A genetic programming model for bankruptcy prediction: Empirical evidence from Iran

Hossein Etemadi, Ali Asghar Anvary Rostamy*, Hassan Farajzadeh Dehkordi

Tarbiat Modares University (TMU), No. 12, Shahid Rahnama Alley, Saidi Street, Lavasani Avenue, P.O. Box 19546, Tehran, Iran

Abstract

Prediction of corporate bankruptcy is a phenomenon of increasing interest to investors/creditors, borrowing firms, and governments alike. Timely identification of firms’ impending failure is indeed desirable. By this time, several methods have been used for predicting bankruptcy but some of them suffer from underlying shortcomings. In recent years, Genetic Programming (GP) has reached great attention in academic and empirical fields for efficient solving high complex problems. GP is a technique for programming computers by means of natural selection. It is a variant of the genetic algorithm, which is based on the concept of adaptive survival in natural organisms. In this study, we investigated application of GP for bankruptcy prediction modeling. GP was applied to classify 144 bankrupt and non-bankrupt Iranian firms listed in Tehran stock exchange (TSE). Then a multiple discriminant analysis (MDA) was used to benchmarking GP model. Genetic model achieved 94% and 90% accuracy rates in training and holdout samples, respectively; while MDA model achieved only 77% and 73% accuracy rates in training and holdout samples, respectively. McNemar test showed that GP approach outperforms MDA to the problem of corporate bankruptcy prediction.

© 2008 Elsevier Ltd. All rights reserved.

Keywords: Bankruptcy prediction; Financial ratios; Genetic Programming; Multiple discriminant analysis; Iranian companies

1. Introduction

Corporate bankruptcy is a very important economic phenomenon. The health and success of the firms are of widespread concern to policy makers, industry participants, investors, and managers (O’Leary, 1998). It also is a problem that affects the economy of every country. The number of failing firms is important for the economy of a country and it can be considered as an index of the development and robustness of the economy (Zopounidis & Dimitras, 1998). The high individual, economic, and social costs encountered in corporate failures or bankruptcies have spurred searches for better understanding and prediction capability (McKee & Lensberg, 2002).

Prediction of corporate bankruptcy is a phenomenon of increasing interest to investors/creditors, borrowing firms, and governments alike. Timely identification of firms’ impending failure is indeed desirable (Jones, 1987). By this time, several methods have been used for predicting bankruptcy. Early research focused primarily on univariate models such as individual financial ratios. Among these studies Beaver (1966) is more noticeable than the others. He introduced a univariate technique for the classification of firms in two groups using some financial ratios. The ratios were used individually and a cut-off score was calculated for each ratio on the basis of minimizing misclassification. The univariate methods were later criticized, in spite of its considerable results, because of the correlation among ratios and providing different signals for a firm by ratios (Dimitras, Zanakis, & Zopounidis, 1996).

Later research turned to multivariate models. Researchers found that corporate bankruptcy can be affected by many different factors at the same time. Altman (1968) introduced a multivariate technique, multiple discriminant
analysis (MDA), for failure prediction. Because this study made use of more than one variable for bankruptcy prediction and applied an advanced statistical technique for determining the relationship among predictor variables, it was of much interest.

MDA provides good predictions but suffers from some limitations. Hence, variety methods introduced to overcome MDA shortcomings and improving accuracy. These methods can be grouped in two categories: statistical and artificial intelligence models. First group consists of Logit (Foreman, 2002; Ohlson, 1980; Zavgren, 1985), Probit (Casey, McGee, & Stinkey, 1986; Theodossiou, 1991), Linear Probability (Stone & Rasp, 1991; Vranas, 1992) Cumulative Sums (Kahya & Theodossiou, 1999), and etc. Neural Networks (Altman, Marco, & Varetto, 1994; Coats & Fant, 1993; Jo, Han, & Lee 1997), Genetic Algorithms (Shin & Lee, 2002; Varetto, 1998), Case Based Reasoning (Park & Han, 2002), Rough Sets (Dimitras, Slowinski, Susmaga, & Zopounidis 1999; McKee & Lensberg, 2002), Support Vector Machine (Min & Lee, 2005), and etc, constitute second group. Some of these models have high predictive accuracy levels but because of absence bankruptcy theory, attempts to establish a generally accepted model for bankruptcy prediction are not successful. Some studies have provided comprehensive surveys on bankruptcy prediction methods such as Dimitras et al. (1996), Jones (1987), and Kumar and Ravi 2007.

