Optimal allocation of financial resources is a critical factor in investment decisions. A proper action in this regard requires the existence of appropriate bases and investing tools and techniques in the capital market. One of these effective methods which is considered as the basis of new investment strategies, in addition to multiple unique features, is called index tracking. By considering the undeniable role of this approach in the future of capital markets, its investigation and implementation have been considered in this research and so the problem of optimal Tehran Exchange Dividend Price Index (TEDPIX) tracker fund selection is studied using a hybrid approach of genetic algorithm and quadratic programming. Neural network was applied to simulate unavailable data. Results illustrate the exactness and acceptable performance of the formed portfolios in several iterations such that achieving similar and even better performance compared to the index is a distinctive characteristic of the proposed algorithm.

Keywords: index tracking, tracker fund, genetic algorithm, quadratic programming, neural network

Introduction

The evolution of investment strategies commenced with the analysis of individual securities and continued with first considering mutual effects of the securities, second trying to build portfolios which imitated the market index, third addressing the challenges between active and index investing, and finally attempting to propose a hybrid investing approach reconciling these two strategies. Implementing this hybrid approach requires experience and insight about both strategies and also awareness of the advantages, deficiencies and barriers of each one. However the index tracking has not been seriously taken into account in Iran.

This paper tries to address index tracking, which will definitely be the foundation of the future investing paradigm in Tehran stock exchange. On the other hand, the advantages of this approach (risk mitigation and transaction cost reduction, etc.) encouraged us to introduce this effective investment approach for unprofessional investors, who seem reluctant to enter Tehran stock exchange due to lack of experience and knowledge or the high risk of investing in this market.

Today computers play an important role in solving problems in numerous scientific fields, especially with the advancement of information technology. Therefore, using computerized simulations, researchers and scientists tried to develop models which were in a greater conformance with real environment conditions. The advent of heuristics, such as genetic algorithm in solving financial optimization problems, is a result of such attempts.
In this study, we try to select the optimal portfolio to track Tehran Exchange Dividend Price Index (TEDPIX), through combining genetic algorithm and quadratic programming. The genetic algorithm will be used to effectively search the feasible solution and select possible subsets of stocks. Quadratic programming will be implemented to determine the optimal weight of the chosen stocks through minimizing the tracking error. The complete process is explained in the next sections of the article which are structured as follows: The literature review is covered in section 2, section 3 will include the research methodology, section 4 demonstrates empirical results, and section 5 concludes.

**Literature Review**

Portfolio construction is defined as selecting a set of assets with minimizing the risk while maximizing the return. The basic strategies, used by portfolio managers to achieve these goals, are addressed in two categories:

1. **Active portfolio management.** The fundamental assumption of this strategy is that the portfolio managers could create value using their experience and knowledge in selecting securities or making appropriate timing decisions.

2. **Passive portfolio management.** In this strategy, the fund managers have much less flexibility. They should meet a set of defined criteria, such as achieving the same return as that of the market index (Beasley, Meade, & Chang, 2003). They have to form a portfolio called the tracker fund. Index tracking is one of the least risky strategies of resource allocation. Many experts believe that this approach leads to a higher level of return in the long run compared to the active strategy (Sharpe, 1991). Positive aspects of index investing were discussed in many researches. The studies done by Brinson, Hood, and Beebower (1986) revealed that asset allocation policies—the major asset classes include stocks, bonds, cash, and real estate—account for 93.6% of the variation in portfolio returns over time (Larsen & Resnick, 1998). Investors should spend more time allocating assets and less time guessing the market direction or fretting over individual stock picking.

Ellis (1975) showed that 85% of active managers have not been able to reach a return higher than S&P 500’s in a 10-year period, and claimed that investing in the stock market is a zero-sum game since all investors achieve a return equal to the market’s. It seems that investors are better off if they chose an index tracking portfolio instead of trying to beat the market.

It is notable that investors will receive market return minus transaction costs and the more active they are, the higher transaction costs, market impact, and tax costs they face (Schoenfeld, 2004).

Different methods of index tracking have been proposed in several studies. In order to categorize the literature, the following clusters are introduced.

