Measuring Retail Company Performance Using Credit Scoring Techniques

Yu-Chiang Hu[a, 1] and Jake Ansell[b]

[a,, b]*Management School and Economics, University of Edinburgh, William Robertson Building, 50 George Square, Edinburgh, EH8 9JY, UK*

Abstract

This paper proposes a theoretical framework for predicting financial distress based on Hunt's (2000) *'Resource-Advantage (R-A) Theory of Competition'*. The study focuses on the US retail market. Five credit scoring methodologies –*Naïve Bayes, Logistic Regression, Recursive Partition, Artificial Neural Network,* and *Sequential Minimal Optimization*— are used on a sample of 195 healthy companies and 51 distressed firms over different time periods from 1994 to 2002.

Analyses provide sufficient evidence that the five credit scoring methodologies have sound classification ability. In the time period of one year before financial distress, logistic regression model shows identical performance with neural network model based on the accuracy rate and shows the best performance in terms of AUROC value. Other models are slightly worse for predicting financial distress, but still present high accuracy rate and AUROC value. Moreover, the methodologies remain sound even five years prior to financial distress with classification accuracy rates above 85% and AUROC values above 0.85 for all five methodologies. This paper also shows external environment influences exist based on the naïve bayes, logistic, recursive partition and SMO models, but these influences are weak.

With regards to the model applicability, a subset of the different models is compared with Moody's rankings. It is found that both SMO and Logistic models are better than the Neural Network model in terms of similarity with Moody's ranking, with SMO being slightly better than the Logistic Model.

Keywords: Resource-Advantage (R-A) Theory of Competition, Naïve Bayes, Logistic Regression, Recursive Partition, Artificial Neural Network, and Sequential Minimal Optimization

1 Corresponding Author. Tel: +44 (0) 131 650 4644; Fax: +44 (0) 131 650 6536; email: Y.A.HU@sms.ed.ac.uk

1. Introduction

There is considerable effort devoted to the performance measurement of companies and being able to forecast their financial distress. The approaches used have covered a wide range of methodologies for example Beaver's (1966) univariate analysis model, Altman's (1968) Z-score model and Ohlson's (1980) logistic model

It has been argued by a number of authors that generic models for all sectors tend to be too general and lack the ability to deal with specific industrial sectors. It was decided to focus on the retail sector, since according to Dawson (2000) retail risk assessment and evaluation will be a critical area of research. The USA retail sector was also chosen because of the clear definitions and reporting of financial distress through Chapters 7 and 11. A sample of 195 healthy companies and 51 distressed firms were selected from 1994 to 2002. Timescale is clearly an issue and in the paper this is explored, the results unsurprisingly find that, for most models, the year before financial distress provides the best prediction, though, up to at least 5 years provide good prediction.

A range of variables that can be assembled to describe the performance of retails companies. In the current research, 170 potential performance measures have been considered which cover both internal and external measures, based on Resource-Advantage theory (Hunt, 2000). Yet obviously with such a large number of variables to choose from there is a danger of over-fitting and so there is a need to reduce the number of variables. After exploring a range of models and taking into account the sample size, five variables were employed in the final analysis.

In this paper, five credit scoring methodologies are used: *Naïve Bayes, Logistic Regression, Recursive Partition, Artificial Neural Network,* and *Sequential Minimal Optimization.* These models were fitted to the data and the all performed well. Since the size of dataset did not allow a hold out sample it was felt that a comparison should be made with an alternative external rating, and this was chosen to be Moody's rating. The results indicated that the most comparable model was the SMO.

The next section will discuss the alternative Credit Scoring modelling approaches considered. Section 3 will initially discuss the measures that could be used to determine financial distress, and then proceed to describe those variables that have been used in the study. Details of the sample selected will be given at the end of the section. Section 4 will describe the approach taken in fitting the models. This is followed by the results of the analysis. Finally there will be a discussion of the results in section 6.

2. Credit Scoring Modelling

Beaver (1966) was a pioneer in financial distress prediction research with a number of authors following his work. He conducted an analysis of likelihood ratios based on a *Bayesian* approach. He argued that the default prediction problem could be regarded as a problem of evaluating the probability of financial distress conditional upon the value of a specific financial ratio. *Naïve Bayesian Approach* provides a simple method to deal with a classification problem. Let H be the healthy samples and let D be the distressed samples. Moreover, let X be a vector of independent variables and let x represent a particular vector of an independent variable. The conditional probability of a financial distress company in terms of a specific financial ratio x can be expressed as:

$$P(D \mid X = x) = \frac{P(D)P(X = x \mid D)}{P(X = x)} \tag{2.1}$$

Beaver (1966) used only a single measure but did limit the performance evaluation, and hence, approach can be generalised. Altman (1968) suggested the *Multiple Discriminant Analysis (MDA)* approach to develop a Z-score bankruptcy prediction model based on five financial ratios. After Altman's (1968) research, a number of studies also use MDA to predict firm's default, including Deakin (1972), Blum (1974), Libby (1975), Altman, Haldeman and Narayanan (1977), and Taffler (1984).

There are limitations to the MDA models such as, the assumption that the covariance matrices of two populations are identical to produce a linear discriminator and both populations need to be described by multivariate normal distribution. These are generally too restrictive with industrial data, (Eisenbeis, 1977). Deakin (1976) contended that even if after performing the normality transforming process, financial ratio data do not follow normal distribution.

