Advances and Applications in Statistics
Volume 11, Number 1, 2009, Pages 29-45
Published Online: April 28, 2009
This paper is available online at http://www.pphmj.com

MULTIVARIATE MODEL FOR BANKRUPTCY PREDICTION

MOHAMMAD AHMAD ALI AL-SALEH and SUHAILA HUMOUD AL-FARHOUD

Department of Statistics
College of Business Studies
PAAET, Kuwait

Abstract

In view of the failure of high profile companies like HIH and Enron, financial distress or bankruptcy prediction has generated lots of interest recently. In this paper, numerous financial ratios of American Health and Medical companies have been analysed to determine which companies will be the best for successful investment. Some guiding discriminate rule is given and a few factors were identified as measures of profitable company.

1. Introduction

Since its inception by Altman [1], the use of financial ratio models to predict failure and success in business firms has continued to generate much discussion in literature. These models are generated and tested using successful and financially distressed firms. The distressed firms are often bankrupt or liquidated companies with often zero or negative cash flows and assets with huge debts.

Using financial ratios and some multivariate techniques, this paper

2000 Mathematics Subject Classification: 62H30.

Keywords and phrases: bankruptcy prediction, financial distress modelling.

Communicated by Kuldeep Kumar

Received April 17, 2008; Revised 16, July 2008

attempts to construct models for predicting the success and failure of companies in the medical industry. This is the main objective of this study.

The remainder of this paper is organised as follows. In Section 2, a brief literature review is discussed. In Section 3, the methodological procedures used are outlined. Section 4 presents and discusses the results. The paper ends with a summary and conclusion.

2. Literature Review

Models employing financial ratios in the assessment of risk, prediction of financial distress, credit rating, and firm failure are widely discussed in literature beginning with the seminal works of Altman [1] and Beaver [3]. More recent studies have extended these models to various prediction scenarios with the most notable by Ohlson [8], who assesses the predictive power of a set of models for decision purposes. The underlying theme common in these studies is the use of historical financial data to predict failure or success among a sample of financially distressed companies.

From the original predictions of success and failure these models have further developed to include distinguishing bankrupt firms from those that are financially distressed (Gilbert et al. [7]), firms that liquidate from those that reorganize (Casey et al. [5] and Campbell [4]), and in prediction of court resolutions on whether firms should acquire, emerge or liquidate (Barniv et al. [2]).

Over the years a variety of statistical analyses techniques has been employed in constructing these models. The origins would probably be traced back to Beaver [3] who employed univariate statistics in his model of distress prediction. Altman [1] constructed a model using multivariate discriminant analysis (MDA) and Ohlson [8] used logistic regression, which is today the most widespread technique employed in firm failure prediction studies. Previous studies have often included an equally matched sample of firms to ensure the robustness of the models and the ability of these models to discriminate between firms that thrive and those that struggle financially or even die. This study will attempt to construct prediction models employing some multivariate techniques and compare the prediction accuracies of these models.

3. Methodology

Failure prediction models were developed using a sample of liquidated and financially healthy firms obtained from US firms associated with the health and medical industry. The developed models are therefore more specific to firms in this market and this industry.

a. Sample selection

Initially a sample of 100 firms was randomly selected. However after 'data examination' the final sample used in this study got reduced to 85 firms. This sample consisted of poor performing firms and successful firms. The firms were all from one industry because having the sample spread across several industries can confound the results.

Financial ratios and annual reports were obtained from the Securities Exchange Commission (SEC) in the United States. A total of 24 financial ratios were obtained for all 85 firms. Some of the ratios were provided while others were calculated from financial reports.

b. Variables of interest

This section briefly discusses the variables and the statistical methods adopted for use in this study. Previous research shows the lack of a well-accepted theory to guide the selection of financial ratios (Casey et al. [5]). No attempt was made to create new financial ratios; rather existing ratios were selected as the independent variables in this study. These ratios were found to be predominantly used and tested in previous models predicting firm failure or success. The 22 ratios are shown in Table 1 of the Appendix.

The dependent variable is the failure or success of a firm and a dummy variable with a binary measure was used where 1 denoted successful firms and 0 represented underperforming and failed firms.

c. Measures and statistical methods

A few multivariate techniques were employed for testing, analysis and construction of the distress prediction models. All techniques were performed using the statistical software package SPSS. These techniques are discussed separately below.

i. Data examination

Before any of the techniques could be applied, the data needed to be tested and checked. A visual inspection of the raw data was the initial approached, then leading to an analysis of frequency tables and descriptive statistics. The data was also checked for the presence of outliers which would affect results leading to incorrect interpretation. Outliers, missing data and unavailability of information led to the removal of 25 firms and leaving a final sample of 85 firms. Normality and linearity testing were preformed to meet the required assumptions of the statistical techniques. A correlation analysis on the 22 independent variables was also undertaken to avoid the possibility that some ratios are redundant.

ii. Logistic regression

Logistic regression is the most commonly used technique in recent literature on studies of financial models predicting firm outcome (Barniv et al. [2]). A logit model using binary logistic regression is shown to be a more optimal method for studies of this nature (Cybinski [6] and Ohlson [8]) and is one of the techniques used here.