**Factor Models**

Rudd (1980) developed a single factor model. He implemented a simple heuristic to track the S&P 500 index while considering transaction costs in the objective function. A variance-minimization model was used. The constraint on this objective function was to make the portfolio beta equal to one. Larsen and Resnick (1998) used Rudd’s model to form tracking portfolios and studied the effect of timing in rebalancing the portfolio composition. Corielli and Marcellino (2006) have proposed a method based on the assumption that stock prices are influenced by a factor model. In their approach the index and its tracking portfolio share the same factor structure. This approach consists of sequencing the factors and adding a stock having maximum correlation with a particular factor.
Mean-Variance Based Approaches

These models apply the fundamental principles of Markowitz mean-variance model for index tracking. Tabata and Takeda (1995) proposed a bi-criteria optimization problem. In the first step the stocks to be in the portfolio are selected, and then an algorithm is used to determine the optimal weights to minimize tracking error. Roll (1992) developed a new method by combining Markowitz approach and factor models through adding a constraint relating to the beta of a tracking portfolio. Rohweder (1998) expanded Markowitz model by including a term relating to transaction cost in the objective function.

Quadratic Programming

Meade and Salkin (1989) studied index tracking by forming a portfolio in which the weights were assigned based on that of market segments in the index. They described the tracking error via a mathematical approximation by using quadratic programming. Jansen and Dijk (2002) tried to solve the problem of minimizing the tracking error with limited number of stocks in the tracking portfolio. They minimized a weighted objective function including both the continuous tracking error and the discrete number of stocks in the portfolio. After stock selection the weights will be optimized using quadratic programming. Coleman, Li, and Henniger (2006) used a similar approach either.

Genetic Algorithm

Beasley et al. (2003) introduced an evolutionary genetic algorithm and considered composition adjustment and transaction costs in addition to forming the tracking portfolio. Oh, Kim, and Min (2005) applied a priority function for each stock. The function was the weighted sum of transaction volume, market value and beta. They applied a simple heuristic which used priority functions to select the stocks for the portfolio. Genetic algorithm was then implemented to define optimal weights. Rafaely and Bennell (2006) proposed a comparative approach based on genetic algorithm and quadratic programming. Results showed the advantage of the new method considering various conditions of tracker fund subset size and portfolio composition update rates. Torrubiano and Suarez (2008) developed a hybrid strategy including an evolutionary algorithm which used the output of quadratic programming as the fitness function.

Research Methodology

Research Question

Index tracking is defined as forming an investment portfolio in order to achieve a performance equal to the market index. In order to reduce the transaction costs, only a subset of stocks included in the index are considered in the tracker fund. In other words, we are looking for a k-element set of stocks which has tracked the index in the \((0, T)\) time interval and is able to perform effectively similar to the index in \((T, T + \varepsilon)\). The objective function is to minimize tracking error of this subset, which is defined as a function of difference between the return of tracking portfolio and that of the index. See Equation (1), objective function—the tracking error.

\[
E = \left[ \frac{1}{T} \left( \sum_{i \in S} \Delta_t \left| r_i - R_t \right|^a \right)^{1/a} \right] / T
\]

where \(T = \) index tracking period; \(r_i = \) return of tracking portfolio; \(R_t = \) return of the index; \(\Delta_t = a\) parameter to
assign different weights to returns according to the time they have been realized; $E$ = tracking error.

In this study the assumptions are: $\alpha = 2$ and $\Delta_t = 1$.

**Data and Return Calculation**

As TEDPIX represents the real return of the market securities, the data used for the return calculation of individual stocks should also reflect their total daily return (Raie & Talangi, 2004). The required data were extracted from “Research Development and Islamic Studies Organization” website. The following equation was used to calculate the return of each stock and that of the index. See Equation (2), continuous return.

$$ R = \ln \frac{I_t}{I_{t-1}} $$

where $I_t$ and $I_{t-1}$ represent the stock value at the end of the period and at the beginning of the period, respectively. Here, the period is assumed to be daily.

**Statistical Samples and Sampling Method**

The statistical samples of this study were registered corporations in Tehran Stock Exchange, which had the most impact on the index. Filtering was utilized to uncover these samples. In other words, stocks with more than 100 trading days in each year (and therefore the most active ones) were chosen.

