

Financial Distress Prediction Using Distress Score as a Predictor

Maryam Sheikhi (Corresponding author)

Management Faculty, Central Tehran Branch, Islamic Azad University, Tehran, Iran

E-mail: sheikhi_ml@yahoo.com

Mirfeiz Fallah Shams

Department of business management, Central Tehran Branch, Islamic Azad University, Tehran, Iran

E-mail: fallahshams@gmail.com

Zeinab Sheikhi

Management Faculty, Central Tehran Branch, Islamic Azad University, Tehran, Iran

E-mail: sheikhi_zeinab@yahoo.com

Received: June 9, 2011

Accepted: August 15, 2011

Published: January 1, 2012

doi:10.5539/ijbm.v7n1p169

URL: <http://dx.doi.org/10.5539/ijbm.v7n1p169>

Abstract

Financial distress can be the reflection of corporation's management condition. Consequently the distress score of corporations should be considered as a new predictor variable in predicting the financial distress.

The analysis of ROC curve, among the models employed to compare the effectiveness of different statistical models, is often used in the fields of psychology and bio-physics in order to summarize the discriminatory of a diagnostic test and also to compare the performance of different models for binary outcomes. Therefore, concerning the topic of this research and the use of ROC curves in predicting the financial distress of corporations, we use logit models to study the financial distress of the manufacturing corporations in Tehran Stock Exchange. We also compare the accuracy of the prediction method with financial distress score variable to the method without this variable.

Concerning the accuracy of prediction and classification, the results of this research show that the accuracy of prediction can be enhanced by using the distressed score, gained from DEA, as a new predictor variable in predicting the financial distress.

Keywords: Financial distress prediction, Logit, Receiver operating characteristic curve analysis, Data Envelopment Analysis, Distress score

1. Introduction

Therefore considering the effects of financial distress of corporations on the beneficiary groups, presenting the models of financial distress prediction, has been one of the most attractive fields of financial and economic researches.

In all financial failure prediction models, variable selection, also called feature selection, is a fundamental problem that has significant impact on the prediction accuracy of the models. Under the assumption that a corporation's financial statements appropriately reflect all of its characteristics, current prediction models select variables directly from various financial ratios defined based on information that appears in the corporation's financial statements (Xu, X. and Wang, Y., 2009). Although financial ratios, originated in a corporation's financial statements, can reflect some characteristics of a corporation from various aspects to a certain extent. It is widely recognized that a main cause of financial failure is poor management (Gestel et al., 2006). Therefore, we believe that the financial distress of corporation can be a reflection of its management condition. As a result the score of the corporation's distress should be selected as a new variable in the prediction model of financial distress of the corporation. Therefore the variables used in this research are divided into five groups: liquidity

ratios, profitability ratios, activity ratios, leverage ratios, and the distress score of the corporations.

In this article we present the financial distress prediction model as the predictor variable by using the financial distress score which is driven from the output-oriented BCC model. So we consider the manufacturing corporations accepted in Tehran stock exchange whose information was available during the years 2001 to 2008, and collected the data related to them in this period of time.

This research is about ROC curves performance in predicting the financial distress of the corporations. We used logit model to study the financial distress and we compare the accuracy of the prediction model with distress score variable to the model without this variable.

The remainder of this paper is organized as follows: Section 2 presents a detailed literature survey that discusses bankruptcy evaluation models. In Section 3, we give a brief exposition of ROC curve analysis. Section 4 describes the data sources, sample selection criteria, and variable selection process. Section 5 presents the empirical results. Section 6 concludes this study.

2. Literature Review

Searching the history of the researches conducted in the financial distress prediction and bankruptcy shows that many researches have been undertaken in this field. These researches have considerable differences in the number and the kinds of predictive factors and the structure of model-making. By presenting a univariate model and multiple discriminant analysis (MDA), Beaver (1966) and Altman (1968) respectively have great effects on the prediction and classification of the distressed and non-distressed corporations. However, the credit and efficiency of this models are dependent on some statistic hypothesizes. For elimination the defects of MDA, Ohlson (1980) has suggested logistic regression (Logit). This traditional method has been used as the best artistic method for classification and prediction (see Barniv et al. (2002), Poon, Firth, and Fung (1999), and West (2000)). Then by achieved improvements in other fields of science like mathematics and computer during the last decades, making use of artificial neural networks, Fuzzy logic, and data envelopment analysis (DEA) for designing models have attracted the researchers.

After it was noted largely in the financial distress prediction in banking industry, Data Envelopment Analysis has increasingly made use of in the investigating the financial condition and prediction of financial distress of corporations.

DEA was first proposed by Charnes et al. (1978). Cooper et al. (2006) provided complete knowledge on recent DEA developments. Gattoufi et al. (2004) listed more than 3000 previous DEA contributions. Instead of designing an independent model based on DEA, Xu&Wang (2009) considered the calculated performance score in this method with other financial ratios as a predictor variable. Initially they designed three models based on discriminant analysis, logit and decision tree by the use of financial ratios to test the ability of prediction of financial ratios. Then the calculated efficiency score for each corporation entered the discriminant analysis models, logit models and decision tree as a new variable. The comparison of the new model results to the elementary model results showed that entering the efficiency score to the former models enhances the ability of prediction in the elementary models.

Studying the past researches shows that different researchers use different composition of inputs and outputs which is one of the defects of DEA. It should be noted that in this article the method of choosing input and output is based on the Premachandra, Bhabra & Sueyoshi' research, (2009). In the context of bankruptcy assessment, the smaller (inferior) values in the financial ratios, which could possibly cause financial distress, are considered to be input variables. In contrast, the larger (superior) values in those ratios, which could cause financial distress, are classified as output variables. This analytical feature due to the selection of inputs and outputs is different from the conventional use of DEA. In the conventional DEA-based production analysis, productive performers consist of an efficiency frontier and insufficient performers exist within a production possibility set shaped by the efficiency frontier. In contrast, this study takes an approach that is opposite to the conventional production analysis because we are interested in distress prediction.

