# On the Predictability of Risk Box Approach by Genetic Programming Method for Bankruptcy Prediction

[1]Alireza Bahiraie, [1]Noor Akma bt Ibrahim and [2]A.K.M. Azhar
[1]Institute for Mathematical Research, University Putra Malaysia,
43400 Serdang, Selangor, Malaysia
[2]Graduate School of Management, University Putra Malaysia,
43400 Serdang, Selangor, Malaysia

**Abstract: Problem statement:** Theoretical based data representation is an important tool for model selection and interpretations in bankruptcy analysis since the numerical representation are much less transparent. Some methodological problems concerning financial ratios such as non-proportionality, non-asymetricity, non-scalicity are solved in this study and we presented a complementary technique for empirical analysis of financial ratios and bankruptcy risk. **Approach:** This study presented new geometric technique for empirical analysis of bankruptcy risk using financial ratios. Within this framework, we proposed the use of a new ratio representation which named Risk Box measure (RB). We demonstrated 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 were the selection of variables (predictors), accuracy of each estimation model and the representation of each model for transformed and common ratios. **Results:** We provided evidence of extent to which changes in values of this index were associated with changes in each axis values and how this may alter our economic interpretation of changes in the patterns and direction of risk components. Results of Genetic Programming (GP) models were compared as different classification models and results showed the classifiers outperform by modified ratios. **Conclusion/Recommendations:** In this study, a new dimension to risk measurement and data representation with the advent of the Share Risk method (SR) was proposed. Genetic programming method is substantially superior to the traditional methods such as MDA or Logistic method. It was strongly suggested the use of SR methodology for ratio analysis, which provided a conceptual and complimentary methodological solution to many problems associated with the use of ratios. Respectively, GP will provide heuristic non linear regression as a tool in providing forecasting regression for studies associated with financial data. Genetic programming as one of the modern classification method out performs by the use of modified ratios. Our new method would be a general methodological guideline associated with financial data analysis.

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

## INTRODUCTION

In classical prediction models a convenient representation of ratios are 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 study presents a complementary perspective on the study of ratios and bankruptcy. 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[17]. 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[1]. Then Altman[1], used Multiple Discriminant Analysis (MDA) and applied to prediction of business failure. Altman[2]

**Corresponding Author:** Alireza Bahiraie, Institute for Mathematical Research, University Putra Malaysia, 43400 Serdang, Selangor, Malaysia Tel: +60-19-2203709

examined railroad bankruptcy propensity and Deakin[9] replicated study Edmister[10] testing the usefulness of financial ratio in order to predict small business failure. Altman, Margaine, Schlosser and Vernimmen[3] developed a model in order to determine the credit worthiness of commercial loan applicants in a cotton and wool textile sector in France. Altman, Haldeman and Narayan[4] developed their classical Z model and named it Zeta Analysis. After DA and Multiple Discriminant Analysis (MDA), the logit and probit models were introduced in Martin[19], Ohelson[23]. 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[6].

Northon and Smith[21] 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[28] 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.

**Problem statement:** According to literature, predictors used in various studies, generally exhibit non-normal distribution and high standard errors[8,20,24]. Some researchers made correction for univariate non-normality and tried to approximate univariate normality by transforming the variables prior to estimation of their models[29,5]. Deakin[9] used log transformation, then square root and log-normal transformation of financial ratios were used by Ooghe[13] and Gu[14]. Other researchers approximate univariate normality by 'trimming' or 'outlier deletion', which involves

segregating outliers by reference to normal distribution[12]. Furthermore rank transformation been used by Perry *et al.*[25] and Kenjegalieva *et al.*[15]. Recently Bahiraie *et al.*[7] used Geometric transformation of ratios, which may become general guidelines concerning the transformation details was discussed.

**Objectives:** Our objective in this study 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, book and market ratio values (X, Y) are used as numerator and denominator of common ratio values and represented as Cartesian coordinates in our constructed modification box in which we derive the isoclines of associated components of bankruptcy risk. This study is regarded as one of the classic studies in this field. We show that Genetic Programming (GP) as one of the modern classification methods outperform by risk box method in compare to ratios.

