix becomes

$$cov(r_n, r_m) = \sum_{k_1,k_2=1}^{K} \chi_{n,k_1} \cdot F_{k_1,k_2} \cdot \chi_{m,k_2} + \Delta_{n,m} \tag{3.8}$$

The symbols are defined as follows,

- F_{k_1,k_2} = covariance of factor k_1 with factor k_2
- $\Delta_{n,m}$ = idiosyncratic covariance of asset n with asset m

Historical estimation

Historical estimation relies on historical variances and covariances. This method does not apply any statistical assumptions, and instead relives history using raw historical data. It requires a large dataset, but it is easy implemented. In this thesis, we will use equal weights for past returns. Returns are calculated continuous as

$$R_n(t) = ln\Big(\frac{P_n(t)}{P_n(t-1)}\Big) \tag{3.9}$$

The covariance matrix is calculated as

$$cov(r_n, r_m) = \frac{\sum_{t=0}^{T-1}(r_n(T-t) - \overline{r}_n)(r_m(T-t) - \overline{r}_m)}{T-1} \tag{3.10}$$

An alternative way of estimating the covariance matrix using historical data is JP Morgan's RiskMetrics^{TM} who use exponentially declining weights for past returns. Consequently, recent observations are more important than old ones.

3.3 Genetic Algorithm

3.3.1 Introduction

As mentioned before, a genetic algorithm is used to search the solution space efficiently for possible subsets of stocks. This section provides a description of how genetic algorithms work. Furthermore, it explains how the algorithm is used in this application. The genetic algorithm (GA) is a randomized search algorithm based on mechanics of natural selection and genetics invented by Holland [1992]. The evolution starts from a population of completely random individuals represented by chromosomes and happens in generations. In each generation, the fitness of the whole population is evaluated, multiple individuals are selected from the current population (based on their fitness) and modified (crossover and mutation) to form a new population, which becomes current in the next iteration of the algorithm. See for more information Mitchell [1996]. Each chromosome represents a possible solution to the general problem. In this case, they take the form of a representation of a particular subset of shares from the index. Typically, binary strings are used as an encoding in order to sample a maximum number of schemata Holland [1992]. An example of a possible chromosome is:

$$[1001010011001000110010010] \tag{3.11}$$

In this investigation, each character represents a stock. For example, the first character is representing ABN AMRO, the second one Aegon etc. If a stock is included in our portfolio it denotes "1", otherwise "0".

3.3.2 Basic Outline

Now we understand how a chromosome might look like, the basic outline of the GA is:

1. (Start) Generate a random population of n chromosomes (solutions for the problem)
2. (Fitness) Evaluate the fitness of each chromosome in the population (In this application, calculate the tracking error of each chromosome)
3. (New population) Create a new population by repeating following steps until the new population is complete
   (a) (Selection) Select two 'parents' chromosomes from the population
   (b) (Crossover) With a crossover probability crossover the parents to form new offspring.
   (c) (Mutation) With a mutation probability mutate new offspring
   (d) (Replace) Replace parents by offspring
4. (Next generation) Use new generated population for a further run of the algorithm until the number of generations is reached

3.3.3 Implementation of operators

Fitness function and hybridization

The fitness function evaluates the strength of the chromosomes. A more fit individual has a higher probability of reproduction over a less fit one. In our research, the fitness function calculates the tracking error of each chromosome. The lower the better, because our objective is to minimize tracking error.

$$FitnessFunction = E(TE_p) = \sqrt{(h_p^T - h_b^T)V(h_p - h_b)} \tag{3.12}$$

In order to calculate the fitness of each chromosome, we do need h_p (the vector defining the weights of the stocks for each chromosome). At this stage, the genetic algorithm is hybridized with the quadratic programming routine. The routine solves the problem for each chromosome and delivers the corresponding tracking error to the fitness function of the genetic algorithm. This approach combines the efficient search procedure of the genetic algorithm in the solution space with the local convergence properties of the quadratic programming solver.

Selection

Selection is the stage of a genetic algorithm in which individuals are chosen from a population for later breeding. In this thesis we use deterministic tournament selection. Tournament selection runs a tournament among a few individuals and selects the winner (the one with the best fitness) Tournament selection has several benefits; it is efficient to code, works on parallel architectures and allows the selection pressure to be easily adjusted. See for more information Miller and Goldberg [June 1995].

Associated with the selection step is the 'elitism' strategy. Elitism is a method that guarantees that a number of best solutions are placed directly into the next generation. In this thesis elitism is used, because that way, the search for a good solution never goes backwards.

Crossover

Crossover operators, which take two parent chromosomes and combine them in such a way as to produce a child, need to be carefully designed. The most common crossover operator is "one point crossover". This crossover operator picks a random point within the chromosomes, and then switches the genes of the two chromosomes at this point to produce two new offspring. If an offspring takes the best parts from each of its parents, the result will likely be a better solution. However, when using this operator the number of ones and zero's in the string of the parents compared to the children can change. This is undesirable; because we don't want that the number of stocks included in the portfolio can change. (Remember, "1" denotes inclusion and "0" non-inclusion)

Parent 1	**11101**00100
Parent 2	01001**01011**
Cut Point	5
Child 1	**1110101011**
Child 2	0100100100

Table 3.1: Example of a single-point crossover

A two point order based crossover is used in this thesis as a result of this constraint. The idea behind the order based crossover is to swap the genes in the order found at the other parent. The number of ones and zeros will remain the same.