Ohlson (1980) was the first to apply the conditional probability model and in particular, the logistic model, to bankruptcy prediction research. Unlike MDA, the logistic model does not require multivariate normality or the equality of covariance matrices of two populations. By logit transformation on odd ratio function, the logistic model can be linearized and used to solve classification problems. A logistic function can be expressed as follow:

$$g(x) = \ln\left[\frac{\pi(x)}{1-\pi(x)}\right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n \qquad (2.2)$$

$$= \beta \times x^T$$

where $\pi(x)$ is the logistic function,

$$\pi(x) = \frac{1}{1+e^{-(\beta \times x^T)}} = \frac{e^{\beta \times x^T}}{1+e^{\beta \times x^T}} \qquad (2.3)$$

Following Ohlson's (1980) study, Mensah (1983), Casey and Bartczak (1985), as well as Gentry et al (1985), also employed the conditional probability models to predict financial distress. In the mid-1980s, *"Recursive Partitioning Analysis (RPA)"* or *"Decision Tree"* was introduced in the financial distress prediction research area. (Frydman et al, 1985; Marais et al, 1984; Carson and Hoyt, 1995) RPA is a non-parametric technique and does not suffer the limitations from MDA or logistic model. Although Fisher's (1936) linear discriminant method is often viewed as the oldest classification technique, Hand (1997) argued that the basic idea of RPA is very straightforward, and hence the oldest conceptually.

RPA can be regarded as a stepwise procedure. The first step is to select an independent variable as the best discriminator and to decide a cutpoint based on the lowest expected misclassification cost. Based on the cutpoint, the second step is to divide both healthy and distressed firms into two sub-nodes. The third step is to select another (or the same) discriminator and further partition the healthy and distressed firms into another two sub-nodes. The same process can be continued, if further splitting is necessary. Thomas et al (2002) mentioned two reasons to stop the partitioning process. First, if the number of samples in a node is too small, then further partition is not appropriate. Second, if the classification results between the old node and new nodes do not have significant differences, then it is also not necessary to split the old node. One of the major problems relative to RPA is overfitting:

the continuous partitioning process is likely to encourage one misclassified case in the terminal node. The overfitting problem can be overcome by a *"Cross-Validation"* procedure.

From the late 1980s, the artificial intelligence (AI) or machine learning techniques, such as *Artificial Neural Networks* (ANN), were successfully applied to financial distress prediction studies. A large number of studies compared ANN's prediction performance with other classification methods and proved that ANN had better prediction performance than other methods. (Coates and Fant, 1993; Zhang et al., 1999) The most popular ANN algorithm in the financial distress prediction domain is the *"Multilayer Perceptron (MLP)"*. A MLP has three main components: input layer, hidden layer and output layer.

The input layer is responsible for receiving information from the outside environment and transferring it to the hidden layer. In the hidden layer, a neuron will assign a series of weights to the inputs, cope with the information via a training process, and then forward the results with weights to the output layer. The training process can be regarded as a weighting determination process. The most frequently used algorithm for training process is the *Back Propagation Algorithm (BPA).*

Thomas et al (2002) pointed out that BPA first calculates the difference between the expected output value and the observed output value (called *error*) in the output layer. The next step is to distribute the error back to the network in terms of a weight and to adjust the weight to decrease the error. The process is repeated for all cases, called an *epoch*. After several epochs training, the learning error will reduce to a minimum level and the training process ends. Trigueiros and Taffler (1996) mentioned some advantages of MLP, such as the independence from statistical distribution assumptions. However, MLP also has some limitations. For example, it does not provide adequate significance tests and requires considerable computer power and skills. (Tam and Kiang, 1992)

In the late 1990s, another machine learning technique, *Support Vector Machine (SVM)*, was introduced to deal with the classification problem. Fan and Palaniswami (2000) applied SVM to select the financial distress predictors. They pointed out that SVM created an optimal separating hyperplane in the hidden feature space in terms of the principle of structure risk minimization and used the quadratic programming to obtain an optimal solution. However, Platt (1999) argued that a large number of quadratic programming in SVM training is time consuming. As a result, he introduced a new algorithm, *Sequential*

Minimal Optimization (SMO), to improve the SVM training time. Unlike the previous SVM training methods, SMO only uses two Lagrange multipliers at each training step. It was found that SMO has better performance than other SVM training methods in terms of many aspects, such as better scaling with training sample size.

Some other methodologies were also applied to the financial distress prediction research area and have shown a good performance, including the *Rough Set Approach* (Dimitras et al., 1999; McKee and Lensberg, 2002) and the *Multidimensional Scaling Approach* (Mar-Molinero and Serrano-Cinca, 2001).

3. Performance Measures Selection and Data Collection

3.1 Previous Research

Most of the academic literature has based on the quantitative financial ratios to predict financial distress, see Altman's (1968) and Ohlson's (1980). Incorporating both financial and non-financial data in a prediction model achieves higher accuracy in a financial distress prediction, (Alves, 1978). Credit-rating companies including Moody's, S&P, and Fitch take into account both quantitative and qualitative factors with emphasis on the qualitative (Moody's, 1998 and 2002; Fitch, 2001; S&P, 2003_a and 2003_b). In this paper a large range of measures are explored. These include measured on industrial sector since many authors, (Williams and Goodman, 1971; Gupta and Huefner, 1972; Bowen et al, 1982; Mensah, 1984), have suggested applying the same variables across different sector produces overly general models that overlook the specific attributes of the sectors. Platt and Platt (1990) used industry-related measures in a bankruptcy model and proved that these industry-relative measures could improve the accuracy of the classification model.

In addition, macro-economical factors have significant impact on financial distress prediction models (Rose et al, 1982), and different macroeconomical environments may affect the accuracy of the bankruptcy predictive model, (Mensah, 1984). Other authors have suggested a company's sustainability must be based on cash flow, rather than on earnings in the accounting statements, for earnings include non-cash items that cannot reflect a company's ability to pay back interests or principal, (S&P, 2003_b). Gentry et al (1985) developed a financial distress prediction model in terms of a cash flow structure. Although their model showed that only one variable, dividends/cash flow, had significant influence to the bankruptcy prediction.

Finally, the lack of theoretical groundwork for variable selection is a common situation in most financial distress prediction studies. Often, financial distress researchers select independent variables for model construction based on the successful prediction performance in previous studies. Obviously, such a variable selection method is limited and fails to provide a holistic framework for research in financial distress. In contrast, the present research develops a theoretical framework based on Hunt's (2000) *'Resource-Advantage (R-A) Theory of Competition'*.