**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[16]. 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[22]. While bankruptcy prediction can be considered as a classification problem, we provide necessary description of GP with emphasis on its application in classification role[16]. 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. Such a GP individual is shown in Fig. 1.
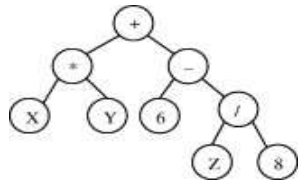
Fig. 1: Tree representation of the program (expression): (X*Y) + 6 - (Z/8)
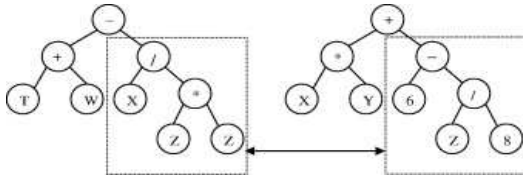


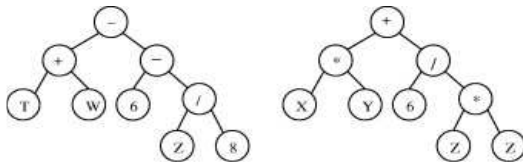Fig. 2: Representation of crossover (parents)



Fig. 3: Representation of crossover (children)

Three genetic operators are mostly used in these algorithms: Reproduction, crossover and mutation. First the reproduction operator simply chooses an individual in the current population and copies it without changes into the new population. 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. Figure 2 and 3 show an example of crossover operators.

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[18]. 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 the depth limit. If this happens mutation just reproduces the original tree into the new generation.
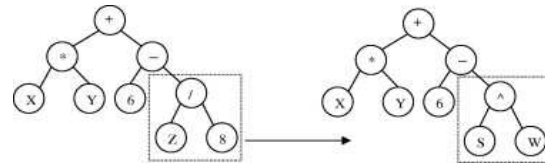


Fig. 4: Representation of mutation

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 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. 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 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 $rf_{max} = n$ 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 = The number of genes and h
T = The head and tail 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. The described procedure is depicted in the flowchart of[27]. Once fitness function is defined, 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. 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 share risk box methodology:** The framework is a two-dimensional box in which associated with ratio values in which pair values of each risk ratios $(X_i, Y_i)$ are represented as Cartesian coordinates. For expositional purposes suppose our proxy for risk chosen is employed by $X_i$ as numerator and $Y_i$ as denominator values of $\frac{X_i}{Y_i}$ ratio. For any number of firms, $\forall i = 1,2,3,\ldots,n$, proposed Share Risk (SRi) is defined as a function of $X_i$ and $Y_i$. Consider a square two-dimensional space that captures all changes in numerator $X_i$ and denominator $Y_i$, for any firm i and any period t where X and Y can be positive, negative or zero (It is applicable to any level of aggregation such as cross-country studies, cross sector and ratios). Assume a hypothetical study of risk covering n years for sector j. For $\forall t = 1,2,3,\ldots,n$, we have: $X_t, Y_t > 0$. 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 which 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.[7], 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 $X_i$ and $Y_i$ value during the period of study. From the definition of TR, NR, OR, SR, we obtain:

$$\max(NR_i) = \max\left(|X_i - Y_i|\right) \le \max(\max X_i - \min Y_i,$$
$$\max Y_i - \min X_i) \le m$$
$$\max(OR_i) = 2\max\left((\min(X_i, Y_i))\right) \le 2m$$
$$\Rightarrow \max SR_i \le 1$$

Each respective risk box will have sides equal to $\max(X_i)$ if for $i \in t$ then $\max(X_i) > \max(Y_i)$ or $\max(Y_i)$ if otherwise. Our exposition of the dimensions of the box is as follows which confirms the elasticity and unit-free nature of SR measure:

**Locus of EQUI TR:** A 45° line from the origin bisects the box into two equal triangles. This positive slope diagonal is the locus of balanced risk where X = Y, TR equals OR, SR equals unity and NR equals zero. This is the risk components' axis of asymmetry[26].

The two triangular planes in the box consists of an upper triangle containing coordinate points $(X_i, Y_i)$ where $X_i > Y_i$ in and points $Y_i > X_i$ in the lower triangle. A fix value TR = TR* implies = 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 asymmetry.