**The Proposed Model for Index Tracking**

The proposed model for index tracking in this study is based on a combination of genetic algorithm and quadratic programming. The steps to solve the research question, according to the model are projected as below:

**Formulating the problem.** In the equations below, $w_i$, $r_{it}$, $R_t$, and $T$ are defined as weight and return of the $i$th stock at time $t$, return of the index at time $t$ and the tracking period respectively.

Minimize:

$$ \sqrt{\frac{\sum_{i=1}^{T} (\sum_{i=1}^{n} z_i w_i - R_i)^2}{T}} $$

Subject to:

$$ \sum_{i=1}^{n} z_i w_i = 1 $$

$$ \sum_{i=1}^{n} z_i = k $$

$$ 0 < w_i < 1, \quad i = 1, 2, \ldots n $$

$$ Z_i \in \{0, 1\} $$

Equation (3) indicates the tracking error which is defined as the objective function. Equations (4) and (6)
refer to the weight constraint of the stocks in the portfolio. Equations (5) and (7) relate to the cardinality constraint. According to the fifth equation, if the investment takes place in the $i$th asset, $Z_i = 1$ otherwise $Z_i = 0$. Parameter $k$ indicates the number of stocks that the investor tends to invest in, based on the third constraint. Hence this constraint ensures the investment in $k$ out of $n$ stocks. With this constraint, the feasible solution will alter to a discrete and non-linear one. This leads to a complex mix of quadratic programming and non-linear integer programming which is a hard problem to solve (Taghavifrad, Mansouri, & Khoshtinat, 2007). On the other hand, the feasible solution expands largely when $n$ or $k$ increases. Therefore the use of heuristic techniques will be of great importance.

Defining the structure of genes and chromosomes. The most important step in solving problems with genetic algorithm, is defining the structure of genes and chromosomes, such that each chromosome represents a potential feasible solution (Holland, 1975). In the proposed model, the designed chromosome is a string of binary numbers. The genes ($s_1, s_2, s_3, s_4, \ldots, s_m$) indicate a set of stocks which impact the index. In other words, they represent all available stocks and their subsets that could be used to form the tracker fund. The initial population generator function randomly selects $k$ stocks to create the portfolio, through assigning the value of one to the corresponding gene in the chromosome. This way the third constraint will be met.

Combining quadratic programming and genetic algorithm. Considering the aforementioned constraints, first, the initial population generator function produces a set of tracking portfolios which constitute the feasible solution. Second, genetic algorithm starts searching the feasible solution. Then, Quadratic programming is called to optimize the weights of stocks in the selected portfolio and to calculate the fitness of this member of the population. So the capability of genetic algorithm in searching and the exactness of quadratic programming in finding the optimal solution are utilized simultaneously. Furthermore the weights of stocks in the final portfolio will be optimized again using quadratic programming. This is because of random nature of the genetic algorithm which may lead to stock weights that can be different from their optimal value.

Applying the genetic operators. After calculating the fitness for each portfolio in the feasible solution, the ones with the highest fitness are selected as parents of the next generation (Mitchel, 1999).

1. Selection operator. The Roulette Selection operator was chosen in our model.
2. Crossover operator. The Single Point Crossover was chosen as the crossover operator. If, as a result of crossover, the number of stocks in the portfolio differs from the pre-determined value, the heuristic algorithm amends this conflict through changing the value of genes, until the cardinality constraint is met. This algorithm ensures the selection of each stock for at least one time, as a part of the potential solution.
3. Mutation operator. The bitwise operator was selected for mutation (Alireza, 2008). In this method, a predefined mutation rate is used to exchange different values of randomly selected genes. It’s clear that the number of 1s in the chromosome remains constant and therefore the cardinality constraint stays met.

Meeting the constraints. The cardinality constraint is met by using the specially designed initial population generator function and crossover and mutation operators. The weight constraint of portfolio elements is met through using the Lsqlin command in Matlab7.

Analyzing the Results

Preparing and Processing the Data

Market Index composition changes with the entry/exit of registered corporations or through weight
adjustment of its elements. Hence data preprocessing is an inevitable part of index investing in order to reflect these amendments. Removing meddlesome factors is an effective technique to improve the accuracy and reliability of the results. Therefore, the time period was chosen from August 30, 2005 to January 27, 2007, in order to minimize the variations in the Index composition. It should be noted that the proposed model can be implemented in any time frame; the mentioned interval was chosen only to reduce the complexity of the calculations. As indicated before, stocks with more than 100 trading days were selected as the ones with the most impact.

