HEB

Predicting the Probability of Corporate Default using Logistic Regression

CASS

*Deepika Verma

*Research Scholar, School of Management Science, Department of Commerce, IGNOU

Address for Correspondence: serviceheb@gmail.com

ABSTRACT

This paper studies the probability of default of Indian Listed companies. Study aims to develop a model by using Logistic regression to predict the probability of default of corporate. Study has undertaken 90 Indian Companies listed in BSE; financial data is collected for the financial year 2010-2014. Financial ratios have been considered as the independent predictors of the defaulting companies. The empirical result demonstrates that the model is best fit and well capable to predict the default on loan with 92% of accuracy.

Access this Article Online	
http://heb-nic.in/cass-studies	Quick Response Code:
Received on 25/03/2019 Accepted on 11/04/2019@HEB All rights reserved	

INTRODUCTION

Credit risk is eternal for the banking and financial institutions, there is always an uncertainty in receiving the repayments of the loans granted to the corporate (Sirirattanaphonkun&Pattarathammas, 2012). Many models have been developed for predicting the corporate failures in advance. To run the economy on the right pace the banking system of that economy must be strong enough to gauge the failure of the corporate in the repayment of loan. Financial crisis has been witnessed across the world from Latin America, to Asia, from Nordic Countries to East and Central European countries (CIMPOERU, 2015). In this paper the Logistic regression shall be applied on the financial data of the listed Indian Companies. Sample of companies consists of both defaulting and non-defaulting. Study shall develop a model to predict the default probability of the corporate.

Prediction of corporate failures has been considered by many research scholars like Beaver in 1966, Altman 1968, Ohlson 1980 etc. Traditionally only Accounting based information was included in the forecasting of the defaults but later on it has been criticized by many authors and statistical methods were combined along with the financial ratios (Agrawal, 2015). The (Altman E. I., Financial Ratios,Discriminant Analysis and the Prediction of Corporate Bankruptcy, 1968) has one of the brilliant papers which has used Financial ratios along with the MDA to predict the bankruptcy of the corporate.

The current study also considered the financial ratios as the independent predictors of the models that shall contribute the most to predict the distress of the firms. A predictive model has been developed by using Logistic Regression Model. Since objective of the study is to find out the default and non default which comes as bounded between 0 and 1 therefore only Logistic regression can resolve this problem (Ohlson, 1980).

Literature Review

Extensive studies have been performed by various renowned authors. Of which Some of the previous studies pertaining to this paper have been covered. (Ohlson, 1980) Studied the empirical results for predicting corporate failure for the period 1970-76 and considered 105 bankrupt and 2058 non bankrupt firms. Study had identified the factors those affects the probability of failure which are size of company, measure of financial structure, measure of performance, measure of current liquidity. (Westgaard&Wijst, 2001) Predicted the probability of default for retail banking and used Logistic regression. Financial ratios covered were categorized into the 4 aspects of accounting like liquidity, solidity, profitability and leverage. Since the purpose of the study was bounded to get default and non default which comes between

0 and 1 therefore only logit can be used. Study had undertaken the sectors such as agriculture, construction, hotel, industry, real estate, retail, transportation and different geographical areas also considered in the model. (Bewick, Cheek, & Ball, Statistical review 14: Logistic Regression, 2005)

Reviewed the statistical implications of the logistic regression and considered the following tests like HosmerLemeshow test to examine the goodness of fit of model, R squared test to verify the usefulness of the model, Wald test used to check the impact of individual coefficient in the model.