Mutation

Mutation is necessary to prevent areas of the search space being discarded, but a too high mutation rate will prevent the desired convergence. The standard mutation operator will choose a single bit at random and swaps its value. (So, a zero becomes a one and vice versa). Again, we don't allow changing the number of ones and zero's. In

Parent 1	100**0011**011
Parent 2	011**0110**010
Cut Point	3 and 7
Child 1	100**0110**011
Child 2	011**1001**010

Table 3.2: Example of a two point order based crossover. The bold part is swapped in the order found at the other parent. The underlined part remains the same.

this thesis we will use mutation inversion. Rather than selecting a single bit to mutate, inversion mutation finds two random characters in the string and reverses them.

Chapter 4

Experimental setup and Results

4.1 Data set

The AEX-index is the best known index of Euronext Amsterdam and is made up of 25 high capitalized stocks. The weights of the stocks are based on market capitalization and are adjusted real time. The index provides a fair representation of the Dutch economy. For more information, see `http://www.euronext.com`. In this thesis we will try to reproduce the performance of this index with a subset of 10 shares.

The composition of the AEX-index changes often. In Figure A.2 we can see that the index has changed 17 times between 01/01/2001-04/05/2004.

We will test our tracking portfolio *out-of-sample* during the period 03/03/2004-03/03/2005. Therefore we will use the composition of the index on 02/03/2004 to determine the benchmark weights and to make the necessary calculations. See Table 4.1 for this composition. The covariance matrix is calculated using daily stock quotes between 02/01/2001-02/03/2004. All quotes are obtained by using DATASTREAM and are adjusted for stock splits, dividends etc. A graph of historical AEX quotes that covers our dataset can be found in Figure A.1.

4.2 Test Environment

The returns and covariance matrix are calculated using Microsoft Excel 2003. This program is also used to evaluate the performance of constructed portfolios and to make the graphs that are presented in the results section. MATLAB 6.5 is used to code the genetic algorithm and it performs the quadratic programming. All computations were performed on an AMD Athlon$^{\mathrm{TM}}$ 64 processor, 1.80 Ghz, 1 GB Ram, personal computer running Windows XP Professional.

Symbol	Company	Quote	Coefficient	Index Contribution	Percentage
AAB	ABN AMRO Holding	18,68	197,00	3679,96	10,22
AGN	Aegon NV	11,98	203,00	2431,94	6,75
AH	Ahold	7,32	206,00	1507,92	4,19
AKZ	Akzo Nobel	31,51	38,00	1197,38	3,32
ASML	ASML	15,45	68,00	1050,60	2,92
BHR	Buhrmann	8,18	18,00	147,24	0,41
DSM	DSM	37,99	13,00	493,87	1,37
FOR	Fortis	18,78	173,00	3248,94	9,02
GUC	Gucci	70,12	3,50	245,41	0,68
GTN	Getronics	2,65	68,00	180,20	0,50
HEI	Heineken	27,06	27,00	730,62	2,03
HGM	Hagemeyer	1,96	66,00	129,36	0,36
ING	ING	19,91	181,00	3603,71	10,01
KPN	KPN	6,49	332,00	2154,68	5,98
MOO	van der Moolen	7,28	5,00	36,40	0,10
NUM	Numico	25,69	23,00	590,87	1,64
PHI	Philips	25,22	146,00	3682,12	10,22
RD	Royal Dutch	40,76	89,00	3627,64	10,07
REN	Reed Elsevier	11,04	95,00	1048,80	2,91
SBMO	SBM Offshore NV	39,97	4,50	179,87	0,50
TNT	TNT	17,96	49,00	880,04	2,44
UN	Unilever	58,85	60,00	3531,00	9,80
VNU	VNU	26,85	34,00	912,90	2,53
VRS	Versatel	2,25	60,00	135,00	0,37
WKL	Wolters Kluwer	15,02	39,00	585,78	1,63
SUM				36012,24	100,00
AEX-Index				360,12	

Table 4.1: Composition of the AEX-Index as of 02/03/2004.

4.3 Parameters of the Genetic Algorithm

One of the difficulties of genetic algorithms is finding the best internal parameters in order to optimize speed and convergence. These parameters are;

- Size of the initial population
- Number of generations
- Crossover probability
- Mutation probability
- Rate of elitism

After extensive trial and error, we have found adequate and robust parameters for our problem. These parameters are presented in Table 4.2.

Size of the initial population	20
Number of generations	100
Crossover probability	0,1
Mutation probability	1,0
Rate of elitism	2

Table 4.2: Internal parameters of the Genetic Algorithm

4.4 Results

We did several experiments and while using the parameters discussed in section 4.3, the hybrid genetic algorithm found the solution presented in Table 4.3. We call this portfolio *GA Tracker*. Both the companies and their weights in the tracking portfolio are given.

