The high social costs associated with bankruptcy have spurred searches for better theoretical understanding and prediction capability. Using an additive super-efficiency data envelopment analysis (DEA) model and genetic algorithm (GA), this paper develops a new assessment index based on two frontiers for predicting corporate failure and success. The proposed approach is applied to 70 bankrupt and 70 healthy matching firms in Tehran Stock Exchange. This sample represents the largest firms that went bankrupt over the period 2004–2010 and represents a full spectrum of industries. Our findings demonstrate that the DEA and GA models is relatively weak in predicting corporate failures compared to healthy firm predictions, and the assessment index improves this weakness by giving the decision maker various options to achieve different precision levels of bankrupt, non-bankrupt, and total predictions.
2. Literature review

The seminal contribution in the literature to address the issue of bankruptcy was written by Altman [20]. Altman [20] was the first to introduce bankruptcy prediction models using discriminant analysis technique. He uses a linear combination (referred to as “Z-score”) of financial variables to obtain a score for each firm in the sample, which discriminates bankrupt firms from non-bankrupt firms using a cutoff point of 0.5. Altman’s model produces adequate results within sample, but its performance out of sample is very poor. Subsequently, Eisenbeis [21] and many others. Grice and Ingram [22] use the C-means cluster model for predicting bankruptcy with a more recent data set and find that some inadequacies in the discriminant analysis approach. Grice and Ingram [22] re-tested Altman’s [20] model on a more recent sample and find that its predictive ability of bankrupt companies fell from 63.5% to 57.8%. Eisenbeis [21] had previously outlined various statistical problems associated with the discriminant analysis approach, but in this study he demonstrates that for matched pair sampling the approach may be adequate. However, in the case of a random sample of firms where the potential failed firms are not around 50%, the predictive ability could seriously be affected. The literature is rich with studies that have used logistic or probit regression in predicting bankruptcy, for example [8, 23, 24, 25, 26]. The traditional approach in using these techniques is to use half of the data sample (estimation sample) for estimating the model and the other half for prediction purposes. These models compute a conditional probability of an observation belonging to a particular category, such as bankrupt or non-bankrupt, and classify off point observation used to classify off point of observations. Some of previous studies have shown that the logistic regression approach provides accurate classification within sample [8], but out-of-sample prediction is very poor. Collins and Green [7] show that logistic regression is superior in predicting failed firms compared to the healthy firms. For a complete review of econometric and operations research methods used to predict financial crises and the logistic regression model refer to Demyanyk et al. [27]. Among other approaches recently developed that play an important role in evaluating corporate failures are neural nets [17], DISUM methodology [28], multidimensional scaling (MDS) techniques [29], the chaos approach [30], and finally DEA [31-33]. In addition, Shannon and Johnson [4] proposed a new approach by integrating the DEA and the principal component analysis (PCA) techniques for ranking of decision making units. Premachandra et al. [33] use an additive DEA model of Charnes et al. [13] to predict bankruptcy. Based on whether the objective function value of the DEA model is positive or not, the firms in the sample were classified as healthy or financially distressed. The authors compared the results from the DEA model with the logistic regression approach and their major findings include (i) the DEA model is superior to the logistic regression approach in predicting financially distressed (bankrupt) firms, whereas (ii) the logistic regression approach is superior to the DEA model in predicting non-bankrupt (healthy) firms. Most bankruptcy prediction models include the DEA and artificial neural networks as dependent and independent ratios of an institution. A recent article by Avkiran [34] investigates to what extent the DEA super-efficiency estimates are associated with key financial ratios in order to address inefficiencies that were not obvious in financial ratio analysis. However, numerous studies have demonstrated that artificial intelligence such as genetic algorithm and support vector machine can be alternative methodology for classification problems to which statistical method have long been applied. Vareto suggested two different models based on genetic algorithm, one of which is a linear model estimating the constant and the variable coefficients of the discriminant function to maximize its discriminant power using genetic algorithm. The other one is a rule based model, which classifies firms according to their respective discriminant scores called GSR (genetic score by rules) using genetic algorithm. Genetic algorithm can also be used to produce a set of rules based on the tests deriving the signs and the cutoff values of selected ratios, and in this regard, Garkaz and Abdollahi suggested a rule based model to improve its prediction power using genetic algorithm. Also, they presented a model beside on genetic algorithm that showed how this algorithm could be used in modeling predicting the bankruptcy. They investigated 528 producing comprised of 264 bankrupted and 265 not bankrupted from 1995 to 1997. Their results showed that the designed model could predict bankruptcy one year in advance its occurrence with the accuracy of 86% [35]. Min and Long suggested a new classification based on the genetic algorithm to predict bankruptcy. The proposed model was also flexible and was able to be applied in other fields such as predicting the purchase of the products or risk management of the project. The financial ratios of 2542 small and middle audited producing companies and the same number not bankrupted one were used as the data of this research [36]. The purpose of this paper is to apply a new approach based on super-efficient additive DEA model designed by Premachandra et al. [8] and on the basis of proposed by Du et al. [37] to improve the predictive ability of the Premachandra et al. [8] model. Finally, the result of DEA model will compare with genetic algorithm proposed [35].

