Full Length Research Paper

Genetic programming approach for testing credit risk box method

Alireza Bahiraie^{1*}, Noor Akma bt Ibrahim² and A. K. M. Azhar³

¹Institute of Mathematical Sciences, Faculty of Science, University of Malaya, 50603, Kuala Lumpur, Malaysia.

²Institute for Mathematical Research, University Putra Malaysia, 43400, Selangor, Malaysia.

³Graduate School of Management, University Putra Malaysia, 43400, Selangor, Malaysia.

Accepted 13 October, 2010

In this paper, Genetic Programming (GP) technique is applied to the empirical analysis of a new geometric approach of credit risk, financial ratios and bankruptcy prediction. Utilizing financial ratios for prediction of corporate bankruptcy and identification of firms' impending failure is indeed desirable for investors, creditors, borrowing firms, and governments. This paper presents new geometric technique for empirical analysis of credit and bankruptcy risk using financial ratios. Within this framework, we propose the use of a new ratio representation which is named Risk Box measure (RB). We demonstrate the application of this geometric approach for variable representation, data visualization and financial ratios at different stages of corporate bankruptcy prediction models based on financial balance sheet ratios. These stages are the selection of variables (predictors), accuracy of each estimation model and the representation of each model for transformed and common ratios. By the time, several methods have been attempted in the use of financial ratios on predicting bankruptcy but some of them suffer from underlying shortcomings. Recently, Genetic Programming (GP) has received great attention in academic and empirical fields of solving highly complex problems. Results of Genetic Programming (GP) as statistical classification methodology are compared for common and transformed ratios and better accuracy is obtained.

Key words: Ratios analysis, risk box, credit risk, bankruptcy prediction, genetic programming.

INTRODUCTION

In classical prediction models, a convenient representation of ratios is in closed form of graphical presentation of data. In contrast, achieving better accuracy often relies on visualization of predictors. It is at this stage when the selection of a proper graphical presentation scheme becomes essential for a correct scaled visualization. Since numerical presentation of ratios cannot be a good representative of characteristics of companies, some other ways of displaying them must be found. Graphical tools give this possibility. This paper presents a complementary perspective on the study of ratios and bankruptcy. Since global increases in bankruptcies, accurate predictions of companies' distress and bankruptcies have been extensive. One of the most

well known anomalies of the risk factors is the effect of some ratios on bankruptcy risk and firm returns. One possible explanation for this effect that is consistent with the "efficient market hypothesis" that ratio is a proxy for risk. Also in banking, the ratios taken to be a proxy for the charter value of banks (Landskroner et al., 2006). Statistical techniques applications to corporate bankruptcy started in the 60's with the development of computers.

The first technique introduced was discriminant analysis (DA) for univariate and multivariate models (Beaver, 1966). Then, Altman (1968), used multiple discriminant analysis (MDA) and applied it to prediction of business failure. Altman (1971) examined railroad bankruptcy propensity and Deakin (1972) replicated the study of Beaver (1966), Edmister (1972) testing the usefulness of financial ratio in order to predict small business failure. Altman et al. (1974) developed a model

*Corresponding author. E-mail: alirezab@um.edu.my.

in order to determine the credit worthiness of commercial loan applicants in the cotton and wool textile sector in France. Altman et al. (1977) developed their classical Z model and named it Zeta Analysis. After DA and Multiple Discriminant Analysis (MDA), the Logit and Probit models were introduced by Martin (1977) and Ohelson (1980). Nowadays, these models are widely used in practice. The solution in the traditional framework is a linear function separating successful and failing companies. A company score is computed as a value of that function (Aziz and Dar, 2006; Chauhan et al., 2009). Northon and Smith (1979) who compared the prediction of bankruptcy using ratios computed from General Price Level (GPL) financial statements to the prediction of bankruptcy, using ratios computed from traditional historical cost financial statements, Taffler (1982) who used linear discriminant analysis for the prediction of bankruptcies in UK with financial ratios. Moreover, recursive partitioning, also known as classification and regression trees (CART) performs classification by dividing the data space.

Moreover, Genetic Programming (GP) is a population of linear classifiers (genes) that are connected with one another in a pre-specified way. The outputs of some of the genes are inputs for others. The performance of GP greatly depends on its structure that must be adapted for solving different problems. However, as there is no widely accepted economic theory, every study has based their model specification on an empirical framework. This results in different accounting ratios used in different models. Generally, these multivariate models are conducted on procedure that is structured in such a way that an equal number of bankrupt and non-bankrupt firms are chosen randomly with respect to company size or industry or large and small samples avoiding matching procedure. According to literature, predictors used in various studies, generally exhibit non-normal distribution and high standard errors (Barnes, 1982; McLeay and Omar, 2000; Ooghe et al., 2005; Andres et al., 2005). Some researchers made correction for univariate non-normality and tried to approximate univariate normality by transforming the variables prior to estimation of their model¹. Deakin (1976) used log transformation, then square root and log-normal transformation of financial ratios were used by Foster (1986) and Gu (2002). Other researchers approximate univariate normality by 'trimming' or 'outlier deletion', which involves segregating outliers by reference to normal distribution (Ezzamel et al., 1987). Furthermore, rank transformation has been used by Perry et al. (1986) and Kane et al. (1996). Recently, Bahiraie et al. (2008) used geometric transformation of ratios, which may become general

guidelines concerning the transformation details was discussed.

