Research Paper

GYAN MANAGEMENT

AN INTERNATIONAL BI-ANNUAL REFEREED JOURNAL OF MANAGEMENT & TECHNOLOGY

Year 2024, Volume-18, Issue-2 (July - December)

ISSN 0974-7621 (Print)

ACS Publisher www.acspublisher.com

Default Risk Prediction in Firms via Statistical Techniques – A review

Sakshi Khurana

Senior Research Fellow, University Business School, Panjab University, Chandigarh.
University Business School, Arts Block III, Madhya Marg, Sector 14, Panjab University, Chandigarh.
Email ID: khurana.2sakshi@gmail.com

ARTICLE INFO

Key Words: *default risk prediction; discriminant analysis, logistic regression*

ABSTRACT

The current study presents the literature review of default risk prediction studies for the period 1930-2022. The rising corporate default is a reason of concern for banks and companies. It has particularly attracted the attention of researchers since the global crisis of 2008-09. The review is based on statistical techniques including discriminant analysis, multiple discriminant analysis, logistic regression, conditional logistic regression and probit regression. The review targets the following aspects of the studies: country origin, period of study, financial indicators used and accuracy of results. Analysis of review depicts that logistic regression (LR) is the most applied technique among statistical techniques to predict corporate failure. However, focus has now been shifting to intelligent techniques in 21st centuary.

Introduction

Prediction of financial health is an important function of business firms. Prediction of corporate failure in advance can help in avoiding bankruptcy situations for a firm. It is a situation where a firm fails to fulfill its financial obligations (Altman, 1968). The situation of business failure has extremely negative consequences for stakeholders. For instance, the collapse of General Motors in 2009 severely affected its employees who were laid off due to this failure. The contagious effects of corporate failure attracted the attention of researchers to come up with mechanisms that can predict default risk before it moves to the insolvency stage.

The literature has provided various default forecasting models to predict corporate failure in real-time

Received 12.05.2023; Accepted 11.06.2023
DOI: 10.48165/gmj.2023.conf6

before it becomes bankrupt. The correct classification by models between firms that go bankrupt and firms that remain insolvent is considered to be an important achievement of modern finance (Scott, 1981). There has been a significant improvement in default risk literature since the formal univariate studies in the 1930s. This paper will cover the review of those research papers where bankruptcy prediction models were developed utilizing statistical techniques during the period 1930-2022. The main objective of this paper is to draw the attention of researchers towards this interesting research problem of failure prediction that will enrich the default risk literature by providing solutions to problems faced by banks and financial institutions.

The study is useful for researchers who are working on developing the default prediction models by providing evidence about the statistical technique that produces the most accurate results. Secondly, it contributes to the literature on accounting and finance by demonstrating the country producing the maximum number of default prediction studies. The current review covers the following aspects in upcoming sections: (i) it mainly focuses on statistical techniques. (ii) It gives a summary of how default prediction models have evolved in nine decades. (iii) It will provide a base to pursue further research work in this study area.

Review methodology

Bankruptcy prediction models are developed by applying two types of techniques: (a) Statistical techniques and (b) Intelligent techniques. The scope of this review is limited to statistical techniques only. The current review has summarized the bankruptcy prediction work done in the last nine decades The methods covered under statistical techniques are (i) Multiple discriminant analysis, (ii) logistic regression, (iii) conditional logistic regression and (iv) probit regression. The study also covers the comparison between statistical and intelligent techniques. Further, the review covers the following characteristics, namely, the time period of the study, sources of data, country of origin, various financial indicators used to predict corporate default, and the results obtained for the review studies. Table 1 presents the details for all papers such as (i) number of ratios considered, (ii) source of data, (iii) time period of the sample, (iv) country of origin, (v) technique applied in the study, (vi) model accuracy.

Summary of bankruptcy prediction studies: 1930 to 2022

The former studies related to default prediction were based on univariate analysis where a single ratio was used to discriminate between a distressed and non-distressed firm. The literature demonstrated that there was a significant difference between the mean ratio characteristics of default and non-default firms. The significance of univariate analysis in corporate default prediction was formally witnessed in the study of Beaver (1966). The results indicated that net profit to total debt had the highest prediction accuracy of 92 percent for one year before default. The first study using multiple ratios, namely, profitability, leverage, solvency, liquidity and activity ratios was conducted on US-based firms by applying multiple discriminant analysis (Altman, 1968). It is one of the most highly used and cited default prediction models in accounting literature with prediction accuracy of 95 percent one year before default. Deakin (1972) found that discriminant analysis had a high predictive capability and was a reliable measure of prediction for at least three years in advance.