3. Methodology

This research is inductive in logic used and applied in purpose and deductive statistical models and methods (cross sectional correlation) are used to carry it. The subjects are all accepted companies in Tehran Stock Exchange (924 companies) from 2004 to 2010. Applying conditions that were necessary to do this research, the number of sample companies is shown according to year, bankruptcy and non-bankruptcy in table 1.

Table 1. The sum number according to year & group

<table>
<thead>
<tr>
<th>Year</th>
<th>Companies</th>
</tr>
</thead>
<tbody>
<tr>
<td>2004</td>
<td>11</td>
</tr>
<tr>
<td>2005</td>
<td>12</td>
</tr>
<tr>
<td>2006</td>
<td>6</td>
</tr>
<tr>
<td>2007</td>
<td>10</td>
</tr>
<tr>
<td>2008</td>
<td>10</td>
</tr>
<tr>
<td>2009</td>
<td>11</td>
</tr>
<tr>
<td>2010</td>
<td>10</td>
</tr>
</tbody>
</table>

According to the information and data available in library of Tehran Stock Exchange, in that span of time only 70 companies were in accordance with Article 141 of Business law. To compare with bankrupted companies, some 70 non-bankrupted companies were also chosen via random sampling. So, the sample consists of 140 companies’ 70 bankrupted and 70 non-bankrupted ones. Those companies were classified into two groups of educational (to design the models) and experimental (to identify the accuracy of the models). The variables are classified into dependent and independent ones. By using of year t-1 data, two models using GA and DEA using the information of those years were designed. The accuracy of the designed models was evaluated with the real conditions of the companies. To do so, modeling has been taken placed with information of t-1 years. In other words, information of the t-1 years has been chosen as educational sample. Educational sample consists of 140 companies including 70 broken companies and 70 non-broken companies established in the t year. In this part the obtained results will be presented. The methodology used in this paper for bankruptcy prediction is based on DEA and GA. DEA was developed by Charnes et al. [13] to assess the efficiency of decision-making units (DMUs) that have multiple inputs and outputs. Compared to the statistical approaches, DEA has the following unique features that make it an excellent tool for predicting corporate failure. GA is developed by Shir and Lee [8] and Abdollahi [3] to classification of bankrupt and non-bankrupt firms. DEA does not require any prior assumptions of the relationship between inputs and outputs. DEA examines each DMU uniquely, by generating individual performance (efficiency) scores that are relative to the entire sample under investigation. DEA development can study the frontier shift over a time horizon. This allows us to explore the dynamic change of corporate failure or success on a time horizon. DEA does not need a large sample size for bankruptcy evaluation, usually required by such statistical and econometric approaches. In fact, DEA has been proven an excellent tool for identifying best practice or corporate success. The current study proposes to use DEA and integrate the DEA best-practice frontier and bankruptcy frontier to develop a discriminant index for corporate failure and success assessments. I use a genetic algorithms (GAs) approach also in this study and illustrate how GAs can be applied to corporate failure prediction modeling. An advantage of this approach is that it is capable of extracting rules that are easy to understand for users like expert systems.