Our objective in this paper is to discuss about new geometric approach to ratios, which involves data transformation and we illustrate the use of this methodology for bankruptcy predictions. For illustration of this new methodology, numerator and denominator of common ratio values are represented as Cartesian coordinates in our constructed modification box in which we derive the isoclines of associated components of bankruptcy and credit risk. This study is regarded as one of the fundamental studies in this field.

MATERIALS AND METHODS

Genetic programming (GP)

Genetic programming (GP) is a search methodology belonging to the family of evolutionary computation (EC). GP can be considered as an extension of genetic algorithms, GA (Koza, 1992). GA is stochastic search techniques that can search large and complicated spaces stemmed on the ideas from natural genetics and evolutionary principle. They have been demonstrated to be effective and robust in searching very large spaces in a wide range of applications. GP is basically a GA applied to a population of computer programs (CP). While a GA usually operates on strings of numbers, a GP has to operate on CP. GP allows, in comparison with GA, the optimization of much more complicated structures and can therefore be applied to a greater diversity of problems (Nwogugu, 2006). While bankruptcy prediction can be considered as a classification problem, we provide necessary description of GP with emphasis on its application in classification role (Koza, 1992). Genetic programming models were inspired by the Darwinian theory of evolution. According to the most common implementations, a population of candidate solutions is maintained, and after a generation is accomplished, the population is fitted better for a given problem. Genetic programming uses tree-like individuals that can represent mathematical expressions. Three genetic operators are mostly used in these algorithms: 1) reproduction, 2) crossover, and 3) mutation.

First, the reproduction operator simply chooses an individual in the current population and copies it without changes into the new population. Reproduction stands for selection of two parents and production of two new children and including them to population. Our study demonstrates the use of standardized quantitative gene expression values from primary individual indicator and data set from different companies' balance sheets as inputs in a genetic programming system to generate classifier rules. In second step, two parent individuals are selected and a sub-tree is picked on each one. Then, crossover swaps the nodes and their relative sub-trees from one parent to the other. 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 (Ravikumar and Ravi, 2007).

The mutation operator can be applied to either a function node or a terminal node which in the tree is randomly selected. If the chosen node is a terminal node it is simply replaced by another terminal and if it is a function and point mutation is to be performed, it is replaced by a new function with the same parity (Lee, 2004). When tree mutation is to be carried out, a new function node is chosen, and the original node together with its relative sub-tree is substituted by a new randomly generated sub-tree. A depth ramp is used to set bounds on size when generating the replacement sub-tree. Naturally, it is to check that this replacement does not violate

¹ For more details about financial ratios properties, Watson (1990) and Tippet (1990).

² Some exposition on some of the weaknesses in the use of common ratios such as scaling, proportionality and symmetric effects are provided in Bahiraie et al. (2008) and, Azhar and Elliott (2006).

Figure 1. Overview of genetic programming process.

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

The last step for obtaining the best fitness function for all classification problems, in order 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” and “0” otherwise. There are many varieties of fitness function such as number of hits, sensitivity/specificity, relative squared error (RSE), mean squared error (MSE), that can be applied for evaluating performance of generated classification rules.

We used “number of hits” as fitness function because of its simplicity and efficiency which 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 $f_i = h$ where h is the number of fitness cases correctly evaluated or number of hits and t is the number of wrong classification by fitness function. So, for this fitness function, maximum fitness $f_{\max}$ is given by $f_{\max} = n$ where n is the

number of fitness cases. Its counterpart uses this fitness measure

f_i as “raw fitness”, rf_i . Thus, in this case, raw maximum fitness $rf_{\max} = n$ and the overall fitness fpp_i , that is, fitness with parsimony pressure is evaluated by $fpp_i = rf_i \times (1 + \frac{1}{5000} \times \frac{S_{\max} - S_i}{S_{\max} - S_{\min}})$ where S_i is the size of

the program, $S_{\max}$ and $S_{\min}$ represent minimum and maximum of program population respectively. Maximum and minimum of program sizes are evaluated by the formulas:

$S_{\max} = G(h + t)$ and $S_{\min} = G$ where G is the number of genes, and h and t are the head (correctly classified) and tail (wrongly classified) sizes. Thus, when $rf_i = rf_{\max}$ and $S_i = S_{\min}$,

with $fpp_{\max} = 1.0002 \times rf_{\max}$ the process will be optimized.

