Pakistan Journal of Engineering and Technology, PakJET
ISSN (p): 2664-2042, ISSN (e): 2664-2050
Volume: 6, Number: 2, Pages: 9-12, Year: 2023

Choosing Banks' Profitability Predictors: A Comparison of Logistic Regression and Discriminant Analysis

Saadut Ullah Khan[1] and Zahid Iqbal[2]

[1]Management Sciences Department, Government College of Management Sciences, Kohat, Pakistan
[2]Department of Statistics, Allama Iqbal Open University, Islamabad, Pakistan

Corresponding author: Saadut Ullah Khan (email: saadut.ullah@gmail.com)

Received: 2023-05-13, Revised: 2023-06-06, Accepted: 2023-06-07

Abstract- - **For researchers, banks' profitability has been a key factor. Various internal and external factors may have an impact on the profitability of banks. Internal factors are unique to each bank and have an impact on management decisions that affect bank profitability. External factors are those that are outside the banks' control that can affect how well they perform. This study examines significant financial variables identifying profitability perspectives and two predictive models of banks' profitability using the banks' data from the period of 2007-2016. Logistic regression and discriminant analyses were used to evaluate their predictive accuracy. Although, logistic regression and discriminant analysis are unusual for such studies as the dependent variable is continuous and these techniques are applicable when the dependent variables are categorical. So, in order to convert the dependent variable i.e., return on assets to a binary variable, '1' is assigned for a positive return on assets and '0' for a negative return on assets. The empirical results described that there is a statistical difference between the prediction results of logistic regression and discriminant analysis. The discriminant models identified that non-interest income, interest rate, operating efficiency, and bank size have a significant effect on banks' profitability. On the other hand, logistic regression showed that there are two significant factors i.e., interest ratio and bank size. In the comparison of the models, the accuracy rate of discriminant analysis is higher than that of logistic regression.**

Index Terms-- **Discriminant analysis, logistic regression, profitability, return on assets**

I. INTRODUCTION

Banks play an important role in the financial sector and are considered the major component of a country's economic growth. If the banks are not involved actively in financial activities, commerce and industries will not be accelerated. The main role of the financial sector is to finance the industrial and economic sectors which are important for their development. So, if there is an established and well-organized banking system, it will allocate funds efficiently and will become the basis for improving the performance and functions of businesses. To achieve this, a strong and developed banking system is required to uphold an environment that is friendly to investors and to facilitate business transactions [1]. Therefore, it is necessary to identify the determinants of banking sector profitability. In this paper, we attempt to categorize the major factors which have had an effect on the profitability of Pakistani banks. There are two objectives of the paper. One objective is to determine the factors for the classification of profitable banks in Pakistan to provide instruction for potential development. The second objective is to compare logistic regression and discriminant analysis. The financial indicators are return on assets (ROA), Size, Non-Interest Income, Liquidity Ratio, Interest Ratio, Non-Performing Loan, Operating Efficiency, Asset Composition, Capital Adequacy, Real Interest Rate, Money Supply, and Exchange Rate. These financial indicators were obtained from the financial statements of the State Bank of Pakistan over ten years. The rest of the paper is ordered as: the second part contains a literature review, the third part explains the methodology, the fourth part relates to the analysis of the data, and results, and the last part provides conclusions.

II. LITERATURE REVIEW

The impact of liquidity risk management on commercial banks' performance in Pakistan was assessed using panel data analysis. Data from the period 2006 to 2019 was taken from the State Bank of Pakistan website. After analyzing the data, it was concluded that there is a direct and important effect of higher liquidity on the banks' profitability. This study can be further enhanced by including more internal variables of the banking sector in Pakistan [2].

Tobit regression was used which indicated that size has a positive effect on the profitability of banks and inflation has a negative relationship with banks' profitability [3]. Applying generalized least square panel data analysis, it was examined that the size of banks has a significant and positive effect on the Malaysian banks' profitability and the liquidity has no impact on banks' profitability [4]. The size and liquidity of the bank have a significant and positive relationship with banks' profitability using the dummy variable model [5]. Using panel data analysis, it was found that one of the vital factors of bank profitability is liquidity [6], [7]. Results of the panel data set examined that size, real interest rate and non-interest income have a significant and positive impact on ROA [8]. By applying panel data, it was observed that non-performing loans have a significant and negative impact on banks' profitability [9]. There is a positive relationship between capital adequacy and banks' profitability by applying regression analysis [10]. The logistic regression and discriminant analysis were evaluated to determine the bankruptcy prediction of the financial firms in Pakistan from 2009 to 2015 and it was concluded that the rate of classification of the logit model was higher than the discriminant model [11]. Banks' performance in Pakistan was investigated from 2010 to 2020 using multiple regression analysis, and it was found that there is some negative and some positive relationship between the variables. Non-performing loans have negatively associated with banks' performance and lending interest rate has a positive impact on banks' profitability [12].