A common approach to bankruptcy prediction is to review the literature to identify a large set of potential predictive financial and/or non-financial variables and then develop a reduced set of variables, through some combination of judgmental and mathematical analysis that will predict bankruptcy (Lensberg, Eilifsen, & McKee 2006). In this study we implemented such approach for variables selection stage. After this, a relatively new technique for bankruptcy prediction, Genetic programming, constructed an accurate classification model for bankruptcy prediction. This model was benchmarked with the MDA, the most common used classification model for bankruptcy prediction. In the rest of the paper, first, we discuss about GP and MDA, two techniques were used for bankruptcy prediction modeling. In the section 4 we explain variable selection process and after that models development, empirical results and conclusion will be discussed.

2. Genetic programming

Genetic programming (GP) is a search methodology belonging to the family of evolutionary computation (EC), GP can be considered an extension of Genetic algorithms (GAs) (Koza, 1992). GAs is stochastic search techniques that can search large and complicated spaces on the ideas from natural genetics and evolutionary principle (Goldberg, 1989; Holland, 1975). They have been demonstrated to be effective and robust in searching very large spaces in a wide range of applications (Colin, 1994; Shin & Han, 1999). GAs has been applied in wide range finan-

Fig. 1. Tree representation of the program (expression): \((X \cdot Y) + 6 - (Z/8)\).

Fig. 2 and 3 show an example of crossover operator.

2.2. Crossover

Two parent individuals are selected and a subtree is picked on each one. Then crossover swaps the nodes and their relative sub-trees from one parent to the other. This operator must ensure the respect of the depth limits. If a condition is violated the too-large offspring is simply replaced by one of the parents. There are other parameters that specify the frequency with which internal or external points are selected as crossover points. Figs. 2 and 3 show an example of crossover operator.

2.3. Mutation

The mutation operator can be applied to either a function node or a terminal node. A node in the tree is
randomly selected. If the chosen node is a terminal it is simply replaced by another terminal. If it is a function and point mutation is to be performed, it is replaced by a new function with the same arity. If, instead, tree mutation is to be carried out, a new function node (not necessarily with the same arity) is chosen, and the original node together with its relative subtree is substituted by a new randomly generated subtree. A depth ramp is used to set bounds on size when generating the replacement subtree. Naturally it is to check that this replacement does not violate the depth limit. If this happens mutation just reproduces the original tree into the new generation. Further parameters specify the probability with which internal or external points are selected as mutation points. An example of mutation operator is shown in Fig. 4.

The major steps of this genetic programming can be formalized as follows:

1. Generate at random an initial population of rules representing potential solutions to the prediction of the bankrupt and non-bankrupt firms classes.
2. Evaluate each rule on the training set by means of a fitness function.
3. Select the rules to undergo the mechanism of reproduction.
4. Apply the genetic operators crossover, reproduction and mutation to produce new rules.
5. Reinsert these offspring to create the new population.
6. Repeat steps (3)–(6) until an acceptable classification rule is found or the specified maximum number of generations has been reached.
7. Repeat steps (2)–(7) until one rule is determined for each class in the database.
8. Assign each example in the training and in the test sets to one and only one class.

2.4. Fitness function

The population is constituted by the possible class descriptions, that is to say sets of conditions on attributes of the objects to classify. For all classification problems, in order to be able to apply a particular fitness function, the learning algorithms must convert the value returned by the evolved model into “1” or “0” using the 0/1 Rounding Threshold. If the value returned by the evolved model is equal to or greater than the rounding threshold, then the record is classified as “1”, “0” otherwise.

There are many varieties fitness function such as number of hits, sensitivity/specificity, relative squared error (RSE), mean squared error (MSE) and etc. that can be applied for evaluating performance of generated classification rules.

We used “number of hits” as fitness function because of its simplicity in efficiency, and is based on the number of samples correctly classified. More formally, the fitness $f_i$ of an individual program corresponds to the number of hits and is evaluated by the formula:

$$f_i = h$$

(1)

where $h$ is the number of fitness cases correctly evaluated (number of hits). So, for this fitness function, maximum fitness $f_{\text{max}}$ is given by:

$$f_{\text{max}} = n$$

(2)

where $n$ is the number of fitness cases.