This procedure is repeated for a given number of steps until the difference between the maximum fitness function values has no major improvement.

The described procedure is depicted in the flowchart of Figure 1 (Tsakonas, 2006). Once fitness function is defined, bankruptcy prediction problem becomes a search problem of the best solution

Figure 2. Total risk isoclines.

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. The implementation of a genetic model is to automatically extract an intelligible classification rule for prediction classes of bankrupt and non-bankrupt firms in a sample by the given values of some financial ratios, called predicting variables. Each rule is constituted by a logical combination of these ratios. The combination determines a class description which is used to construct the classification rule. Given a number of variables describing each firm and their related domains, it is easy to understand bankruptcy prediction problems by the number of possible solutions obtained which is enormous.

The risk box

Here, we will present more details of the existing new graphical technique for ratio transformation. The framework is a two-dimensional box in which pair values of risk ratios numerator (X) and denominator (Y) are represented as Cartesian coordinates. Assume a hypothetical study of risk covering n years for sector j ². For expositional purposes suppose our proxy for risk chosen is employed by (X) and (Y) values. $\forall t = 1, 2, 3, \dots, n$, we have: $X_t, Y_t, TR_t, NR_t, OR_t, SR_t > 0$, all risk components measure indices such as, Total Risk $TR = X + Y$, Net Risk $NR = |X - Y|$, Overlapping Risk $OR = (X + Y) - |X - Y|$, and lastly the proposed Share Measure of Risk (SR) as we define below, are linear functions of X and Y.

$$X + Y = TR = NR + OR$$

$$SR = \frac{OR}{TR} = \frac{(X + Y) - |X - Y|}{(X + Y)} = \frac{2 \min(X, Y)}{(X + Y)}$$

Following Bahiraie et al. (2008), we can construct a two dimensional box that encapsulates all of these variables for n years. The dimensions of the risk box are generated by the maximum

value of either the X or Y value during the period of study. From the definition of TR, NR, OR, SR, we obtain:

$$\max(NR_t) = \max(|X_t - Y_t|) \leq \max(\max X_t - \min Y_t, \max Y_t - \min X_t) \leq m$$

$$\max(OR_t) = 2 \max(\min(X_t, Y_t)) \leq 2m$$

$$\max SR_t \leq 1$$

Each respective risk box will have sides equal to $\max(X_t)$ if for $i \in t$, $\max(X_t) > \max(Y_t)$ or $\max(Y_t)$ if otherwise. Our exposition of the dimensions of the box is as follows which confirms the elasticity and unit-free nature of SR measure. A 45° line from the origin bisects the box into two equal triangles and will be measure of $\tan \theta$. This positive slope diagonal (as demonstrated in next sections) is the locus of balanced risk where $X_t = Y_t$, TR equals OR, SR equals unity, and NR equals zero. This is the risk components' axis of symmetry³. The two triangular planes in the box consists of an upper triangle containing coordinate points (X_t, Y_t) where $X_t > Y_t$ in the net X value (NX) plane and points $Y_t > X_t$ in the lower triangle, defined as the net Y value (NY) plane.

Locus of Equi TR

A fix value $TR = TR^*$ implies $X = TR^* - Y$. Comparing with $y = mx + c$, we have the gradient m equals minus unity. Hence, locus of equi TR is perpendicular to the axis of symmetry (Figure 2). With respect to the vertical intercepts in Figure 2,

$$\forall t > t-1 \Rightarrow TR_t = (X + Y)_t > (X + Y)_{t-1} = TR_{t-1}$$

Locus of Equi NR

Recall that Net Risk $NR = |X - Y|$. The line 45° can be regarded as the contour of the value $NR^* = 0$. For positive value $NR^* > 0$, we have below the central 45° line, $Y - X = NR^*$ so $X = Y - NR^*$, which also slopes upward at 45° , meeting the (horizontal) Y axis at NR^* . Above the 45° line through the origin we have another segment of same contour, namely the line $X - Y = NR^*$ or $X = Y + NR^*$. These two 45° lines from the contour corresponding to NR^* . Increasing the value of constant NR^* moves both segments higher up their respective axis, away from the central NR^* line.

From (Figure 3), we have the following:

² For expositional simplification, sectoral subscripts are dropped from the rest of the paper.

³ See Schattsneider (1978) for a broad discussion on plane symmetries.

Figure 3. Net risk isoclines.