From the above literature, it can be observed that logistic regression and discriminant analysis were not used to assess the determinants of banks' profitability in Pakistan. So, in this study, both of these techniques are used for the first time to assess the banks' profitability and evaluate the performance of these techniques in the banks' profitability area.

III. DATA AND METHODOLOGY

This part of the paper consists of data, variable selection, and methodology employed.

A. DATA

In this study, the secondary data were obtained from the state bank of Pakistan and the economic survey of Pakistan for the period 2007 to 2016. There are total 62 banks which have negative return on assets during 2007-2016, so 62 more banks which have positive return on assets with the matching period of failure banks from 2007 to 2016 were added to the sample. The total sample consists of 122 banks. 30% of 122 banks' data were set aside for testing purpose to check the accuracy of the models.

B. VARIABLES

This study includes the following accounting ratios as variables

1. Return on Asset (ROA): ROA measures how much a bank earns from each unit of asset investment
2. Non-interest income: Income earned by banks other than mark-up
3. Interest Ratio: Interest ratio is the ratio of interest paid and interest received
4. Liquidity Ratio: It is the percentage of assets available in the form of liquid asset
5. Credit Risk: Ratio of non-performing loan to gross advances
6. Size: It indicates the total assets of the bank
7. Operating Efficiency: It measures how management efficiently plans to decrease the expenses
8. Capital Adequacy: Ratio of total assets financed by equity
9. Asset Composition: It indicates how much assets were financed by debts
10. Real Interest Rate: Real Interest Rate = Nominal Interest Rate - Inflation (Expected or Actual)
11. Exchange Rate: An exchange rate is a rate at which one currency will be exchanged for another. It is also regarded as the value of one country's currency with another currency
12. Money Supply: The money supply is the entire stock of currency and other liquid instruments circulating in a country's economy at a particular time.

C. DATA ANALYSIS MODELS

Profitability prediction has remained a warm topic over the last few decades. In this area, wide work has been completed and numerous models are proposed by different studies of profitability. This study models profitability by using both logistic and discriminant analysis to determine significant factors affecting profitability and to discover the model with a high rate of classification.

Discriminant Model:

Discriminant model was developed for the identification of important factors of banks' profitability [13]. The discriminant model that is developed, takes the following structure:

$$D = \beta 0 + \beta 1X1 + \beta 2X2 + \beta 3X3 + ... + \beta kXk \quad (1)$$

Where D is in general Z score and βi are coefficients of discriminant and X_1, X_2, X_3….X_k are the predictor variables. We select the best factors by looking at Wilks lambda and F statistics to see the statistical significance of every individual.

Logistic Model:

Logistic regression is a statistical technique that is used to examine the relationship between a categorical binary response variable and one or more explanatory variables. Logistic regression cannot be used to apply when the dependent variable is continuous, in such case linear regression analysis is applied.

Let p represent the probability of an event depends on k independent variables. Then multiple logistic regression equation is given as:

$$\log it(p) = \log(p/(1-p)) = \beta 0 + \beta 1X1 + ... + \beta kXk \quad (2)$$

$\beta 0 + \beta 1\ X1 + \cdots + \beta k\ Xk$ are coefficients of the linear regression line. Taking antilog of both sides of (2), the obtained (3) is used to predict the probability of occurrence of the interest outcomes as:

$$p = \exp^{(\beta 0 + \beta 1X1 + ... + \beta kXk)} / (1 + \exp^{(\beta 0 + \beta 1X1 + ... + \beta kXk)}) \quad (3)$$

The self-descriptive block diagram of logistic regression is as under:

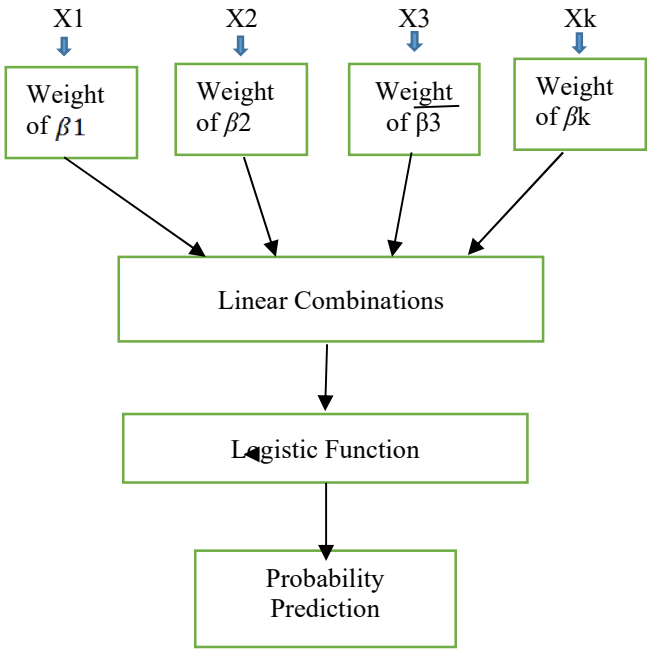

FIGURE 1. Diagram of the logistic regression.