Unlike the traditional perfect competition theory which focuses on factors of production, the R-A theory includes significant qualitative issues such as entrepreneurship and a company's relationship with its suppliers. The theory holds that demand is not only heterogeneous across industries but within them. It also holds that information is imperfect and costly and so that maximising profit is not a viable proposition, one can only seek superior financial performance. Given that companies' resources are different and imperfectly mobile, then Hunt and Morgan (1997) argued that a comparative advantage in resources provides also a comparative advantage in the market place and hence a superior financial performance. The theory suggests seven categories of measures, see Table 3.1.

Table 3.1 Internal Resources

Internal Resources	**Examples**
Financial Resource	Cash reserves and access to financial markets
Physical Resource	Plant, raw materials, and equipment
Legal Resource	Trademarks and licenses
Human Resource	The skills and knowledge of individual employees and the entrepreneurial skills
Organizational Resource	Controls, routines, cultures, and competences for entrepreneurship
Informational Resource	Knowledge about the market segment, competitors and technology
Relational Resource	Relationships with competitors, suppliers and customer

Source: Modified from Hunt (2000) A General Theory of Competition, Thousand Oaks: Sage pp.128

Based on past findings[2] and the R-A theory, 170 potential retail performance measures were obtained. These variables were then studied and classified as either *'Quantifiable Measure - Available Data'*, *'Quantifiable Measure - No Available Data'*, and *'Difficult to*

2 Include relative literature survey and 24 semi-structured phone interviews and one semi-structured face-to-face interview between the end of June and the end of August 2004. The interviewees included: 1) retail companies' management (14 interviews); 2) bank managers in business loan departments (7 interviews); 3) industrial analysts in investment institutions (4 interviews).

Quantify'. Obviously, the analysis focussed on the 67 performance measures in the category *'Quantifiable Measure - Available Data'*. These are combined into two groups 'Internal Resources Group' (G1), and 'External Factors Group' (G2) and are presented in Appendix A. In order to detect external influences, these factors will be re-grouped as G1 and G12 (G1+G2).

3.2 Sample Selection

3.2.1 Sample Selection Criteria for Non-defaulting Companies

In connection with the sample selection of non-defaulting companies, five criteria were considered. Only publicly listed companies were chosen. Given that listed companies had to abide by regulations in the financial market, their financial information tended to be more open and transparent than that of private companies. In addition, small companies were included based on the SBA size standards[3]. This is an improvement from previous studies using the *Wall Street Journal Index* and *Compustat* database. (E.g. Ohlson, 1980; Hamer, 1983; Frydman, Altman and Kao, 1985) These data sources are likely to exclude small companies despite the fact that small companies are likely to face financial distress.

Although Edmister (1972) argued that new firms had great probability to face financial distress and should be considered in any bankruptcy prediction model, in the present study, only those public sample companies that had been listed for at least three years were considered. There are two reasons to support this criterion. First, a newly listed company may not a new company. Second, studies show that newly listed stocks have abnormal returns after the public announcements of listing, (Sanger and McConnell, 1986). In order to avoid the influences from the newly listed companies, especially for some market relevant measures, no healthy companies listed after December 2000 are included. Furthermore, this research does not consider e-retailers[4] because their performance measures are different from those of traditional retailers. Finally, even if a sample company satisfied the previous four criteria, it is excluded if its data is not complete.

---

3 SBA's size standards define whether a business entity is small and, thus, eligible for Government programs and preferences reserved for "small business" concerns. Size standards have been established for types of economic activity, or industry, generally under the North American Industry Classification System (NAICS). Information available at: http://ecfr.gpoaccess.gov/

4 An e-retailer is defined as that the company's primary business line is the e-business or e-commerce.

As a result of applying the five criteria above, 67 different retail performance measures are collected from a dataset of 195 non-defaulting US retail companies over the time period of 1998 to 2002.

3.2.2 Sample Selection Criteria for Companies Facing Financial Distress

The USA retail sector was chosen because of the clear definitions and reporting of financial distress through Chapters 7 and 11. Based on the US federal bankruptcy law, a financially distressed company might use the bankruptcy code of *'Chapter 11'* to reorganize its financial structure and try to recover from distress, or that of *'Chapter 7'*, to go into liquidation and stop all business operations. Drawing on this insight, any company filing for the bankruptcy code of *'Chapter 11'* or *'Chapter 7'*, were deemed to be under financial distress and selected for the research.

An important issue is the timing of failed firms' data. Ohlson (1980) suggested that the financial statements prior to the financial distress year should be viewed as the last report, since reports *after* financial distress would usually include the adjustments from auditors in light of the bankruptcy filing. Adopting Ohlson's (1980) viewpoints, data prior to the financial distress year was considered as the last report. Overall, data were collected from 51 financially distressed firms and these companies were divided into five groups in terms of different time scales. (See table 3.2)

Table 3.2 Descriptions of Time Scales of Failed Firms' Data

Group	**Number of Failed Firms**	**Financial Distress Year**	**Data Collection Time Scale**
A	5	2003	From 1998 to 2002
B	13	2002	From 1997 to 2001
C	15	2001	From 1996 to 2000
D	12	2000	From 1995 to 1999
E	6	1999	From 1994 to 1998
Total	51		

4. Methodology

As with any data analysis there need to clean the data to remove outliers. This was done using standard approaches (10-means cluster analysis) and reduced the samples by about 5%. Given large number of variables, 67, for consideration would tend towards overfitting. Prior to model construction, a cross-validation process is performed to resolve overfitting. Moore

(2001) compared three cross-validation methods: the test set method, the leave one out method and the 10-folders method. He argued that the 10-folders cross-validation process only wasted 10% of total data and the training cost was much lower than the leave one out method. Drawing on this insight, the 10-folder method is selected for cross-validation.