Figure 4. Overlapping risk isoclines.

$$\forall X = Y \text{ then } NR = 0$$

$$\forall X > Y \text{ then } NR = |X - Y| = X - Y \text{ i.e. } X = Y + NR$$

$$\forall X < Y \text{ then } NR = |X - Y| = Y - X \text{ i.e. } X = Y - NR$$

Comparing with $y = mx + c$, we have for a net book value, $m = 1$ with a vertical intercept $c = NR$. Since the central line balanced is the axis of symmetry for NR, $m = 1$ and $c = NR$ (see Figure 3). Consequently, locus of equi NR values is perpendicular to lines of equi TR (that is $m_{TR} = \frac{-1}{m_{NR}}$).

Locus of Equi OR

Recall that Overlapping Risk $OR = 2 \min(X, Y)$. Below the central 45° line, $OR = 2X$ which, remains constant for constant

X. Above the line $OR = 2Y$ which remains constant for constant Y. Thus the equi corresponding to constant overlapping risk OR^* is L-shaped (Figure 4), the kink occurring along the central 45° line. As OR^* increases, the kink moves up the line, away from the origin.

From (Figure 4), we have the following;

$$\forall X = Y \Rightarrow OR = (X + Y) = TR$$

$$\forall X > Y \Rightarrow OR = (X + Y) - (X - Y) = 2Y$$

$$\forall X < Y \Rightarrow OR = (X + Y) - (Y - X) = 2X$$

Proposed share measure of risk and locus of Equi SR

Consider our proposed unit-free share measure of risk $SR = \frac{2 \min(X, Y)}{X + Y}$, the followings are obtained:

(i) Below the line, $Y > X$ and thus $SR = \frac{2X}{X + Y}$.

corresponding to a constant value SR^* is defined by the relation $SR^*(X + Y) = 2X$, which can be solved for X to yield $X = \frac{SR^*}{2 - SR^*}Y$. Thus this segment of the equi is a ray from the

origin with constant slope $\gamma = \frac{SR^*}{2 - SR^*}$. Since $0 \leq SR \leq 1$, we have $0 \leq \gamma \leq 1$, showing that the ray passes between the central 45° line and the horizontal axis.

(ii) Above the central 45° line on the other hand we have $SR = \frac{2Y}{X + Y}$. Given a constant value SR^* we obtain $X = \gamma^{-1}Y$, whose slope γ^{-1} satisfies $1 \leq \gamma^{-1} < \infty$.

Thus the equi corresponding to a particular value SR^* consists of two rays in the positive quadrant meeting at the origin, with slopes γ and γ^{-1} . In Figure 9 these rays are shown as OC and OB.

Note that the symmetry of the diagram about the central 45° line implies that the angles θ_1 and θ_2 are equal. In Figure 5, relationships between the four risks measures and slopes γ and γ^{-1} , consider rays OB and OC subtending the angles θ_1 or θ_2 measured from the symmetry axis. We measure θ_1 if the point is at North West half side of the box and θ_2 for any point at South East half of the box. Let A, B, C, and D represent points on the risk plane with A, B and C sharing equal total risk values, TR^* . In addition, B, C, and D share equal OR values, OR^* .

$$OA = TR^*, \text{ and } TR^* - OR^* \stackrel{\text{defn}}{=} NR^* = AB$$

Figure 5. Share risk isoclines.

Hence;

$$\tan \theta_1^{defn} = \frac{AB}{OA} = \frac{TR^* - OR^*}{TR^*} = 1 - \frac{OR^*}{TR^*} = 1 - SR^*$$

$$\Rightarrow SR^* = 1 - \tan \theta_1$$

These will confirm that SR values are constant along any ray from origin and the two extreme case the two extreme cases are (i)

θ_1 or $\theta_2 = 45^\circ$, in which case $SR=0$ and either the Y value or the

X value is zero, and (ii) θ_1 or $\theta_2 = 0$, in which case $SR=1$ and

$X = Y$.

EMPIRICAL RESULTS

The database used in our illustrative empirical study consists of 200 Malaysian companies from Kuala Lumpur Stock Exchange (KLSE) which 60 companies went bankrupt under section 167 of Malaysian companies law act 1965 which a firm is bankrupt when its total value of retained earning is equal or greater than 50% of its listed capital and 140 companies are "matched" companies from the same period of listing. In this study on the basis of the financial ratios successfully identified by past studies and availability, 40 indices have been built by using balance-sheet data. Ratios and significances on mean differences for each group is tested and presented in Table 1. Since normality of the data is not met, we employ a non-parametric method to examine the disparity in financial ratios between bankrupt and non-bankrupt groups. To examine the difference in means of a variable between two groups, this study chooses the Mann-Whitney test.