3.1. Genetic Algorithm

The genetic algorithm is an artificial intelligence procedure based on the theory of natural selection and evolution. GA uses the idea of survival of the fittest by progressively accepting better solutions to the problems. It is inspired by and named after biological processes of inheritance, mutation, natural selection, and the genetic crossover that occurs when parents mate to produce offspring [15]. GA differs from conventional non-linear optimization techniques in that it searches by
maintaining a population (or data base) of solutions from which better solutions are created rather than making incremental changes to a single solution to the problem. GA simultaneously possesses a large number of candidate solutions to a problem, called a population. The key feature of GA is the manipulation of a population whose individuals are characterized by possessing a chromosome. Therefore, GA is distinct from many conventional search algorithms in the following ways:

1. GA consider not a single point but many points in the search space simultaneously reducing the chance of converging to local optima;
2. GA work directly with strings of characters representing the parameter set, not the parameters themselves;
3. GA use probabilistic rules, not deterministic rules, to guide their search.

Two important issues in GA are the genetic coding used to define the problem and the evaluation function, called the fitness function. Each individual solution in GA is represented by a string called the chromosome. The initial solution population could be generated randomly, which evolves into the next generation by genetic operators such as selection, crossover and mutation that means GAs perform the search process in four stages: initialization, selection, crossover, and mutation. Fig. 1 shows the basic steps of GAs. In the initialization stage, a population of genetic structures, called chromosomes that are randomly distributed in the solution space is selected as the starting point of the search. After the initialization stage, each chromosome is evaluated using a user-defined fitness function. The role of the fitness function is to numerically encode the performance of the chromosome. For real-world applications of optimization methods such as GAs, choosing the fitness function is the most critical step. The solutions coded by strings are evaluated by the fitness function. The selection operator allows strings with higher fitness to appear with higher probability in the next generation. Crossover is performed between two selected individuals, called parents, by exchanging parts of their strings starting from a randomly chosen crossover point. This operator tends to enable to the evolutionary process to move toward promising regions of the search space. The mating convention for reproduction is such that only the high scoring members will preserve and propagate their worthy characteristics from generations to generation and thereby help in continuing the search for an optimal solution. The chromosomes with high performance may be chosen for replication several times whereas poor performing structures may not be chosen at all. Such a selective process ensures the best-performing structures may not be chosen for replication several times whereas poor performing chromosomes may not be chosen at all. Such a selective process ensures the best-performing chromosomes in the population to occupy an increasingly larger proportion of the population over time. The crossover forms a new offspring between two randomly selected ‘good parents’. The crossover operates by swapping corresponding segments of a string representation of the parents and extends the search for new solution in farther reaching direction. The crossover occurs only with some probability, which is called the crossover rate. There are many different types of crossover that can be performed: the one-point, the two-point, and the uniform type [3]. The mutation is a GA mechanism where we randomly choose a member of the population and change one randomly chosen bit in its bit string representation. Although the reproduction and the crossover produce many new strings, they do not introduce any new information into the population at the bit level. If the mutant member is feasible, it replaces the member, which was mutated in the population. The presence of mutation ensures that the probability of reaching any point in the search space is never zero. Mutation is used to search for further problem space and to avoid local convergence of GA. GA has been extensively researched and applied to many combinatorial optimization problems. Furthermore, GA has been increasingly applied in conjunction with other AI techniques such as NN and CBR. Various problems of neural network design have been optimized using GA. GA has also been used in conjunction with CBR to select relevant input variables and to tune the parameters of CBR (Figure 1).

