

Application of Genetic Algorithm in Development of Bankruptcy Predication Theory Case Study: Companies Listed on Tehran Stock Exchange

Mohsen Hajiamiri*

Department of Industrial Engineering,
Zahedan Branch, Islamic Azad University,
Zahedan, Iran

Mohammad Reza Shahraki

Department of Industrial Engineering,
University of Sistan and Baluchistan,
Zahedan, Iran

Seyyed Masoud Barakati

University of Sistan and Baluchistan,
Zahedan, Iran
smbaraka@ece.usb.ac.ir

Abstract. The bankruptcy prediction models have long been proposed as a key subject in finance. The present study, therefore, makes an effort to examine the corporate bankruptcy prediction through employment of the genetic algorithm model. Furthermore, it attempts to evaluate the strategies to overcome the drawbacks of ordinary methods for bankruptcy prediction through application of genetic algorithms. The sample under investigation in this research includes 70 pairs of bankrupt and non-bankrupt companies during 2001-2011. Having examined the obtained data from financial statements of the companies under study, 5 financial independent variables were identified so as to be used in the model. The results indicated that employment of genetic algorithm in

Received: February (2014); Final Revision: May (2014)

*Corresponding author

predicting financial bankruptcy is highly effective, to the extent it managed to correctly predict the financial bankruptcy of companies two years before the base year, one year before the base year and the base year at accuracies of 96.44, 97.94 and 95.53, respectively.

Keywords: Bankruptcy; bankruptcy prediction; multiple discriminant analysis; logistic regression; neural networks; genetic algorithm.

1. Introduction

Prediction is important in many aspects of life, i.e. any weak prediction can lead to inefficient decisions. In fact, any planning, decision-making and other key tasks associated with managers would face failure without proper predictions. Generally, the purpose of prediction is to reduce risk in decision-making. Since prediction cannot be completely eliminated, it is essential for the decision-making process to explicitly consider the results of remaining uncertainties in prediction [1].

From the perspective of macroeconomic theories, the level of economic development of a society correlates with the level of investments made in it. If such investments are not made in the right opportunities or directed in inefficient ways, the national economy will be damaged [14].

One of the strategies to assist investors is to offer prediction models about the financial status of companies. The closer predictions are to reality, the more appropriate the basis of decisions become. The bankruptcy prediction models are regarded as a tool for estimating the future performance of companies. Investors and creditors extremely tend to predict the bankruptcy of businesses or else, great costs are imposed on them. There are advantages and disadvantages to employment of each prediction model [3].

Selection of a model according to the consumer needs for financial information and their environmental circumstances is complicated. In fact, national wealth can be preserved in the form of physical and human capital if the probability of corporate bankruptcy in businesses is correctly predicted and the corporate affairs are adjusted through detecting the problems to be solved. Furthermore, such model can provide an ideal

guideline for financial decision-makers, i.e. investment firms, banks and government [8].

2. Problem Statement

The growing competition among businesses has restricted the chances of gaining profits while making bankruptcy more probable. Decision-making on financial issues has always involved risk and uncertainty. One of the strategies to assist investors is to offer prediction models about the overall outlook of companies. The closer predictions are to reality, the more appropriate the basis of decisions become. As Beaver argues: “Prediction is possible without decision-making. The smallest decisions, however, cannot be made without prediction” (Beaver, 1996). As one of the strategies for predicting the future status of companies, a bankruptcy prediction model estimates the probability of bankruptcy through combining a group of financial ratios. Being a telltale sign of misallocation of resources, the ability to financially and commercially predict is considered vital from the perspective of private investors and also from social perspective. The early warning of possible bankruptcy enables managers and investors to take preventive measures and distinguish desirable investment opportunities from those undesirable [11].