Its counterpart with “parsimony pressure” uses this fitness measure $f_i$ as “raw fitness” $rf_i$ and complements it with a parsimony term. Parsimony pressure puts a little pressure on the size of the evolving solutions, allowing the discovery of more compact models. Thus, in this case, raw maximum fitness $rf_{\text{max}} = n$. And the overall fitness $fpp_i$ (that is, fitness with parsimony pressure) is evaluated by the formula:

$$fpp_i = rf_i \times \left( 1 + \frac{1}{5000} \times \frac{S_{\text{max}} - S_i}{S_{\text{max}} - S_{\text{min}}} \right)$$

(3)

where $S_i$ is the size of the program, $S_{\text{max}}$ and $S_{\text{min}}$ represent, respectively, maximum and minimum program sizes and are evaluated by the formulas:

$$S_{\text{max}} = G(h + t)$$

(4)

$$S_{\text{min}} = G$$

(5)

where $G$ is the number of genes, and $h$ and $t$ are the head and tail sizes (note that, for simplicity, the linking function was not taken into account). Thus, when $rf_i = rf_{\text{max}}$ and $S_i = S_{\text{min}}$ (highly improbable, though, as this can only
happen for very simple functions as this means that all the sub-ETs are composed of just one node), \( fpp_i = fpp_{\text{max}} \), with \( fpp_{\text{max}} \) evaluated by the formula:

\[
fpp_{\text{max}} = 1.0002 \times r_{f_{\text{max}}}
\]  

(6)

Usual termination criterions appear to be the accomplishment of a number of generations, the achievement of a desired classification error, etc. The described procedure is depicted in the flowchart of Fig. 5. (Tsakonas, 2006).

2.5. Genetic programming for solving the bankruptcy prediction problem

The aim is the implementation of a genetic model able to automatically extract an intelligible classification rule for prediction classes of bankrupt and non-bankrupt firms in a sample, given the values of some financial ratios, called predicting variables. Each rule is constituted by a logical combination of these ratios. This combination determines a class description which is used to construct the classification rule.

Once defined a fitness function, bankruptcy prediction problem becomes a search problem of the best solution in the search space of all the possible solutions, that is to say an optimization of the fitness function for which optimization techniques can be used.

Given a number of variables describing each firm and their related domains, it is easily understandable that for bankruptcy prediction problems, the number of possible solutions is enormous. An exhaustive search by enumerating all the possible solutions is computationally impracticable. Hence, we appeal to GP which is a powerful search method inspired by natural selection. It does not guarantee to find the global optimum; nonetheless it usually allows retrieving a suboptimal solution in a reasonable computation time.

The evolving population is constituted by individuals or ‘programs’ representing the classification rules in the form of trees whose sizes are intrinsically variable in length. Each rule is constituted by a number of conditional clauses, in which conditions on or between certain variables are set, and by a predictive clause representing the firms class. A class together with its description forms a classification rule of type “If \{solution\} then \{bankrupt/ non-bankrupt\}”. The conditional part of the rule is formed by all the active conditional clauses.

A population composed by a set of these candidate rules is maintained and gradually improved by constructing new fitter ones until rules of sufficient quality are found or other stopping criteria are satisfied. The model allows attaining rules covering the different classes by running the evolutionary algorithm as many times as the number of the classes to predict. In practice the model will analyze one class at a time.

To construct the bankruptcy prediction model, data is partitioned into two sets: the training and the test sets. The training set contains the known firms used during the evolution process to find an explicit classification rule able to separate an instance of a class of bankrupt firms from instances of non-bankrupt class, while the test set is used to evaluate the generalization ability of the rule found (De Falco, Della Cioppa, & Tarantino, 2002).

3. Multiple discriminant analysis (MDA)

MDA is a statistical technique used to classify an observation into one of several a priori groupings dependent upon the observation’s individual characteristics. Therefore, MDA allows the researcher to study the difference between two or more groups of objects with respect to several variables simultaneously (Klecka, 1985). It is used primarily to classify and/or make predictions in problems where the dependent variable appears in qualitative. In the case of two groups consisting of bankrupt firms on one hand and non-bankrupt firms on the other, MDA attempts to establish a linear combination of characteristics (financial ratios) which “best” discriminates between the groups (Altman & Narayanan, 1997). The discriminant function is specified as

\[
Z_i = \alpha_1 X_{1i} + \alpha_2 X_{2i} + \ldots + \alpha_n X_{ni}
\]  

(7)

where \( \alpha_1, \alpha_2, \ldots, \alpha_n = \text{discriminant coefficients} \), \( X_{1i}, X_{2i}, \ldots \), \( X_{ni} = \text{classifying variables (financial ratios) of firm} \ i \).