3.1.1. Rule extraction using GA

Although numerous experimental studies reported the usefulness of NNs in classification studies, there is a major drawback in building and using a model in which the user cannot readily comprehend the final rules that NN models acquire. An advantage of present approach using GAs is that it is capable of extracting rules that are easy to understand for users like expert systems. In extracting bankruptcy rule, I use the similar approach that Bauer [38] and Mahmoud and Mani [39] suggest in their stock selection applications. I apply GA to find thresholds (cut-offs) for one or more variables, above or below which a company is considered ‘dangerous’. For instance, if the model’s structure consists of two variables representing a particular company’s quick ratio and a debt ratio, the final rule the GA returns might look like the following:

IF [Debt ratio > 1.50 and Quick ratio < 0.35] THEN Dangerous.

In many cases, the simplistic rule like the above example is insufficient to model relationships among financial variables. Our rule structure contains five conditions using ‘AND’ relations. The general form of the rule that GAs generate is as follows:

IF [The VAR1 is GREATER THAN OR EQUAL TO (LESS THAN) C1, AND the VAR2 is GREATER THAN OR EQUAL TO (LESS THAN) C2, AND ...
AND the VAR9 is GREATER THAN OR EQUAL TO (LESS THAN) C9] THEN Prediction is Dangerous.

If all of the nine conditions are satisfied, then the model will produce ‘dangerous’ signal for an evaluated company. C1 to C9 denote the cut-off values, which are found through genetic search process. The cutoff values range from zero to one, and represent the percentage of the data source’s range. This allows the rules to refer to any data source, regardless of the values it takes on. Above rule structure is summarized in Table 2. In the table, ‘which data’ means data source the rule refers to:

<table>
<thead>
<tr>
<th>Condition</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Which data</td>
<td>VAR1, VAR2, VAR3, VAR4, VAR5, VAR6, VAR7, VAR8, VAR9</td>
</tr>
<tr>
<td>Less than</td>
<td>0.35, 1.50, 0.35, 0.35, 0.35, 0.35, 0.35, 0.35, 0.35</td>
</tr>
<tr>
<td>Greater than</td>
<td>0.35, 1.50, 0.35, 0.35, 0.35, 0.35, 0.35, 0.35, 0.35</td>
</tr>
<tr>
<td>Cutoff value</td>
<td>C1, C2, C3, C4, C5, C6, C7, C8, C9</td>
</tr>
</tbody>
</table>

The string encoded for the experiments is as follows:

String [VAR1>VAR2>VAR3>VAR4>VAR5>VAR6>VAR7>VAR8>VAR9
L/G>1.50;L/G>1.50;L/G>1.50;L/G>1.50;L/G>1.50;L/G>1.50;L/G>1.50;L/G>1.50;L/G>1.50;C1>C2>C3>C4>C5>
C6>C7>C8>C9].

The GA maintains a population of strings which are chosen at random. This initialization allows the GAs to explore the range of all possible solutions, and this tends to favor the most likely solutions. Generally, the population size is determined according to the size of the problem, i.e. bigger population for larger problem. The common view is that a larger population takes longer to settle on a solution, but is more like-
ly to find a global optimum because of its more diverse gene pool. We use 100 strings in the population. The task of defining a fitness function is always application specific. In this study, the objective of the system is to find a rule which would yield the highest ratio if rules are fired across the company. Thus, we define the fitness function to be the ratios of the rule. The genetic operators such as crossover and mutation which are described in Section 3 are used to search for the optimal solutions. Several parameters must be defined for the above operators, and the values of these parameters can greatly influence the performance of the algorithm. The crossover rates range from 0.6 to 0.8 and the mutation rate ranges from 0.08 to 0.14 for our experiment. As a stopping condition, we use 3000 trials. These processes are done by the GAs toolbox in MATLAB software.