Five credit scoring methodologies are employed for model construction: Naïve Bayes, Logistic Regression Model, Recursive Partitioning, Artificial Neural Network[5] and Sequential Minimal Optimization (SMO). An initial interest of the study was the timescale effect, whether on should use data just prior to the potential financial distress or some time before. Hence a series of models were fitted to M1 to M5 to allow evaluation of prediction performance from one to five years before.

Selection of the variables was via two stage model. Hosmer and Lemeshow (2000) suggest that one should initially use univariate analysis to identify the potential variables for the modelling using a p-value of 0.25. This was followed by use of forward stepwise model for each approach. The top five variables with higher appearance frequency in each variable group are selected for final model construction, as can be seen in table 4.1.

Table 4.1 Stepwise Variable Selection Results

Variable Group	**Key Performance Measures**
G1	V6: Net Profit Margin V15: Receivable Turnover V10: Return on Total Assets V28: Total Debt / (Total Debt + Market Capitalization) V36: Operation Cash Flow
G12	V6: Net Profit Margin V15: Receivable Turnover V28: Total Debt / (Total Debt + Market Capitalization) V36: Operation Cash Flow V56: The Five Years Correlation Coefficient between Government Debt and Total Sales

The Net Profit Margin (V6), Receivable Turnover (V15), Total Debt / (Total Debt + Market Capitalization) (V28) and Operation Cash Flow (V36) are significant variables, since they are common across the models.

Model performance was evaluated in terms of two approaches: the *Classification Accuracy Rate* approach, see Hand (1997), and the *Area under the 'Receiver Operating Characteristics Curve' (AUROC)* approach, see Thomas et al (2002). In this research, AUROC is applied to

[5] With regards to the artificial neural network, a three layers back propagation network is selected.

the Naïve Bayes, Logistic Regression Model and Artificial Neural Network. Both the accuracy rate and AUROC are employed for subsequent analyses.

Given the size of the sample available for study it was not possible, and probably it would not have been informative, to employ a hold out sample. Hence, the above methodology will result in potentially overly optimistic results. To overcome this problem for the best modelling approaches, it was decided to compare the results from the study with a standard rating system; in this case Moody's rating. In retailing there are only 8 rating grades given Aa to C in Moody's system. Hence the data was ranked according to score and divided into 8 groups. Unfortunately, Moody's ratings were only available for a limited number of companies, since firms undergo the credit rating process due to special circumstances, such as corporate bond issuing. Therefore, the sample size for comparative analysis varies year on year. Logistic regression, neural network and SMO models in the time period one year before financial distress in the variable group (G12) are selected for the ranking comparison analysis. Again, a range of measures for comparison were used, Kolmogorov-Smirnov (K-S) test, Distance Analysis, and Weighted Kappa Analysis and finally Graphical Bubble Charts.

5. Empirical Analysis

5.1 Time Scale Analysis

As mentioned in section 4, a five-year time scale analysis can be carried out in this research by comparing the performance of models from five different time periods (M1, M2, M3, M4 and M5). M1 is designed for evaluating a model's performance one year before financial distress; M2 is designed for assessing a model's utility two years before the financial distress, and so on. An arrangement of accuracy rate and AUROC results in terms of the five models are expressed in table 5.1.

Table 5.1 shows that regardless of the groups of performance measure, M1 has the best classification performance in terms of the naïve bayes, logistic regression and neural network models. With regards to SMO and recursive partitioning models, although some results indicate that M2 has better performance than M1, the difference is very small and is only around 1%. In addition, even if the time period is five years prior to financial distress, the accuracy rates are above 85% and the AUROC values are above 0.85 among all five methodologies. The results suggest that the overall performance of these five modelling methodologies is sound, even if the time period chosen is as long as five years before financial distress. Furthermore, these results also showed that the key variables selected are effective to predict financial distress.

Table 5.1 Model Performance Evaluation

G1 (Internal Resources Group)							
Methodology	**Performance Measures**	**M1**	**M2**	**M3**	**M4**	**M5**	**Average**
Naïve Bayes	Accuracy Rate	87.40%	86.59%	84.55%	83.74%	84.55%	*85.37%*
	Average AUROC	0.9291	0.8894	0.8265	0.8017	0.8484	*0.8590*
Logistic Model	Accuracy Rate	91.06%	86.18%	82.52%	84.15%	82.11%	*85.20%*
	Average AUROC	0.9354	0.8984	0.8398	0.8172	0.8176	*0.8617*
Neural Network	Accuracy Rate	93.50%	91.46%	89.02%	87.40%	88.21%	*89.92%*
	Average AUROC	0.9454	0.8974	0.8485	0.8366	0.8577	*0.8771*
SMO	Accuracy Rate	88.21%	89.02%	86.18%	84.15%	79.67%	*85.45%*
Recursive Partitioning	Accuracy Rate	88.21%	89.43%	88.62%	86.18%	86.18%	*87.72%*
G12 (External Factors Group)							
Methodology	**Performance Measures**	**M1**	**M2**	**M3**	**M4**	**M5**	**Average**
Naïve Bayes	Accuracy Rate	89.84%	88.21%	87.40%	85.77%	88.21%	*87.89%*
	Average AUROC	0.948	0.9127	0.8837	0.8881	0.9069	*0.9079*
Logistic Model	Accuracy Rate	93.50%	89.84%	87.40%	86.59%	86.59%	*88.78%*
	Average AUROC	0.9537	0.8963	0.8839	0.8786	0.9011	*0.9027*
Neural Network	Accuracy Rate	92.68%	89.02%	87.80%	87.40%	87.40%	*88.86%*
	Average AUROC	0.9415	0.8997	0.8400	0.8656	0.8899	*0.8873*
SMO	Accuracy Rate	89.43%	88.21%	86.18%	86.59%	86.99%	*87.48%*
Recursive Partitioning	Accuracy Rate	88.62%	89.02%	88.62%	89.84%	86.99%	*88.62%*

When comparing the performance of different methodologies, logistic model proves to have the best performance one year before the financial distress (based on both average accuracy rate and average AUROC value) in the G12 and neural network model shows to have best performance one year before the financial distress (based on both average accuracy rate and average AUROC value) in the G1. However, the same cannot be concluded for different time periods. For example, in the time periods of three years, four years and five years before financial distress, the neural network showed the best performance in the G12 in terms of average accuracy rate, but not in terms of the average AUROC value.