Company	Weight
ABN AMRO	0,126271
Aegon NV	0,075173
Ahold	0,04673
Akzo Nobel	0,078453
Fortis	0,087774
ING	0,084059
KPN	0,075942
Philips	0,141958
Royal Dutch	0,139015
Unilever	0,144623
SUM	1
$E(TE_1)$	0,032607

Table 4.3: The portfolio constructed by the genetic algorithm

We observe that the stocks selected for our GA Tracker coincide with the 10 largest stocks of to the AEX-index. (Compare with Table 4.1) This is very reasonable because these heavy weighted stocks do influence the index most. In addition we note that the minimum weight in the tracking portfolio (that of Ahold shares) is around 4,7% and the maximum weight (that of Unilever shares) somewhat less than 14,5%.

In order to estimate the quality of this tracking portfolio found, we performed several other experiments. First of all, we measured the in section 4.1 described out-of-sample performance. Figure A.3 shows the out-of-sample performance of our tracking portfolio compared to the performance of the true AEX-index. We observe that initially, both performances are almost equal while gradually small differences emerge (due to, among other things, changes in the weights of the shares composing the AEX-index). We further note that the up and down movements of both portfolios are quite similar.

To better assess the quality of the tracking portfolio found, we also constructed other portfolios:

- Random: 10 random selected stocks, equal proportions: Figure A.4
- High Cap.: the 10 highest capitalized stocks, equal proportions: Figure A.5
- Low Cap.: the 10 lowest capitalized stocks, equal proportions: Figure A.5

- Full Replication: All the shares making up the AEX are held in their respective market weights (figure not shown, because the graph becomes one line)

The corresponding out-of-sample tracking errors as defined in Chapter 3 are listed in Table 4.4.

	GA Tracker	Random	High Cap.	Low Cap.	Full Replication
TE_1	0,001170	0,005197	0,001611	0,005449	0,000122
TE_2	0,001534	0,005229	0,001905	0,006034	0,000089

Table 4.4: Out-of-sample tracking errors of different portfolios

We observe that generally TE_1 is slightly larger than TE_2, but the pattern between the portfolios is very similar. As might be expected, we note that the tracking performance of the high capitalized portfolio is quite good, but not as good as the performance of the GA Tracker. Probably because the former is equally weighted and the latter is optimal weighted. Furthermore, the performance of the GA Tracker clearly dominates randomly chosen portfolios and portfolios consisting of low capitalized stocks. The full replication method performs even better, but this strategy is accompanied with high transaction costs. (as discussed in section 1.1.3)

Chapter 5

Conclusions and Further Research

5.1 Conclusions

In this thesis we have considered the index tracking problem and presented a heuristic approach for its solution. A genetic algorithm has been used to search the solution space efficiently for possible subsets of stocks in order to create a tracking portfolio. For each subset quadratic programming has been applied to determine the weights of the chosen stocks by minimizing tracking error. We have employed this hybrid approach in order to track the Dutch AEX-index as closely as possible.

We conclude that the performance of the tracking portfolio found is much better than that of randomly selected portfolios and that of low capitalized portfolios which can be derived from the AEX-index. In addition, it is shown that the performance of high capitalized portfolios is similar although still slightly worse than the optimal GA tracking portfolio. This research shows that even a small stock index like the AEX-index can be tracked quite well by a relatively subset of its composing stocks.

On the other hand, we have to remark that we tested our portfolios only 1 year out-of-sample. In this period the AEX-index was a rather dull market without big movements up or down. (See Figure A.1 2004-2005) A longer and more turbulent period is preferable in order to test the robustness of our portfolios.

Another weakness in this project is the design of the crossover operator. It turns out that our genetic algorithm performs best without this “order based crossover operator” or when it is applied with a low probability. (See Table 4.2)

Nevertheless, the algorithm finds the optimal solution considerable fast. (Our GA generally needs about 2000 calculations out of 3.268.760 possibilities) We suspect that this is possible because the ultimate cooperation between the mutation operator and the elitism strategy.

5.2 Further Research

In this thesis we have focused on building a tracking portfolio with minimal tracking error. However, in practice we are also confronted with the revision of an existing tracking portfolio. It is interesting to analyze the frequency needed for re-balancing the optimal tracking portfolio in an attempt to further improve the performance in the long run of the GA tracker. Moreover, when re-balancing we are also faced with the trade-off between tracking error and transaction costs. For more information, see Beasley et al. [2003].

Other interesting issues are the effects of index composition changes (due mergers, liquidations etc.) and the effect of dividend payments. This is by our means not yet fully investigated.

Bibliography

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

A. Corana, M. Marchesi, C. Martini, and S. Ridella. Minimizing multimodal functions of continuous variables with the simulated annealing algorithm. *ACM Transactions on Mathematical Software*, 13:262–280, 1987.

U. Derichs and N. Nickel. On a local-search heuristic for a class of tracking error minimization problems in portfolio management. *University of Cologne*, 2003.

C. Dose and S. Cincotti. Clustering of financial time series with application to index and enhanced-index tracking portfolio. 2005.

D. Eddelbütter. A hybrid genetic algorithm for passive management. 1996.

A. Frino and D.R. Gallagher. Tracking s & p 500 index funds. *The Journal of Portfolio Management*, 28(1):44–55, 2001.

M. Gilli and E. Kllezi. Threshold accepting for index tracking. *University of Geneva*, 2001.

R.C. Grinold and R.N. Kahn. *Active Portfolio Management: a quantative approach for providing superior returns and controlling risk*. McGraw-Hill, 1999.