In Tehran Stock Exchange, the criterion for bankruptcy and removal of companies from the list is Article 141 of the Iranian commercial code, which states: “If at least half of the capital of a company is lost due to the caused damages, the board of directors shall immediately call on the stockholders for an extraordinary general meeting in order to discuss whether the company should be dissolved or continue operating. In case the board does not vote for dissolution, the corporate capital shall, at the same meeting under Article 6 of the mentioned law, be reduced to sum of the currently available capital. In case the board of directors, against the provisions of this Article refuses to call for the extraordinary general meeting or the members invited fail to gather, the stakeholders can individually request the dissolution from a competent court of law”. [6].

3. Background

The first studies leading to development of a model for bankruptcy prediction were those done by William Beaver back in 1966. In his studies, Beaver concluded that validity of a financial ratio reflects the degree of its success in classifying a company as bankrupted or non-bankrupted, i.e. the low error of classification represents a higher validity of a ratio. In 78% of the cases, Beaver's model was able to correctly predict financial bankruptcies five years in advance [4].

In subsequent studies, great focus was turned on to multivariable models. In 1968, Edward Altman began his study on devising a comprehensive multivariable model. Selecting a total of 22 financial ratios and analyzing them statistically through multiple detection analysis, he proposed Z-Score composed of 5 financial ratios. The accuracy of Altman's model for one year before bankruptcy and two years before bankruptcy were approximately 95% and 83%, respectively [3].

As one of the first researchers using the Logit Analysis, Martin conducted a study in 1977 selecting as sample a total of 577 non-bankrupted banks and 58 bankrupted during 1970-1976. In his final analysis, Martin employed six variables. The accuracy of his model for bankrupted firms and non-bankrupted firms was 87-96% and 89%, respectively [10].

Another method for examining the bankruptcy prediction models is neural networks, employed first by Odom & Sharda in 1990 for designing their models. The findings suggested that neural networks were more accurate and powerful in comparison with the multiple discriminant analysis [13].

In 2000, Shah & Mortaza proposed a model using the neural networks. In this study, the data from 60 bankrupted and 54 non-bankrupted firms during 1992-1994 were used. Shah & Mortaza employed eight financial ratios, which had been selected based on previous studies and consultation with financial experts. The accuracy of prediction in this model was estimated to be 73% [15].

In 2004, Wallace designed a mode using the neural networks method, in which the values of key financial ratios previously reported in bankruptcy studies as best ratios were employed. The overall accuracy of Wallace's

model was 94% evaluating 65 different financial ratios in the previous studies [17].

It should be noted, however, additional studies have been done for comparing various neural networks. In 2010, for instance, Khashman A. compared different neural networks in order to predict credit risk. In his research, Khashman put different structures of neural networks into comparison [7].

One technique employed in analysis of financial crisis is genetic algorithm. In 1998, Varetto was the first scholar strictly employing the genetic algorithm for bankruptcy prediction. The sample in his study consisted of 500 firms; a total of 236 bankrupted and 264 non-bankrupted firms. The results suggested an accuracy of 93% for one year before bankruptcy and 91.6% for two years before bankruptcy [16].

In 2006, Mein et al simultaneously employed the genetic algorithm and support vector machine, dubbing it GA-SVM. The results of their study indicated an accuracy of 86.53% for the training set one year before bankruptcy [12].

In a 2006 study using genetic programming, Lensberg identified 6 out of 28 potential variables of bankruptcy previously examined as significant [9].

4. Methodology

In the present study, the genetic algorithm model was employed so as to offer a solution to tackle weaknesses of ordinary methods for bankruptcy prediction. In terms of methodology, therefore, it is a mathematical-analytic research. Moreover, it can be regarded as a case study in terms of research type and developmental in terms of objective.

There are several fundamental steps taken in conducting the present study as below:

- 1-Identification of financial ratios in order to predict bankruptcy.
- 2-Calculation of financial ratios and other required parameters as independent variables used in the tested model.
- 3-Classification of firms into bankrupted and non-bankrupted under Article 141 of the Iranian commercial Code.