The developed model was developed where the probability of failure is given by the logistic function:

$$\pi_i = \frac{e^{Zi}}{1 + e^{Zi}},$$

where $Zi = b_0 + b_1 X_1 + \cdots + b_p X_p$,

π_i – probability the ith case experiences the event of interest which in our case is 1 for success and 0 for failure with cut-off at 0.5,

Zi – value of the variable for the ith case,

b_0 – coefficient of constant,

b_1 to b_p – coefficients of independent variables in model. p stands for the number of independent variables,

X_1 to X_p – values of financial ratios.

iii. Linear discriminant analysis

Linear discriminant analysis is another technique useful for building a predictive model. The procedure generates a discriminant function based on linear combinations of the predictor variables that provide the best discrimination between the groups. The functions are generated from a sample for which group membership is known.

Unlike binary logistic regression the grouping variable can have more than two values. However as with before we only need to discriminate between two outcomes and the 1 (success) and 0 (failure) were used as the integer codes. Discriminant analysis allows the estimation of the coefficients of the linear discriminant function. For example using coefficients a, b and c, the discriminant function can be represented as:

$$d_{ik} = b_{0k} + b_{1k}x_{i1} + \cdots + b_{pk}x_{ip},$$

where

d_{ik} is the value of the kth discriminant function for the ith case,

p is the number of predictors,

b_{jk} is the value of the jth coefficient of the kth function,

x_{ij} is the value of the ith case of the jth predictor.

There are several assumptions that must be met to use discriminant analysis properly. These are:

– The predictors are not highly correlated with each other.

– The mean and variance of a given predictor are not correlated.

– The correlation between two predictors is constant across groups.

– The values of each predictor have a normal distribution.

iv. Principal component and factor analysis

Principal component and factor analysis are somewhat related techniques and are discussed together. These techniques attempt to identify underlying variables, or components, that explain the pattern of correlations within a set of observed variables. Factor analysis is often used in data reduction to identify a small number of factors that explain most of the variance that is observed in a much larger number of manifest variables. Factor analysis can also be used to generate hypotheses regarding causal mechanisms or to screen variables for subsequent analysis (for example, to identify collinearity prior to performing a linear regression analysis). These techniques are employed to see whether the number of independent variables (financial ratios) can be reduced to a smaller number without loss of much information.

v. Cluster analysis

Cluster analysis is a multivariate statistical technique which assesses the similarities between cases of interest, based on the occurrence or non-occurrence of specific artifact types or other components within them. These artifacts are often termed ‘outliers’. However, it can also be used to search for natural groupings in the variables of interest and if a pattern is identified, then cluster analysis can be used for further research and generation of hypothesis testing. Using cluster analysis here attempts to test whether based on the financial ratios of failed and successful companies, there is a clustering pattern that can be identified. Wards minimum variance method is used here as this is the most appropriate technique to this data set.

4. Results and Discussion

This section presents and discusses the results obtained via analysis of our sample using SPSS.

a. Data examination

A visual inspection of the raw data for the 85 firms did not reveal any serious errors or missing values. Descriptive statistics of the independent variables employed in this study are found in Table 2 in the Appendix.

Together with frequency tables, the data did not reveal any mistakes or 'out-of-range' responses. Histograms and box plots showed normality of data. T-tests for differences in the means in ratios between successful and unsuccessful companies showed some ratios to be significantly different and others not. These are presented in Table 3 of the Appendix. A correlation analysis showed some independent variables have a high correlation to each other and suffer from a possible redundancy of ratios.

b. Logistic regression model

A general rule of thumb for regression is a sample size of at least 10 times the number of independent variables is appropriate. This is satisfied here with 17 independent variables and a sample size of 85 companies. Using stepwise procedure the results of running logistics regression on the financial ratios are shown in Table 1. Only 3 ratios of the 17 ratios initially entered into the logistics regression were found to be significant in predicting distress in companies.