3.2. Proposed DEA model

This section, we introduce our methodology for predicting corporate failure and success based upon DEA. Specifically, the DEA model we based upon is the additive model of Charnes et al. [13]. Suppose we have a set of N DMU's (e.g., firms). Each DMU, i (i = 1,..., n) has m inputs and s outputs. \( x_i \) (i=1,..., m) and \( y_i \) (i=1,..., s), respectively. Then, the additive model for a specific DMU can be written as

\[
\rho_i^* = \max \rho_i \quad \text{s.t.} \quad \sum_{j=1}^{n} \lambda_j x_{ij} + s^*_m = x_i, \quad i = 1,2,\ldots,m \\
\sum_{j=1}^{n} \lambda_j y_{ij} - s^*_r = y_i, \quad r = 1,2,\ldots,s \\
\sum_{j=1}^{n} \lambda_j = 1 \\
\lambda_j, s^*_m, s^*_r \geq 0, \quad j = 1,2,\ldots,n; \quad i = 1,2,\ldots,m; \quad r = 1,2,\ldots,s
\]

(1)

Where \( S^*_m \) and \( S^*_r \) represent input and output slacks for DMU \( i \) under evaluation. A DMU is efficient or on the DEA frontier if and only if \( S^*_m = S^*_r = 0 \) at optimality. The additive DEA model (1) determines inefficiency in each input and each output in a single model. Unlike the standard radial DEA model of Charnes et al. [13], the model presented above the model presented above (1) does not yield an efficiency score in-between [0,1]. We, therefore, develop the following index as the efficiency score based upon model (1).

\[\text{Let an optimal solution to model (1). Then we can define.} \]

\[
\sigma^*_i = \frac{1 - (1/m) \sum_{j=1}^{m} \frac{1}{x_{ij}} + s^*_m}{1 + (1/s) \sum_{j=1}^{s} \frac{1}{y_{ij}} - s^*_r}
\]

(2)

As the additive efficiency score for DMU \( i \). It can be verified that \( \sigma^*_i \) falls between zero and one, and is unit-invariant and monotone decreasing in input/output slacks. DMU \( i \) is called additive efficient if and only if \( \sigma^*_i = 1 \), indicating that all optimal slacks are zero. In order to discriminate the performance of efficient DMU \( i \), we can employ the related super-efficiency model. However, as noted in [37], to obtain the super-efficiency of an efficient DMU under model (1), we cannot simply modify additive model (1) by removing DMU \( i \) from the reference set. If we do that, the resulting model may not have a feasible solution. Therefore, for an additive efficient DMU under model (1), we need to adopt the following super-efficiency model proposed by Du et al. [37]:

\[
\beta_i^* = \min \beta_i \quad \text{s.t.} \quad x_i + t_{im} \geq \sum_{j=1}^{n} \lambda_j x_{ij}, \quad i = 1,2,\ldots,m \\
y_i - t_{ir} \leq \sum_{j=1}^{n} \lambda_j y_{ij}, \quad r = 1,2,\ldots,s \\
\sum_{j=1}^{n} \lambda_j = 1 \lambda_j, t_{im}, t_{ir} \geq 0, \\
j = 1,2,\ldots,n; \quad i = 1,2,\ldots,m; \quad r = 1,2,\ldots,s
\]

(2)

It can be seen that after DMU \( i \) is removed from the reference set of model (1), it is removed from the reference set of model (1), we need to modify the constraints and the objective function of model (1). The constraints should be modified because we need to increase the inputs and decrease the outputs for DMU \( i \) to reach the frontier constructed by the remaining DMU. We change the objective from maximization to minimization, so that the resulting model is bounded. We divide each slack by its corresponding input/output in the objective to make the resulting model unit invariant. Let

\[\{f^*_i, x^*_i, j = 1,2,\ldots,n; \quad f^*_i = 0; \quad f^*_i, x^*_i, j = 1,2,\ldots,m; \quad t^*_i, r = 1,2,\ldots,s\} \]

be an optimal solution to model (2). Then we can define

\[
\sigma^*_i = \frac{(1/m) \sum_{j=1}^{m} f^*_i + x^*_i}{(1/s) \sum_{j=1}^{s} f^*_i + y^*_i} \geq 1
\]