**Locus of EQUI NR:** Recall that Net Risk NR = |X-Y|. The line 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 Y-X = NR* or X = Y+NR*. These two 45° lines from the contour are corresponding to NR*. Increasing the value of constant NR* moves both segments higher up their respective axis, away from the central NR* line:

$$\forall X = Y \text{ then } NR = 0$$
$$\forall X > Y \text{ then } NR = |X-Y| = X-Y \text{ i.e., } x = Y+NR$$
$$\forall X < Y \text{ then } NR = |X-Y| = Y-X \text{ i.e., } X = Y-NR$$

Increasing the value of constant NR* moves both segments higher up their respective axis, away from the central NR* line. 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 (Fig. 5). Consequently, locus of EQUI NR values is perpendicular to lines of EQUI TR ( $m_{TR} = \frac{-1}{m_{NR}}$ ).
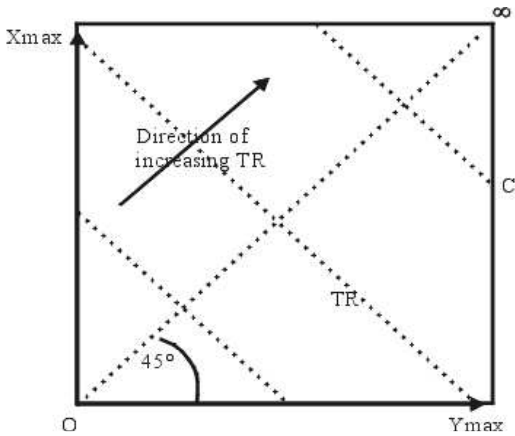
Fig. 5: Total risk isoclines

**Locus of EQUI OR:** Recall that overlapping risk OR = 2 min(X,Y), below the central 45° line, OR = 2X that remains constant for constant X. Above the line OR = 2Y which remains constant for constant Y.

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

Thus the EQUI corresponding to constant overlapping risk OR* is L-shaped (Fig. 6), the kink occurring along the central 45° line. As OR* increases, the kink moves up the line, away from the origin.

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

- Below the line, Y>X and thus $SR = \dfrac{2X}{X+Y}$. The EQUI corresponding to a constant value SR* is defined by the relation SR*(X+Y) = 2B, which can be solved for X to yield $X = \dfrac{SR^*}{2-SR^*} Y$. Thus this segment of the EQUI is a ray from the origin with constant slope $\gamma = \dfrac{SR^*}{2-SR^*}$. Since 0≤SR≤1, we have 0≤γ≤1, showing that the ray passes between the central 45° line and the horizontal axis

- Above the central 45° line on the other hand we have $SR = \dfrac{2Y}{X+Y}$. Given a constant value SR* we obtain $X = \gamma^{-1}Y$, whose slope $\gamma^{-1}$ satisfies 1≤$\gamma^{-1}$<∝
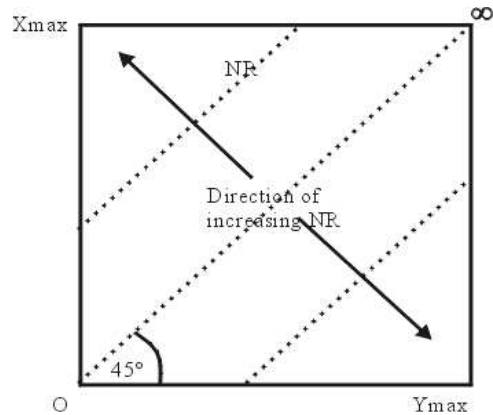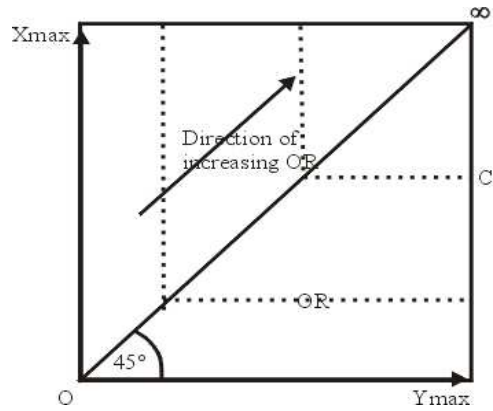


Fig. 6: Net risk isoclines



Fig. 7: Overlapping risk isoclines

Thus the EQUI corresponding to a particular value SR* consists of two rays in the positive quadrant meeting at the origin, with slopes $\gamma$ and $\gamma^{-1}$. In Fig. 8 these rays are shown as OC and OB. Note that the symmetry of the diagram about the central 45° line implies that the angles $\theta_1$ and $\theta_2$ are equal.