The value of the dividing point of the two groups is calculated as
\[ D = \frac{Z_b + Z_{ab}}{2} \] 

where \( Z_b \) is the mean value of the \( Z \) in the bankrupt group and \( Z_{ab} \) is the mean of the \( Z \) in the non-bankrupt group. After coefficients \( x_1 \) through \( x_n \) are estimated and the discriminant function is established, the \( Z \) value of a particular firm can be calculated and compared to the dividing point \( D \). The classification depends on whether the \( Z \) value is greater or smaller than the dividing point \( D \). The quality of the model is measured by the degree of success in classifying the two groups of firms correctly (Fernandez, 2003). Altman (1993) has provided complete description of MDA and its financial applications.

As a model for dichotomous events, MDA has its shortcomings. A major limitation of MDA is its restrictive assumption of normal distribution of each independent variable (Hamer, 1983). Other assumptions are:

- The predictors are not highly correlated with each other.
- The mean and variance of a given predictor are not correlated.
- The correlation between two predictors is constant across groups.

MDA is prevalent technique in bankruptcy prediction (Aziz & Dar, 2006). In terms of classification or prediction ability among traditional models, while some studies, have found Logit model superior to MDA (Press & Wilson, 1978), other studies (Gu, 2002; Aziz & Dar, 2006) have shown that the two models are equally efficient.

4. Development of models

4.1. Data collection

The data set used for this research consists of 144 Iranian companies. All of them were or still are listed on the Tehran Stock Exchange (TSE). 72 companies went bankrupt under paragraph 141 of Iran Trade Law\(^1\) from 1998 through 2005. The other 72 companies are “matched” companies, from the same period of listing on the TSE. Because of small population, we could not match two groups in each of industries. Also size of the firms as a potential explanatory variable has been considered in variable selection phase.

4.2. Variables selection

We apply a three stages predictive variable selection process. At the first stage, bankruptcy prediction literature was reviewed and 65 variables among more than 250 financial ratios were selected as predictive variables. These financial ratios were chosen based on popularity in literature. In the second stage, 43 variables were selected based on availability of the necessary data. Table 1 shows the selected variables.

In the third stage, using stepwise discriminant analysis (SDA) was used to select final variables. SDA is a prevalent procedure for reducing dimensions of a problem and selecting the most significant variables that can among widespread set of variables. SDA select variable based on a cut-off F value for pre-specified statistical significance level until no variable is qualified for selecting. With significance level set at the 0.05 level, the discriminant stepwise procedure selected 5 variables from the 43 candidates for the model which could best discriminate the bankrupt firms from the non-bankrupt firms.

These selected financial ratios are: (1) Operational income to sales ratio (X36), (2) Total liability to total assets (X9), (3) Sales to current assets ratio (X43), (4) Interest expense to gross profit (X25), and (5) Quick assets to total assets (X20).

Operational income to sales (margin of profit) ratio reflects profitability of firm that shows the ability of a firm in covering all production costs and providing some reasonable margin of profit. A higher ratio can be translated as better coverage production costs and lower default risk. Therefore, margin of profit has a negative relation to bankruptcy. This ratio had the highest influence in segregation of two groups.

Total liability to total assets (Debt ratio) measure the degree of indebtedness and the ability of paying off debt interests and principal and is categorized as solvency ratios. Solvency may reveal fundamental causes of bankruptcy rooted in a firm’s financing policy. This ratio has a negative relation with bankruptcy occurrence.

Sales to current assets (current assets turnover) ratio, measuring revenue created per dollar of current assets is used as efficiency ratios. A higher efficiency means higher profitability and better liquidity and finally to lower default risk. However, as Gu and Gao (2000) found, a high efficiency ratio accompanied by poor profitability may indicate fast sales growth without proper costs control, thus signaling a high default risk. Analyzing balance sheet statements of firms in our sample specify that, in average, current assets constitute over than 60% total assets, hence, inclusion of current assets in the model can improve it’s the performance.