As the additive super-efficiency score for an efficient DMU. Using the same set of inputs and outputs and the above proposed DEA approach, the current study identifies two frontiers: one for failure assessment and the other for success assessment. For example, as in Premachandra et al. [8] bankruptcy assessment, the smaller values in the financial ratios, which could possibly cause financial distress, are considered to be input variables, and the larger values in those ratios, which could cause financial distress, are classified as output variables. This input-output classification will identify the failure frontier, and indicate those firms which are about to fail and tend to have an efficiency score equal to or close to one. In contrast, if we swap the inputs and outputs, namely, the larger values in those financial ratios are classified as inputs and smaller values are classified as outputs, we identify the frontier for firm success. Under the context of bankruptcy, the failure frontier is the bankruptcy frontier, and the success frontier is the non-bankruptcy frontier. The current study will then construct an assessment index based upon the two frontiers, and further demonstrate that the newly constructed index (discriminant index) improves upon the standard DEA approach used in Premachandra et al. [8]. Denote the DEA score (from models (1) and (2)) for identifying the failure frontier as \( \theta^f \) and the corresponding score for identifying the success frontier as \( \theta^s \). Namely, \( \theta^f \) is associated with the bankruptcy (failure) frontier model, and \( \theta^s \) is associated with the non-bankruptcy (success) frontier model. We then define our prediction or assessment index as

\[
\lambda \theta^f - (1-\lambda) \theta^s \geq 0
\]

(3)

Where \( \lambda \) is a user-specified weight reflecting the relative emphasis on the two frontiers. Note that negative \( \theta^s \) used in (3), as one is a bankruptcy frontier and the other is a success frontier. For \( \theta^f \) and \( \theta^s \) we used normalized values, as the skewness of the distributions of the original values of \( \theta^f \) and \( \theta^s \) is substantially different. Specifically, when a DMU is inefficient, \( \theta^f \) represents the efficiency score.
Based on model (1). When a DMU is efficient under model (1), then $\theta_i$ represents the super-efficiency score

$$\sigma^* = \frac{1}{(1/n) \sum_{i=1}^{n} \left( \frac{1}{x_{io}} + \frac{t_{io}}{x_{io}} \right) / x_{io}}$$

Based upon model (2). $\theta_2$ is obtained in the same manner. The difference between $\theta_1$ and $\theta_2$ lies in the fact that different sets of inputs and outputs are related to $\theta_1$ and $\theta_2$, the inputs and outputs related to $\theta_1$ and $\theta_2$ are defined and discussed in the follow.

### 4. Data

Following Premachandra et al. [8], we use the same nine financial variables that proxy for the financial strength/weakness and potential insolvency of firms. These financial ratios have commonly been used in past bankruptcy literature and are considered to be the most efficient ones. These variables include measures of liquidity, market-based growth, leverage, and variables related to short-term debt obligations. The financial ratios which need to be maximized are considered to be output variables in DEA model, and those to be minimized are considered to be input variables in the bankruptcy (failure) frontier model, and the corresponding efficiency score is denoted by $\theta^*$. The non-bankruptcy (success) frontier DEA model is constructed by simply swapping the input and output variables of the bankruptcy frontier model. Let the efficiency score of the non-bankruptcy frontier DEA model be $\theta^*$. In GA model, the data set is split into two subsets, a training set and a validation (holdout) set of 90% and 10% of entire data, respectively. The training data are used for learning rules, and the validation data which have not been used to develop the system are used to test the results. Note that there is not input and output in GA model. This variables show in table 3.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Abbreviation</th>
<th>DEA Input</th>
<th>DEA Output</th>
<th>GA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total debt/total assets</td>
<td>TDTA</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Current liabilities/total assets</td>
<td>CLTA</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Cash flow/total assets</td>
<td>CFTA</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Net income/total assets</td>
<td>NITA</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Working capital/total assets</td>
<td>WCTA</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Current assets/total assets</td>
<td>CATA</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Earnings before interest and taxes/total assets</td>
<td>EBTA</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Market value of equity/book value of common equity</td>
<td>MVCE</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Variables of Research