Table 1. Results of logistic regression

Ratios	B	S.E.	Wald	df	Sig.	Exp(B)
ROA1	3.77	1.09	11.94	1	0.001	43.543
QATA	2.89	1.24	5.41	1	0.020	17.910
TDTA	−3.13	1.55	4.09	1	0.043	0.044
Constant	−0.40	0.74	0.30	1	0.586	0.669

The derived estimated equation model for distress prediction is written as:

$$Z = -0.40 + 3.77(\text{ROA}\,1) + 2.89(\text{QATA}) - 3.13(\text{TDTA}).$$

Using a cut-off value of 0.5, the model was able to correctly predict 81% of the unsuccessful firms and 73% of the successful firms. The overall predictability accuracy of the logistic regression model was 77.6%. These are shown in Table 2.

Table 2. Prediction classification accuracy

Observed		**Predicted**		
		Outcome		% Correct
		0	1	
Outcome	0	39	9	81.3
	1	10	27	73.0
Overall %				**77.6**

The Hosmer-Lemeshow test of goodness of fit is 0.278. For the model to be a poor fit the shown statistic must have a significance value that is less than 0.05. For this model the value is greater than 0.05 and therefore the logistic model passes the goodness of fit test and is acceptable.

An application of the prediction model is illustrated

Using the logistic regression prediction model a practical example of the application of the model is presented. A sampled firm gave the following estimated parameter coefficients:

ROA1: −0.16,

QATA: −0.86,

TDTA: 0.27.

Using the model formula:

$$\begin{aligned} Z &= (-0.40) + (3.77 * -0.16) + (2.89 * 0.86)(3.13 * 0.27) \\ &= 0.64. \end{aligned}$$

Using the Z score the prediction outcome π is then calculated

$$\pi = \frac{e^{0.64}}{1 + e^{0.64}} = 0.65.$$

With the cut-off value at 0.5 we conclude that this firm has a greater propensity of success than failure, given those ratio values.

c. Linear discriminant analysis

The Box-M test statistic satisfied the assumption that the covariances across the two groups are equal. The results of the linear discriminant analysis are given in Table 3. The model correctly predicts 71% of failed firms and 78.83% of successful firms.

Table 3. Classification function coefficients

Ratios	**Outcome**	
	Failure (0)	Success (1)
PTM	−0.45	−0.86
ROA1	−3.73	0.34
CR1	1.76	1.46
QR	−1.38	−0.84
CR2	0.47	0.28
CTA	0.37	0.73
CATA	19.12	18.86
QATA	−4.08	−2.33
TDTA	30.55	21.46
GR	−0.38	−0.34
DTIC	−13.48	−9.12
CETIC	0.34	0.21
TETA	1.50	1.53
AR	1.47	1.40
SR	3.83	3.53
SNW	0.46	0.79
(Constant)	−13.70	−13.31
No. of obs.	**48**	**37**
Correctly classified	**34**	**29**
% classification	**70.83**	**78.38**

Using stepwise discriminant analysis the ratios ROA1, QATA and TDTA were found significant in predicting financial distress. The two equations derived were:

Success (1)

$$Y = -0.63 - 0.63(\text{ROA1}) + 12.72(\text{QATA}) - 10.17(\text{TDTA}),$$

Failure (0)

$$Y = -4.60 - 3.77(\text{ROA1}) + 10(\text{QATA}) - 13.32(\text{TDTA}).$$

Using these equations the discriminant function predictability increased to 72.9% for failed companies and 83.8% for successful companies.

d. Principal component and factor analysis

The principal components method of extraction begins by finding a linear combination of variables (a component) that accounts for as much variation in the original variables as possible. A correlation matrix was constructed to check if the financial ratios are highly correlated. The correlation matrix is shown in Table 4 of the Appendix. It can be seen that only a few correlations showed high values (> 0.7).

Table 4 shows that the first four principal components explain approximately 75% of the total variation in the system.

Table 4. Eigen vectors, eigen values and cumulative percentage

	Component			
	PC1	PC2	PC3	PC4
CTA	−0.8745	−0.0908	0.1668	0.2726
CR2	−0.7977	0.0263	0.3142	0.0337
QATA	−0.7702	0.3111	0.0934	0.4122
DTIC	0.7656	−0.2874	0.4372	0.2143
TDTA	0.7508	−0.2826	0.4972	0.1710
CATA	−0.7330	0.3163	−0.0941	0.3456
SR	0.7177	−0.1017	0.1823	0.4558
QR	−0.6786	0.2816	0.5200	0.1556
SNW	0.6435	0.4788	−0.1586	0.2564
CR1	−0.5896	0.3068	0.4883	0.0348