J.H. Holland. *Adaptation in Natural and Artificial Systems (Second ed.)*. MID Press, 1992.

G.A. Larsen-Jr. and B.G. Resnick. Empirical insights on indexing. *The Journal of Portfolio Management*, 25(1):51–60, 1998.

B. Malkiel. Returns from investing in equity mutal funds 1971 to 1991. *The Journal of Finance*, 50(2):549–572, 1995.

H.M. Markowitz. Portfolio selection. *The Journal of Finance*, 7:77–91, 1952.

H.M. Markowitz. *Portfolio Selection: Efficient Diversification of Investments*. Wiley, 1959.

N. Meade and J.E. Beasley. An evaluation of passive strategies to beat the index. *The Tanaka Business School, London*, 2004.

N. Meade and G.R. Salkin. Developing and maintaining an equity index fund. *The Journal of Operational Research Society*, 41(7):599–607, 1990.

N. Meade and G.R. Salkin. Index funds-construction and performance measurement. *The Journal of Operational Research Society*, 40(10):871–879, 1989.

B.L. Miller and D.E. Goldberg. Genetic algorithms, tournament selection, and the effects of noise. *Complex Systems*, pages 193–212, June 1995.

M. Mitchell. *An Introduction to Genetic Algorithms*. MIT Press, 1996.

A. Petajisto. Selection of an optimal index rule for an index fund. *Yale School of Management, International Centor for Finance*, 2004.

R. Roll. A mean/variance analysis of tracking error. *The Journal of Portfolio Management*, 18(4):57–66, 1992.

J. Shapcott. Index tracking: Genetic algorithms for investment portfolio selection. 1992.

W. Sharpe. The arithmetic of active management. *The Financial Analyst Journal*, 47 (1):7–9, 1991.

E.H. Sorenson, K.L. Miller, and V. Samak. Allocating between active and passive management. *Financial Analysts Journal*, 54(5):18–31, 1998.

Y. Tabata and E. Takeda. Bicriteria optimization problem of designing an index fund. *The Journal of the Operational Research Society*, 46(8):1023–1032, 1995.