In Fig. 7, relationships between the four risks measures and slopes $\gamma$ and $\gamma^{-1}$, consider rays OB and OC subtending the angles $\theta_1$, $\theta_2$ measured from the symmetry axis. 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*:

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

Hence:

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

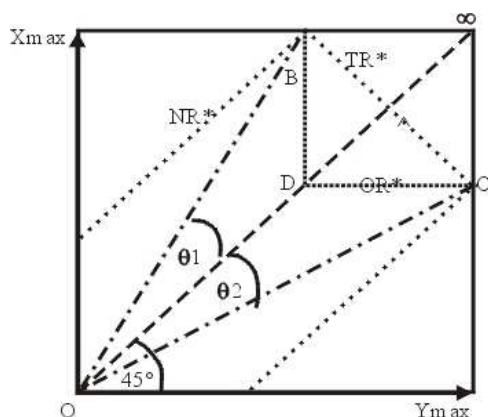$$\Rightarrow SR^* = 1 - \tan\theta_1$$

Fig. 8: Share risk isoclines

These will confirm that SR values are constant along any ray from origin and the two extreme case the two extreme cases are (i) $\theta_1 = \theta_2 = 45°$, in which case SR = 0 and either the Y value or the X value is zero and (ii) $\theta_1 = \theta_2 = 0$, in which case SR = 1 and X = Y.

### RESULTS

**Data collection:** The database used in our illustrative empirical study consists of 200 Malaysian companies from Kuala Lumpur Stock Exchange (KLSE) which 60 companies went bankrupt and 140 companies are non-bankrupt companies from the same period of listing.

**Variables:** 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.

**Significance mean test:** Ratios and significances on mean differences for each group is tested and presented in Table 1. These indices reflect different aspects of firm structure and performance and have been calculated as one-year ratios prior to bankruptcy.

**Genetic programming:** variable selection using Genetic Programming (GP) is to illustrate that this new transformation will produce more accurate prediction statistically and can be used as an alternative for common ratios. Following recent research by Etemadi *et al.*[11] 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.
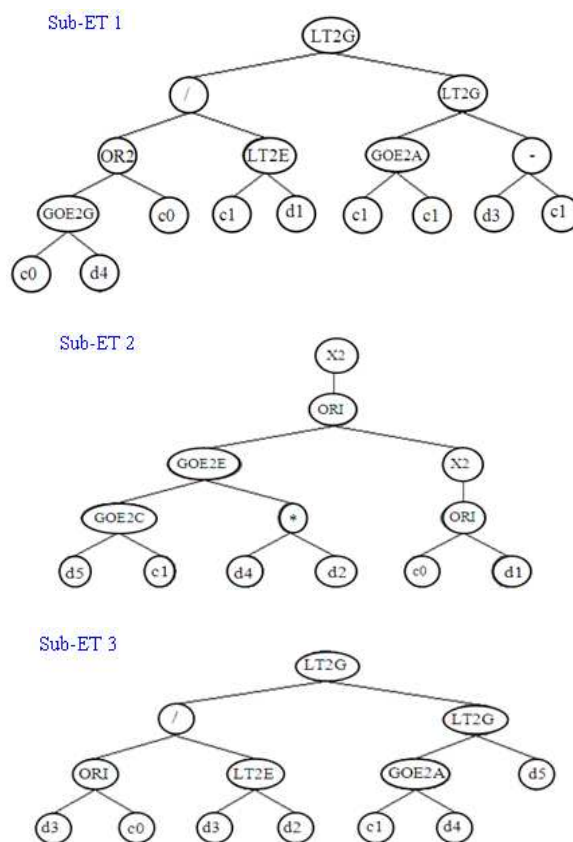


Fig. 9: The best GP model obtained for SR method

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, GeneXproTools software version 4.1 was used. Crossover and mutation operators were set as 0.44 and 0.05 respectively.

Figure 9 and 10 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 Fig. 9, decision trees for SR approach with 95% accuracy rate obtained.

From the classification sub-trees in Fig. 10, decision trees for common ratios approach with 89% accuracy level. Variables, which are found significant in each sub-trees are represented in Table 2.