Ohlson (1980) later applied logistic regression to overcome the difficulties related to discriminant analysis. The results indicated nine financial ratios to be significant contributors to default prediction. Scott (1981) developed a new theory for default prediction. He found that existing theories explain the success of empirical models. However, new theories of corporate default are required to be developed for the changing time periods and financial conditions.

Zmijewsiki (1984) developed a default forecasting model for US-based firms by applying the probit regression technique. The study mentioned two estimation biases in a study that arise when sample data is collected non-randomly. The first bias results from the 'oversampling' of distressed firms and falls under the topic of choice-based sample biases. The second results from using a 'complete data' sample selection criterion and is included under the head of sample selection biases.

Table 1: Brief review of some related articles

Reference	Author's Name and Year of Publication	Country of Origin	Samples	Techniques Used / Independent Variables (Financial Ratios)	Sample Period	Model Accuracy
[1][1]	Altman (1968)	USA	66	MDA / 5	1946-65	For initial sample- 95% for one year prior to bankruptcy
[2]	Aggarwal (2019)	INDIA	270	MDA, Logistic Regression / 1	2000-12	Estimated sample Hold out Sample LR 81.7% 65.6% MDA 80% 63.3%
[3]	Bandopadhyay (2006)	INDIA	104	MDA, Logistic Regression / 5	1998-03	MDA- 91%
[5]	Bhunia And Sarkar (2011)	INDIA	64	MDA / 16	1996-05	Model accuracy: 86-96% for each of 5 years prior to failure
[4][2]	Beaver (1966)	USA	158	UA / 30	-	Model accuracy: 50% to 92%
[6]	Chen (2011)	TAIWAN	240	LDA, LR, C5.0, CART, SVM, DTC / 8	2000-08	SVM had the highest accuracy with 86.79% level
[7]	Deakin (1972)	USA	-	MDA / 14	1969-75	Model accuracy: For failed Firms (1 year prior to failure)- 77% For non-failed firms (1 year prior to failure) - 82%
[8]	Doumpos and Zopounidis (1999)	Greece	118	M.H. DIS, DA, LOGIT / 8	1986-90, 1988-92	
[10]	Jo, Han and Lee (1997)	Korea	544	MDA, NN, CBF / 61	1991-93	Average accuracy of three models- 81.5 to 83.8%
[11]	Jones and Hensher (2004)	Australia	6241	Mixed Logit Model, MNL / 7	1996-2000	Mixed Logit- 98.73%, MNL- 90%
[13]	Karels and Prakash (1987)	-	14 in 1972, 11 in 1976	MDA / 5	1972,76	Bankrupt firms- 54.5% Non-bankrupt firms- 96.0%
[14][3]	Kim (2011)	Korea	33	Logistic Regression, SVM, ANN, MDA / 17	1995-02	MDA- 72.6%, LR- 80.00%, ANN- 91.6%, SVM- 95.95%
[15]	Lee and Choi (2013)	Korea	229	BNN, MDA / 6	2000-09	For general model- BNN- 81.43% MDA- 74.82
[16]	Lennox (1999)	UK	949	Probit / 9	1987-94	
[17]	MatusMihalovic (2016)	SLOVAK REPUBLIC	236	DA, Logistic Regression / 5	2014	Training Data Test Data Discriminant Function 61.86% 64.41% Logit Function 73.73% 68.64%
[19]	Ohlson (1980)	USA	2163	Logistic / 9	1970-76	For 1 Year prior to bankruptcy Model 1 Model 2 Model 3 96.12% 95.55% 92.84%

[1]**MDA**- Multiple Discriminant Analysis; LR- Logistic Regression
[2]**UA**- Univariate Analysis; **LDA**- Linear Discriminant Analysis; **LR**- Logistic Regression; **SVM**- Support Vector Machine; **DTC**- Decision Tree Classification; **CBF**- Case Based Forecasting; **MNL**- Multinominal Logit; MDA- Multiple Discriminant Analysis; '-' :Indicates Data Not Available
[3]**LR**- Logistic Regression; **SVM**- Support Vector Machine;**MDA**- Multiple Discriminant Analysis;**ANN**- Artificial Neural Networks; BNN- Back Propagation Neural Networks