TDTA indicates the long-term financial obligations of a firm. An increase in this variable would lead to higher probability of financial distress. A firm with high CLTA indicates that the firm does not have sufficient cash flow to continue its day-to-day operations, i.e., such firms would find it difficult to fulfill short-term debt obligations and therefore would become financially distressed. The following input variables are considered in such a way that smaller values for these variables would result in a firm to become more probable for bankruptcy. It can be seen that the following input variables are functions of liquidity measures, such as net income, working capital, current assets, cash flows, operating earnings, etc. Firms with smaller values for such liquidity measures are more probable for bankruptcy. The above inputs and outputs are used in models (1) and (2) to compute the efficiency score $\theta_1$ and the corresponding bankruptcy frontier. To obtain $\theta_2$ and identify the success frontier, we use TDTA and CLTA as the two inputs. Firms with small TDTA and CLTA are less likely to go bankrupt and would appear close to the success frontier. The inputs of the bankruptcy frontier model are used as outputs in the success frontier model, i.e., in this case, large values on CFTA, NITA, WCTA, CATA, EBTA, EBIE, and MVCE indicate better financial performance of a firm. According to the information and data available in library of Tehran Stock Exchange, in that span of time only 70 companies were in accordance with article 141 of Business law over the period 2004-2010 across a full spectrum of industries. To compare with bankrupt companies, some 70 non-bankrupt companies were also chosen via random sampling.

### 5. Results and Discussion

In this section, I will present the result of GA and DEA in prediction of bankruptcy in bankrupt and non-bankrupt firms. I use the follow phrases to show the result in the models:

(i) $P(PBR/BR)$: percentage of bankrupt firms predicted as bankrupt.

(ii) $P(PNBR/NBR)$: percentage of non-bankrupt firms predicted as non-bankrupt.

(iii) Percentage of total correct predictions: the mean of (i) and (ii).

#### 5.1. GA result

Our genetic search process finally extracts one bankruptcy prediction rule. This rule generated and the corresponding description is illustrated in Table 4.

<table>
<thead>
<tr>
<th>Variable code</th>
<th>Con 1</th>
<th>Con 2</th>
<th>Con 3</th>
<th>Con 4</th>
<th>Con 5</th>
<th>Con 6</th>
<th>Con 7</th>
<th>Con 8</th>
<th>Con 9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cutoff</td>
<td>0.8746</td>
<td>0.595</td>
<td>0.082</td>
<td>0.4261</td>
<td>0.3112</td>
<td>0.3104</td>
<td>0.631</td>
<td>0.6653</td>
<td>0.4535</td>
</tr>
<tr>
<td>Description</td>
<td>IF current liabilities to total assets greater than or equal to 0.8746 AND earnings before interest and taxes to total assets greater than or equal to 0.595 AND current assets to total assets less than or equal to 0.082 AND market value of equity to book value of common equity less than or equal to 0.4261 AND total debt to assets greater than or equal to 0.3112 AND cash flow to total assets less than or equal to 0.3104 AND net income to total assets less than or equal to 0.631 AND earnings before interest and taxes to interest expense greater than or equal to 0.6653 AND working capital to total assets less than or equal to 0.4535</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4. The description of rule generated

The general goal in optimization is to find the best solution to a problem. Since GAs try to find out the optimal or near optimal combination of above searching parameters, the final solution is one. The average ratio if the rules are fired is 88.2% of training and validation sets, respectively. This means if the financial variables of a company are within the feature ranges of derived rules, the probability of bankruptcy is about 80% of cases. Table 5 show the result of GA in training set and validation set.