Appendix A

List of Figures

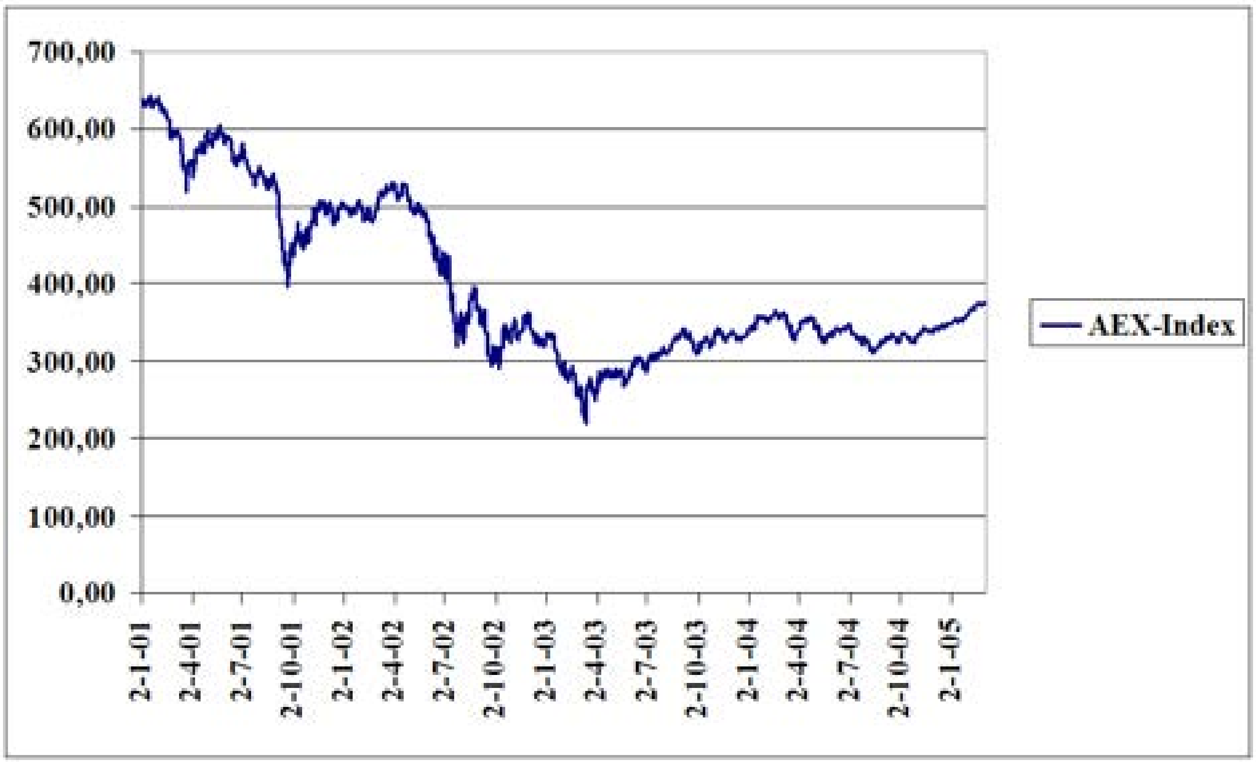
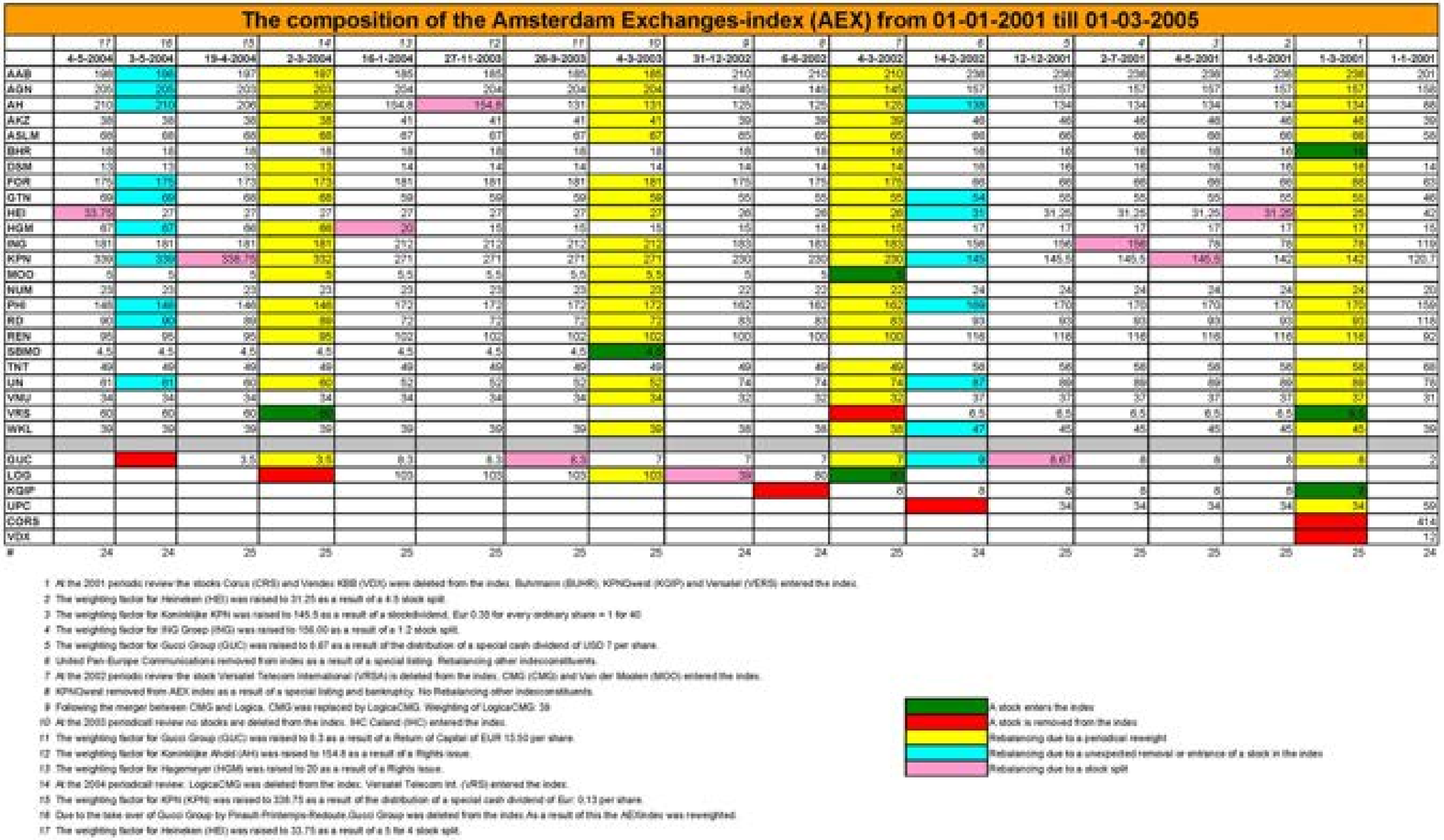


Figure A.1: Historical AEX-Index quotes



The composition of the Amsterdam Exchanges-index (AEX) from 01-01-2001 till 01-03-2005