Interest expense to gross profit measures firms’ ability to pay interest expense. When a firm has good profitability then creditors are sure that their interests in this firm can be achievable. This ratio has positive relation with bankruptcy occurrence; the lower interest expense to gross profit ratio, the lower bankruptcy risk.

Quick assets to total assets ratio is one of the liquidity ratios and shows firms’ ability of paying short term debts. It can be interpreted as closeness of firms’ assets to cash. Creditors prefer that firms hold assets with high liquidity because these assets have lower risk than the others and firms can quickly convert them to money in extraordinary circumstances. Thus liquidity ratios have a momentous role

---

\(^1\) Under paragraph 141 of Iran Trade Law, a firm is bankrupt when its total value of retained earnings is equal or greater than 50% of its listed capital.
in debt contracts and cost of capital. In general, firms with higher liquidity have a lower bankruptcy risk.

4.3. Genetic programming model

The formerly mentioned procedure applied for GP model. For this study, the Number of Hit, as described in the pervious section, was used as the fitness function. For testing predictive power of model, data set was randomly separated into two groups. Training sample consists of 104 (51 bankrupt and 53 non-bankrupt) firms and Holdout sample contains 40 (21 bankrupt and 19 non-bankrupt) firms.

For implementing GP process and developing bankruptcy model, GeneXproTools software version 4.0 was used. Crossover and mutation operators were set as 0.6 and 0.06, respectively.

The representation of a solution for the problem provided by the GP algorithm is a decision tree. Each node of this tree is a function node taking one of the values from the set \{+, -, *, ^, NOT, and LT\}: “+” adds up two variables, “-” subtracts them, “*” multiply them, “^” is symbol of the power, “NOT” means one minus a variables, and “LT” returns 1 when the first variable is lesser than the second, else 0.

When return of the GP model (decision tree) for a training or test firm is greater or equal 0.5 (Rounding Threshold), this firm is marked as “bankrupt firm”. Alternatively, when return of the GP model for a training or test firm is lesser than 0.5, this firm is classified as “non-bankrupt firm”. Comparing real class of firms with predicted class by the GP model, in the training and test sample, determines accuracy of the model. Fig. 6 Shows the best GP model obtained. This model has been divided in three sub-trees (each tree represents a Gene that’s means the model is a Chromosome consisted of three genes). Sum of the returns of sub-trees for a firm should be compared with “Rounding Threshold” for determining class of the firm.

4.4. Multiple discriminant analysis model

To determine the relative prediction effectiveness of GP model, we benchmarked it against a MDA model,