4-Evaluation of accuracy in the genetic algorithm prediction model for bankruptcy prediction. The statistical population in this study includes the entire companies listed on Tehran Stock Exchange (TSE) during 2001-2011. The quality and accessibility of information regarding financial statements were two factors contributing to selection of such a population.

The statistical population was divided into two categories, the first of which covers bankrupted companies. The criterion taken into account for bankruptcy was Article 141 of the Iranian Commercial code. The second category consisted of survived company's not encountered bankruptcy. For data collection, effort was made to select non-bankrupted companies similar in terms of industry and size, except for cases it was impossible due to extremely small industry size, which can be regarded as one of the study's restrictions. Since financial information used regarding each company covers two years prior to bankruptcy, it can generally be stated that corporate information between 1999 and 2009 has been employed. The base year (t) regarding bankrupted companies refers to the year at which a company faces financial crisis or bankruptcy. With regard to non-bankrupted companies, the base year refers to the year at which information from two previous years have been collected.

Throughout the examinations done in the present study, a total of 82 companies over the defined period were subject to the mentioned law. A few of companies, however, were different from other samples in terms of financial ratios, which led to poor performance and accuracy of prediction models. Consequently, a number of bankrupted companies and their selected pairs were removed from the research reducing the count of remaining pairs used in the study to 70 (i.e. 70 bankrupted companies and 70 non-bankrupted companies).

From a systematic viewpoint, it is highly essential to have appropriately valid inputs in order to achieve the right outputs. Since accurate and correct information was required more than anything else, the data regarding the tested financial ratios were obtained from the public archive of TSE financial statements in CD form as well as a software called Rahavard-e-Novin.

In order to refine the collected data from financial statements of the

sample companies, MS Excel was employed. Then SPS was used to statistically analyze the refined information. Furthermore, MATLAB was employed since there were non-linear relationships among the financial data and the objective was to predict the bankruptcy of companies listed on TSE.

Since these financial ratios have been widely employed in previous studies, a total of 15 basic ratios as telltale signs of bankruptcy in a few previously proposed models for financial bankruptcy prediction were selected taking into account the fact that ratios were gathered from every major analytical perspective such as liquidity, profitability, liquidation, etc.

The initial analysis of variable was conducted through 7 computer operations on genetic programming algorithm. For each operation, the results of variable were reported after every 4000 periods. This led to 50 reports which consisted of 15 variables examined in order to determine whether or not it contribute to classification capability of the best program at the operation time. If the variable left non-zero impact on the classification capability of the best program, 1 value was added to it. In other conditions, it received 0 values.

Finally, six financial ratios were used as independent variables enumerated below:

- 1) Immediate ratio, i.e. immediate assets divided by current debt.
- 2) Debt ratio, i.e. total debts divided by total assets.
- 3) Return on assets ratio, i.e. net profit divided by total assets.
- 4) Profit to revenue ratio, i.e. profit divided by total income.
- 5) Gross profit ratio, i.e. gross profit divided by total income.
- 6) Shareholder's return on equity, i.e. net profit divided by shareholder's equity.

The genetic programming algorithm is a technique allowing the researcher to find a solution to problem without the need to predetermine the model. It implies that solution can be any model mathematically describable. The purpose is to allow the data to as much as possible represent the facts, so that minimize the level of previous structure offering functional forms and statistical methods of selection [9].

Basically, the genetic programming algorithm is supposed to take the

structure and function when it receives the relevant corporate data, so as to be able to make decisions regarding progress or non-progress of the company toward bankruptcy. In short, genetic programming should be able to resolve the divisions of samples sorting into two categories. The first category consists of companies that will get bankrupted and the second category includes companies that will remain profitable. In this procedure, the data was initially needed to be divided into two groups; training and testing.