Furthermore, based on the five years average performance, regardless of G1 or G12, neural network shows the best performance in terms of five years average accuracy rate. However, the same conclusion cannot be achieved, if the model performance is evaluated by using five years AUROC value. Drawing on this insight, it is difficult to conclude which modelling methodology has the *'absolute'* best performance in time scale comparison analysis.

5.2 External Influences Detection Analysis

As mentioned in section 3, the external influences can be detected by comparing the performance of G1 and G12 models. If G12 performs better than G1, external factors have significant impacts on the model classification utility.

In table 5.1, most G12 models have better classification power than G1 models founded on the five years average accuracy rate and the five years average AUROC value. The only exception is the neural network method. However, the same as the time scale analysis, the performance differences among these models are small. For example, the difference of the accuracy rate is below 3% and AUROC value is below 0.05 for all models. As a result, it can be concluded that external environment influences exist based on the naïve bayes, logistic, SMO, and recursive partitioning models, but these influences are weak.

5.3 Modelling Methodology Comparison Analysis

Based on the findings above, the G12 model (except the neural network model) in the time period of one year before the financial distress has the best performance. With regards to the neural network method, the G1 model replaces G12 model, since the G1 model shows better classification ability than G12 model. Hence, five final models are selected for the purpose of modelling methodology comparison analysis and the results are illustrated in table 5.2.

Table 5.2 Modelling Methodology Comparison Analysis

Performance Measures	Naïve Bayes	Logistic Model	Neural Network	SMO	Recursive Partitioning
Accuracy Rate	89.84%	93.50%	93.50%	89.43%	88.62%
AUROC	0.948	0.9537	0.9454		

From table 5.2, it is clear that the logistic model and neural network model has identical performance better than other models in terms of accuracy rate. Moreover, logistic model has the best performance based on the AUROC value. The other models are slightly worse for predicting financial distress, but still present high accuracy rate and AUROC value in the time period one year prior financial distress.

5.4 Test of Significance

The Kolmogorov-Smirnov test assesses whether two datasets differ significantly. Results of the Kolmogorov-Smirnov Two Sample Test are shown in table 5.3.

Table 5.3 Two-sample Kolmogorov-Smirnov (K-S) test

Modelling Methodology	K-S	2002	2001	2000	1999	1998
Logistic Model	Z Value	1.417	1.407	1.633	1.773	1.076
	p-value	0.036	0.038	0.01	0.004	0.197
Neural Network	Z Value	1.417	1.738	1.796	1.128	1.407
	p-value	0.036	0.005	0.003	0.157	0.038
SMO	Z Value	1.167	1.407	1.306	1.37	1.159
	p-value	0.131	0.038	0.066	0.047	0.136

The highlighted *p*-values in table 5.3 are not significant at 5% level of significance and indicate when a proposed model provides rankings similar to Moody's. SMO has similar rankings in years 1998, 2000, and 2002. The other two models only have similar rankings to Moody's in a specific year. Significance testing is useful for determining whether or not there is similarity in ranking. The following techniques attempt to assess the level or degree of similarity.

5.5 Distance Analysis

The most straightforward approach to compare the degree of similarity between two ordinal data sets is distance analysis. The smaller the distance between the rankings from Moody's and the present study, the better the practical applicability of the study's proposed model. To calculate distances, each cell in the crosstabulation is presented as a proportion of the total sample size. (This allows for year on year comparison, as the sample size of each year is different.) The cell value is then multiplied by the value in the distance matrix. Finally, the resulting values are summed up. This gives an overall distance between Moody's model and each of the proposed models. Results are shown in table 5.4.

Table 5.4 Overall Distances for Each Modelling Methodology

Modelling Methodology	**2002**	**2001**	**2000**	**1999**	**1998**	***Average Distance***
Logistic Model	1.1528	1.2466	1.3867	1.4935	1.3562	*1.3271*
Neural Network	1.3611	1.4521	1.5200	1.5455	1.6027	*1.4963*
SMO	1.0833	1.2603	1.3733	1.4416	1.3562	*1.3029*

Amongst the three models, the Neural Network model has the highest average distance between 1998 and 2002, and the highest distances each year. The best model is SMO based on average distance over the five years. The Logistic Model has similar performance to SMO, and although the average distance is higher.

5.6 Measure of Agreement

Weighted Kappa can be used to measure the concordance between two raters. Weighted Kappa is an extension of Cohen's Kappa (1960) suitable for ordinal data and for measuring relative concordance. The values of weighted Kappa are shown in table 5.5.

Table 5.5 Weighted Kappa Analysis

Modelling Methodology	**2002**	**2001**	**2000**	**1999**	**1998**	***Average Weighted Kappa***
Logistic Model	0.3844	0.3431	0.3253	0.3219	0.3869	*0.3523*
Neural Network	0.2314	0.2107	0.2147	0.2030	0.1980	*0.2116*
SMO	0.4038	0.3183	0.3386	0.3353	0.3925	*0.3577*

As with distance analysis, average weighted Kappa results suggest that SMO is the better performing model amongst the three models, closely followed by Logistic Model. Neural Network still shows lowest performance in terms of agreement with Moody's.

5.7 Bubble Chart Analysis

In this research, graphical analysis using the bubble chart was developed to facilitate interpretation of similarity. The bubble chart enables a visualization of crosstabulation tables with clear localization of frequencies and a graphical representation of the observations through bubble size.

Bubble charts are interpreted as follows. First, the closer the bubbles are to the diagonal line, the more similar the rankings are. Second, if the bubbles that are close to the diagonal line are large in size, then it can be concluded that the degree of similarity between rankings is higher. Third, if the bubbles are gathered in the upper left hand corner and in the lower right hand corner, then the degree of similarity between the compared rankings is low.