<table>
<thead>
<tr>
<th>#</th>
<th>Definition of variables</th>
<th>Means of non-bankrupt companies (72)</th>
<th>Means of bankrupt companies (72)</th>
<th>Tests of equality of group means (Sig level)</th>
<th>#</th>
<th>Definition of variables</th>
<th>Means of non-bankrupt companies (72)</th>
<th>Means of bankrupt companies (72)</th>
<th>Tests of equality of group means (Sig level)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>EBIT/TA</td>
<td>0.2176</td>
<td>0.05754</td>
<td>0.000</td>
<td>2</td>
<td>LTD/SE</td>
<td>2.54796</td>
<td>2.85370</td>
<td>0.851</td>
</tr>
<tr>
<td>3</td>
<td>RE/SC</td>
<td>0.47895</td>
<td>0.02016</td>
<td>0.000</td>
<td>4</td>
<td>MVE/TL</td>
<td>2.09833</td>
<td>0.39374</td>
<td>0.008</td>
</tr>
<tr>
<td>5</td>
<td>MVE/SE</td>
<td>3.68678</td>
<td>1.73033</td>
<td>0.032</td>
<td>6</td>
<td>MVE/TA~</td>
<td>5.22278</td>
<td>4.87388</td>
<td>0.001</td>
</tr>
<tr>
<td>7</td>
<td>Ca/TA</td>
<td>0.05437</td>
<td>0.03103</td>
<td>0.009</td>
<td>8</td>
<td>Size/log(TA)</td>
<td>2.28985</td>
<td>2.51011</td>
<td>0.874</td>
</tr>
<tr>
<td>9</td>
<td>CL/TL</td>
<td>0.68384</td>
<td>0.78209</td>
<td>0.000</td>
<td>10</td>
<td>CL/SE</td>
<td>0.68362</td>
<td>0.83518</td>
<td>0.234</td>
</tr>
<tr>
<td>11</td>
<td>(R+Inv)/TA</td>
<td>0.53664</td>
<td>0.62095</td>
<td>0.041</td>
<td>12</td>
<td>(Ca+STI)/CL</td>
<td>1.05794</td>
<td>0.05352</td>
<td>0.000</td>
</tr>
<tr>
<td>13</td>
<td>R/Inv</td>
<td>14.5339</td>
<td>0.92670</td>
<td>0.292</td>
<td>14</td>
<td>R/S</td>
<td>0.44839</td>
<td>0.35287</td>
<td>0.445</td>
</tr>
<tr>
<td>15</td>
<td>SE/TA</td>
<td>0.34665</td>
<td>0.24232</td>
<td>0.000</td>
<td>16</td>
<td>CA/CL</td>
<td>1.17647</td>
<td>1.13654</td>
<td>0.567</td>
</tr>
<tr>
<td>17</td>
<td>QA/CL</td>
<td>0.66072</td>
<td>0.50096</td>
<td>0.002</td>
<td>18</td>
<td>*CA/TA</td>
<td>0.56170</td>
<td>0.44675</td>
<td>0.001</td>
</tr>
<tr>
<td>21</td>
<td>FA/(SE+LTD)</td>
<td>0.64560</td>
<td>0.78329</td>
<td>0.410</td>
<td>22</td>
<td>FA/TA</td>
<td>0.23400</td>
<td>0.23308</td>
<td>0.976</td>
</tr>
<tr>
<td>23</td>
<td>CA/TA</td>
<td>0.66938</td>
<td>0.71408</td>
<td>0.201</td>
<td>24</td>
<td>CA/CL</td>
<td>0.10662</td>
<td>0.04838</td>
<td>0.002</td>
</tr>
<tr>
<td>25</td>
<td>IE/GP</td>
<td>0.41801</td>
<td>1.68147</td>
<td>0.211</td>
<td>26</td>
<td>S/Ca</td>
<td>45.2066</td>
<td>122.611</td>
<td>0.245</td>
</tr>
<tr>
<td>27</td>
<td>S/TA</td>
<td>0.88878</td>
<td>0.78572</td>
<td>0.107</td>
<td>28</td>
<td>WC/TA</td>
<td>0.07149</td>
<td>0.05773</td>
<td>0.696</td>
</tr>
<tr>
<td>29</td>
<td>PIC/SE</td>
<td>0.44833</td>
<td>0.52932</td>
<td>0.199</td>
<td>30</td>
<td>S/WC</td>
<td>17.8260</td>
<td>5.19366</td>
<td>0.213</td>
</tr>
<tr>
<td>31</td>
<td>RE/TA</td>
<td>0.03836</td>
<td>-0.02026</td>
<td>0.000</td>
<td>32</td>
<td>NI/SE</td>
<td>0.53836</td>
<td>0.32618</td>
<td>0.072</td>
</tr>
<tr>
<td>33</td>
<td>NI/S</td>
<td>0.20475</td>
<td>-0.00552</td>
<td>0.000</td>
<td>34</td>
<td>NI/TA</td>
<td>0.17323</td>
<td>0.01155</td>
<td>0.000</td>
</tr>
<tr>
<td>35</td>
<td>D/NI</td>
<td>0.79203</td>
<td>-1.0847</td>
<td>0.311</td>
<td>36</td>
<td>*O/U/S</td>
<td>0.22485</td>
<td>0.01508</td>
<td>0.000</td>
</tr>
<tr>
<td>37</td>
<td>OI/TA</td>
<td>0.19046</td>
<td>0.02571</td>
<td>0.000</td>
<td>38</td>
<td>EBIT/IE</td>
<td>33.00551</td>
<td>8.02330</td>
<td>0.078</td>
</tr>
<tr>
<td>39</td>
<td>EBIT/S</td>
<td>0.21757</td>
<td>-0.0035</td>
<td>0.000</td>
<td>40</td>
<td>GP/S</td>
<td>0.29107</td>
<td>0.10676</td>
<td>0.000</td>
</tr>
<tr>
<td>41</td>
<td>S/SE</td>
<td>3.06774</td>
<td>3.09217</td>
<td>0.987</td>
<td>42</td>
<td>S/FA</td>
<td>6.03377</td>
<td>6.17452</td>
<td>0.893</td>
</tr>
<tr>
<td>43</td>
<td>*S/CA</td>
<td>1.40478</td>
<td>1.13062</td>
<td>0.006</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CA: Current assets
Ca: Cash
CL: Current liabilities
PIC: Paid in capital
EBIT: Earnings before interest & taxes
FA: Fixed assets
GP: Gross profit
IE: Interest expenses
Inv: Inventory
LA: Liquid assets
LTD: Long term debt
MVE: Marked value of equity
NI: Net income
OE: Operational income
QA: Quick assets
R: Receivables
RE: Retained earnings
SC: Stock capital
SE: Shareholders’ equity
STI: Short term investments
TA: Total assets
TL: Total liabilities
WC: Working capital