From Table 1, Mann-Whitney test shows whether there is a significant mean difference between groups for each specific variable. These indices reflect different aspects of firm structure and performance: liquidity, turnover, operating structure and efficiency, capitalization and finally profitability. The indices have been calculated as one-year ratios prior to bankruptcy. Subsequent testing of variable selection using Genetic Programming (GP) is done to illustrate that this new transformation will produce more accurate prediction statistically and can be used as an alternative for common ratios.

Genetic programming

Following recent research by Etemadi et al. (2008) we tested these selected variables with Genetic Programming (GP) to obtain fitness function tree and to illustrate that this new transformation will predict more accurate and can be used as an alternative for common ratios even with GP. In the final regressions with fewer significant variables in different classification trees where as expected and we observed that different variables were identified as significant indicators for each procedure from the selected list. For implementing GP process and developing bankruptcy model, newly 2008, released software GeneXproTools software version 4.1 with C# output was used. Crossover and mutation operators were set as 0.44 and 0.05 respectively. Tables 2 and 3 show the best GP model obtained for each approach. These models have been divided in three sub-trees which each tree representing a Gene meaning the model is a chromosome consisting of tree genes. Sum of the returns of sub-trees for a firm should be compared with "Rounding Threshold" for determining the class of the firm. From the classification sub-trees depicted in Table 2, decision trees for SR approach with 95 % accuracy rate obtained.

From the classification sub-trees in Table 3, decision trees for common ratios approach with 89 % accuracy level. For decision making of whether a firm is bankrupt or non-bankrupt through the genetic programming decision tree, a benchmark value of 0.5 is used. If the value for specific training or test firm is greater or equals 0.5, then this firm is marked as "bankrupt firm". If the value of the GP model for a training or test firm is less than 0.5, then this firm is classified as "non-bankrupt firm". Comparison of real class of firms with predicted class by the GP model will determine the accuracy of the model as reported in Table 4.

DISCUSSION

Misclassification error

An alternative to error rate is a misclassification error which is simply a number that is assigned as a penalty for

Table 1. Variables used and comparison of means in two groups.

#	Definition of variables	Original ratios			Transformed ratios		
		Means of non-bankrupt companies	Means of bankrupt companies	TEGM (Sig level)	Means of non-bankrupt companies	Means of bankrupt companies	TEGM (Sig level)
1	EAIT/TA	0.21767	0.05754	0.000	-0.46251	-0.40127	0.125
2	TD/SE	2.54796	2.85370	0.851	0.45211	0.32475	0.043
8	R/S	0.47895	0.02016	0.000	-0.48291	-0.32476	0.023
4	TD/TA	0.41801	1.68147	0.211	-0.35923	-0.41897	0.000
5	CL/SE	2.28985	2.51011	0.874	0.34801	0.42379	0.211
6	CL/TD	0.86624	0.83518	0.234	-0.46399	-0.42776	0.323
7	OA/TA	0.66938	0.71408	0.201	-0.48277	-0.32476	0.783
8	R/S	0.44839	0.35287	0.445	-0.37242	-0.32476	0.527
9	R/Inv	14.5389	0.92670	0.292	0.12682	-0.00443	0.000
10	SE/TD	0.15794	0.05352	0.000	-0.48291	-0.37836	0.025
8	R/S	0.53664	0.62095	0.041	-0.29156	-0.32286	0.000
12	CA/CL	1.17647	1.13654	0.567	0.03709	-0.07049	0.000
13	QA/CL	0.66072	0.50096	0.002	0.33256	0.34225	0.311
8	R/S	0.56170	0.44675	0.001	0.25676	0.48553	0.000
15	NFA/TA	0.23400	0.23308	0.976	0.44839	0.35287	0.445
16	WC/TA	0.07149	0.05773	0.696	0.79395	0.55703	0.313
17	CL/TA	0.68384	0.78209	0.000	0.40458	0.13032	0.000
18	POC/SE	0.44833	0.52932	0.199	0.67863	0.03372	0.008
19	RE/TA	0.05836	-0.02026	0.000	0.08645	-0.19154	0.078
20	EAIT/SE	0.64560	0.78329	0.410	-0.18379	-0.23001	0.000
21	EAIT/S	0.21757	-0.0035	0.000	0.35475	0.33072	0.000
22	EBIT/TA	0.34665	0.24232	0.000	0.23106	0.15496	0.000
23	D/EAIT	0.79203	-1.0847	0.311	0.01546	-0.22849	0.072
24	OI/S	0.22485	0.01508	0.000	0.28985	0.51011	0.874
25	MVE/TA	2.09833	0.59374	0.008	-0.00790	0.05308	0.006
26	EBIT/IE	33.00551	8.02330	0.078	-0.54796	-0.85370	0.213
27	OI/TA	0.19046	0.02571	0.000	0.21672	0.36189	0.107
28	Ca/S	0.20475	-0.00552	0.000	0.22037	0.20962	0.000
29	GP/S	0.29107	0.10676	0.000	0.41375	-0.19657	0.214
30	S/SE	0.53856	0.23618	0.072	0.40633	0.38439	0.844
31	S/NFA	6.03377	6.17452	0.893	0.06507	0.310689	0.234
8	R/S	1.40478	1.13062	0.006	0.000178	0.33498	0.000
33	S/WC	17.8260	5.19366	0.213	0.33448	0.32529	0.008
8	R/S	0.88878	0.78572	0.107	-0.66072	-0.50096	0.002
35	S/Ca	45.2066	122.611	0.245	0.04343	0.00932	0.000
36	IE/GP	3.06774	3.09217	0.987	0.38344	0.25299	0.405
37	Ca/CL	0.10662	0.04838	0.002	0.45923	0.49223	0.292
38	Ca/TA	0.05437	0.03103	0.009	-0.48291	-0.32476	0.023
39	S/GP	0.17323	0.01155	0.000	0.46251	0.40127	0.125
40	BVD/MVE	3.68678	1.73033	0.032	0.45211	0.32475	0.043