Based on the distance and measure of agreement analyses, SMO shows the highest degree of similarity with Moody's ratings in the year 2002. (See tables 5.4 and 5.5) In contrast, the neural network model presents the worst performance in the comparative study in the year 1998. In order to illustrate the utility of bubble chart analysis, these two counterexamples are presented in figures 5.1 and 5.2, respectively.

Figure 5.1 SMO vs. Moody's, 2002

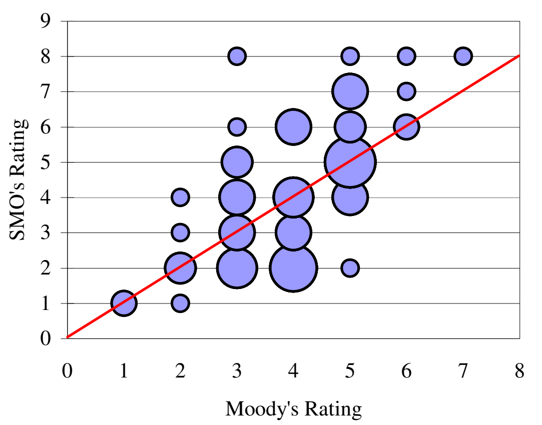
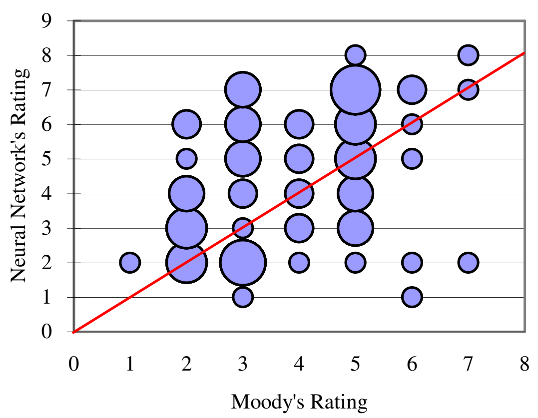


Figure 5.2 Neural Network vs. Moody's, 1998



The bubble chart analysis is a quick way of comparing the degree of similarity between different ranking methods. Comparing the two figures above, it can be argued that the better performing model, SMO, has a narrower bubbles area than the worse performing model, neural network. Also, the neural network vs. Moody's diagram has a greater number of large bubbles away from the diagonal line.

Overall, it can be concluded that SMO's ability to rank company performance is slightly better than Logistic Model and relatively better than the Neural Network model. This is true for significance testing using Kolmogorov-Smirnov test, distance analysis, and measure of agreement using Weighted Kappa. Moreover, the bubbles distribution is a very useful graphical method to detect the similarity between two ordinal datasets.

8. Discussions and Further Research

This paper proposed a theoretical framework for predicting financial distress based on Hunt's (2000) 'Resource-Advantage (R-A) Theory of Competition'. 170 measures were drawn from literature on performance measurement and interviews with outside stakeholders. After a regrouping process, 67 variables are chosen out of the 170 for model construction. Out of the 67 variables, key variables were found via cluster analysis, univariate analysis, and forward step-wise approach.

The USA retail sector was also chosen because of the clear definitions and reporting of financial distress through Chapters 7 and 11. Five credit scoring methodologies –*Naïve Bayes, Logistic Regression, Recursive Partition, Artificial Neural Network,* and *Sequential Minimal Optimization*— were used on a sample of 195 healthy companies and 51 distressed firms over different time periods from 1994 to 2002.

The time scale analysis showed unsurprisingly that M1 model has the best classification performance in terms of the naïve bayes, logistic regression and neural network models. Although some results indicate that M2 has better performance than M1 based on the SMO and recursive partitioning models, the difference is very small and is only around 1%. In addition, even if the time period is five years prior to financial distress, the accuracy rates are above 85% and the AUROC values are above 0.85 among all five methodologies.

The external influences exist in the naïve bayes, logistic, SMO, and recursive partitioning models, but these influences are weak. Furthermore, in the time period one year before financial distress, logistic model and neural network model has identical performance and their performance is better than other models in terms of accuracy rate. Moreover, logistic model has the best performance based on the AUROC value. The other models are slightly worse for predicting financial distress, but still present sound performance.

The above results are potentially overly optimistic, since the limits of sample size. To overcome this problem, a series of comparison analysis from the study with Moody's rating were performed. Using the Kolmogorov-Smirnov significance test, distance measure, and weighted Kappa measure, it was found that SMO's ability to rank company performance according to Moody's rankings is slightly better than logistic model and both SMO and logistic models are better than the neural network model. The bubbles distribution was also introduced in this research for detecting the similarity between two ordinal datasets.

One may question whether Moody's ranking should be used as a standard to measure the applicability of new models. Indeed, it seems paradoxical that logistic model and the neural network showed better performance than SMO in terms of sample classification, but that SMO seemed to be better in terms of practical applicability using Moody's as a benchmark. An explanation is that logistic and neural network fit closely to the sample and hence overfitting, whilst SMO does not.

Finally, it must be noted that the scope of this study was limited to publicly listed firms and the US retail market. The study can be extended to non-listed firms as well as other markets in retail in order to ensure each model's theoretical utility and practical applicability.