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

| | Original ratios | | | Transformed ratios | | |
|---|---|---|---|---|---|---|
| Definition of variables | Means of non-bankrupt companies | Means of bankrupt companies | TEGM (Sig level) | Means of non-bankrupt companies | Means of bankrupt companies | TEGM (Sig level) |
| EAIT/TA | 0.21985 | 0.05165 | 0.00000 | 1.39008 | 1.47417 | 0.02500 |
| TD/SE | 2.32591 | 2.99969 | 0.05100 | 0.17897 | 0.33310 | 0.04300 |
| R/S | 0.53916 | 0.01808 | 0.00000 | 1.29721 | 1.49609 | 0.02300 |
| TD/TA | 0.64600 | 0.78450 | 0.01100 | 1.17700 | 1.10775 | 0.00000 |
| CL/SE | 2.07355 | 2.60760 | 0.87400 | 0.13713 | 0.28837 | 0.21100 |
| CL/TD | 0.87258 | 0.83419 | 0.23400 | 1.06371 | 1.08290 | 0.32300 |
| OA/TA | 0.54037 | 0.62549 | 0.20100 | 1.22981 | 1.18725 | 0.08300 |
| R/S | 0.64792 | 0.40207 | 0.44500 | 1.28176 | 1.31233 | 0.52700 |
| R/Inv | 64191.96287 | 60.03362 | 0.00000 | 1.00444 | 1.12682 | 0.00000 |
| SE/TD | 0.81727 | 0.33380 | 0.00000 | 1.17897 | 1.33310 | 0.02500 |
| E/TA | 0.37868 | 0.24421 | 0.04100 | 1.31066 | 1.37789 | 0.00000 |
| CA/CL | 1.37059 | 1.13940 | 0.56700 | 0.07046 | 0.03709 | 0.00000 |
| QA/CL | 0.88108 | 0.49283 | 0.00200 | 1.14017 | 1.25456 | 0.31100 |
| QA//CA | 0.59121 | 0.44456 | 0.00100 | 1.20439 | 1.27772 | 0.00000 |
| NFA/TA | 0.22169 | 0.22309 | 0.97600 | 1.38916 | 1.38846 | 0.00500 |
| WC/TA | 0.11022 | 0.06320 | 0.69600 | 1.44489 | 1.46840 | 0.31300 |
| CL/TA | 0.56389 | 0.65641 | 0.00000 | 1.21806 | 1.17179 | 0.00000 |
| POC/SE | 0.53201 | 0.57998 | 0.19900 | 1.23447 | 1.10467 | 0.00800 |
| RE/TA | 0.06492 | -0.02391 | 0.00000 | 1.46754 | 1.51196 | 0.07800 |
| EAIT/SE | 0.53080 | 0.17283 | 0.41000 | 1.24864 | 1.46834 | 0.00000 |
| EAIT/S | 0.27192 | -0.04296 | 0.00000 | 1.36405 | 1.50608 | 0.00000 |
| EBIT/TA | 0.17862 | 0.00639 | 0.00000 | 1.41069 | 1.49680 | 0.00000 |
| D/EAIT | 2.02476 | 0.92434 | 0.31100 | 1.11523 | 0.24383 | 0.07200 |
| OI/S | 0.28441 | -0.01012 | 0.00000 | 1.35780 | 1.49572 | 0.87400 |
| MVE/TA | 0.04992 | 0.05746 | 0.00800 | 1.47504 | 1.47127 | 0.00600 |
| EBIT/IE | 4496.20577 | -43.01149 | 0.00000 | 0.59907 | 0.55253 | 0.21300 |
| OI/TA | 0.19620 | 0.02240 | 0.00000 | 1.40190 | 1.48880 | 0.10700 |
| Ca/S | 0.18568 | 0.05238 | 0.00000 | 1.43579 | 1.47381 | 0.00000 |
| GP/S | 0.35047 | 0.09577 | 0.00000 | 1.32476 | 1.45211 | 0.21400 |
| S/SE | 3.01240 | 3.06662 | 0.07200 | 0.20837 | 0.29016 | 0.84400 |
| S/NFA | 10.53526 | 5.98830 | 0.89300 | 0.33491 | 0.31069 | 0.03400 |
| S/CA | 1.37378 | 1.07683 | 0.00600 | 0.06508 | 0.00171 | 0.00000 |
| S/WC | 14.68814 | 5.10868 | 0.21300 | 0.40842 | 0.44656 | 0.00800 |
| S/TA | 0.88013 | 0.75620 | 0.10700 | 1.08629 | 1.12527 | 0.00200 |
| S/Ca | 37.35053 | 121.39542 | 0.00500 | 0.43579 | 0.47381 | 0.00000 |
| IE/GP | -0.32201 | -1.87164 | 0.08700 | 1.57508 | 1.60523 | 0.40500 |
| Ca/CL | 0.17422 | 0.05219 | 0.00200 | 1.41614 | 1.47391 | 0.29200 |
| Ca/TA | 0.08993 | 0.03416 | 0.00900 | 1.45503 | 1.48292 | 0.02300 |
| S/GP | 4.81397 | 24.35715 | 0.00000 | 0.32476 | 0.45211 | 0.12500 |
| BVD/MVE | 81.75837 | 73.27468 | 0.03200 | 0.46128 | 0.46254 | 0.04300 |