As a result of this filtering, the number of selected stocks reached 169. The stocks daily returns from August 30, 2005 to March 13, 2006 (the training period) were used to select the optimal portfolio and those returns corresponding to the time period from March 13, 2006 to October 2, 2006 (the testing period) were used to evaluate the performance of the selected portfolio. The designed algorithm requires daily data of stocks throughout the study period. Shiraz Petrochemical stock entered the market during the training period (November 2, 2005). Therefore, the requisite unavailable data were generated using neural network, so it would be possible for the stock to be chosen in the portfolio. It was assumed that the stock existed from the beginning of the training period with the blank data being generated. To simulate the data from August 30, 2005 to December 2, 2005, the correlation between the return of Shiraz Petrochemical and other 169 stocks was measured from November 3, 2005 to March 13, 2006. In fact we were looking for the stocks with the highest correlation with the target stock to prepare the input for neural network. Pearson correlation coefficient calculated by SPSS was used and the following results were obtained, see Table 1.

<table>
<thead>
<tr>
<th>Stock symbol</th>
<th>Correlation coefficient</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>V-Sanat</td>
<td>0.829</td>
<td>0.00</td>
</tr>
<tr>
<td>V-Bank</td>
<td>0.775</td>
<td>0.00</td>
</tr>
<tr>
<td>Kh-kaveh</td>
<td>0.772</td>
<td>0.00</td>
</tr>
<tr>
<td>Fama</td>
<td>-0.771</td>
<td>0.00</td>
</tr>
<tr>
<td>Chafsat</td>
<td>-0.736</td>
<td>0.00</td>
</tr>
</tbody>
</table>

As seen in Table 1, the correlation coefficient between Shiraz Petrochemical and five other stocks is statistically significant. Hence the return of these stocks was used to simulate unavailable data. Neuro Solution software was utilized to generate the data, which benefits from a Generalized Feed Forward Neural Network with the following specifications:

(1) One hidden layer;
(2) Five perceptrons in the hidden layer;
(3) Tanh Axon transfer function in the hidden layer;
(4) Axon transfer function in the output layer.

Genetic algorithm was applied to train the neural network, which demonstrated a better result than other methods such as multi-layer training. After training the neural network, its performance was evaluated using test data. The following results were obtained in Figure 1.

The performance measure of the neural network, mean square error, was equal to 0.000076137 which indicates its accuracy in simulating and forecasting the daily return of Shiraz Petrochemical stock. In the next
step, the developed neural network properly simulated the daily data, using the learnt relationship between this group of stocks return and that of Shiraz Petrochemical in the training stage. The next step is to apply the proposed algorithm and to test it with different parameters in order to determine their optimal values, before solving the research question.

**Figure 1.** Real and expected output.

### Exploited Parameters

Excel 2007 was employed for primary data gathering and stock filtering. Matlab7 was used for coding the algorithm. A PC with an Intel® Pentium® M 1.86 processor and 512 megabytes of RAM was utilized for the calculations. Due to random nature of the genetic algorithm, calculations were done five times for each sample problem and the best results of the objective function, its mean value and the mean solving time were recorded as performance indicators (Emami, 2007)

**Genetic algorithm parameters.** In order to obtain the best results for crossover and mutation rates, a multistage procedure was implemented. First, some random numbers between 0.65 and 0.85 were generated for the crossover rate. Based on the best result, the next random number was selected to be around the last optimal value. The same method was implemented for the mutation rate in the [0, 0.35] interval. This process was done using Matlab7. This sequence resulted in optimal values for the parameters. Literature review showed that similar studies reached almost the same optimal values for the parameters as Table 2 (Beasley et al., 2003; Rafaely & Bennell, 2006).

<table>
<thead>
<tr>
<th>Genetic Algorithm Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of feasible solutions in the initial population</td>
</tr>
<tr>
<td>Crossover rate</td>
</tr>
<tr>
<td>Mutation rate</td>
</tr>
<tr>
<td>Iteration numbers</td>
</tr>
</tbody>
</table>

### Choosing the Optimal Tracking Portfolio, Using Historical Data

The main objective of forming a tracking portfolio is to achieve risk and return similar to the index in an
indicated time frame. Tracking error is the relevant performance measure (Meade & Salkin, 1990) which in this study is considered to be the root mean squared error. Therefore the proposed algorithm was applied on the processed data which was addressed in the previous section. The results are shown in Table 3.