The frontier used in this study is a "distress frontier" (not an efficiency frontier found in the conventional use of DEA-based production analysis) which contains many distress corporations (poor performers). Non-distressed (healthy) corporations are expected to exist inside a "distress possibility set", which is shaped by the distress frontier (See Premachandra et al (2009), Xu, X. and Wang, Y. (2009)).

The mathematical definition of distress possibility set is like the production possibility set in economics. However, the nature of outputs in predicting the financial distress is the opposite point of the production analysis outputs meaning that in production analysis the higher the outputs, the better; on the contrary in the financial

distress prediction the lower the outputs, the better. In addition, an opposite description can be applied to the use of inputs.

The analysis of ROC curve, among the models employed to compare the efficiency of different statistic models, is often used in the fields of psychology and bio-physics in order to measure the accuracy of a diagnosis test and also to compare the operation of different models of two-category results (Lloyd, 1998; Marzban,1998; Pepe,2000).

3. An Overview of ROC Curve Analysis

For non-probabilistic or categorical prediction, the four elements of a 2x2 contingency table provide a complete representation of performance for any number of classes (Wilks, 1995). Although the table has four elements, there are only two degrees of freedom, i.e. distressed&non-distressed firms. The contingency table, in turn, can be reduced to a host of scalar measures of performance, but it is difficult to display and interpret (Tang and Chi, 2005).

		Condition	
		Distressed (+)	Non-Distressed (-)
Test outcome	Positive results	TP	FP
	Negative results	FN	TN

$TP + FN = 100\%$

$TN + FP = 100\%$

True positive (TP): the distressed corporation predicted distressed according to the model

True negative (TN): the non-distressed corporation predicted non-distressed according to the model

False negative (FN): the distressed corporation falsely predicted non-distressed according to the model

False positive (FP): the non-distressed corporation falsely predicted distressed according to the model

The aim is minimizing errors (FN & FP). There is a relation between FN and FP: if FP decreases, FN increases and vice versa. ROC curve indicates that minimization of FN and FP independently is impossible.

3.1 Sensitivity and Specificity

In order to preclude any loss data (due to the decrease from two degrees of freedom to one), the context of two scalar measure is employed in this paper. The true positive rate (TPR) or the 'sensitivity' (the probability of correctly identifying a positive) and true negative rate (TNR) or the 'specificity' (the probability of correctly identifying a negative) for functions are two common measures. Their complements are the false negative rate (FNR) and the false positive rate (FPR) for functions. In fact, they estimate two varied aspects of any dichotomous test (Tang and Chi, 2005).

Because the denominators are different, their measures are not complements of one another. In our case, the sensitivity and specificity of the dichotomous test are defined as follows:

$$\text{Sensitivity} = \frac{n_{TP}}{n_{TP} + n_{FN}} \tag{1}$$

$$\text{Specificity} = \frac{n_{TN}}{n_{TN} + n_{FP}} \tag{2}$$

Generally TPR indicates the ability of the test to predict all the distressed corporations rightly. The higher the sensitivity, the lesser the amount of the false negative rate. TNR shows the ability of the test to predict rightly all the corporations which are not distressed. The higher the TNR, the lower the FPR. A good discriminatory test should have high sensitivity and high specificity.

3.2 The Area under the ROC Curve

The area under the ROC curve (AUC) is an important index of a general measure of features of the underlying distribution of forecasts. This measure is equivalent to the Gini coefficient (Thomas et al. 2002) and also to the Mann-Whitney-Wilcoxon two-independent sample non-parametric test statistic (Hanley & McNeil, 1982) and is referred to in the literature in many ways, including AUC, the c-index or c-statistic, and $\hat{\theta}$.

Let D and N represent the measurements for distress and non-distress subject, with high values suggesting a positive result and low values a negative result. Then, the AUC is $\hat{\theta} = \text{Prob}(D > N)$. The larger the area under the ROC curve, the more the accuracy of the model prediction. In the case of perfect prediction, the area beneath will equal 1.0, while an area of 0.5 reflects random forecasts. In Equation (3), the non-parametric approximation of $\hat{\theta}$ is

$$\hat{\theta} = \frac{1}{N_D N_N} \sum_{(D,N)} \mu_{D,N} \quad (3)$$

In Equation (3), N_D and N_N are the number of distressed and non-distressed firms, respectively, and

$$\mu_{D,N} = \begin{cases} 1 & \text{if } S_D > S_N \\ 1/2 & \text{if } S_D = S_N \\ 0 & \text{if } S_D < S_N \end{cases} \quad (4)$$

With S_D , S_N being the score of a distress and non-distress subject, respectively (see Bamber, 1975; Tang and Chi, 2005).

4. Research Design

In this section, our data sources, sample selection criteria, and the process of selecting the variable are going to be discussed. The design of the research is shown in figure 1 (see Tang and Chi, 2005).

4.1 Research Data

The field of study in this research is the manufacturing corporations accepted in Tehran Securities & Exchange Organization; therefore the needed data has been collected from the database of Tehran Securities & Exchange Organization. Regarding the studied period of time (2001-2008) the data of the 304 corporations, 79 of which were distressed and 225 of which non-distressed, was available.