Mathematical Estimation of Parameters and the likelihood function for logistic regression is:

$$l(\beta 1, \beta 2, \beta 3, \ldots \beta k) = \Pi(\ \theta(xi)^{yi})[1-\theta(xi)]^{(a-yi)} \quad (4)$$

Where θ is the predicted probability of the Y=1 and Π notation is used for the product. To estimate the β values which maximize the likelihood function, mathematically it is simple to take the logarithm of (4).

$$L((\beta 1, \beta 2, \beta 3, \ldots \beta k) = \ln(l(\beta 1, \beta 2, \beta 3, \ldots \beta k)) \\ = \Sigma\{yi * \ln[\theta(xi)] + (1-yi)\ln[1-\theta(xi)]\} \quad (5)$$

Now to find the β values which maximize $L((\beta 1, \beta 2, \beta 3, \ldots \beta k)$, differentiating (5) with respect to $\beta 1, \beta 2 \ldots, \beta k$ and the resulting sets are equal to zero.

IV. EMPIRICAL RESULTS

This section contains the observed estimates of the logit model and discriminant analysis.

Results of discriminant analysis before using the discriminant method, assumptions of normality, homogeneity of covariances and multicollinearity were checked. Normality assumptions were tested by the Shapiro wilk test and it was found that all the variables were non-normal. Some outliers were found using the box plot method, removed, and some log transformation was applied to make the data normal. Homogeneity was tested by Using Box M test and it was found that homogeneity of equal covariances exist. Multicollinearity was tested by Variance Inflation Factor (VIF). The VIF score of two variables, money supply and exchange rate, were more than 5 and these two variables were dropped to remove the multicollinearity problem. Now by applying discriminant analysis, it was found that four predictors out of nine have a significant effect on the banks' profitability. These four predictors are shown in table1 which are size (log of size), Interest Ratio (Interest), Non-interest income (NII), and operating efficiency (log_op).

TABLE 1
VARIABLES IN THE ANALYSIS

Variables	Wilks Lambda	P-value
Log(size)	0.450	0.000
Interest Ratio	0.448	0.000
Non-Interest Income	0.383	0.000
Log (operating efficiency)	0.381	0.000

The significant model will take the following form based on the coefficients shown in Table II.

$$D = -7013 + 0.86NII + 1.455\text{logsize} \\ -2.552\log op - 0.046\,\text{Interest} \quad (6)$$

TABLE II
CANONICAL DISCRIMINANT FUNCTION COEFFICIENTS

Variables	Function
Non-Interest Income	0.86
Log size	1.455
Log Operating efficiency	-2.552
Interest Ratio	-0.046
Constant	-7.13

The banks are considered to be profitable if the bank's Z score is above 0 and is considered to be a failure if the Z score is below 0. From the below table III of the discriminant technique, banks' Z score above 0.796 is to be considered profitable and the banks' Z score below -2.270 is classified as a failure.

Table IV shows the classification results from the testing data, the 2×2 table shows the classification of actual and predicted outcomes. The overall accuracy of classification is 100%. It is obvious that the four main financial ratios have a vital role in classification and have the power to forecast the profitability of banks.

TABLE III
FUNCTION AT GROUP CENTROIDS

ROA	Function
Failure	-2.27
Profitable	0.796

Results of the logit model:

The logit model provides a probabilistic explanation that is why it is very popular in prediction. Table V reveals that the estimated signs are similar to the projected ones. Two predictors were found significant i.e., Interest ratio and log of size. When the sign of interest ratio is negative, banks with a lower interest ratio are more likely to be profitable. If the sign of log of size is positive, it indicates that with increasing the total assets banks are more likely to be profitable.

TABLE IV
CLASSIFICATION RESULT

ROA		Predicted Group Membership		Total
		Non-Profitable	Profitable	
Original	Non-Profitable	32	0	32
	Profitable	0	5	5
Overall Percentage				100

TABLE V
SIGNIFICANT VARIABLES

Variable	B	S.E.	Wald	df	Sig.	Exp(B)
Interest	-0.2	0.09	6.64	1	0.01	0.78
Logsize	6.1	2.06	8.84	1	0.00	456.12
Constant	-33.8	12.5	7.34	1	0.01	0.00

Table VI shows the classification results of the testing data. Out of 5 profitable banks 5 were classified as profitable, and similarly out of 32 non-profitable 29 were classified as non-profitable banks which indicates that the overall accuracy rate of the logit model is 92%.