[21]	Shumway (2001)	USA	300	Hazard / 13	1962-92	Hazard model outperformed all other statistical models
[22]	Singh And Mishra (2016)	INDIA	208	MDA, LOGIT, PROBIT / 4	2006-14	Re-estimated model accuracy Altman- 96.923 Ohlson- 95.938 Zmijewsiki- 89.231
[24]	Wu, Liang and Yang (2008)	Taiwan	48	BPNN, MDA / 7	1998-02	BPNN- 81.25% MDA- 48%
[25]	Zmijewsiki (1984)	USA	1681	Probit / 6	1972-78	Bankrupt firms- 20%, Non-Bankrupt firms- 99.5%

According to Grice and Dugan, (2001), the most cited default forecasting studies include Altman (1968), Ohlson (1980) and Zmijewsiki (1984). Karel and Prakash (1987) conducted their research for three main objectives. (i) whether the financial ratios used in previous studies fulfilled normality conditions (ii) the construction of financial ratios that were either completely normal or near to normal (iii) comparing the results of newly constructed ratios using a discriminant model with the results of previous studies. The results reported a classification accuracy of 96% for healthy firms and 54.5% for failed firms.

Since then, several studies have been conducted drawing comparisons between classification methods such as discriminant analysis, logistic regression and probit regression. Lennox (1999) found that logistic and probit regression more efficiently predicted the defaulting firms. The results further identified the important indicators of corporate default, namely, profitability, leverage, cash flow and size. Bandopadhyay (2006) developed a new default prediction model for the Indian bond market using MDA and logistic regression. The results indicated that the new default forecasting model outperformed the original Altman's Z-Score (1968) model. Jones and Hensher (2004) compared the efficiency of mixed logit to that of standard logistic regression. The results indicated that the new mixed logit outperformed the standard logit model by massive margins.

Mihalovic (2016) compared the results of logistic regression versus discriminant analysis using data from companies listed in the Slovak Republic. The results suggested that the logit model had an incremental predictive accuracy than multiple discriminant analysis. Chen (2011) compared statistical methods for predicting corporate failure with intelligent techniques such as neural networks and decision tree classification. The empirical results showed that statistical techniques were more accurate for predicting financial distress for large sample size while intelligent techniques showed better accuracy for small datasets. Lee and Choi (2013) outlined the importance of industry factors in predicting corporate default. The results further indicated that the back-propagation neural network had a higher accuracy than multivariate discriminant analysis. Similarly, Jo et al., (1997) concluded that neural networks outperformed the multiple discriminant analysis by significant margins.

Singh and Mishra (2016) developed a revised default forecasting model using data from Indian bankrupt and non-bankrupt companies. The study re-estimated the original Altman's Z-score, Ohlson's X-score and Zmijewsiki's Y-score models and found that new models outperformed the original models with significant margins. Kim (2011) conducted a study to develop an optimal bankruptcy model for Korean hotels by investigating the functional characteristics of Multivariate Discriminate Analysis, Logistic, Artificial Neural Networks, and Support Vector Machine (SVM). When taking into account both Type 1 and Type 2 errors, the study found that the error rate of ANN was lower than that of SVM and ANN should be considered for prediction of hotel bankruptcy.

Similarly, Du Jardin (2010) found that the set of variables chosen for representing the financial characteristics of healthy companies plays a role in reducing Type I error. Recently, Agrawal and Maheshwari (2019) assessed the impact of industry beta on the firm's chances of default. They applied logistic regression and MDA on a sample of Indian firms and found that industry beta is significant in predicting corporate failure. Gupta and Jain, (2021) applied multinomial logistic regression and indicated that their model is robust in predicting the default probability of firms and can be used by banks and financial institutions for credit risk assessment.

Discussion

Table 1 where other dimensions of the review are presented indicates that maximum studies of bankruptcy prediction have been conducted in the USA. The study which brought revolution and laid down a benchmark for forecasting the bankruptcy of firms found its roots in the USA. The world's most applicable model Altman Z-Score was also developed in the same country using data from US firms. The table also depicts that among the statistical techniques, logistic regression is the most popular technique applied by researchers followed by MDA and Probit. Some studies where the comparison between statistical and intelligent techniques has been done indicate that the latter shows high accuracy over the statistical techniques. This review also shows that most studies have targeted the manufacturing industry for the development of default prediction models. Some studies have also developed industry-specific models. Overall, the findings indicate that statistical technique, namely, logistic regression is widely used for the prediction purpose. However, logistic regression overcomes the limitations of DA and is one of the prominent statistical techniques.