5. Training and Testing Sets

In order to apply the genetic programming algorithm to prediction problems, the data sets are divided into two subgroups, i.e. the training set and the testing set both selected randomly. In case the database is imbalanced (i.e. the low ratio of bankrupted companies to the non-bankrupted), it will be necessary to take such ratio into account for selection of the training set. In order to divide the data into training and testing sets, 50 pairs of bankrupted and non-bankrupted firms were selected to be examined for training the model. Rest of the data included a 20 pairs of bankrupted and non-bankrupted firms dedicated to the network testing.

The sorting is done as follow: Assuming $X = x_0, x_N$ which covers the corporate information. $f(x)$ is a function defined by single tree structure in genetic programming.

The y value from $f(x)$ is determined based on input vector X .

$$y = f(x_0, x_1 \dots x_N) \quad (1)$$

Where X can be inserted as an input for the genetic programming tree, so as to calculate the y output. The results of sorting will be as below:

$$y > 0, x \in \quad (2)$$

$$y \leq 0, x \in \quad (3)$$

B represents the set including bankrupted companies while represents the set including the profitable companies.

In other words, if the evaluation of genetic programming tree leads to a numerical value higher than zero, then the examined company falls under the category of companies progressing toward bankruptcy. If the value is lower than or equal to zero, the company falls under the category of profitable.

Since these financial ratios have been widely employed in the employed database might be extremely imbalanced that only 5 to 6% of the available companies are bankrupted, which should be taken into account in order to design the fitness function. Otherwise, the assessment might turn into a convergent structure sorting the entire companies as profitable. In fact, they are not sorted from the first place and the obtained success rate becomes highly favorable. There are three ways to tackle such issue:

- Under sampling of the larger set
- Oversampling the smaller set
- Change in value (weight) regarding missing of the positive and negative set to compensate for the imbalanced ratio. For instance, if the imbalanced ratio of 1/10 is in favor of the negative set, then the outcome of sorting the positive sample should be 10 times greater [2].

Therefore, the fitness function can be formulated as below:

$$fitness = \sum_{i=1}^n u_i$$

where $u =$

0: incorrect sorting

1: Correctly sorted bankrupted companies

$\frac{n_b=0}{n_b=1}$: Correctly sorted profitable companies

$n_b = 0$ Is the number of bankrupted companies in the training set, while $n_b = 1$ is the number of profitable companies in the training set.

Table 1 illustrates the major parameters taken into account for assessment.

Table 1. GP parameters

Ramped half and half	Initial method
A generation with elitism (0.2%)	Assignment operator
Tournament selection	Selection operator
10	Size of the tournament group
0.05	Cloning rate
Tree incoherent integration	Incoherent integration
0.9	Internal node selection rate
0.5	Integration rate
0.45	Integration rate for volume control
7	Tree initial maximum depth
18	Tree maximum depth
500	Size of population
20	Number of operations
50 generations	Criterion of initiation and termination

Table 2. The results obtained from examining each year

Status	Training			Testing			Combination		
	Success (%)	TP (%)	TN (%)	Success (%)	TP (%)	TN (%)	Success (%)	TP (%)	TN (%)
Year T-2	98.12	99.00	88.00	94.92	96.47	84.83	96.44	97.48	84.32
Year t-1	99.69	100.00	89.10	96.94	98.59	83.31	97.94	98.04	85.28
Year T	97.50	98.50	86.40	93.04	96.00	82.57	95.53	96.67	83.72

The above table suggests the result obtained from GP through training was highly desirable. The best GP structure achieved a successful percentage of approximately 99.7%. Furthermore, the best GP structure in the testing set achieved a successful percentage of approximately 97%. In the model designed through genetic programming algorithm, a successful percentage of approximately 98% was achieved, indicating there is a high potential of genetic algorithm in prediction the bankruptcy of companies listed on Tehran Stock Exchange.

6. Discussion and Conclusion

Financial bankruptcy is a crucial issue affecting the economies throughout the world. The extravagant social costs suffered by various stakeholders in connection with bankrupted companies leads to an inquiry for empowerment of prediction and better understanding of this theory.