GR	0.5286	−0.3266	0.4406	0.2585
TETA	−0.5192	0.0864	0.1117	−0.2723
AT	0.4794	0.7911	−0.2313	0.1934
AR	0.4785	0.7907	−0.2315	0.1943
PTM	0.4697	0.5674	0.3147	−0.3322
ROA1	0.2847	0.5266	0.5238	−0.4175
CETIC	−0.1078	−0.2387	−0.4938	0.1046
Eigen values	6.74	2.80	2.08	1.25
Cumulative %	39.65	56.13	68.37	75.73

In factor analysis the factors are rotated to maximally explain the variation in the system. The data was tested for its factorability and shown to be significant at 5% level of significance. The Kaiser-Meyer-Olkin measure of sampling adequacy gave a value of 0.748. The principal axis factoring method was used with varimax rotation. Only factors loading with an eigenvalue greater than 1 and with correlations greater than 0.3 were considered. These results are presented in Table 5.

Table 5. Factor results of financial data

Total	**% of variance**	**Cumulative %**
3.91	22.97	22.97
3.23	18.99	41.96
2.96	17.41	59.37
1.83	10.76	70.12

	Factor			
	1	2	3	4
QATA	**0.86**			
QR	**0.84**			
CTA	**0.79**		−0.36	

CR2	**0.68**		−0.37	
CATA	**0.67**	−0.40		
CR1	**0.67**			
TETA	**0.32**	−0.31		
TDTA		**0.92**		
DTIC		**0.91**		
SR		**0.70**	0.30	
GR		**0.66**		
AT			**0.96**	
AR			**0.96**	
SNW			**0.69**	
ROA1				**0.84**
PTM			0.35	**0.70**
CETIC				**−0.34**

Extraction method: Principal axis factoring.
Rotation method: Varimax with Kaiser normalization.

Like principal component analysis, factor analysis returned 4 factors with eigenvalue greater than 1 that explained approximately 70% of the variation. Factor 1 loads with ratios QATA, QR, CTA, CR2, CATA, CR1 and hence this factor can be called '*liquidity measures*'. Factor 2 loads with ratios TDTA, DTIC, SR and GR and these can be collectively referred to as 'Debt management measures'. Factor 3 loads with AT, AR and SNW, and can be renamed as 'Sales measures'. The 4th factor loads with ROA1, PTM and CETIC. These can be grouped as 'Profitability measures'.

e. Cluster analysis

Cluster results are presented in table in the Appendix. The dendrogram confirms the grouping obtained by PCA and factor analysis. Three distinct groups can be seen and possibly represent successful firms, failed firms and those firms around the centre.

5. Discussions and Conclusions

This study applied multivariate techniques to financial ratio data for constructing prediction models and analysis. The effect on the output predictability of the estimated logistic regression model and linear discriminant models were estimated. These models gave 77.6% and 78.38% prediction accuracies.

Factor analysis, principal components analysis and cluster analysis gave some meaning descriptions of the data and results that can be used to generate further hypothesis and research.

A second limitation of this study is availability of data and the study was therefore limited to publicly traded companies. It was also conducted on a particular industry only and therefore the results can suffer from external validity. The ratio data figures for the sampled firms were obtained from online databases with no checks as to its accuracy. Adjusting for these limitations should be noted in further research.

References

[1] E. I. Altman, Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, J. Fin. 23 (1968), 589-609.

[2] R. Barniv, A. Agarwal and R. Leach, Predicting bankruptcy resolution, J. Bus. Fin. Account. 29 (2002), 497-520.

[3] W. H. Beaver, Financial ratios as predictors of failure, J. Account. Res. 4 (1966), 71-111.

[4] S. V. Campbell, Predicting bankruptcy reorganization for closely held firms, Accounting Horizons 10 (1996), 12-25.

[5] C. J. Casey, V. E. McGee and C. P. Stickney, Discriminating between reorganized and liquidated firms in bankruptcy, Account. Rev. 61 (1986), 249-262.

[6] P. J. Cybinski, A discrete-valued risk function for modelling financial distress in private Australian companies, Account. Fin. 35 (1995), 17-32.

[7] L. R. Gilbert, K. Menon and K. B. Sscwartz, Predicting bankruptcy for firms in financial distress, J. Bus. Fin. Account. 17 (1990), 161-171.

[8] J. A. Ohlson, Financial ratios and the probabilistic prediction of bankruptcy, J. Account. Res. 18 (1980), 109-131.