Conclusion

The current review has summarized the bankruptcy prediction work done in the last nine decades based on statistical techniques including Discriminant Analysis, Logistic regression and Probit Regression. The main focus of the review is on statistical techniques employed besides some other dimensions such as source of data collection, country of origin, number of variables employed have been covered. The comparison of performance between these techniques has been made in terms of prediction accuracy. Analysis of models from the table suggests that LR is the most employed technique by researchers. However, hybrid models in integration with other techniques can further increase the predictive accuracy. Therefore, for future directions, it is suggested that researchers should investigate other combinations of statistical and intelligent techniques for developing new models. Since some industries have more failure rate than others, therefore, industry-specific models in a particular country can be developed.

References

Agrawal, K., & Maheshwari, Y. (2019). Efficacy of industry factors for corporate default prediction. *IIMB Management Review*, *31*(1), 71-77.

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The journal of finance*, *23*(4), 589-609.

Bandyopadhyay, A. (2006). Predicting probability of default of Indian corporate bonds: logistic and Z-score model approaches. *The Journal of Risk Finance*.

Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of accounting research*, 71-111.

Bhunia, A., & Sarkar, R. (2011). A study of financial distress based on MDA. *Journal of Management Research*, *3*(2), 1-11.

Chen, M. Y. (2011). Bankruptcy prediction in firms with statistical and intelligent techniques and a comparison of evolutionary computation approaches. *Computers & Mathematics with Applications*, *62*(12), 4514-4524.

Deakin, E. B. (1972). A discriminant analysis of predictors of business failure. *Journal of accounting research*, 167-179.

Doumpos M.; Zopoudinis C. (1999). A multicriteria discrimination method for the prediction of financial distress: the case of Greece. Multinational Finance Journal 3(2): 71–101

Du Jardin, P. (2010). Predicting bankruptcy using neural networks and other classification methods: The influence of variable selection techniques on model accuracy. *Neurocomputing*, *73*(10-12), 2047-2060.

FitzPatrick, P. 1932. A comparison of ratios of successful industrial enterprises with those of failed companies. *The Certified Public Accountant* (October, November, December): 598-605, 656-662, and 727-731, respectively.

Grice JS, Dugan M (2001) The limitations of bankruptcy prediction models: Some cautions for the researcher. Rev Quant Finan Acc 17:151–165.

Gupta, P. K., & Jain, K. K. (2022). Credit default prediction for micro-enterprise financing in India using ensemble models. *Global Business and Economics Review*, *26*(1), 84-98.

Jo, H., Han, I., & Lee, H. (1997). Bankruptcy prediction using case-based reasoning, neural networks, and discriminant analysis. *Expert Systems with Applications*, *13*(2), 97-108.

Jones, S., & Hensher, D. A. (2004). Predicting firm financial distress: A mixed logit model. *The Accounting Review*, *79*(4), 1011-1038.

Karels, G. V., & Prakash, A. J. (1987). Multivariate normality and forecasting of business bankruptcy. *Journal of Business Finance & Accounting*, *14*(4), 573-593.

Kim, S. Y. (2011). Prediction of hotel bankruptcy using support vector machine, artificial neural network, logistic regression, and multivariate discriminant analysis. *The Service Industries Journal*, *31*(3), 441-468.

Lee, S., & Choi, W. S. (2013). A multi-industry bankruptcy prediction model using back-propagation neural network and multivariate discriminant analysis. *Expert Systems with Applications*, *40*(8), 2941-2946.

Lennox, C. (1999). Identifying failing companies: a re-evaluation of the logit, probit and DA approaches. *Journal of economics and Business*, *51*(4), 347-364.

Merwin, C. L. (1942). *Financing small corporations in five manufacturing industries, 1926-1936*. National Bureau of Economic Research, New York.

Mihalovic, M. (2016). Performance comparison of multiple discriminant analysis and logit models in bankruptcy prediction. *Economics & Sociology*,+ *9*(4), 101.

Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of accounting research*, 109-131.

Scott, J. (1981). The probability of bankruptcy: A comparison of empirical predictions and theoretical models. *Journal of banking & finance*, *5*(3), 317-344.

Shumway, T. (2001). Forecasting bankruptcy more accurately: A simple hazard model. *The journal of business*, *74*(1), 101-124.

Singh, B. P., & Mishra, A. K. (2016). Re-estimation and comparisons of alternative accounting based bankruptcy prediction models for Indian companies. *Financial Innovation*, *2*(1), 1-28.

Smith, R. F., & Winakor, A. H. (1935). Changes in the financial structure of unsuccessful industrial corporations. *University of Illinois.*

Wu, D. D., Liang, L., & Yang, Z. (2008). Analyzing the financial distress of Chinese public companies using probabilistic neural networks and multivariate discriminate analysis. *Socio-Economic Planning Sciences*, *42*(3), 206-220.

Zmijewski, M. E. (1984). Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting research*, 59-82.