Data collection of the distressed and non-distressed corporations has been done in the T, T-1 and T-2 years. The T, T-1 and T-2 years for the distressed corporations are defined respectively as the year of the occurrence of the financial distress (the corporations have lost at least half of their capital in 2 successive years), one and two years before that. Regarding the non-distressed corporations, respectively the year the corporation had the most profit and one and two years before that. The selected set for analysis is divided randomly into two set: Training (80 percent) and testing (20 percent) samples. The former is for making and estimating the models and the parameters related to prediction methods, and the latter is used for studying the righteousness and accuracy of the prediction of the model.

4.2 Variable Selection

In the models of the prediction of financial distress, the statistic models of the predictive variables are specifically selected by the use of some statistic tests to make sure of their being normal, independent and etc.

In this article the predictor variables are divided into two parts, one of them is the distress score of corporations gained from using the output-oriented BCC model, the other includes financial ratios derived from financial statements. It is worth noting that the selected ratios should reflect the specificities of profitability, liquidity, activity, and leverage. We have chosen 20 financial ratios whose performances in the former researches of financial distress prediction have been proved, as the potential prediction variables. These variables are shown in table 1.

5. Results

5.1 Statistic Results of the Variables

In order to select the financial ratios, which are informative and closer to the financial condition of the corporation, we used Mann-Whitney non-parametric test in order to check the significance of the differences of each ratio in non-distressed and distressed corporations. Table 2 shows that most of the variables except X_6 , X_9 , X_{11} , X_{16} , X_{18} , X_{20} are significant at 0.95 confidence level.

After investigating the meaningfulness of the average difference of financial ratios of non-distressed and distressed corporations at 0.95 confidence level, it is necessary to eliminate those of which have not high ability to discriminate between distressed and non-distressed corporations. Generally speaking, when the goal of the researcher is to know the relations among a large number of independent variables to enter to the relation, the forward stepwise selection (conditional) is a proper one. 6 variable for one year prior (T-1): Quick ratio, Net working capital/ Total assets, Total debt/ Total assets, Net profit/Total assets, Earnings before interest and taxes/Net sales, Net sales/Fixed assets and 4 variables for 2 years prior (T-2): Net working capital/ Total assets, Total debt/ Total assets, Earnings before interest and taxes/ Total debt, Gross profit/ Net sales were selected. The tables 3 and 4 show a summary of the results of estimating the Logit models (see fallah shams et al, (2011).

5.2 Results of Logit Models without Using Distress Score

The expenses of falsely classifying a corporation, approaching distress in reality, as a non-distressed (type I error) is much more than the time a non-distressed corporation is classified as mistakenly as a distressed (type II error). In the prediction of financial distress models, there must be a balance between these 2 errors (FP and FN) which are done by optimal cutoff point of the model. Optimal cutoff point is a point of model performance in which the error of false classification of the model (including error I and II) are minimized.

Different cutoffs have been used in prior studies for measuring classification accuracy. For example, Altman (1968) and Deakin (1972) use cutoffs which minimize misclassification accuracy; Ohlson (1980) and Palepu (1986) use the cutoffs where the distributions of the two groups intersect; and Barniv et al. (2002) and Frydman et al. (1985) use cutoffs that minimize the number of misclassifications. Searching the prior researches shows that most of the researchers have chosen the figure 0.5 as cutoff point, however this research is going to use optimal cutoff point, which is achieved by the largest area under the ROC curve, in order to reach the highest overall prediction accuracy of the model prediction. Tables 5 and 6 present the results of the Logit models at the 0.1-0.9 cutoffs, respectively, on the training set, which has the θ , or the overall prediction accuracy ranging from 0.738 to 0.914 for T-1 and from 0.588 to 0.807 for T-2.

Optimal cutoff points shown in figure 2 and 3, are equal to 0.4 in T-1 model and 0.3 in T-2 model, is determined by the largest area under the ROC curve. θ is equal to 0.914 for T-1 model and 0.807 for T-2 model.

5.3 The Results of Logit Models with Using Distress Score

As shown in table 7, the sensitivity and specificity of logit models by using the distress score in cutoff point which has the most accuracy, is respectively equal to 0.905 and 0.950 for T-1 model, and 0.841 and 0.822 for T-2 model. The table indicates that this model has high accuracy in recognizing the distress and non-distress corporations. As it is seen, the amount of θ is high and is approximately equal to 0.928 ± 0.042 (0.844, 1.012) for T-1 model, and 0.832 ± 0.080 (0.672, 0.992) for T-2 model.

5.4 Comparative Models Performance

Roc curves and the area under it (AUC), describes the balance between sensitivity and specificity, besides comparing the performance of different tests on specific classification activity (Tang, T. C and Chi, L.C, 2005). ROC curves in logit model by using the distress score, whose graph drawn in figure 4 and 5, is higher and in the left side of logit model curve without using the distress score. In figure 6 and 7, we also see that this model has higher sensitivity and specificity in the training set. Consequently it could be said that in the classification of distress and non-distress corporations, the logit model with using the distress score operates better than the logit model without using distress score.

5.5 Prediction Accuracy of Models

To recognize the accuracy of prediction of T-1 and T-2 models, the testing set was investigated. Table 8 and figure 8 and 9 show the results of both two models. In T-1 model without using the distress score, sensitivity, specificity, error I and II are respectively equal to 0.938, 0.840, 0.062, 0.160, and for T-2 model they are respectively equal to 1.000, 0.378, 0.000, and 0.622. The amount of θ for T-1 model is equal to 0.889 ± 0.028 (0.833, 0.945) and is 0.689 ± 0.030 (0.630, 0.748) for T-2 model. In T-1 model with using the distress score, the sensitivity, specificity, error I and II are respectively equal to 0.938, 0.889, 0.062, 0.111 and are equal to 1.000, 0.422, 0.000 and 0.578 for T-2 model. The amount of θ for T-1 model is equal to 0.913 ± 0.026 (0.861, 0.965) and is equal to 0.711 ± 0.027 (0.657, 0.765) for T-2 model.