BVD: Book Value of Dept. ; CA: Current assets; EAIT: Earning After Income and Taxes; GP: Gross Profit; Inv: Inventory; MVE: Marked Value of Equity; NI: Net Income; OI: Operational Income; QA: Quick Assets; RE: Retained Earnings; SC: Stock Capital; TA: Total Assets; 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; R: Receivables; S: Sales; SE: Shareholders' Equity; TEGM: Test of Equity of Group Mean

Table 2: Predictors used by Gp

| Under SR method | | Under original ratios | |
|---|---|---|---|
| T1C0 | -4.22037 | T1C1 | -5.57389 |
| TIC1 | 1.28492 | T2C0 | 4.37029 |
| T2C0 | -1.98109 | T3C1 | -2.49165 |
| T2C1 | 3.37945 | | |
| T3C0 | 4.36542 | | |
| T3C1 | -2.31471 | | |
| d1 | R/S | d0 | R/S |
| d2 | SE/TD | d1 | SE/TD |
| d3 | QA/CA | d3 | QA/CA |
| d4 | POC/SE | d5 | OI/S |
| d5 | OI/TA | d6 | OI/TA |

The representation of a solution for the problem provided by the GP algorithm is in the form of decision sub-tree. Each node of this tree is a function node taking one of the values from the set +, -, *, ^, EXP and etc. Some of operators which were used in our study are shown in Table 3. 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".
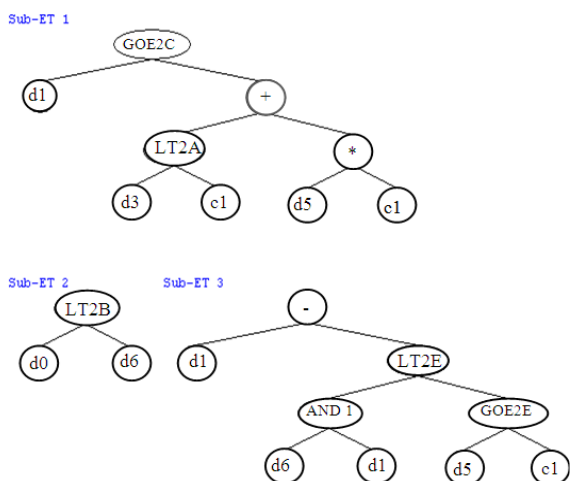
Fig. 10: The best GP model obtained for original ratios

Table 3: Function nodes reported in decision trees in Fig. 9 and 10

| Representation | Name | Representation | Name |
|---|---|---|---|
| + | Addition | Exp | Exponential |
| - | Subtraction | E | e |
| * | Multiplication | Pi | Π |
| / | Division | Log | Logarithmic |
| Sin | Sine | Asin | Arcsine |
| Cos | Cosine | Acos | Arccosine |
| Tan | Tangent | Atan | Arctangent |
| Cot | Cotangent | Acot | Arc cotangent |
| Sec | Secant | Asinh | Arcsine hyperbolic |
| Csc | Cosecant | Acosh | Arccosine hyperbolic |
| Sinh | Sine hyperbolic | Atanh | Arctangent hyperbolic |
| Cosh | Cosine hyperbolic | Acoth | Arc cotangent hyperbolic |
| Acsch | Arc cosecant Hyperbolic | Csch | Cosecant hyperbolic |
| Asech | Arc secant hyperbolic | Sech | Secant hyperbolic |
| 4RT | $X^{1/4}$ | X3 | $X^3$ |
| 5RT | $X^{1/5}$ | X4 | $X^4$ |

Table 4: Possible classification response

| Symbol | Actual | Prediction |
|---|---|---|
| 11 | 1: Dsitress | 1: Dsitress |
| 10 | 1: Dsitress | 0: Non-distress |
| 01 | 0: Non-distress | 1: Dsitress |
| 00 | 0: Non-distress | 0: Non-distress |