References

Altman, Edward I. (1968), Financial ratios, discriminate analysis and the prediction of corporate bankruptcy, *Journal of Finance*, Vol. 23, p589-609

Altman, Edward I., Robert G. Haldeman, and P. Narayanan (1977) ZETA™ ANALYSIS, *Journal of Banking and Finance*, Vol. 1, Issue 1, p29-54

Alves, J. R. (1978) The Prediction of Small Business Failure Utilizing Financial and Non-financial Data, PhD Thesis, University of Massachusetts

Beaver, William H. (1968) Alternative accounting measures as predictors of failure, *Accounting Review*, Vol. 43, Issue 1, p113-122

Blum, Marc. (1974) Failing company discriminant analysis, *Journal of Accounting Research*, Vol. 12, Issue 1, p1-25

Bowen, Robert M., Lane A. Daley, and Charles C. Huber, Jr. (1982) Evidence on the existence and determinants of inter-industry differences in leverage, *Financial Management (Financial Management Association)*, Vol. 11, p10-20

Carson, James M. and Robert E. Hoyt (1995) Life insurer financial distress: classification models and empirical evidence, *Journal of Risk and Insurance*, Vol. 62, Issue 4, p764-775

Casey, Cornelius and Norman Bartczak (1985) Using operating cash flow data to predict financial distress: some extensions, *Journal of Accounting Research*, Vol. 23, Issue 1, p384-401

Coates, P. and L. Fant (1993) Recognizing financial distress patterns using a neural network tool, Financial Management, Vol. 22, p142-155

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, Vol.20, p37-46.

Comrey, A. L., & H. B. Lee (1992). A first course in factor analysis, 2nd ed., Lawrence Erlbaum Associates, Inc, NJ

Dawson, John (2000) Retailing at century end: some challenges for management and research, *Internal Review of Retail, Distribution and Consumer Research*, Vol. 10, Issue 2, p119-148

Deakin, Edward B. (1972) A discriminant analysis of predictors of business failure, *Journal of Accounting Research*, Vol., 10 Issue 1, p167-179

Deakin, Edward B. (1976) Distributions of financial accounting ratios: some empirical evidence, *Accounting Review*, Vol. 51, Issue 1, p90-96

Dimitras, A.L., R. Slowinski, R. Sumaga and C. Zopounidis (1999) Business failure prediction using rough sets, *European Journal of Operational Research*, Vol. 114, Issue 2, p263-280

Edmister, Robert O. (1972) Financial ratios as discriminant predictors of small business failure, *Journal of Finance*, Vol. 27, Issue 1, p139-140

Eisenbeis, R. A. (1977), Pitfalls in the application of discriminant analysis in business, finance and economics, *Journal of Finance*, p875-900

Fan, A. and M. Palaniswami (2000) Selecting bankruptcy predictors using a support vector machine approach, *Proceedings of the Internal Joint Conference on Neural Networks.*

Fisher, R. A. (1936) The use of multiple measurements in taxonomic problems, *Annals of Eugenics*, Vol. 7, p179-188

Fitch (2000) Assigning Credit Ratings to European Retails, Fitch press, New York

Fitch (2001) Corporate: Corporate Rating Methodology, Fitch press, New York

Frydman, Halina, Edward l. Altman, and Duen-Li Kao (1985) Introducing recursive partitioning for financial classification: the case of financial distress., *Journal of Finance*, Vol. 40, Issue 1, p269-291

Gentry, James A., Paul Newbold, and David T. Whitford (1985) Classifying bankrupt firms with funds flow components, *Journal of Accounting Research*, Vol. 23, Issue 1, p146-160

Gupta, Manak C. and Ronald J. Huefner (1972) A cluster analysis study of financial ratios and industry characteristics, *Journal of Accounting Research*, Vol. 10, p77-95

Hamer, Michelle M. (1983) Failure prediction: sensitivity of classification accuracy to alternative statistical methods and variable sets, *Journal of Accounting & Public Policy*, Vol. 2, Issue 4, p289-307

Hand, David J. (1997) Construction and Assessment of Classification Rules, John Wiley & Sons Ltd, England

Harrison, Jeffrey S. (2003) Strategic Management of Resources and Relationships: Concepts and Cases, 1st edition, Wiley press, New York

Hasty, Ronald W. and James Reardon (1997), Retail Management, 1st edition, The McGraw-Hill Companies press, New York

Hosmer D.W., Lemeshow S. (2000) Applied Logistic Regression. Wiley, New York

Hunt, Shelby D. (2000) A General Theory of Competition, Sage Publications, Inc., California

Hunt, Shelby D. and Dennis B. Arnett (2001) Competition as a process and antitrust policy, *Journal of Public Policy and Marketing*, Vol. 20, p15-26

Hunt, Shelby D. and Dennis B. Arnett (2003) Resource-Advantage theory and embeddedness: explaining R-A theory's explanatory success, *Journal of Marketing, Theory and Practice*, Vol. 11, p1-17

Hunt, Shelby D. and Robert M. Morgan (1997) Resource-Advantage theory: a snake swallowing its tail or a general theory of competition?, *Journal of Marketing*, Vol. 61, p74-82

Libby, Robert (1975) Accounting ratios and the prediction of failure: some behavioral evidence, *Journal of Accounting Research*, Vol. 13, Issue 1, p150-161

Mar-Molinero, Cecilio and Carlos Serrano-Cinca (2001) Bank failure: a multidimensional scaling approach, *European Journal of Finance*, Vol. 7, Issue 2, p165-183

Marais, M. Laurentius, James M. Patell, and Mark A. Wolfson (1984) The experimental design of classification models: An application of recursive partitioning and bootstrapping to commercial bank loan classifications, *Journal of Accounting Research*, Vol. 22, p87-114

McKee, Thomas E. and Terje Lensberg (2002) Genetic programming and rough sets: A hybrid approach to bankruptcy classification, *European Journal of Operational Research*, Vol. 138, Issue 2, p436-451

Mensah, Yaw M. (1983) The differential bankruptcy predictive ability of specific price level adjustments: some empirical evidence, *Accounting Review*, Vol. 58, Issue 2, p228-246

Mensah, Yaw M. (1984) An examination of the stationarity of multivariate bankruptcy prediction models: A methodological study, *Journal of Accounting Research*, Vol. 22, Issue 1, p380-395

Moody's Investors Service Inc. (1998) Rating Methodology: Industrial Company Rating Methodology, Moody's Investors Service Inc. Press, New York

Moody's Investors Service Inc. (2002) Retail Rating Methodology: Moody's Approach to Assessing Key Credit Issues in Retailing, Moody's Investors Service Inc. Press, New York