(Sarlija, Bensic, &Susac, 2006) Assessed the significance of different variables to predict the default of amongst the customers using logistic regression, survival analysis, and neural network credit scoring models. Variable considered for the models were segregated into Demographic data, socioeconomic data and Behaviour data. Sample considered were 44087 cases and for validation purpose 9208 samples were taken. (Bandyopadhyay A. , 2007) Developed a hybrid logistic model, examining the capacity of market value of assets, asset volatility, and firm's leverage structure to predict probability of default in future. And also explored the ability of accounting ratios to predict. 150 NSE listed Indian Companies, for the period 1998-2005 considered as sample. Financial data of companies obtained from CMIE prowess database; Panel was used. Model incorporated information from the structural model as well as profitability of firms, liquidity risk, other firm specific supplementary information and macroeconomic factors to predict real world corporate distress potential through a multivariate analysis. Result depicted that default may be occurred by both low asset value and liquidity shortage. (Kim &Gu, 2010) Developed a Logit model by applying forward stepwise selection procedure to predict bankruptcy in the hospitality industry for up to 2 years in advance. Study recognized that operating cash flow to total liabilities distinguish the bankrupt set from the non bank set. (Sirirattanaphonkun & Pattarathammas, 2012) used MDA and Logit model to develop the predicting model therefore study explored the financial ratios to apply them in Thai loan applications. Study also compared them; Buggakupta and kiatkajornvong models also evaluated. Result revealed that Logit regression provided more accurate result of 85.5% than other two models. (Bartual, Garcia, Guijarro, & Romero-Civera, 2012) Data consisted of Spanish firms like agriculture, stock farming, forestry, manufacturing, food processing, and soft drinks obtained from SABI data base for the financial year 2007 consist of 622 companies analyzed out of which 49 companies were defined as insolvent. Principal component analysis and Kaiser criterion were used. Conclusion conveyed that study required paying special attention while selecting the database. (Soureshjani&Kimiagari, 2013) Tests performed on Logit models were: evaluation of model, check the statistical significance of individual variables, check the goodness of fit of the model and validation of predicted probabilities. Study found that logistic regression can estimate the probability of defaulting. The best cutoff point in both logistic regression and neural network is calculated that have minimum errors and capable to categorize the companies into defaulting and non-defaulting.

(Suleiman, Issa, Suleman, Usman, & O, 2014) Study aimed to improve the predictive power of Linear Discriminant and Logistic regression model using Principal components and Linear Discriminant Analysis. keisermeyerolkin's and bar lett's test of sampling adequacy and box m test for equality of covariance matrices were applied. Result indicated that principal component as input enhanced the linear discriminant and logistic regression models prediction. (Sharma, Sing, &Upadhyay, 2014) Paper studied the default probabilities of 47 Indian firms for the period 2007 to 2013. option based method is used to predict the probability of default. Study estimates that the market value of assets, asset volatility, risk neutral default probability and real default probability of firms and recognized the factors that have impact on the default probabilities. The predictive efficiency of the model was 80%. (Kliestik, Kocisova, &Misankova, 2015) According to the study Logit and Probit models are similar to each other. But distribution of Logit model has flatter tails and probit model has steeper slope. Probit model is the alternative of logit but it assumes normal distribution of random variables whereas the Logit model contains extreme values. (Memic, Assessing credit default using logistic regression and multiple discriminant analysis: empirical evidence from bosnia and herzegovina, 2015) Study examined the accuracy of default prediction of both Logit and MDA. Financial data for four year before the default was collected from AFIP data base of bosnia and Herzegovina. Result discovered that the created models have high predictive abilities. Return on asset is the statistically significant; having highest impact on the model's ability to predict default. (Kwofie, Ansah, &Boadi, Predicting the probability of loan - Default : An Application of Binary Logistic Regression , 2015) Age, marital status, gender number of years of education, no. of years in business and base capital were used as predictors in performing binary logistic regression. Predictors that impacted the most in the model were marital status, no. of years in business and base capital. (Raei, Kousha, Fallahpour, &Fadaeinejad, 2016) A hybrid model is developed by consolidatingLogit and neural network model to get the advantages of both linear and non linear models. Step wise and Swap wise least square method were used. Financial data was collected from the companies listed in tehran stock exchange for the period 2008-2014. Sample of 175 companies were considered of which 125 samples were used for model estimation and 50 were considered for model testing. Result displayed that the projected hybrid model better than the logit model and neural network. (Kamau, Muthoni, &Odhiambo, 2018) Multiple Logistic Regression model has been developed, data set taken from the Kenya’s higher education loans board (helb). Sample of 5100 clients considered and these independent variable were considered such as loan amount, overdue days, age, interest rate, employer, gender, marital, father's education level etc. result demonstrate that logistic and log logistic model performed well under concentrated outliers and have minimal type II error. (Agrawal, 2015) Study used logistic regression, among the individual predictors in the model, change in leverage is found to be statistically significant in predicting default and increases in leverage associated with higher probability of the defaults, the Piotroski's F-score is found to be statistically significant in predicting defaults. (Altman&Sabato, Modeling Credit Risk for SMEs: Evidence from the US Market, 2007) authors developed a distress prediction model using Logit for SMEs and compare it to the general corporate model. Financial ratios have been used that classify into liquidity, profitability, leverage, coverage, activity. Logarithimic transformation is used for all five selected ratios in order to reduce the range of possible values. MDA is also been used on the data to discriminate. Result discovered that prediction power is 30% more than the generic model. (CIMPOERU, 2015) Logistic model applied on two panel data groups like advanced and emerging economies to develop a new early warning system for identifying systematic banking risk to recognize the macroeconomic indicators which can predict the financial trouble in the economic environment. Two separate models have been proposed for both the economies. Macroeconomic variables covered were cash deficit, GDP growth, exports, stocks, inflation, output gap and debt for advanced economy, for emerging the M2 growth, GDP growth, stocks, inflation. Outcome defined that the GDP growth or debt level as main causes for financial crises.