<table>
<thead>
<tr>
<th>Samples&lt;sup&gt;*&lt;/sup&gt;</th>
<th>Real group membership</th>
<th>Predicted group GA model</th>
<th>0</th>
<th>1</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training Set</td>
<td>1 (Count)</td>
<td>14</td>
<td>56</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0 (Count)</td>
<td>68</td>
<td>2</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 (%)</td>
<td>20</td>
<td>80</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0 (%)</td>
<td>97.1</td>
<td>2.9</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Validation Set</td>
<td>1 (Count)</td>
<td>16</td>
<td>54</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0 (Count)</td>
<td>69</td>
<td>1</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 (%)</td>
<td>22.8</td>
<td>77.2</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0 (%)</td>
<td>98.6</td>
<td>1.4</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

<sup>*</sup>Group 1: bankrupt firms. Group 0: non-bankrupt firms.

Table 5. Test result of GA model

#### 5.2. DEA result

Super-efficiency and additive efficiency scores are calculated for the 140 firms in the sample using models (1) and (2), and then the proposed assessment index for each firm is calculated using (3). Then the following percentages of correct predictions are obtained for a cutoff point. First, the firms in the sample were arranged in the descending order of the super-efficiency values from the DEA model, and the firms whose super-efficiency values are greater than or equal to the cutoff point are considered to be predicted as bankrupt firms and firms whose efficiency scores are below this cutoff point are considered to be predicted as non-bankrupt firms. The actual number of bankrupt and non-bankrupt firms above and below this cutoff point is also counted. Next, the firms were arranged in the descending order of the value of the assessment index. In logistic regression, a cutoff point of 0.5 is traditionally used to classify bankrupt and non-bankrupt firms, with a cutoff point of 0.5 one can achieve 61.4% of BR, 88.6% of NBR, and 75.7% of total correct predictions. The table 6 shows the compare of GA and DEA results in bankruptcy prediction.
6. Conclusion

Bankruptcy is a highly significant worldwide problem that affects the economic well-being of all countries. The high social costs incurred by various stakeholders associated with bankrupt firms have spurred searches for better theoretical understanding and prediction capability. Recently genetic algorithm has been applied to the problem of bankruptcy prediction. The GA-based model has been compared with other methods such as the neural network and logistic regression, and has shown good results. This paper revisits the use of GA and DEA in corporate failure assessment and improves upon the work of Garkaz and Abdollahi, and Premachandra et al. The findings show that the models can be used in Tehran Stock Exchange and the classification performance of the GA model is significantly higher than that of DEA model. The result of DEA showed that since the traditional cutoff point of 0.5 may not be appropriate in classifying the firms as bankrupt or non-bankrupt based on the standard DEA efficiency scores, I propose to use a newly developed additive super-efficiency model. Experimental results indicate that the super-efficiency DEA and GA models overall is weaker in predicting bankrupt firms correctly compared to non-bankrupt firms. The assessment index based on two frontiers improves this weakness. It is suggested that the genetic algorithm and support vector machines be used with caution in Iran.

Table 6. Test result of DEA model

<table>
<thead>
<tr>
<th>Real group membership</th>
<th>Predicted group GA model</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Sample 1 (Count)</td>
<td>27 43 70</td>
</tr>
<tr>
<td>0 (Count)</td>
<td>62 8 70</td>
</tr>
<tr>
<td>1 (%)</td>
<td>38.6 61.4 100</td>
</tr>
<tr>
<td>0 (%)</td>
<td>88.6 11.4 100</td>
</tr>
</tbody>
</table>

*Group 1: bankrupt firms; Group 0: non-bankrupt firms.

References