The ROC curve has been presented for two models in figure 8 and 9. In T-1 model with using the distress score, the ROC curve is higher and in the left side of T-1 model without using the distress score. It is the same in T-2 model. As a result logit model with using the distress score on the testing set operates better than the logit model without using the distress score.

Generally, the classification accuracy in the training set and the prediction accuracy in the testing set of the A models for T-1 and T-2 are respectively equal to 91.4% , 80.7%, 88.9% and 68.9% and in the B models are equal to 92.8% , 83.2%, 91.3% and 71.1%. Actually the results show that the logit models with using the distress score outperform the logit model without using distress score.

6. Conclusion

The financial distress and consequently the bankruptcy of the economic units can sustain enormous losses in large and small scales. In large scale, the financial distress of corporations' results in the decrease of Gross Domestic Product (GDP), the increase of unemployment, wastes the country's resources, and so on. In small scale, the beneficiaries, economic agencies, such as share-holders, investigators, creditors, managers, employees, raw material distributors and customers sustain loss. Therefore the prediction of future financial condition of corporations has largely attracted the financial researchers and finding the alerting indexes of the occurrence of financial distress has turned to be one of the most attractive and significant fields of financial and economic researches. Studying the history of researches show that many of researchers have focused on financial ratios to predict the financial distress. It is worth noting that the score of the distress of the corporations is beneficial for studying their operation condition. Therefore, in this article the distress score is gained from DEA and then it is used as predictor variable beside the other financial ratios. Moreover, this article also compares the performance of the models by using the ROC curve analysis.

The studied fields of this research are the corporations accepted in Tehran Stock Exchange; therefore the needed data has been collected from the database of Tehran Securities & Exchange Organization. Regarding the studied period of time (2001-2008), the data of the 304 corporations, 79 of which were distressed and 225 of which non-distressed, was available.

Data collection of the distressed and non-distressed corporations has been done in the T, T-1 and T-2 years. The T, T-1 and T-2 years for the distressed corporations are defined respectively as the year of the occurrence of the financial distress (the corporations have lost at least half of their capital in 2 successive years), one and two years before that. Regarding the non-distressed corporations, respectively the year the corporation had the most profit and one and two years before that. The selected set for analysis is divided randomly into two sets: Training (80 percent) and testing (20 percent) sets. The former is for making and estimation of the models and the parameters related to prediction methods, and the latter is used for studying the righteousness and accuracy of the prediction of the model.

The results show that the general performance of logit model with using the distress score as the predictor variable is better than logit model without using the distress score. Therefore, it could be concluded that we can enhance the accuracy of prediction by using the distress score gained from DEA as a predictor variable in financial distress prediction.

References

- Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance*, 23, 589–609. <http://dx.doi.org/10.2307/2978933>
- Bamber, D. (1975). The area above the ordinal dominance graph and the area below the receiver operating graph. *Journal of Mathematical Psychology*, 12, 387–415. [http://dx.doi.org/10.1016/0022-2496\(75\)90001-2](http://dx.doi.org/10.1016/0022-2496(75)90001-2)
- Barniv, R., Agarwal, A., & Leach, R. (2002). Predicting bankruptcy resolution. *Journal of Business Finance and Accounting*, 29(3), 497–520. <http://dx.doi.org/10.1111/1468-5957.00440>
- Beaver, W. H. (1966). Financial ratios as predictors of failure. Supplement to *Journal of Accounting Research: Empirical Research in Accounting-Selected Studies*, 4, 71–111. <http://dx.doi.org/10.2307/2490171>
- Charnes, A., Cooper, W.W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operation Research*, 2, 429–444. [http://dx.doi.org/10.1016/0377-2217\(78\)90138-8](http://dx.doi.org/10.1016/0377-2217(78)90138-8)
- Cooper, W.W., Seiford, L., & Tone, K. (2006). *Introduction to Data Envelopment Analysis and its Use: With DEA-Solver Software and References*. Springer Science, New York.
- Deakin, E. B. (1972). A discriminant analysis of predictors of failure. *Journal of Accounting Research*, 167–179. Spring.
- Fallah Shams, M., Sheikhi, M., & Sheikhi, Z. (2011). Financial distress prediction: Comparisons of logit models using receiver operating characteristic (ROC) curve analysis. *African Journal of Business Management*, 30, 12164-12173. <http://dx.doi.org/10.5897/AJBM11.1969>