	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	
	4-5-2004	3-5-2004	19-4-2004	2-3-2004	16-1-2004	27-11-2003	26-9-2003	4-3-2003	31-12-2002	6-6-2002	4-3-2002	14-2-2002	12-12-2001	2-7-2001	4-5-2001	1-5-2001	1-3-2001	1-1-2001
AAB	196	🟦 196	197	🟨 197	185	185	185	🟨 185	210	210	🟨 210	238	238	238	238	238	🟨 238	201
AGN	205	🟦 205	203	🟨 203	204	204	204	🟨 204	145	145	🟨 145	157	157	157	157	157	🟨 157	156
AH	210	🟦 210	206	🟨 206	154.8	🟪 154.8	131	🟨 131	125	125	🟨 125	🟦 138	134	134	134	134	🟨 134	68
AKZ	38	38	38	🟨 38	41	41	41	🟨 41	39	39	🟨 39	46	46	46	46	46	🟨 46	39
ASLM	66	66	66	🟨 66	67	67	67	🟨 67	65	65	🟨 65	66	66	66	66	66	🟨 66	58
BHR	18	18	18	18	18	18	18	18	18	18	🟨 18	18	18	18	18	18	🟩 18	
DSM	13	13	13	🟨 13	14	14	14	14	14	14	🟨 14	16	16	16	16	16	🟨 16	14
FOR	175	🟦 175	173	🟨 173	181	181	181	🟨 181	175	175	🟨 175	66	66	66	66	66	🟨 66	63
GTN	69	🟦 69	68	🟨 68	59	59	59	🟨 59	55	55	🟨 55	🟦 54	55	55	55	55	🟨 55	46
HEI	🟪 33.75	27	27	27	27	27	27	🟨 27	26	26	🟨 26	🟦 31	31.25	31.25	31.25	🟪 31.25	🟨 26	42
HGM	67	🟦 67	66	🟨 66	🟪 20	15	15	15	15	15	🟨 15	17	17	17	17	17	🟨 17	15
ING	181	181	181	🟨 181	212	212	212	🟨 212	183	183	🟨 183	156	156	🟪 156	78	78	🟨 78	119
KPN	339	🟦 339	🟪 338.75	🟨 332	271	271	271	🟨 271	230	230	🟨 230	🟦 145	145.5	145.5	🟪 145.5	142	🟨 142	120.7
MOO	5	5	5	🟨 5	5.5	5.5	5.5	🟨 5.5	5	5	🟩 5							
NUM	23	23	23	23	23	23	23	🟨 23	22	22	🟨 22	24	24	24	24	24	🟨 24	20
PHI	146	🟦 146	146	🟨 146	172	172	172	🟨 172	162	162	🟨 162	🟦 169	170	170	170	170	🟨 170	158
RD	90	🟦 90	88	🟨 88	72	72	72	🟨 72	83	83	🟨 83	93	93	93	93	93	🟨 93	118
REN	95	95	95	🟨 95	102	102	102	🟨 102	100	100	🟨 100	116	116	116	116	116	🟨 116	92
SBMO	4.5	4.5	4.5	4.5	4.5	4.5	4.5	🟩 4.5										
TNT	49	49	49	49	49	49	49	49	49	49	🟨 49	58	58	58	58	58	🟨 58	68
UN	61	🟦 61	60	🟨 60	52	52	52	🟨 52	74	74	🟨 74	🟦 87	89	89	89	89	🟨 89	76
VNU	34	34	34	34	34	34	34	🟨 34	32	32	🟨 32	37	37	37	37	37	🟨 37	31
VRS	60	60	60	🟩 60							🟥	6.5	6.5	6.5	6.5	6.5	🟩 6.5	
WKL	39	39	39	39	39	39	39	🟨 39	38	38	🟨 38	🟦 47	45	45	45	45	🟨 45	39
GUC		🟥	3.5	🟨 3.5	8.3	8.3	🟪 8.3	7	7	7	🟨 7	🟦 9	🟪 8.67	6	6	6	🟨 6	2
LOG				🟥	103	103	103	🟨 103	🟪 39	80	🟩 80							
KQIP										🟥	8	8	8	8	8	8	🟩 8	
UPC												🟥	34	34	34	34	🟨 34	59
CORS																	🟥	414
VDX																	🟥	17
#	24	24	25	25	25	25	25	25	24	24	25	24	25	25	25	25	25	24

1 At the 2001 periodic review the stocks Corus (CRS) and Vendex KBB (VDX) were deleted from the index. Buhrmann (BUHR), KPNQwest (KQIP) and Versatel (VERS) entered the index.
2 The weighting factor for Heineken (HEI) was raised to 31.25 as a result of a 4:5 stock split.
3 The weighting factor for Koninklijke KPN was raised to 145.5 as a result of a stockdividend, Eur 0.30 for every ordinary share = 1 for 40.
4 The weighting factor for ING Groep (ING) was raised to 156.00 as a result of a 1:2 stock split.
5 The weighting factor for Gucci Group (GUC) was raised to 8.67 as a result of the distribution of a special cash dividend of USD 7 per share.
6 United Pan-Europe Communications removed from index as a result of a special listing. Rebalancing other indexconstituents.
7 At the 2002 periodic review the stock Versatel Telecom International (VRSA) is deleted from the index. CMG (CMG) and Van der Moolen (MOO) entered the index.
8 KPNQwest removed from AEX index as a result of a special listing and bankruptcy. No Rebalancing other indexconstituents.
9 Following the merger between CMG and Logica, CMG was replaced by LogicaCMG. Weighting of LogicaCMG: 39
10 At the 2003 periodicall review no stocks are deleted from the index. IHC Caland (IHC) entered the index.
11 The weighting factor for Gucci Group (GUC) was raised to 8.3 as a result of a Return of Capital of EUR 13.50 per share.
12 The weighting factor for Koninklijke Ahold (AH) was raised to 154.8 as a result of a Rights issue.
13 The weighting factor for Hagemeyer (HGM) was raised to 20 as a result of a Rights issue.
14 At the 2004 periodicall review: LogicaCMG was deleted from the index. Versatel Telecom Int. (VRS) entered the index.
15 The weighting factor for KPN (KPN) was raised to 338.75 as a result of the distribution of a special cash dividend of Eur: 0.13 per share.
16 Due to the take over of Gucci Group by Pinault-Printemps-Redoute Gucci Group was deleted from the index As a result of this the AEXindex was reweighted.
17 The weighting factor for Heineken (HEI) was raised to 33.75 as a result of a 5 for 4 stock split.