BVD: Book Value of Debt

CA: Current assets

EAIT: Earning after income and taxes

GP: Gross profit

Inv: Inventory

MVE: Market value of equity

NI: Net income

OI: Operational income

Ca: Cash Flow

CL: Current liabilities

EBIT: Earnings before interest and taxes

IE: Interest expenses

LA : Liquid assets

NFA: Net Fixed assets

OA: Operating asset

POC: Paid on capital

Table 1. Contd

QA: Quick assets	R: Receivables
RE: Retained earnings	S: Sales
SC: Stock capital	SE: Shareholders' equity
TA: Total assets	TD: Total debt
TL: Total liabilities	TEGM: Test of equity of group mean

Table 2. The best GP model obtained for RB method.

```

const double ROUNDING_THRESHOLD = 0.5;

const double G1C0 = -8.625672;
const double G1C1 = -2.830445;
const double G2C0 = -8.625672;
const double G2C1 = 4.159759;
const double G3C0 = -7.803864;
const double G3C1 = 8.105743;

double dblTemp = 0.0;

1- dblTemp = d[4];

2- dblTemp += pow(((sin(d[2]) >= (((d[0] < 0) || (d[0] < 0)) ? 1 : 0) ? (sin(d[2]) + (((d[0] < 0) || (d[0] < 0)) ? 1 : 0)) :
atan(sin(d[2]) * (((d[0] < 0) || (d[0] < 0)) ? 1 : 0))) * ((G2C0 * d[0]) * (d[0] < d[4] ? (d[0] + d[4]) : atan(d[0] * d[4]))), 3);

3- dblTemp += (((((1 - d[0]) * d[4]) >= (G3C0 >= d[1] ? 1 : 0) ? (((1 - d[0]) * d[4]) + (G3C0 >= d[1] ? 1 : 0)) : (((1 - d[0]) * d[4]) - (G3C0 >=
d[1] ? 1 : 0))) >= (d[2] >= (G3C1 * d[0]) ? d[2] : (G3C1 * d[0])) ? (((1 - d[0]) * d[4]) >= (G3C0 >= d[1] ? 1 : 0) ? (((1 - d[0]) * d[4]) + (G3C0
>= d[1] ? 1 : 0)) : (((1 - d[0]) * d[4]) - (G3C0 >= d[1] ? 1 : 0))) + (d[2] >= (G3C1 * d[0]) ? d[2] : (G3C1 * d[0])) : atan((((1 - d[0]) * d[4]) >=
(G3C0 >= d[1] ? 1 : 0) ? (((1 - d[0]) * d[4]) + (G3C0 >= d[1] ? 1 : 0)) : (((1 - d[0]) * d[4]) - (G3C0 >= d[1] ? 1 : 0))) * (d[2] >= (G3C1 * d[0]) ?
d[2] : (G3C1 * d[0]))));

return (dblTemp >= ROUNDING_THRESHOLD ? 1 : 0);

```

making a particular type of a mistake. An average rate of misclassification can be obtained by weighing each of the errors by the respective error rate. In Table 4 possible classifications and misclassifications are shown and Table 5 shows the comparison accuracy by each classification model respect to different data representations. Table 5, exhibits the summarized accuracy level for GP procedures and clearly the results improved under data transformation procedure. Due to better performance observation of this new transformation, data set is not collected from particular industry type or similar firm size or any outlier deletion applied. Thus, our process is free of any potential explanatory effect errors, which may caused by independent variable's distribution⁴. The new model properties are briefly explained in methodology.