- Frydman, H., Altman, E. I., & Kao, D. (1985). Introducing recursive partitioning for financial classification: The case of financial distress. *The Journal of Finance*, 40, 269–291 (March).
- Gattoufi, S., Oral, M., & Reisman, A. (2004). Data envelopment analysis: A bibliography update (1951–2001). *Socio-Economic Planning Science*, 38, 159–229. [http://dx.doi.org/10.1016/S0038-0121\(03\)00023-5](http://dx.doi.org/10.1016/S0038-0121(03)00023-5)
- Gestel, T. V., Baesens, B., Suykens, J., Poel, D. V., Baestaens, D. E., & Willekens, M. (2006). Bayesian kernel based classification for financial distress detection. *European Journal of Operational Research*, 172, 979–1003. <http://dx.doi.org/10.1016/j.ejor.2004.11.009>
- Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristics curve. *Radiology*, 143, 29–36.
- Lloyd, C. J. (1998). The use of smoothed ROC curves to summarize and compare diagnostic systems. *Journal of the American Statistical Association*, 93, 1356–1364.
- Marzban, C. (1998). Bayesian probability and scalar performance measures in Gaussian models. *Journal of Applied Meteorology*, 37, 72–82. [http://dx.doi.org/10.1175/1520-0450\(1998\)037<0072:BPASPM>2.0.CO;2](http://dx.doi.org/10.1175/1520-0450(1998)037<0072:BPASPM>2.0.CO;2)
- Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, 109–131. Spring.
- Palepu, K. G. (1986). Predicting takeover targets: A methodological and empirical analysis. *Journal of Accounting and Economics*, 8(1), 3–35. [http://dx.doi.org/10.1016/0165-4101\(86\)90008-X](http://dx.doi.org/10.1016/0165-4101(86)90008-X)
- Pepe, M. S. (2000). An interpretation for the ROC curve and inference using GLM procedures. *Biometrics*, 56, 352–359. <http://dx.doi.org/10.1111/j.0006-341X.2000.00352.x>
- Poon, P. H., Firth, M., & Fung, H. G. (1999). A multivariate analysis of the determinants of Moody's bank financial strength ratings. *Journal of International Financial Markets, Institutions and Money*, 9, 267–283. [http://dx.doi.org/10.1016/S1042-4431\(99\)00011-6](http://dx.doi.org/10.1016/S1042-4431(99)00011-6)
- Premachandra, I.M., Bhabra, G.S., & Sueyoshi, T. (2009). DEA as a tool for bankruptcy assessment: A comparative study with logistic regression technique. *European Journal of Operational Research*, 193, 412–424. <http://dx.doi.org/10.1016/j.ejor.2007.11.036>
- Tang, T. C., & Chi, L.C. (2005). Predicting multilateral trade credit risks: comparisons of Logit and Fuzzy. Logic models using ROC curve analysis. *Expert Systems with Applications*, 28, 547–556. <http://dx.doi.org/10.1016/j.eswa.2004.12.016>
- Thomas, L. C., Edelman, D. B., & Cook, J. N. (2002). Credit scoring and its applications. Philadelphia, PA. *Society for Industrial and Applied Mathematics*. <http://dx.doi.org/10.1137/1.9780898718317>
- West, D. (2000). Neural network credit scoring models. *Computers and Operations Research*, 27, 1131–1152. [http://dx.doi.org/10.1016/S0305-0548\(99\)00149-5](http://dx.doi.org/10.1016/S0305-0548(99)00149-5)
- Wilks, D. S. (1995). *Statistical methods in the atmospheric sciences*. New York: Academic Press. 467pp.
- Xu, X., & Wang, Y. (2009). Financial failure prediction using efficiency as a predictor. *Expert system with Applications* 36, 366–373. <http://dx.doi.org/10.1016/j.eswa.2007.09.040>

Table 1. A list of predictor variables

Category	Variables	Description
Liquidity	X ₁	Current ratio
	X ₂	Quick ratio
	X ₃	Cash Flow/ Current debts
	X ₄	Net Working Capital
	X ₅	Net Working Capital / Total Assets*100
	X ₆	Net Working Capital / Equity*100
Profitability	X ₇	Net Profit / Total Assets*100
	X ₈	Net Profit / Net Sales*100
	X ₉	Return on Equity (ROE)
	X ₁₀	Earnings Before Interest, Taxes/ Net Sales*100
	X ₁₁	Retained Earnings / Equity*100
Activity	X ₁₂	Gross Profit / Net Sales*100
	X ₁₃	Inventory Turnover
	X ₁₄	Net Sales / Fixed Assets
	X ₁₅	Net Sales / Total Assets
	X ₁₆	Net Sales / Equity
Financial Leverage	X ₁₇	Total Debt / Total Assets*100
	X ₁₈	Long-Term Debt / Equity*100
	X ₁₉	Earnings Before Interest, Taxes/ Total Debt*100
	X ₂₀	Total Debt / Equity*100

Table 2. Mann-Whitney non-parametric test

Variables	Description	Mann-Whitney U	Z	Asymp. Sig. (2-tailed)
X ₁	Current ratio	3063.000	-8.666	.000*
X ₂	Quick ratio	2684.000	-9.230	.000*
X ₃	Cash Flow/ Current debts	3651.000	-7.807	.000*
X ₄	Net Working Capital	3404.000	-8.158	.000*
X ₅	Net Working Capital / Total Assets*100	3154.000	-8.530	.000*
X ₆	Net Working Capital / Equity*100	8082.000	-1.198	.231
X ₇	Net Profit / Total Assets*100	1004.000	-11.729	.000*
X ₈	Net Profit / Net Sales*100	198.000	-12.928	.000*
X ₉	Return on Equity (ROE)	7874.000	-1.508	.132
X ₁₀	Earnings Before Interest, Taxes/ Net Sales*100	263.000	-12.831	.000*
X ₁₁	Retained Earnings / Equity*100	8751.000	-.203	.839
X ₁₂	Gross Profit / Net Sales*100	887.000	-11.903	.000*
X ₁₃	Inventory Turnover	5166.500	-5.536	.000*
X ₁₄	Net Sales / Fixed Assets	7298.500	-2.364	.018*
X ₁₅	Net Sales / Total Assets	6924.000	-2.921	.003*
X ₁₆	Net Sales / Equity	7694.000	-1.776	.076
X ₁₇	Total Debt / Total Assets*100	617.500	-12.304	.000*
X ₁₈	Long-Term Debt / Equity*100	8337.500	-.818	.413
X ₁₉	Earnings Before Interest, Taxes/ Total Debt*100	114.000	-13.053	.000*
X ₂₀	Total Debt / Equity*100	8583.000	-.453	.651

*It is significant at the 0.05 level

Table 3. Summary of the Estimated Logit Model (T-1)