The major problems tackled in the present study were imbalance between the number of companies progressing toward bankruptcy and the number of profitable companies as well as the amount of unavailable information in the database used for analysis. The approach adopted for solving this problem was normalization of data and employment of a fitting function solving the imbalance problem. The obtained results were highly favorable. As it was mentioned earlier, the best GP structure achieved successful percentages of approximately 99.7 and 97 in the training set and the testing set, respectively.

The results obtained from examining each year has been shown in Table 2, the first row of which indicates the results of two years before the base year, the second row indicates the results of one year before the base year, and finally the third row indicates the results of the base year. Each table illustrates the obtained results from training, testing and the combination. The first column shows the percentage of achieved successes (i.e. number of correct predictions), the second column shows the percentage of true positives (TP, i.e. the number of companies correctly sorted as bankrupted), and the third column shows the true negatives (TN, i.e. the number of companies correctly sorted as non-bankrupted).

References

- [1] Kurdestani, A. M. (1996-97), "Profitability used for predicting the cash flow and future profits", *Journal of Accounting and Auditing Reviews* 18 & 19, P42-55.
- [2] Alfaro, E. and Sharman, k. (2007), "A Genetic Programming Approach for Bankruptcy Prediction Using a Highly Unbalanced Database". *European Journal of Evolutionary Computing*, 93, 132-143.

- [3] Altman, E. (2000), "Predicting Financial Distress of Companies". Retrieved on September 4th, working paper.
- [4] Beaver, W. (1996), "Financial Ratios as Predictors of Failure". *Journal of Accounting Research*, 666-16.
- [5] Dimitras, A., Zanakis, S., and Zopudinis, C. (1996), "A survey of business failures with an emphasis on failure prediction methods and industrial applications". *European Journal of Operational Research*, 90 (3), 487-513.
- [6] Jahangir, M. (2000), "Commercial Code with Cheques act. The amended registration regulation of non-commercial organizations", Tehran, Didar Publications.
- [7] Khashman, A. (2010), "Neural networks for credit risk evaluation: Investigation of different neural models and learning schemes". *Expert Systems with Applications*, (37), 6233-6239.
- [8] Lee, K. C., Han, I., and Kwon, Y. (1996), "Hybrid neural network models for bankruptcy predictions, *Decision Support Systems*", (18), 63-72.
- [9] Lensberg, T., Eilifsen, A., and McKee, T. E. (2006), "Bankruptcy theory development and classification via genetic program". *European Journal of operational research*, 169, 677-697.
- [10] Martin, D. (1977), "Early warning of bank failures: A logit regression approach". *Journal of Banking and Finance*, 1, 249-276.
- [11] Mehrani S., Bahramfar, N., and Ghayur, F. (2005), "A Study on the Correlation between the Traditional Liquidity Ratios and Ratios of Cash Flow Statement for Assessing the Continuity of Corporate Activities", *Journal of Accounting and Auditing Reviews*, 40, 3-17.
- [12] Min, S. H., Lee, J., and Han, I. (2006), "Hybrid genetic algorithms and support vector machines for bankruptcy prediction". *Expert systems with applications*, 31: 652-660.
- [13] Odom, M. D. and Sharda, R. (1990), "A Neural Network Model for Bankruptcy Prediction". *IJCNN International Joint Conference on Neural Networks*. San Diego: CA, 2: 163-168.
- [14] Reese, W. (1995), "Financial Analysis (second ed.)". London: Prentice Hall.

- [15] Shah, J. and Murtaza, M. (2000), "A neural network based clustering procedure for bankruptcy prediction". *American Business Review*, 18 (2), 80-86.
- [16] Varetto, F. (1998), "Genetic Algorithms application in the analysis of insolvency risk".
- [17] Wallace Wanda, A. (2004), "Risk assessment by internal auditors using past research on bankruptcy applying bankruptcy models".