Moore, Andrew W. (2001) Cross-validation for detecting and preventing overfitting, *Statistical Data Mining Tutorials*, available at: http://www.autonlab.org/tutorials/

Ohlson, James A. (1980), Financial ratios and the probabilities prediction of bankruptcy, *Journal of Accounting Research*, Vol. 18, p109-131

Pinches, George E., Kent A. Mingo and J. Kent Caruthers (1973) The stability of financial patterns in industrial organizations, *Journal of Finance*, Vol. 28, Issue 2, p389-396

Platt, Harlan D. and Marjorie B Platt. (1990) Development of a class of stable predictive variables, *Journal of Business Finance & Accounting*, Vol. 17 Issue 1, p31-51

Plat, J. C. (1998) Fast training of support vector machines using sequential minimal optimization In B. Scholkopf, C. Burges and A. Smola, Advances in Kernel Methods: Support Vector Machines, MA:MIT press, Cambridge

Rose, Peter S., Wesley T. Andrews and Gary A. Giroux (1982) Predicting business failure: A macroeconomic perspective, *Journal of Accounting, Auditing & Finance*, Vol. 6, Issue 1, p20-31

Sanger, Gary C. and John J. McConnell (1986), Stock exchange listings, firm value, and security market efficiency: The impact of NASDAQ, *The Journal of Financial and Quantitative Analysis*, Vol. 21, p1-25

Standard and Poor's (2003a), Standard and Poor's 2002 Corporate Rating Criteria, The McGraw-Hill Companies press, New York

Standard and Poor's (2003b), Standard and Poor's 2003 Corporate Rating Criteria, The McGraw-Hill Companies press, New York

Taffler, Richard J. (1984) Empirical models for the monitoring of UK corporations, *Journal of Banking and Finance*, Vol. 8, Issue 2, p199-228

Tam, Kar Yan and Melody Y. Kiang (1992) Managerial applications of neural networks: the case of bank failure predictions, *Management Science*, Vol. 38, Issue 7, p926-947

Thomas, Lyn C., David B. Edelman and Jonathan N. Crook (2002) Credit Scoring and its Applications, 1st edition, Philadelphia: Society for Industrial & Applied Mathematics

Trigueiros, Duarte and Richard Taffler (1996) Neural networks and empirical research in accounting, *Accounting & Business Research*, Vol. 26, Issue 4, p347-355

Wheelen, Thomas L. and J. David Hunger (2004) Concepts in Strategic Management and Business Policy, 9th edition, Pearson Prentice Hall, N.J.

Williams W. H. and M. L. Goodman (1971) A statistical grouping of corporations by their financial characteristics, *Journal of Financial and Quantitative Analysis*, Vol.6, p1095-1104

Zhang, G. P., M. Y. Hu, B. E. Patuwo, and D. C. Indro (1999) Artificial neural networks in bankruptcy prediction: general framework and cross-validation analysis, *European Journal of Operational Research*, Vol. 116, p16-32

Appendix A: Performance Measures Arrangement

Internal Resources Group

Resources	Principle	Main Measures
Financial Resources	• Profitability	1. EBIT margin 2. EBITDA margin 3. EBITDAR margin 4. Pre-tax profit margin 5. Pre-tax profit on capital 6. Net profit margin 7. Gross profit margin 8. SG&A as % of net sales 9. EBIT on capital 10. Return on total assets 11. Return on total equity 12. Operating margin 13. Dividend payout ratio
	• Liquidity	14. Current ratio 15. Acid ratio 16. Cash ratio
	• Sustainability	17. Net operating cash flow / gross capex 18. Cash dividend cover 19. Fixed charge cover 20. Interest cover 21. Funds from operations / total debt 22. EBITDA / interest 23. Total debt / discretionary cash flow
	• Leverage	24. Gearing ratio 25. Debt / EBITDA 26. Leased-adjusted net debt / EBITDAR 27. Net debt / market capitalization 28. Total debt / (total debt + market capitalization) 29. Debt to equity ratio

Internal Resources Group (Continue)

Resources	Principle	Main Measures
Financial Resources	• Market Measure	30. P/E ratio
	• Financial Scale	31. Net sales 32. Total assets 33. Market share by retail sector (based on sales) 34. Market share by retail sector (based on gross margin) 35. Total capital employed 36. Operation cash flow
Physical Resources	• Reach Ability	37. Store numbers
Legal Resources	• Brand Strength	38. Market capitalization / net assets
Human Resources	• Human Resource Quality	39. Sales per employee 40. EBIT per employee
	• Human Resource Management	41. Number of payrolls
Organizational Resources	• Actability	42. Total assets turnover 43. Fixed assets turnover
	• Growth Power Analysis	44. Sales growth 45. Market value growth 46. Capital growth 47. EBIT growth 48. Number of stores growth 49. The operating income growth 50. Number of payrolls growth
	• Financial Management	51. Net cash cycle
Informational Resources	• Market Segment Risk management	52. Main market sales as percentage total sales
Relational Resources	• Customer Relations Management	53. Receivable turnover
	• Supplier Relations Management	54. Inventory turnover 55. Payables turnover

External Environmental Factors

Factors	Principle	Main Measures
Societal Resources, Societal Institutions and Actions from Government	• Political Environmental Factors	56. The correlation coefficient between government debt / GDP and total sales 57. The correlation coefficient between government avenue / GDP and total sales 58. The correlation coefficient between government expense / GDP and total sales
	• Economic Environmental Factors	59. The correlation coefficient between real GDP and total sales 60. The correlation coefficient between average interest rate and total sales 61. The correlation coefficient between unemployment rate and total sales 62. The correlation coefficient between disposable income and total sales
	• Technological Environmental Factors	63. The correlation coefficient between total government spending for R&D and total sales
	• Socio-cultural Environmental Factors	64. The correlation coefficient between birth rate and total sales 65. The correlation coefficient between death rate and total sales 66. The correlation coefficient between age structure ratio (0-14years old) and total sales 67. The correlation coefficient between age structure ratio (65 years and above) and total sales