<i>Model Summary</i>						<i>value</i>
-2 log likelihood (-2LL)						92.326
Cox & Snell R ²						.534
Nagelkerke R ²						.784
<i>Omnibus Test of Model Coefficients</i>				<i>Chi-Square</i>	<i>df</i>	<i>Sig.</i>
Step				9.501	1	.002
Block				185.803	6	.000
Model				185.803	6	.000
<i>Hosmer & Lemeshow Test</i>				<i>Chi-Square</i>	<i>df</i>	<i>Sig.</i>
				3.576	8	.893
<i>Variable in Equation</i>	<i>(B)</i>	<i>S.E</i>	<i>wald</i>	<i>df</i>	<i>Sig.</i>	<i>Exp(B)</i>
Quick ratio	-4.119	1.432	8.274	1	.004	.016
Net Working Capital / Total Assets	-.123	.029	17.815	1	.000	.131
Total Debt / Total Assets	.216	.049	19.414	1	.000	1.241
Net Profit / Total Assets	-.132	.058	5.230	1	.022	.877
Earnings Before Interest, Taxes/ Net Sales	-.112	.034	11.024	1	.001	.894
Net Sales / Fixed Assets	-.270	.085	10.177	1	.001	.763
Constant	-12.327	3.201	14.828	1	.000	.000

Table 4. Summary of the Estimated Logit Model (T-2)

<i>Model Summary</i>						<i>value</i>
-2 log likelihood (-2LL)						174.766
Cox & Snell R ²						.346
Nagelkerke R ²						.508
<i>Omnibus Test of Model Coefficients</i>				<i>Chi-Square</i>	<i>df</i>	<i>Sig.</i>
Step				8.396	1	.004
Block				103.362	4	.000
Model				103.362	4	.000
<i>Hosmer & Lemeshow Test</i>				<i>Chi-Square</i>	<i>df</i>	<i>Sig.</i>
				6.777	8	.561
<i>Variable in Equation</i>	<i>(B)</i>	<i>S.E</i>	<i>wald</i>	<i>df</i>	<i>Sig.</i>	<i>Exp(B)</i>
Net Working Capital / Total Assets	-.043	.013	11.444	1	.001	.044
Total Debt / Total Assets	.042	.016	7.535	1	.006	1.044
Earnings Before Interest, Taxes/ Total Debt	-.068	.021	10.449	1	.001	.934
Gross Profit / Net Sales	-.055	.024	5.302	1	.021	.946
Constant	-2.251	1.283	3.076	1	.079	.105

Table 5. Statistics of the estimated ROC curve parameters (training set), in the Logit model (T-1)

<i>Cutoff</i>	<i>TPR</i>	<i>TNR</i>	<i>FPR</i>	<i>FNR</i>	$\hat{\theta}$	$\hat{\delta}_{\theta}$	<i>p</i>	<i>95% Confidence interval</i>
0.1	1	0.756	0.244	0	0.878	0.058	0.830	[0.762, 0.994]
0.2	0.905	0.867	0.133	0.095	0.886	0.058	0.570	[0.770, 1.002]
0.3	0.889	0.900	0.100	0.111	0.895	0.057	0.080	[0.781, 1.009]
0.4	0.889	0.939	0.061	0.111	0.914	0.057	0.001	[0.800, 1.028]
0.5	0.873	0.950	0.050	0.127	0.912	0.059	0.066	[0.794, 1.030]
0.6	0.794	0.967	0.033	0.206	0.881	0.059	0.171	[0.763, 0.999]
0.7	0.730	0.978	0.022	0.270	0.854	0.059	0.312	[0.736, 0.972]
0.8	0.603	0.994	0.006	0.397	0.799	0.058	0.487	[0.683, 0.915]
0.9	0.476	1	0	0.524	0.738	0.058	0.589	[0.622, 0.854]

Table 6. Statistics of the estimated ROC curve parameters (training set), in the Logit model (T-2)

<i>Cutoff</i>	<i>TPR</i>	<i>TNR</i>	<i>FPR</i>	<i>FNR</i>	$\hat{\theta}$	$\hat{\delta}_{\theta}$	<i>p</i>	<i>95% Confidence interval</i>
0.1	1	0.544	0.456	0	0.772	0.087	0.623	[0.598, 0.946]
0.2	0.889	0.700	0.300	0.111	0.795	0.086	0.517	[0.623, 0.967]
0.3	0.825	0.789	0.211	0.175	0.807	0.086	0.001	[0.635, 0.979]
0.4	0.714	0.889	0.111	0.286	0.802	0.088	0.054	[0.626, 0.978]
0.5	0.540	0.917	0.083	0.460	0.729	0.088	0.124	[0.553, 0.905]
0.6	0.381	0.967	0.033	0.619	0.674	0.088	0.271	[0.498, 0.850]
0.7	0.302	0.978	0.022	0.698	0.640	0.088	0.342	[0.464, 0.816]
0.8	0.222	0.994	0.006	0.778	0.608	0.087	0.587	[0.434, 0.782]
0.9	0.175	1	0	0.825	0.588	0.087	0.679	[0.414, 0.762]

Table 7. Testing summary on the training set

<i>Logit</i>	<i>Cutoff</i>	<i>TPR</i>	<i>TNR</i>	<i>FPR</i>	<i>FNR</i>	$\hat{\theta}$	$\hat{\delta}_{\theta}$	<i>p</i>	<i>95% Confidence interval</i>
T-1	0.4	0.889	0.939	0.061	0.111	0.914	0.057	0.001	[0.800, 1.028]
T-2	0.3	0.825	0.789	0.211	0.175	0.807	0.086	0.001	[0.635, 0.979]
T-1+DS	0.4	0.905	0.950	0.050	0.095	0.928	0.042	0.000	[0.844, 1.012]
T-2+DS	0.3	0.841	0.822	0.178	0.159	0.832	0.080	0.000	[0.672, 0.992]