According to the literature, the logit model has a higher rate of classification of financial firms' profitability and non-profitability than discriminant analysis [11]. This study revealed that the discriminant analysis has more classification rate than the logistic regression and has more potential to be applied to banks' profitability classification.

Besides the above comparison technique, the following three more techniques i.e., precision, specificity and F1 score were applied to evaluate the performance of discriminant and logistic regression. Table VII gives us the results of precision, specificity and F1 score for discriminant and logistic regression. It is clear that the performance of the discriminant model is higher than the logistic regression on the basis of high specificity and F1 score.

TABLE VI
CLASSIFICATION

Observed		Predicted		Percentage Correct
		ROA		
		Non-Profitable	Profitable	
ROA	Non-Profitable	29	3	90
	Profitable	0	5	100
Overall Percentage				92

TABLE VII
PRECISION, SPECIFICITY AND F1 SCORE

Performance Measure	Discriminant Model	Logistic Model
Precision	1.00	1.00
Specificity	1.00	0.63
F1 Score	0.40	0.16

V. CONCLUSION

According to the previous studies, the logit model is better to be used in contrast to discriminant analysis. The financial sector is one of the most important factors in the economic development of Pakistan. Using the discriminant procedure, four financial ratios namely non-interest ratio, bank size, interest ratio, and operating efficiency of the discriminant model were found. The model obtained 100% accuracy rate on the testing data. The estimates represent that a bank having a Z score above 0.796 is to be considered a profitable bank and a bank having a Z score below -2.27 will be considered as non-profitable. The logit model discovered two significant variables i.e., bank size and interest ratio that are more parsimonious than discriminant and obtained an accuracy rate of 92%. These ratios are among the four ratios which were included in the discriminant model. This study presents an evaluation of two classification techniques in the area of banks' profitability and found that discriminant models have higher classification rate, precision, specificity and F1 score than the logistic regression model. This study proposes that stakeholders should continue these significant financial ratios in order to review the financial health of the bank. This study can be further extended. First, to include more macroeconomic variables like gross domestic product (GDP) to assess the impact on banks' profitability. Second, to include banks of other countries.

FUNDING STATEMENT

The authors received no specific funding for this study.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

[1] A. Raza, M. Farhan, & M. Akram, "A comparison of financial performance in investment banking sector in Pakistan". *International Journal of Business and Social Science*), 2011, pp.72-81.

[2] W. Alim, A. Ali, & M. R. Metla, "The Effect of liquidity risk management on financial performance of commercial banks in Pakistan". Journal of Applied Economics and Business, 2021, pp. 109-128.

[3] M. Abbas, T. Azid, & M. H. A. H. Besar, "Efficiency, effectiveness, and performance profile of Islamic and conventional banks in Pakistan". *Humanomics*, *2016,* pp. 2-18.

[4] I. Asma, A. Fadli, & T. Noor, "Determinant of Islamic banking institutions' profitability in Malaysia", *Word Applied Sciences Journal,* 2011*,* pp. 1818-4952.

[5] S. Haron, B. Shanmugam, & H. Doran, "Determinants of Islamic bank profitability". *Global Journal of Finance and Economics,* 1997, working paper series 002.

[6] A. Bashir, "Determinants of profitability and rates of return margins in Islamic banks: Some evidence from the Middle East", *Grambling State University,* Mimeo*,* 2000.

[7] P. Bourke, "Concentration and other determinants of bank profitability in Europe". *Journal of Banking and Finance*, 1989, pp. 65-80.

[8] A. Anbar, & D. Alper, "Bank specific and macroeconomic determinants of commercial bank profitability: Empirical evidence from Turkey".*Business and Economics Research Journal*, *2011,* pp. 139-152.

[9] C. T. Albulescu, "Banks' profitability and financial soundness indicators: A macro-level investigation in emerging countries". *Procedia Economics and Finance*, 2015, pp. 203-209.

[10] B. Capraru, & I. Ihnatov, "Bank's profitability in selected central and eastern European countries". *Procedia Economics and Finance,* 2014, 587-591.

[11] U. E. Khan, "Bankruptcy prediction for the financial sector of Pakistan: Evaluation of logit and discriminant analysis approaches". *Pakistan Journal of Engineering Technology & Sciences,* Pakistan, 2018, pp. 210-220.

[12] S. A. Rasool, & M. A. Khan, "Relationship between Risk Determinants and Banks' Profitability: Empirical Evidence from Pakistan". *Global Economics Review,* 2022*,* pp. 112-123

[13] R. Ante, & K. Ana, "Discriminant analysis of bank profitability levels". *Croatian Operational Research Review,* Croatia, 2013, pp. 139-152.