Appendix

Table 1. Financial ratios

Abbrev.	Variable name and definition
ROA1	Return on assets (EBIT / TA) We expect a positive estimated coefficient in the logistic regression because as this value increases, the probability of failure decreases
ROA2	Return on assets after tax (EAT/TA)
GR	Gross gearing (Debt / Equity) We expect a negative estimated coefficient in the logistic regression model because as this value increases, the likelihood of failure also increases
CR1	Current ratio (Current assets / Current liabilities) We expect a positive estimated coefficient in the logistic regression because as this value increases, the probability of failure decreases
QR	Quick ratio – cash plus accounts receivable / Current liabilities Liquidity measure
CR2	Cash ratio – cash / Current liabilities Liquidity measure
CTA	Cash to total assets
CATA	Current assets to total assets
QATA	Quick assets to total assets
OM	Operating margin
PTM	Pre-tax margin
ATM	After tax margin
OP	Operating profitability
PR	Profitability ratio
RICBT	Return on invested capital before tax
RICAT	Return on invested capital after tax
TDTA	Total debt to total assets
DTIC	Debt to total invested capital
CETIC	Common equity to total invested capital
TETA	Total equity to total assets
AR	Activity ratio
SR	Solvency ratio
AT	Asset turnover – sales / fixed assets
SNW	Sales to net worth

Table 2. Descriptive statistics of the financial ratios used in this exercise

Ratios	Mean	S.D.	Range	Variance	Skew	Kurtosis
OM	−0.67	1.18	4.48	1.39	−1.35	0.58
PTM	−0.68	1.24	6.39	1.54	−1.97	4.05
ATM	−0.72	1.26	6.01	1.58	−1.90	3.51
RICBT	−0.30	0.63	3.37	0.40	−2.28	5.88
RICAT	−0.33	0.65	3.45	0.42	−2.52	7.09
ROA1	−0.17	0.36	1.77	0.13	−1.18	0.95
ROA2	−0.21	0.37	1.76	0.14	−1.51	1.93
CR1	2.51	1.63	7.81	2.67	1.47	1.92
QR	2.03	1.70	7.43	2.89	1.55	1.79
CR2	1.28	1.65	7.14	2.73	1.85	3.41
CTA	0.26	0.29	0.93	0.08	1.03	−0.28
CATA	0.57	0.26	0.98	0.07	−0.20	−1.07
QATA	0.42	0.26	0.93	0.07	0.41	−1.02
TDTA	0.20	0.22	0.80	0.05	0.92	−0.37
GR	0.88	2.08	16.76	4.32	5.76	41.19
DTIC	0.26	0.28	0.94	0.08	0.80	−0.74
CETIC	0.47	1.60	10.74	2.56	4.73	24.83
TETA	0.72	1.03	8.69	1.07	6.31	45.40
OP	−0.63	1.25	7.39	1.56	−1.88	4.16
PR	−0.68	1.35	9.61	1.83	−1.10	3.88
AR	0.84	0.66	3.12	0.43	1.20	1.40
SR	1.92	0.94	4.74	0.88	1.21	1.78
AT	0.84	0.65	3.09	0.43	1.18	1.33
SNW	1.59	1.29	4.72	1.65	0.83	−0.14

Table 3. Cluster results

	Clusters		
	1	2	3
PTM	−0.53	−0.06	−1.10
ROA1	−0.17	0.02	−0.20
CR1	1.78	1.30	4.34
QR	1.20	0.70	4.08
CR2	0.45	0.04	3.32
CTA	0.14	0.01	0.55
CATA	0.50	0.24	0.77
QATA	0.33	0.11	0.66
TDTA	0.23	0.71	0.08
GR	0.76	9.26	0.14
DTIC	0.30	0.88	0.09
CETIC	0.62	0.12	0.16
TETA	0.59	0.10	1.08
AR	1.01	0.77	0.47
SR	2.14	2.55	1.34
AT	1.01	0.77	0.47
SNW	2.01	1.33	0.66

Table 4

	PTM	ROA1	CR1	OR	CR2	CTA	CATA	QATA	TDTA	GR
PTM	1.000									
ROA1	0.713	1.000								
CR1	–0.042	0.154	1.000							
OR	–0.060	0.127	0.837	1.000						
CR2	–0.242	–0.122	0.626	0.673	1.000					
CTA	–0.443	–0.271	0.462	0.643	0.709	1.000				
CATA	–0.260	–0.171	0.404	0.513	0.521	0.692	1.000			
QATA	–0.249	–0.106	0.474	0.666	0.619	0.855	0.793	1.000		
TDTA	0.275	0.260	–0.275	–0.297	–0.449	–0.489	–0.634	–0.538	1.000	
GR	0.138	0.112	–0.223	–0.231	–0.264	–0.299	–0.398	–0.361	0.687	1.000