SIGNIFICANCE OF STUDY

Public sector banks have lent about 37% of their total credit to the industry sector, the corporate and industry loan which accounted for the 73% of the total NPAs of the banking sector in 2016-17 (https://indianexpress.com, 2018). Study has incorporated the data of many companies which are listed in the Bombay Stock Exchange; the sectors covered were Electronics, Airlines, Energy, Cement, Steel, Manufacturing etc. The study shall contribute both academically and professionally particularly to the banking sectors and the corporate which shall be able to predict the default event in advance and can take corrective action.

OBJECTIVE OF STUDY

- To predict the probability of default of the Indian companies listed in the BSE.
- To assess the efficiency of the Logistic model in predicting the default probability.
- To examine the statistical significance of the proposed model developed in this paper.

DATA

A panel data has been used from the sample of 90 BSE Listed companies. Total 444 cases considered of which 417 cases selected after removing the extreme values in the data. Data covered for the financial year 2010 to 2014. A model is developed using SPSS version 23.Data extracted from the financial statements of the companies which is available on the moneycontrol.com & the data for share prices obtained from the BSE Website.

The Independent variables considered in the model are:

- WC/TA: Working capital to total asset .
- RE/TA: Retained earnings to total asset.
- EBIT/TA: Earnings before interest and tax to total asset.
- MVE/TBD: Market value of equity to book value of total debts.
- SALES/TOTAL ASSETS: Turnover of the company in terms of total assets.
- NI/TA: Net income to total assets.
- NP/TE: Net profit to total equity.
- TBD/TA: Total value of debts to total assets.
- EBIT/INT: Earnings before interest and tax to interest expenses.

- OCFR: Operating cash flow ratio.
- GRTA: growth rate of total assets.

METHODS

Logistic Regression Model

The Logistic Regression is the statistical method applied when the dependent variable is binary, which can take only two values like in this study the event shall occur if the value p is 0 to 1 (Kliestik, Kocisova, &Misankova, 2015). The Enter Selection method of logistic regression is used in this study. According to the Logistic Regression Model the probability of default can measured by the below equation.

$$p = \frac{e \sum \sum_{k=1}^{K} \beta_k x_k}{1 + e \sum_{k=1}^{K} \beta_k x_k}$$

Where

p = probability of the event occurring

K = Independent variables

x_k = each weighted by a coefficient β

(Frade, 2008)