<table>
<thead>
<tr>
<th>Cardinality constraint</th>
<th>RMSE</th>
<th>Best RMSE</th>
<th>Running time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>$1.54 \times 10^{-4}$</td>
<td>$1.44 \times 10^{-4}$</td>
<td>56.19</td>
</tr>
<tr>
<td>10</td>
<td>$1.16 \times 10^{-4}$</td>
<td>$1.09 \times 10^{-4}$</td>
<td>65.24</td>
</tr>
<tr>
<td>15</td>
<td>$9.41 \times 10^{-5}$</td>
<td>$8.70 \times 10^{-4}$</td>
<td>79.23</td>
</tr>
<tr>
<td>20</td>
<td>$7.41 \times 10^{-5}$</td>
<td>$6.89 \times 10^{-5}$</td>
<td>99.57</td>
</tr>
</tbody>
</table>

*Note.* The calculations are done using historical data from August 30, 2005 to March 13, 2006.

The third column shows the best result of the objective function (the least tracking error of the optimal portfolio considering the corresponding cardinality constraint) out of the five trials. It can be inferred from the results of Table 3 that by increasing the number of stocks in the portfolio (the first column of the table), mean tracking error and the optimal value of the objective function (second and third column) will decrease. The reason is that the portfolio composition becomes more similar to that of index and also the diversity of stocks increases. As expected, the time required to solve the problem enlarges when the number of stocks increases on account of more requisite calculations. As a result of better index tracking performance in parallel with increasing the number of stocks, we will discuss the best portfolio with 20 stocks in more details.

As shown in Figure 2, the portfolio tracks the index with a notable accuracy. The correlation coefficient between the time series data of the index return and that of the portfolio is measured to be 97.45%.

---

**Performance Measurement of the Optimal Tracking Portfolio Using Test Period Data**

Two general strategies exist for maintaining the selected portfolio: (1) maintaining and adjusting the portfolio such that the defined stock weights remain constant; (2) maintaining the portfolio and automatically...
adjusting the weights based on price variations (Torrubiano & Suarez, 2008). According to the researches, the latter benefits more advantages. The first approach requires active portfolio management and involves more transaction costs to keep the weights constant, while the second approach adjusts the weights based on price variations (as a result of changes in relative values of the stocks in the portfolio), and there is no need to decide on selling/buying in order to keep the weights constant. In this study, both approaches were implemented, but the results of the second one will be presented. The initial investment is assumed to be USD 1,000, so the number of each stock in the portfolio could be calculated by dividing the investment on each stock by its spot price. This could be used to calculate daily variations of the weights in the portfolio. As it can be seen in Figure 3, the tracking error for test data has increased to 0.0008248.

![Figure 3. Return of the tracking portfolio compared to that of the index.](image)

Performance of the portfolio based on the test data could be scrutinized from two aspects:

1. Composition of the tracking portfolio should be revised on a regular basis, as a result of corporations entering or leaving the market and changes in the composition of the index. The required time frame for this rebalancing could not be defined with certainty; it should be identified by a fundamental analysis of market information. According to Beasley et al. (2003), this time frame should be less than six months. But it should be considered that the aforementioned time frame is more suitable for efficient capital markets with fewer fluctuations. Therefore a shorter time period is suggested for Tehran stock exchange. According to Figure 3, the tracking accuracy remains significant and the correlation coefficient is 72.3% for about three months after forming the portfolio. Hence it can be inferred that the portfolio composition should be updated after an initial period. This is an inevitable step in the process of index tracking.

2. The increase in tracking error alongside with the decrease in correlation coefficient was a positive outcome; in other words, this incontinency happened due to better performance of the tracker fund compared to that of the index. This could be seen by measuring the discrepancy between return of index and that of portfolio in the mentioned time frame. The results demonstrate a 0.14% excess return.

According to data analysis, the hybrid genetic algorithm is able to precisely track the index throughout the training period. Additionally the proposed method demonstrates a high capability in index tracking for the first three months of test period and a better performance compared to the index during the whole test period.
Optimal TEDPIX Tracker Fund

Composition of the optimal tracking portfolio and the relative weights are shown in Table 4.