*Final variables selected by SDA.
as a more prevalent methodology. Accordingly, for comparison purposes, we developed a MDA model using same five variables (ratios) and same training and test data set.

A basic assumption of MDA is the multivariate normality of the classifying variables (Kleinbaum, Kupper, & Muller, 1988) but it is applicable just when the sample size is less than 50 observations (Gu, 2002). Since in this study the sample size was 140 (firms) then the normality test was not necessary. We used SPSS software version 13 for constructing MDA model. Significance levels were determined at 0.1 values. The estimated MDA model is as:

\[
Z = 3.764X_{20} - 3.609X_{9} + 0.102X_{36} - 0.034X_{52} - 0.251
\]  

To estimate the model, SPSS software adjusted the dividing point to a cut-off value of zero using constant \(-0.251\). Z score is calculated by above equation for each firm. Companies with Z scores above zero were put into non-bankrupt group.

To assess the classification accuracy of the discriminant model, we used a cross validation procedure. In this procedure, \((n - 1)\) firms out of \(n\) in training sample are used. MDA estimates discriminant functions based on \((n - 1)\) firms and applies them to classify a firm left out. This cross validation is performed for each of \(n\) training firms. The classification accuracy rate for each group is estimated from the proportion of sample firms that are classified correctly.

5. Experimental results

5.1. GP model

Table 2 shows obtained result from GP model. This model could classify firms in the training sample with 94% overall accuracy rate. In detailed view, the GP model has achieved 96% accuracy rate for correct classifying bankrupt firms and 93% accuracy rate for correct classifying non-bankrupt firms in the training sample. In addition, the GP model was applied to the holdout sample for testing. This model could correctly classify 90% firms in the holdout sample. That this rates for correct classifying bankrupt and non-bankrupt firms are 95% and 84%, respectively.

Table 2  
Results of genetic programming model

<table>
<thead>
<tr>
<th>Samples</th>
<th>Real group membership</th>
<th>Predicted group “GP model”</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 (count)</td>
<td>0</td>
</tr>
<tr>
<td>Learning sample</td>
<td>49</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>0 (count)</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>1 (%)</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td>0 (%)</td>
<td>7</td>
</tr>
<tr>
<td>Holdout sample</td>
<td>20</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>0 (count)</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>1 (%)</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>0 (%)</td>
<td>16</td>
</tr>
</tbody>
</table>

*Group 1: bankrupt firms; Group 0: non-bankrupt firms.
An important contribution of this paper is identification of the effective predictive financial ratios. The financial ratios of OI/TA, S/CA, TL/TA, QA/TA, and IE/GP are identified as highly predictive factors by SDA method. These five ratios cover most significant financial aspects of a firm. In addition According to Dimitras et al. (1996), these financial ratios are members of most popular predictive variables in bankruptcy prediction literature.

Using these variables, a GP model was developed. The GP model was 94% and 90% accurate on training and holdout samples, respectively. Also, a MDA model was developed using the same five variables and same data set. The model achieved only 77% and 73% accuracy in training and holdout samples, respectively.

In summary, GP produced what can be deemed as a highly accurate bankruptcy prediction model considering both the nature of sample companies and prediction period. As a final conclusion, it can be claimed that GP model was more accurate than traditional MDA model. The model provides insight into the complex interaction of bankruptcy related factors and suggests avenues for future research.

References