K-fold cross validation

In order to confidently lesson the effects of biasness, we conduct the K-fold cross validation procedure. Each one of the subsets is then in turn as testing set after all other sets combined have been training set on which a tree has been built. This cross validation procedure allows mean error rates to be calculated which gives a useful insight into classifiers decision. This technique is simply k-fold cross validation whereby k is number of data instances. This has advantage of allowing the largest amount of training data to be used in each run and conversely means that the testing procedure is deterministic. With large data sets this is computationally infeasible however and in certain situations the deterministic nature of testing

⁴ Deakin (1967) found that financial ratios might be more normally distributed within a specific industry groups.

⁴ Level of significance of a discriminant function can be proved through different measures like Walks' Lamba, Hotelling's Trace and Pillai's criterion. Results are available from the authors upon request.

Table 6. The transformed ratios still outperform original ratios.

Items	Original ratios (%)	RB method (%)
1	93.94	96.97
2	100	100
3	84.85	90.91
4	100	100
5	75.76	96.97
Average	90.91	96.97
S.E	10.49	3.71

favorable to active companies but also turnover indices are higher for active firms. Assets to operating income ratio are higher for failed firms because of their reduced capital resources. Earning indices, display greater solvency for active firms, even though debts have increased for those firms with respect to go bankrupt. Operating structure ratios for active companies have a lower incidence of interest charges on sales and value added, and higher depreciation charges over gross fixed assets for failed ones. Results suggest that some indicators like earnings to total debt traditionally considered in the empirical analysis but is not being significant in each of the three considered models. Profitability ratios emphasize the overall higher profitability of active enterprises. Finally, additional indices such as market share holders' dividend, sale, return and operating assets are significantly higher for healthy companies.

Conclusions

In this paper we demonstrated the application of new graphical geometric approach for variable representation, data visualization and financial ratios. We believe that graphical analysis will have an increased importance as becoming more and more popular. On the other hand graphical ratio representation can facilitate the acceptance of prediction models in various areas, e.g. finance, medicine, sound and image processing, etc. This will contribute to the development of those areas since better represent reality and provide higher forecasting accuracy. Within our new transformation methodology each company is described by a set of variables X_i , such transformed financial ratios instead of original ratios. Financial ratios, such as debt ratio (leverage) or interest coverage (earnings before interest and taxes) characterize different sides of company operation. They are constructed on the basis of balance sheets and income statements. We used 40 ratios (predictors) computed using the company statements from their corporate bankruptcy data base. The predictors and basic statistics are given in Table 1.

Initially, an unknown classifier function $f: x \rightarrow y$ is

estimated on a training set of companies (x_i, y_i) , $i = 1, 2, \dots, n$. The training sample classification regression represents prediction for companies which are unknown to be survived or gone bankrupt for testing sample. As was demonstrated, this paper presented a complementary perspective on the study of risk and bankruptcy with use of financial ratios. In this paper, a new dimension to risk measurement, bankruptcy, and ratio transformation with the advent of the share risk was proposed. We briefly derived the respective properties of new risk approach components of which were over come of using common ratios limitations. Our simple methodology, called Risk Box index, provided a geometric illustration of our new proposed risk measure and transformation behavior. Our study employed 60 distressed companies with matched sample of another 140 non-failed companies listed in Kuala Lumpur Stock Exchange (KLSE). We found a rise in classification accuracy on application of this new independent variables transformation using Genetic Programming (GP).

According to provided properties of new Share Risk method discussed in section 3 and better numerical results in compare to other studies in section 4, it is strongly suggested the use of this new methodology for ratio analysis, which provided a conceptual and complimentary methodological solution to many of problems associated with the use of ratios. Alternatively, the Share Risk model (Risk Box) can be employed as a tool of analysis in providing a crucial first stage for analysing studies associated with changes in risk patterns, in particular those assumed to be linked with potential bankruptcies. The adaptability of our proposed methodology is emphasised by its applicability for any number of years on sectoral or cross-country studies on risk and bankruptcy studies.

Since previous studies used one and two year prior to bankruptcy, consequently, generalize ability of model with expansion for an additional year is recommended for further studies. Furthermore, as reported by IMF, to undertake such research to understand the capital structures and other financial indicators such as macro and micro economic variables simultaneously that might be effect on firms' performance and eventually can

improve prediction is necessitate, therefore testing above model respect to this issue will be important to be continued.