Table 4

<table>
<thead>
<tr>
<th>Stock</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teraktor Sazi Iran</td>
<td>0.021143</td>
</tr>
<tr>
<td>Saipa</td>
<td>0.101188</td>
</tr>
<tr>
<td>Iran Khodro</td>
<td>0.102596</td>
</tr>
<tr>
<td>Damleran</td>
<td>0.021442</td>
</tr>
<tr>
<td>Informatics Services</td>
<td>0.080098</td>
</tr>
<tr>
<td>Tehran Cement</td>
<td>0.025366</td>
</tr>
<tr>
<td>Arak Petrochemicals</td>
<td>0.014857</td>
</tr>
<tr>
<td>Isfahan Petrochemicals</td>
<td>0.013897</td>
</tr>
<tr>
<td>Farabi Petrochemicals</td>
<td>0.000493</td>
</tr>
<tr>
<td>Iran Mineral Processing</td>
<td>0.008941</td>
</tr>
<tr>
<td>Iran Zinc Mine Development</td>
<td>0.01436</td>
</tr>
<tr>
<td>Golgohar Ironstone</td>
<td>0.116089</td>
</tr>
<tr>
<td>Melli Bank Investment</td>
<td>0.07216</td>
</tr>
<tr>
<td>Parsian Bank</td>
<td>0.041241</td>
</tr>
<tr>
<td>Petrochemical Investment</td>
<td>0.032878</td>
</tr>
<tr>
<td>DarooPakhsh</td>
<td>0.133181</td>
</tr>
<tr>
<td>Rena Investment</td>
<td>0.058087</td>
</tr>
<tr>
<td>Civic Pension Fund</td>
<td>0.096052</td>
</tr>
<tr>
<td>Karafarin Bank</td>
<td>0.014273</td>
</tr>
<tr>
<td>Mine and Metal Development Investment</td>
<td>0.031657</td>
</tr>
<tr>
<td>Sum</td>
<td>1</td>
</tr>
</tbody>
</table>

It can be seen from Table 4 that the DarooPakhsh, GolGohar ironstone, Iran Khodro, Saipa and civic pension fund investment stocks, account for the highest weights in the portfolio which are calculated to be 13.31%, 11.6%, 10.25%, 10.11%, and 9.6%, respectively. This was predictable due to the high market value of GolGohar ironstone, Iran Khodro, Saipa, and civic pension fund investment stocks in Tehran stock exchange. Therefore the index is highly affected by the variations of aforementioned stocks.

According to Tehran stock exchange Brokerage Corporation, different performance indicators of this market experienced a decreasing trend in 2005. This trend lingered from the last season of 2004 and was a result of diverse domestic and foreign events and irregular growth of indexes in previous years. Therefore DarooPakhsh, as a stock with a decreasing return, was selected to make the portfolio consistent with the index. As it can be seen, a significant weight is assigned to this stock.

The categorization of the portfolio stocks based on their corresponding industry is depicted in Figure 4. Figure 4 shows that industries such as chemicals, metal mineral extractions, auto and part makers account for the highest share in the portfolio. Industries such as multi-functional corporations, banks and financial intermediaries form the next important part of the portfolio. Above mentioned industries have the highest impact on the index in the considered time frame. The categorization of portfolio stocks to different industrial
groups is done based on standard definitions announced by Tehran stock exchange. In this categorization, the investment companies are clustered in different industrial groups based on their field of activities. If we incorporate all investment companies as a single industrial group, this group will represent the most significant part of the portfolio.

The Proposed Approach for Index Investing in Tehran Stock Exchange

1. **Step 1, gathering time-series data for stock and index return.** First, the training period, used to select the optimal tracker fund, should be determined and accordingly the return of the stocks and that of index have to be calculated. Some stocks have less trading days than the index. This may be a result of closure of the trading symbol or thin trading. So generating the missing data is inevitable according to literature. On the other hand, type of the index has to be considered. Ignoring the type could make the results unreliable. For instance, as TEDPIX reflects the capital and dividend gain, stock price changes and in general the total return, the stock data should be adjusted in the tracking process based on the aforementioned parameters.