**Misclassification cost:** An alternative to error rate is a misclassification cost which is simply a number that is assigned as a penalty for making a particular type of a mistake. An average cost of misclassification can be obtained by weighing each of the costs by the respective error rate. Computationally this means that errors are converted into costs by multiplying an error by its misclassification cost. 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: Comparison accuracy of GP trees

| Items | Original ratios | SR method |
|---|---|---|
| 11 | 51 | 56 |
| 10 | 9 | 4 |
| 00 | 127 | 134 |
| 01 | 13 | 6 |
| Total accuracy (%) | 89 | 95 |

Table 6: The transformed ratios still outperform original ratios

| Items | Original ratios (%) | SR method (%) |
|---|---|---|
| 1 | 93.94 | 96.97 |
| 2 | 100.00 | 100.00 |
| 3 | 84.85 | 90.91 |
| 4 | 100.00 | 100.00 |
| 5 | 75.76 | 96.97 |
| Average | 90.91 | 96.97 |
| SE | 10.49 | 3.71 |

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 form 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 distribution Deakin[9].

**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 results in weir errors. Further, k-fold crosses validation primary method for estimating turning parameters, dividing the data into k equal parts. For each k = 1,2…, k fit the model with parameters to the other k-1 parts and the kth part as testing sample. In our experiment we set our sample to 5-fold accuracy results. Table 6 represents the comparison of 5-fold accuracy results.

Description results highlight the following evidences that under transformation process better classification accuracy results achieved. While the pattern of not only liquidity variation is alternatively 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. Capitalization ratios clearly reflect the superior growth of active versus failed firms. 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.

## DISCUSSION

In this study we demonstrated the application of new graphical geometric approach for variable representation and data visualization. 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. 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→y is estimated on a training set of companies $(x_i, y_i)$, I = 2,…,n. The training sample classification regression represents prediction for companies which are unknown to be survived or gone bankrupt for testing sample.

## CONCLUSION

This study presented a complementary perspective on the study of risk and bankruptcy with use of financial ratios. In this study, 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). 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.

## REFERENCES

1. Altman, E.I., 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. J. Finance, 23: 589-609. http://www.defaultrisk.com/pa_score_01.htm
2. Edward I. Altman, 1971. Railroad bankruptcy propensity. J. Finance, 26: 333-345. http://ideas.repec.org/a/bla/jfinan/v26y1971i2p333-45.html
3. Altman, E.I., M. Margaine, M. Schlosser and P. Vernimmen, 1974. Financial and statistical analysis for commercial loan evaluation: A french experience. J. Financ. Quant. Anal., 9: 195-211. http://www.jstor.org/pss/2330096
4. Altman, E.I., R.G. Haldeman and P. Narayan, 1977. ZETA$^{TM}$ analysis a new model to identify bankruptcy risk of corporations. J. Bank. Finance, 1: 29-54. DOI: 10.1016/0378-4266(77)90017-6
5. Azhar, A.K.M. and R.J.R. Elliott, 2006. On the measurement of product quality in intra industry trade. Rev. World Econ., 142: 476-495. DOI: 10.1007/s10290-006-0077-5
6. Aziz, M.A. and H.A. Dar, 2006. Predicting corporate bankruptcy: Where we stand?. Corporate Governance, 6: 18-33. DOI: 10.1108/14720700610649436
7. Bahiraie, A.R., N.A. Ibrahim, M. Ismail and A.K.M. Azhar, 2008. Financial ratios: A new geometric transformation. Int. Res. J. Finance Econ., 20: 164-171. http://www.citeulike.org/user/pgeymueller/article/1560516

8.  Barnes, P., 1982. Methodological implications of non-normality distributed financial ratios. J. Bus. Finance Account, 9: 51-62. http://www3.interscience.wiley.com/journal/11956 3525/abstract?CRETRY=1&SRETRY=0

9.  Deakin, E., 1976. On the nature of distribution of financial accounting ratios: Some empirical evidence. Account. Rev., 51: 90-97. http://www.jstor.org/pss/245375

10. Edmister, R.O., 1972. An empirical test of financial ratio analysis for small business failure prediction. J. Financ. Quant. Anal., 7: 1477-1493. http://www.jstor.org/pss/2329929

11. Etemadi, H., A.A.A. Rostamy and H.F. Dehkordi, 2009. A genetic programming model for bankruptcy prediction: Empirical evidence from Iran. Expert Syst. Appli., 36: 3199-3207. http://portal.acm.org/citation.cfm?id=1465259