Note: Distress score (DS)

Table 8. Testing summary on the testing set

<i>Logit</i>	<i>Cutoff</i>	<i>TPR</i>	<i>TNR</i>	<i>FPR</i>	<i>FNR</i>	$\hat{\theta}$	$\hat{\delta}_{\theta}$	<i>p</i>	<i>95% Confidence interval</i>
T-1	0.4	0.938	0.840	0.160	0.062	0.889	0.028	0.001	[0.833, 0.945]
T-2	0.3	1.000	0.378	0.622	0.000	0.689	0.030	0.001	[0.630, 0.748]
T-1+DS	0.4	0.938	0.889	0.111	0.062	0.913	0.026	0.000	[0.861, 0.965]
T-2+DS	0.3	1.000	0.422	0.578	0.000	0.711	0.027	0.000	[0.657, 0.765]

Note: Distress score (DS)

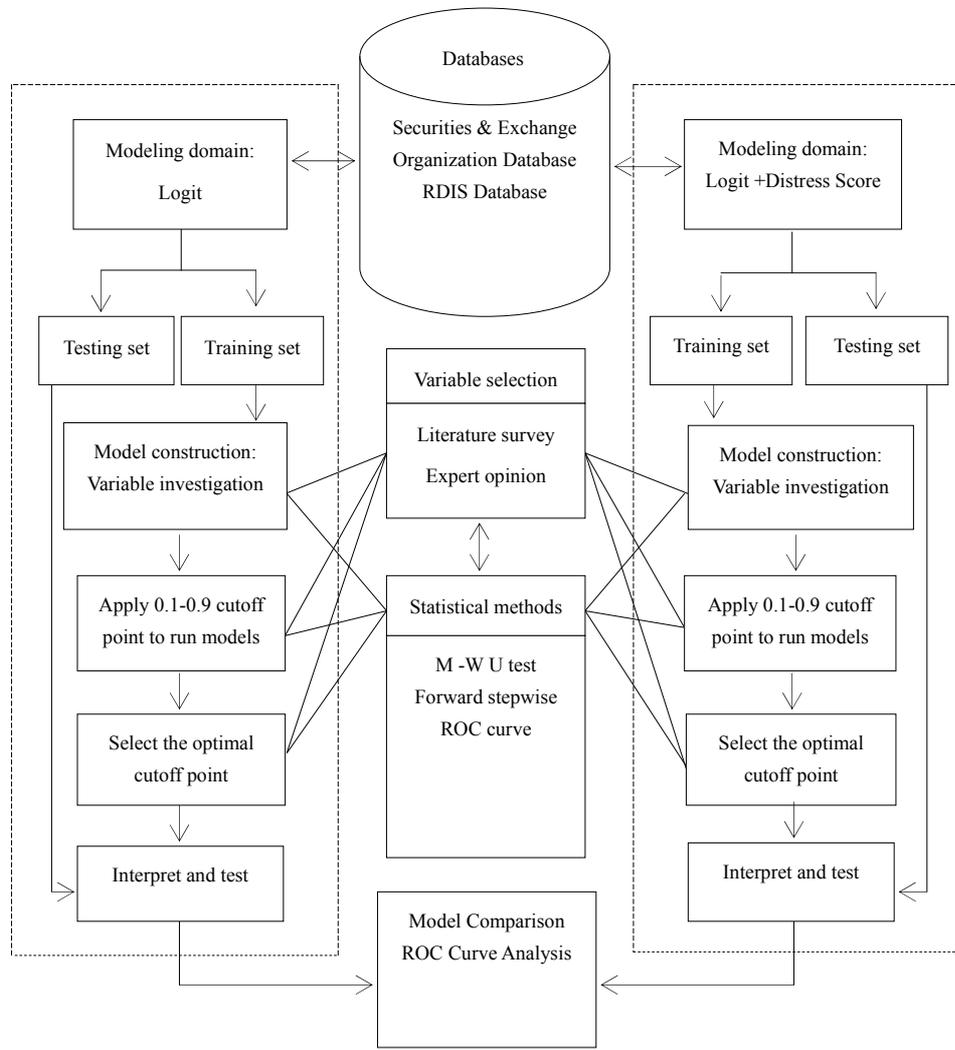


Figure 1. Research framework

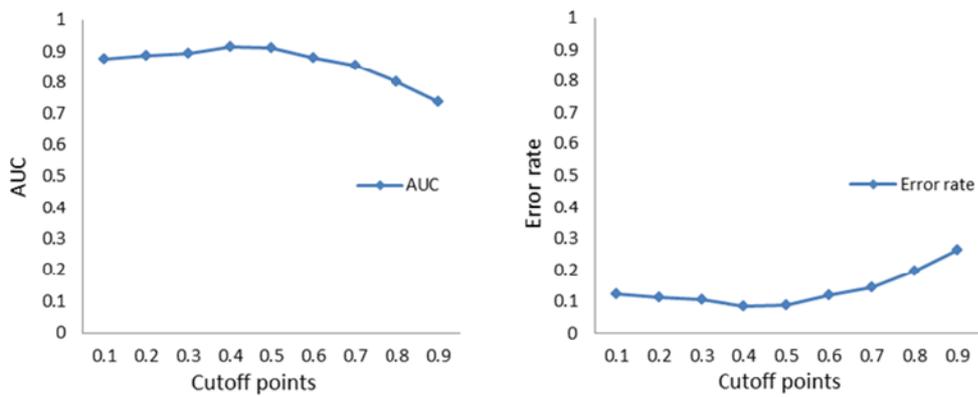


Figure 2. The performance of the T-1 model (training set)