In this paper the following equations have been prepared:

$$L\,(odds) = \beta 0 + \beta 1X1 + \beta 2X2 + \cdots \ldots \ldots . . + \beta KXK$$

$$\text{p} = \frac{1}{1+\text{EXP}(-\text{L})}$$

EMPIRICAL RESULTS AND ANALYSIS

Table 1

Omnibus Tests of Model Coefficients				
		Chisquare	Df	Sig.
Step 1	Step	247.421	14	0
	Block	247.421	14	0
	Model	247.421	14	0

Source: Author (SPSS output)

Table 1 consists of Omnibus Tests Model Coefficients; Chi Square of the omnibus test signifies overall fitness of the model in predicting the default (Kim &Gu, 2010). This is a complete model in which all the financial ratios have been incorporated as the independent predictors of the model. This statistics tests whether the model is capable to predict the default; in this case as the P value ie. Significant value of the Chi Square < .05 which indicates the model is best fit to required conditions for which it is applied. As the value of the Chi square is significant which says that this model is an improvement of the earlier model in which no independent predictors have been considered.

Table 2

Model Summary			
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	148.452	0.479	0.74

Source: Author (SPSS output)

In table 2 there are 3 statistical tests which explain the result in the following manner:

-2 log likelihood: it measure that how fit the model is in explaining variation in the result by adding the variables. To test that the study shall check the P value of the chi squared distribution; -2 Log Likelihood has a chi square distribution, which is appears as 0 in table 1 that indicate that model is statistically significant.

Negelker R square is scaled from 0 to 1.0 is termed as the pseudo r square statistics for the logit model. As the pseudo R square has same attributes like the r square which explain the variation in the dependent variable by the independent variable. Table 2 displays .740 in the column of Nagelkerke R square statistics which conveys that the variation in the probability of the default is 74% explained by the predictors.

Cox &snell R squareis as maximum value of .75,Cox and Snell R square is similar to the R square based on "likelihood" but its maximum value usually is less than 1 (Kwofie, Ansah, &Baodi, 2015); here the value of Cox and Snell R square is .479 i.e. .48 which indicates that 48% of variation in the dependent variable is explained by logistic model.

Table 3

Hosmer and Lemeshow Test			
Step	Chisquare	Df	Sig.

1	4.986	8	0.759

Source: Author (SPSS output)

The strength of the model is evaluated using Hosmer and Lemeshow goodness of fit test. The test statistics is a chi square statistics with a desirable outcome of non significance which indicates that there is no difference between the observed and the predicted values (Kamau, Muthoni, &Odhiambo, 2018). In the above table the P value of the chi square is .759 which is greater than .05 that conveys that the developed model is best fit to predict the defaults.

Table 4

Contingency Table for Hosmer and Lemeshow Test					
	Z = non default		Z = default		Total
Categories	Observed	Expected	Observed	Expected	
1	38	38	0	0	38
2	38	38	0	0	38
3	38	38	0	0	38
4	38	37.998	0	0.002	38
5	37	37.785	1	0.215	38
6	37	36.649	1	1.351	38
7	35	32.359	3	5.641	38
8	25	24.935	13	13.065	38
9	10	11.99	28	26.01	38
10	1	1.284	36	35.716	x37

Source: Author (SPSS output)

In this table the predicted probabilities have been segregate into ten categories; a comparison of expected and observed probabilities is given. If the difference between the expected and observed values is larger, the predictive capacity of the model shall be lesser. Here the study compares the observed and expected value of the tenth category that is 36 and 35.716 respectively, the resulted difference is .29 which is very nominal therefore, It can be said here that the model is good fit to predict the default.

Table 5

Classification Table							
Null Model				Developed Model			
	Non Default	Default	Total		non default	Default	Percentage Correct
Non Default	297	0	100	non default	286	11	96.3
Default	82	0	0	Default	17	65	79.3
overall percentage			78.4	Overall Percentage			92.6

Source: Author(SPSS output)

Table 4 demonstrates two models one is Null model with no independent predictors just the intercept and second is the developed model using Enter selection method with all independent predictors/variables.