🟩	A stock enters the index
🟥	A stock is removed from the index
🟨	Rebalancing due to a periodical reweight
🟦	Rebalancing due to a unexpected removal or entrance of a stock in the index
🟪	Rebalancing due to a stock split

Figure A.2: Composition changes of the AEX-Index

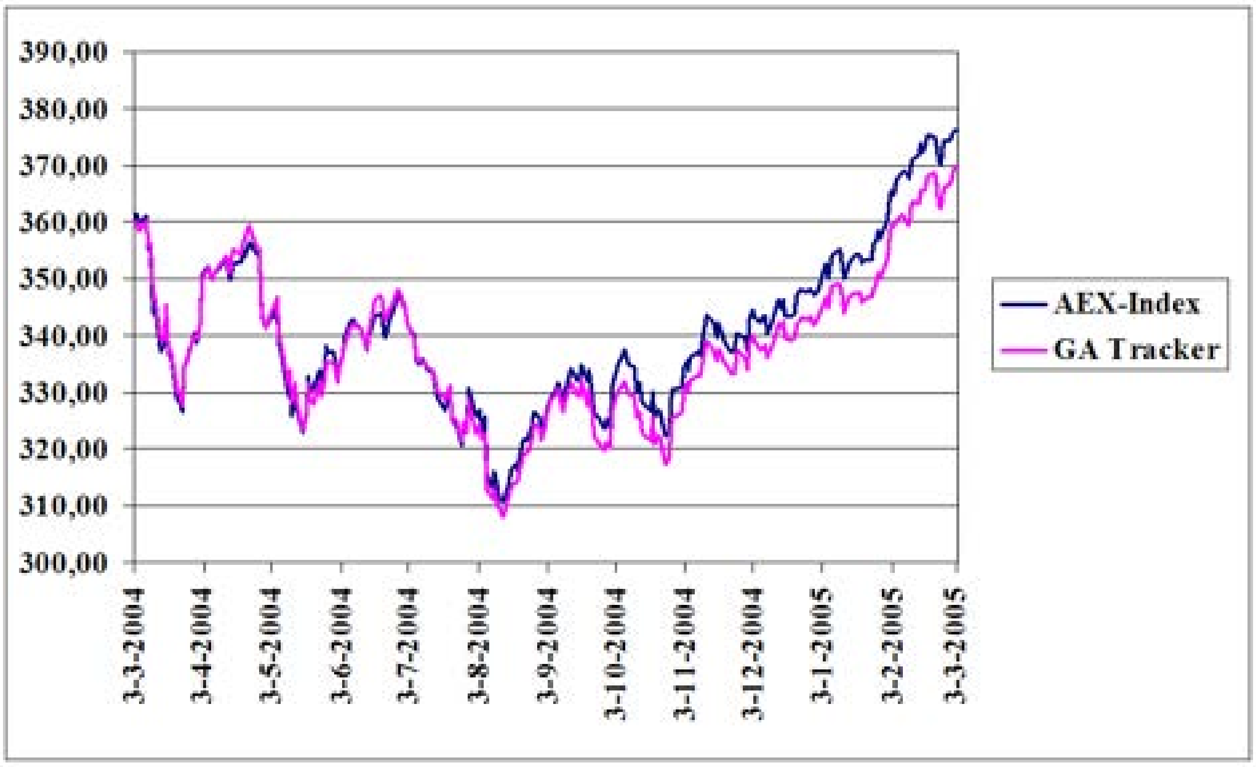


Figure A.3: Tracking performance of a portfolio (10 stocks) constructed by the Hybrid Genetic Algorithm