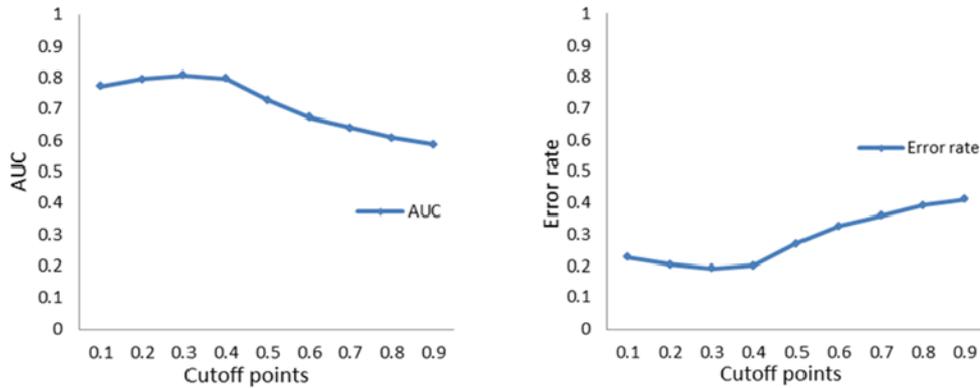


Figure 3. The performance of the T-2 model (training set)

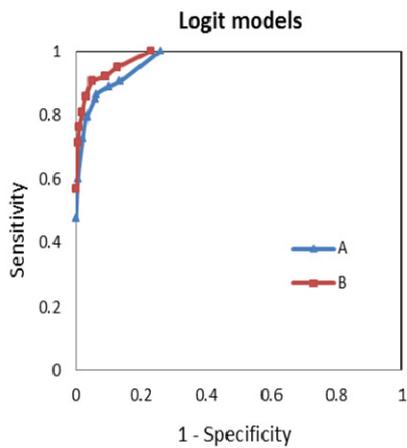


Figure 4. ROC curves for the results T-1 models (training set)

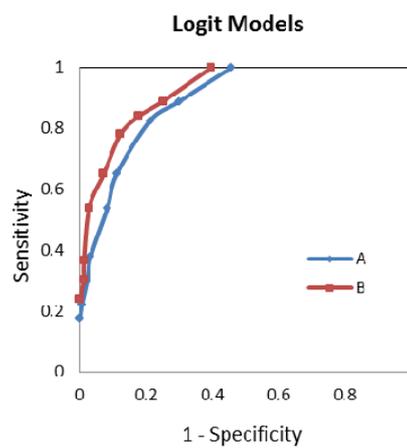


Figure 5. ROC curves for the results of T-2 models (training set)

Notes:

A: The logit model without using the distress score

B: The logit model with using the distress score

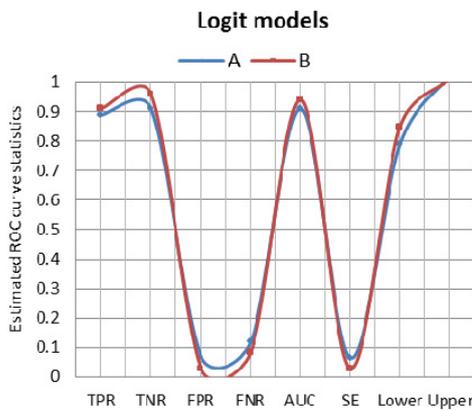


Figure 6. Comparisons of A and B models(training set) T-1

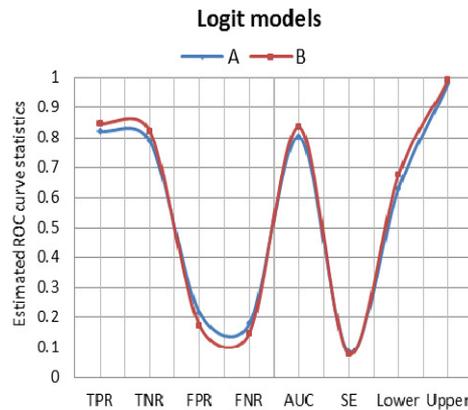


Figure 7. Comparisons of A and B models(training set) T-2

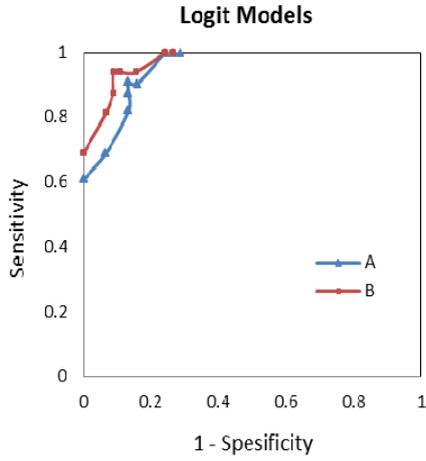


Figure 8. ROC curves for the results of T-1 models (testing set)

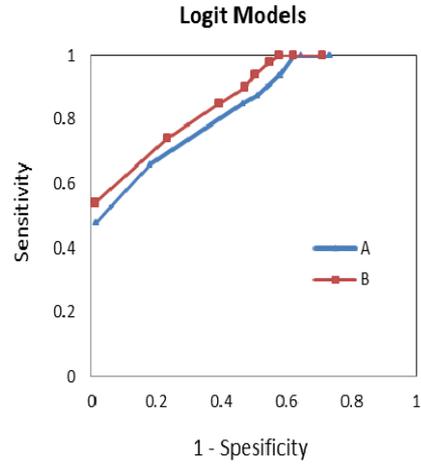


Figure 9. ROC curves for the results of T-2 models (testing set)