12. Ezzamel, M., C. Mar-Molinero and A. Beecher, 1987. On the distributional properties of financial ratios. J. Bus. Finance Account., 14: 463-481. http://www3.interscience.wiley.com/journal/11947 8082/abstract

13. Ooghe, H., C.H. Speanjers and P. Vandermoere, 2005. Business failure prediction: Simple-intuitive models versus statistical models. Working Paper, Vlerick Leuven Gent. http://ideas.repec.org/p/rug/rugwps/05-338.html

14. Gu, Z., 2002. Analyzing bankruptcy in resturant industry: A multiple discriminant model. Hospital. Manage., 21: 25-42. DOI: 10.1016/S0278-4319(01)00013-5

15. Kenjegalieva, K., R. Simper, J.T. Weyman and V. Zelenyuk, 2009. Comparative analysis of banking production frameworks in eastern european financial markets. Eur J. Operat. Res., 198: 326-340. DOI: 10.1016/j.ejor.2008.09.002

16. Koza, J.R., 1992. Genetic Programming: On the Programming of Computers by Means of Natural Selection. 1st Edn., MIT Press, Cambridge, MA., ISBN: 10: 0262111705, pp: 840.

17. Landskroner, Y., D. Ruthenberg and S. Pearl, 2006. Market to Book Value Ratio in Banking: The Israeli Case. http://www.bankisrael.gov.il/deptdata/pikuah/paper s/dp0606e.pdf

18. Maronna, R., D. Martin and V. Yohai, 2006. Robust Statistics. 1st Edn., Wiley and Sons, New York, ISBN: 10: 0470010924, pp: 436.

19. Martin, D., 1977. Early warning of bank failure: A logit regression approach. J. Bank. Finance, 1: 249-276. http://ideas.repec.org/a/eee/jbfina/v1y1977i3p249-276.html

20. McLeay, S. and A. Omar, 2000. The senility of prediction models to the non-normality of bounded and unbounded financial ratios. Br. Account. Rev., 32: 213-230. DOI: 10.1006/bare.1999.0120

21. Northon, C.L. and R.E. Smith, 1979. A comparison of general price level and historical cost financial statement in the prediction of bankruptcy. Account. Rev., 54: 72-87. http://www.jstor.org/pss/246235

22. Nwogugu, M., 2006. Decision-making, risk and corporate governance: new dynamic models/algorithms and optimization for bankruptcy decisions. Applied Math. Comput., 179: 386-401. http://cat.inist.fr/?aModele=afficheN&cpsidt=1808 4436

23. Ohlson, J.A., 1980. Financial ratios and the probabilistic prediction of bankruptcy. J. Account. Res., 18: 109-131. http://www.jstor.org/pss/2490395

24. Ooghe, H., C. Speanjers and P. Vandermoere, 2005. Business failure prediction: Simple-intuitive models versus statistical models. http://ideas.repec.org/p/rug/rugwps/05-338.html

25. Perry, L.G. and T.P. Cronan, 1986. A note on rank transformation discriminant analysis. J. Bank. Finance, 10: 605-610. http://app.cul.columbia.edu:8080/ac/bitstream/1002 2/AC:P:123/1/fulltext.pdf

26. Schattsneider, D., 1978. The plane symmetry groups: Their recognition and notation. Am. Math. Monthly, 85: 439-450. http://www.eric.ed.gov/ERICWebPortal/custom/po rtlets/recordDetails/detailmini.jsp?_nfpb=true&_& ERICExtSearch_SearchValue_0=EJ187452&ERIC ExtSearch_SearchType_0=no&accno=EJ187452

27. Tsakonas, A., 2006. A comparison of classification accuracy of four genetic programming-evolved intelligent structures. Inform. Sci., 176: 691-724. ftp://ftp.cs.bham.ac.uk/pub/authors/W.B.Langdon/ biblio/gp-html/Tsakonas_2006_IS.html

28. Taffler, R.J., 1982. Forecasting company failure in the UK using discriminant analysis and financial ratio data. J. R. Stat. Soc., 145: 342-358. http://www.jstor.org/pss/2981867

29. Watson, C.J., 1990. Multivariate distributional properties, outliers and transformation of financial ratios. Account. Rev., 65: 682-695. http://www.jstor.org/pss/247957