REFERENCES

- Altman EI (1968). "Financial ratios, Discriminant analysis and the Prediction of Corporate Bankruptcy". *J. Finan.*, 23(4): 589-609.
- Altman E (1971). "Railroad Bankruptcy Propensity". *J. Finan.*, 26(2): 333-345.
- Altman E, Margaine M, Schlosser M, Vernimmen P (1974). "Financial and Statistical Analysis for Commercial Loan Evaluation: A French Experience". *J. Finan. Quantit. Anal.*, 9(2): 195-211.
- Altman E, Haldeman R, Narayan P (1977). "A new model to identify bankruptcy risk of corporations". *J. Bank. Finan.*, 35(2): 29-54.
- Andres JD, Landajo M, Lorca P (2005). "Forecasting business profitability by using classification techniques: a comparative analysis based on a Spanish case". *Eur. J. Operat. Res.*, 167: 518-542.
- Azhar AKM Robert JRE (2006). "On the Measurement of Product Quality in Intra industry Trade". *Rev. World Econ.*, 142(3): 476-495.
- Aziz MA, Dar HA (2006). "Predicting Corporate Bankruptcy: Where we stand?". *Corporate Governance*, 6(1): 18-33.
- Bahraie AR, Ibrahim NA, Ismail M, Azhar AKM (2008). "Financial ratios: A new geometric transformation". *Int. Res. J. Finan. Econ.*, 20: 164-171.
- Barnes P (1982). "Methodological Implications of Non-Normality Distributed Financial Ratios". *J. bus. Finan. Account.*, spring, 9: 51-62.
- Chauhan JN, Ravi V, Chandra DK (2009). "Differential Evolution trained Wavelet Neural Network: Application to bankruptcy prediction in banks". *Expert Systems With Applications*, 36(4): 7659-7665.
- Deakin E (1976). "On the Nature of Distribution of Financial Accounting Ratios: Some Empirical Evidence". *Account. Rev.*, 51: 90-97.
- Edmister R (1972). "An Empirical Test of Financial Ratio Analysis for Small Business Failure Prediction". *J. Finan. Quantitat. Anal.*, 7(2): 1477-1493.
- Etemadi H, Rostamy AA, Farajzadeh H (2008). "A genetic Programming Model for Bankruptcy Prediction: Empirical Evidence from Iran". *Expert Syst. Appl.*, 36(2): 3199-3207.
- Ezzamel M, Mar-Molinero C, Beecher A (1987). "On the Distributional Properties of Financial Ratios". *J. Bus. Finan. Account.* 14: 463-481.
- Foster G (1986). "Financial Statement Analysis", Second edition. Prentice-Hall: Englewood Cliffs, NJ.
- Gu Z (2002). "Analyzing Bankruptcy in Restaurant Industry: A Multiple Discriminant Model". *Hospitality Management*, 21: 25-42.
- Kane GD, Richardson F, Meade N (1996). "Rank Transformations and The Prediction of Corporate Failure". *Working paper* University of Delaware, Newark: DE.
- Koza JR (1992). "Genetic programming: On the programming of computers by means of natural selection". MIT Press, Cambridge, MA.
- Landskroner Y, Ruthenberg D, Pearl D (2006). "Market to Book Value Ratio in Banking: the Israeli Case". *Bank of Israel banking, Supervision Department*, Research Unit. pp. 1-36.
- Lee WC (2004). "Empirical Study for KMV Model in the Financial Distress (in Chinese)". *Sin Pac Finan. J.*, 26: 97-138.
- Martin D (1977). "Early Warning of Bank Failure: A Logit Regression Approach". *J. Bank. Finan.* (1): 249-276.
- McLeay S, Omar A (2000). "The Senility of Prediction Models to the Non-Normality of Bounded and Unbounded Financial Ratios". *British Account. Rev.*, 32: 213-230.
- Northon L, Smith E (1979). "A Comprasion of General Price Level and Historical Cost Financial Astatment in the Prediction of Bankruptcy". *Account. Rev.*, 54(1): 72-87.
- Nwogugu M (2006). "Decision-making, Risk and Corporate Governance: New Dynamic Models/Algorithms and optimization for bankruptcy decisions". *Appl. Math. Comput.*, 179: 386-401.
- Ohelson J (1980). "Financial Ratios and the Probabilistic Prediction of Bankruptcy". *J. Account. Res.*, 34: 109-131.
- Ooghe H, Speaners CH, Vandermoere P (2005). "Business Failure Prediction: Simple-Intuitive Models versus Statistical Models". *Working Paper Series*. Vlerick Leuven Gent, p. 22.
- Perry LG Cronan TP (1986). "A Note on Rank Transformation Discriminant Analysis". *J. Bank. Finan.*, 10: 605-610.
- Ravikumar P, Ravi V (2007). "Bankruptcy Prediction in banks and firms via Statistical and Intelligent Techniques: A Review". *Eur. J. Operat. Res.*, 180(1): 1-28.
- Tsakonas A (2006). "A comparison of classification accuracy of four genetic programming-evolved intelligent structures". *Inf. Sci.*, 176(3): 691-724.
- Taffler RJ (1982). "Forecasting Company Failure in the UK Using Discriminant Analysis and Financial Ratio Data". *J. Royal Stat. Soc.*, 145(3): 342-358.