In Null Model there are 297 non default and 82 default companies this model assuming that there is no default and the resulted accuracy level is 78.4%. on the other hand in developed model there are 286 companies are predicted as non default and 65 companies are predicted to default on repayment of loan amount. In the first model there were 82 cases who defaulted whereas in developed model there were only 17 cases defaulters, which reveals that the earlier model failed to predict 65 defaulted cases. The corrected prediction has accuracy of 79.3% given in the table which indicates 79.3% cases defaulted on loan did not predicted by the null model.

And the overall accuracy level of prediction has jumped to 92.6% from 78.4% that means the model is improved by 14% by incorporating the independent variables.

Table 6

Variables in the Equation						
Financial Ratios	B	S.E.	Wald	Df	Sig.	Exp(B)
WCTA	0.159	1.169	0.019	1	0.891	1.173
RETA	-4.42	1.308	11.422	1	0.001	0.012
EBITTA	-1.427	5.425	0.069	1	0.792	0.24
MVETBD	-0.817	0.317	6.633	1	0.01	0.442

SalesTA	-1.558	0.529	8.676	1	0.003	0.211
CACL	0.164	0.146	1.261	1	0.261	1.178
NITA	-0.01	0.055	0.03	1	0.861	0.991
NPTE	-0.046	0.056	0.678	1	0.41	0.955
TBDTA	0.235	1.02	0.053	1	0.818	1.264
EBITInt	-0.517	0.223	5.371	1	0.02	0.596
OCFR	0.516	0.408	1.601	1	0.206	1.676
GRTA	-1.151	1.195	0.929	1	0.335	0.316
InventoryTurnover	0.008	0.005	2.489	1	0.115	1.008
FixedAssetsTurnover	0.076	0.036	4.339	1	0.037	1.079
Constant	1.838	0.949	3.751	1	0.053	6.285

Source: Author(SPSS output)

L= (1.838-4.42*RE/TA-0.817*MVE/TBD-1.558*SALES/TA-0.517*EBIT/INT+0.076*FIXES ASSETS TURNOVER)

Table 5 demonstrates the unstandardized beta weights and constant. The logistic regression equation is prepared using the constant and the weights of significant financial ratios. The value of RETA has -4.42 which is least amongst other variables it defines that least value of the RETA is associated with greater probabilities of default. Therefore lower value of "B" is associated with a higher probability of defaulting.

CONCLUSION

A logistic regression analysis is performed to build a predictive model, 90 BSE listed companies were considered to predict the default. Paper 4 sections: Introduction, Literature review, Method and Empirical Analysis of result. Literature review witnessed the applications of financial ratios as an independent predictor in the logistic regression in the previous studies. Likewise this study prepared the logistic regression model using significant ratios that were RETA, MVE/TBD, SALES/TA, EBIT/INT, FIXED ASSETS TURNOVER ratios. Logistic regression is performed on SPSS program. And the output of the various statistical tests revealed that the model is best fit to predict the default and the accuracy level it achieved is 92.6% which is pretty good in comparison to (Sharma, Singh, &Upadhyay, 2014) in which the author used option based method and the accuracy level of the developed model was 80%.

REFERENCES

Agrawal, K. (2015). Default Prediction Using Piotroski's F-score. Global Business Review , 16 (5s), 175S-186S.

Altman, E. I. (1968). Financial Ratios, Discriminant Analysis and the prediction of corporate bankruptcy. The journal of Finance , 23 (4), 589-609.

Altman, E. I., &Sabato, G. (2007). Modeling Credit Risk for SMEs: Evidence from the US Market. A Journal of Accounting, Finance and business studies , 43 (3), 332-357.

Baixauli, J. S., Alverez, S., &Modica, A. (2012). combining structural models and accounting based models for measuring credit risk in real estate companies. International Journal of Managerial Finance, 8 (1), 77-95.

Bandyopadhyay, A. (2007). Mapping Corporate drift towards default : Part 2 : a hybrid credit scoring model. The journal of Risk Finance , 8 (1), 46-55.

Bartual, C., Garcia, F., Guijarro, F., & Romero-Civera, A. (2012). Probability of Default using the logit model: the impact of explanatory variables and data base selection. international Scientific conference" Whither our economies", (pp. 118-124). Spain.