2. **Step 2, checking possible entry of new stocks in the training period.** The designed heuristic algorithm requires complete data for return of each stock during the whole trading days in the study period to run properly. If an effective stock enters the market, it should have the same opportunity to be selected in the optimal portfolio as other stocks. For this to happen, it is assumed that the stock was available in the market from the beginning of the period and the requisite unavailable data is generated using neural network algorithm.

3. **Step 3, filtering the stocks based on 100 trading days in each year.** The proposed algorithm is capable of analyzing and selecting a portfolio from an unlimited number of stocks. To increase the convergence speed, stocks with minimum effects on the index, are filtered based on 100 trading days in each year.

4. **Step 4, applying the genetic algorithm and selecting the optimal tracking portfolio.**

5. **Step 5, evaluating the portfolio performance based on specific measures.** The portfolio performance is evaluated using tracking error and correlation coefficient criteria. These measures are calculated based on training data.
Step 6, maintaining the optimal portfolio and comparing its performance with that of index. After selecting the optimal portfolio, it is maintained in the test period and its return is studied by comparing it with that of the index. Indicators such as tracking error and excess return provide the investors with sufficient data about the performance of the portfolio in the test period.

Step 7, rebalancing the portfolio. Entry/exit of corporation and the need to adjust the weights in the portfolio coerce us to update the composition of the tracker fund.

Maximum time for maintaining the portfolio is defined around three months (31 trading days) according to the followings:

1. Analysis of data;
2. Comparing the fluctuations of return of the index and that of the tracking portfolio in the test period.

After three months, new weights should be calculated and requisite adjustments should be applied by making necessary sell/buy decisions. In addition, if an effective new stock enters the market during the test period, the composition of the portfolio should be rebalanced and the possibility of selecting the new stock should be provided.

Step 8, returning to step 6 and implementing next steps. After rebalancing, the portfolio should be maintained and fluctuations of the portfolio return have to be compared with those of the index.

Conclusions and Further Research

A suitable investment strategy is based on reducing the risk, minimizing transaction volumes, and reducing transaction costs and taxes. These objectives are the main rationale behind index investing. In fact, this strategy could be considered as the most efficient method to achieve the return of the market. A minimum level of inefficiency in the market is necessary to use active managers, but it is not sufficient. Skillful managers should be employed to take advantage of these inefficiencies. Without required skills, inefficient markets demonstrate a zero-sum game (and even a negative one when considering the associated costs). In other words, investors who do not benefit from such managers should take an index investing position.

The primary objective of this study is to provide the possibility of investing for the ones who seem reluctant to enter capital market due to lack of experience and expertise or inaccessibility to skilful managers. In this regard, the combination of genetic algorithm and quadratic programming was applied to form the optimal tracking portfolio and neural network was implemented to prepare and simulate missing data. Demonstrating similar or even sometimes better performance could be regarded as a unique characteristic of our proposed strategy, while considering that many analogous studies (Beasley et al., 2003; Jeurissen, 2005; Rafaely & Bennell, 2006) achieved the same/lower performance compared to the market index. Furthermore the accuracy of tracking is better than previous studies. The accuracy of the hybrid genetic algorithm is around 0.00001, while this parameter was measured between 0.00001 and 0.001 in similar studies. Evaluation of the index performance during the test period emphasizes the importance of rebalancing the composition of the portfolio.

Selection of Daroopakhsh, GolGohar ironstone, Iran Khodro, Saipa and civil pension fund Investment Company stocks supports the fact that securities with higher market capitalization have more impact on the index.
Structure of the recommended approach is shown in Figure 5.

The followings are suggested for further research:

1. Index investing should not be considered as a replacement for active management. It should be utilized as an efficient tool in line with a general investment strategy. Hence studying fundamental indexing such as core-satellite investing is suggested.

2. Considering transaction costs and trying to minimize them, in addition to defining the portfolio rebalancing period, should be taken into account as critical parts of portfolio management. So, applying required adjustments in the model through scenario planning and analyzing portfolio management from various aspects are strongly suggested.

3. Genetic algorithm as a heuristic method was applied to solve the specified problem. Further optimization methods such as fuzzy logic, particle swarm optimization, etc. which are infrequently used for index tracking are recommended for future studies.
References