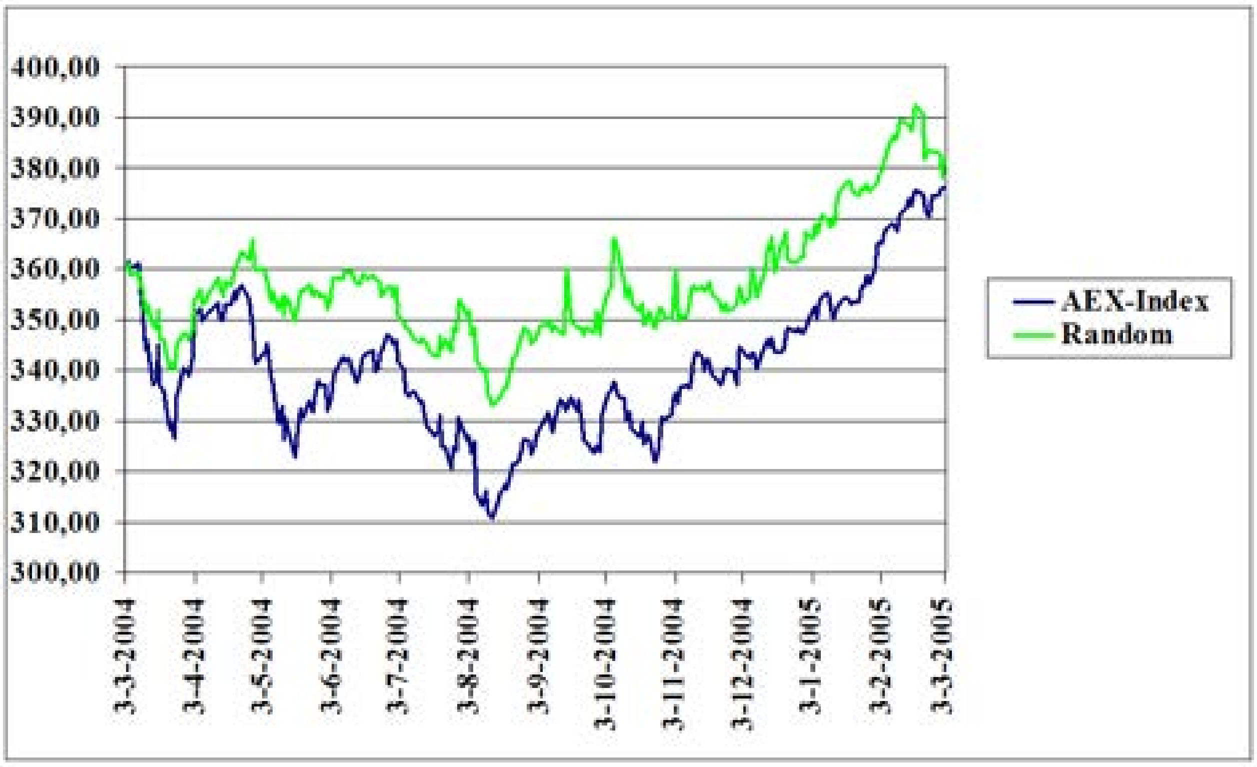


Figure A.4: Tracking performance of a typical random drawn portfolio (10 stocks)

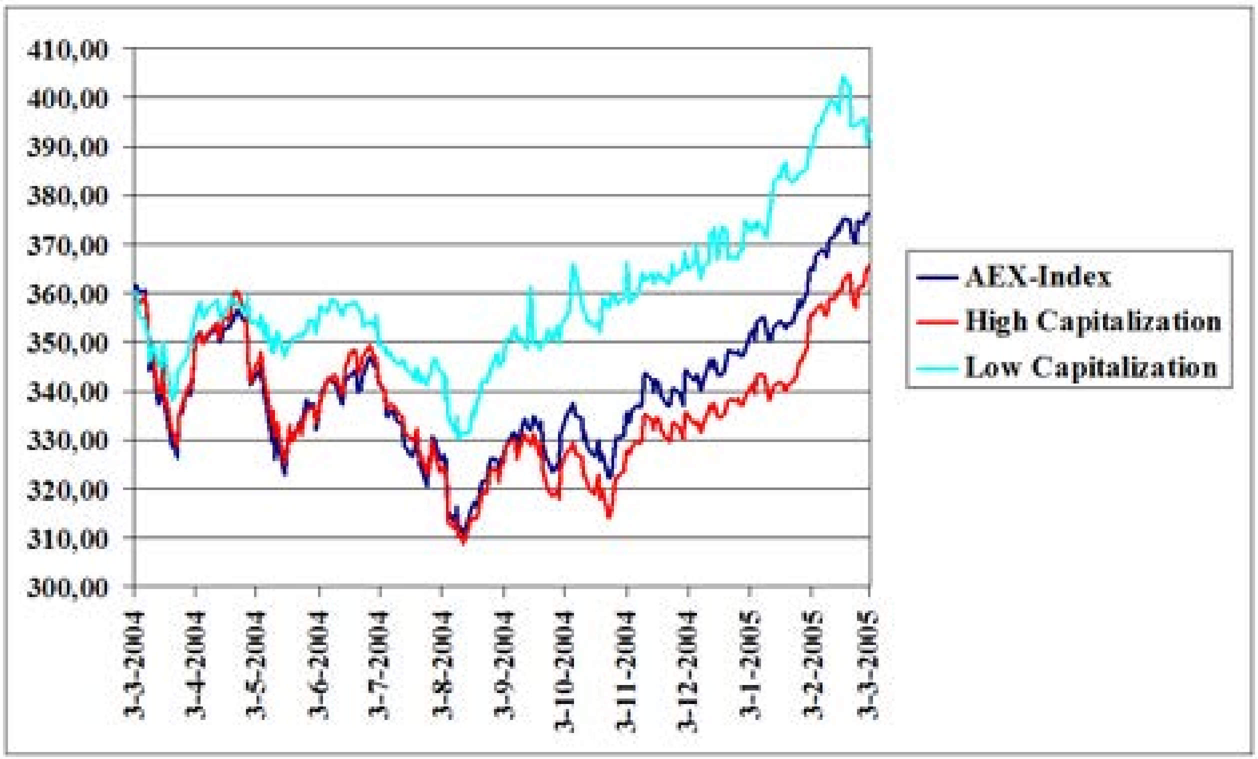


Figure A.5: Tracking performance of respective high and low capitalized portfolios (10 stocks)