Beaver, W. H. (1966). Financial Ratios as Predictors of Failure. Journal of Accounting Research , 4, 71-111.

Bewick, V., Cheek, L., & Ball, J. (2005). Statistical review 14: Logistic Regression. Critical Care , 9 (1), 112-118.

CIMPOERU, S. (2015). A logistic model on panel data for systemic risk assessment - evidence from advanced and developing economies. Journal of Applied Quantitative Method , 10 (1), 15-29.

Frade, J. (2008, December). Credit Risk Modeling: Default Probabilities.

http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.465.8402&rep=rep1&type=pdf .

https://indianexpress.com. (2018, march 31). Retrieved January 28, 2019, from

Kamau, P. N., Muthoni, L., &Odhiambo, C. (2018). Modelling factors affecting probability of loan default: A quantitative Analysis of the Kenyan Student's Loan. International Journal of Statistical Distribution and Applications , 4 (1), 29-37.

Khemais, Z., Nesrine, D., & Mohamed, M. (2016). Credit scoring and Default Risk Prediction : A comparative Study between Disciminant Analysis & Logistic Regression. International Journal of Economics and Finance , 8 (4), 39-53.

Kim, N. (2018, February 5). bloombergquint. Retrieved April 15, 2018, from bloombergquint: www.bloombergquint.com

Kliestik, T., Kocisova, K., &Misankova, M. (2015). Logit and Probit Model used for prediction of financial health of company. 2nd Global Conference on Business, Economics, Management and Tourism (pp. 850-855). Prague, Czech Republic: Elsevier.

Kwofie, C., Ansah, C. O., &Baodi, C. (2015). Predicting the Probability of Loan - Default : An Application of Binary Logistic Regression. Research journal of Mathematics and Statics , 7 (4), 46-52.

Kwofie, C., Ansah, C. O., &Boadi, C. (2015). Predicting the probability of loan - Default : An Application of Binary Logistic Regression . Research journal of Mathematics and statitics , 7 (4), 46-52.

Memic, D. (2015). Assessing Credit Default Using Logistic Regression and Multiple Discriminant Analysis: Empirical Evidence From Bosnia And Herzegovina. Interdisciplinary Description of Complex System , 13 (1), 128-153.

Ohlson, J. A. (1980). Financial Ratios and the probabilistic prediction of bankruptcy. Journal of Accounting Research (Spring, 1980) , 18 (1), 109-131.

Sarlija, N., Bensic, M., &Susac, M. Z. (2006). Modeling customer revolving credit scoring using logistic regression, survival analysis and neural networks. WEAS international conference on neural network, 7, pp. 164-169. cavtat, croatia.

Sharma, C. S., Singh, R. K., &Upadhyay, R. K. (2014). Predicting Probability of Default. Primax International Journal of Commerce and Management Research , II (3), 5-13.

Sirirattanaphonkun, W., &Pattarathammas, S. (2012). Default Prediction for Small Medium Enterprise in Emerging Market: Evidence from Thailand. Seoul Journal ofBusiness , 18 (2), 26-54.

Smaranda, C. (2014). Scoring functions and bankruptcy prediction models - case study for Romanian companies. Procedia Economics and Finance (Elsevier) , 10, 217-226.

Soureshjani, M. H., &Kimiagari, A. M. (2013). Calcuating the best cut off point using logistic regression and neural network on credit scoring problem- A case study of a commercial bank. African journal of business management, 7 (16), 1414-1421.

Suleiman, S., Issa, Suleman, Usman, U., & O, S. Y. (2014). Predicting an Applicant status using principal component, discriminant and logistic regression analysis. International Journal of Mathematics and Statitics invention, 2 (10), 05-15.

Suo, W., & Wang, W. (2005, July). Assessing Default Probabilities from structural credit risk models. Queen University.

Westgaard, S., &Wijst, N. V. (2001). Default probabilities in a corporate bank portfolio: A logistic model approach. European Journal of Operational Research , 135, 338-349.