Credit Scoring Model for Retail Banking Sector in Pakistan

Arif Hussain[*],Muhammad Khan[†],Shams Ur Rehman[‡], Aaiya Khattak[§],

Abstract

Credit risk evaluation is getting importance due to competition and also the Basel requirements. This study is based on the evaluation of credit scoring models for retail bank loan customers of Askari Bank Limited in Pakistan. Credit risk models further brings efficiency in the evaluation of a loan application which ultimately reduces the overall risk of bank and also reduces the probability of default on loans. This study has used three credit scoring models of logistic regression, discriminant analysis and probit analysis for the evaluation of credit applicants. The ranking criteria for ranking among the various models is based on the average correct classification rate. Therefore, the ranking shows that probit analysis techniques has the highest average correct classification rate while Logistic regression technique has the lowest average correct classification rate. These statistical credit scoring approach further augments the judgmental approach of credit application evaluation procedure. Further studies can be carried out on an extended data sample of retail loans. Other important variables can be included in the model to further improve the validity of the credit evaluation procedure according to the regulatory requirements. The statistical procedures of decision trees and neural networks can also be applied in credit scoring procedures to further make it more authentic and reliable.

Keywords: credit scoring, retail banking, discriminant analysis, probit analysis, logistic regression

Introduction

The ‘credit’ is a Latin word and it refers ‘to trust upon’. In a credit transaction the lender has to trust upon the borrower for the credit obligations created there in. Access to finance or credit creates certain obligations like, borrowers pay back with respect to the agreed terms, borrowers should also pay for creating trust and borrowers also pay the risk premium in case of default. The whole process of advancing credit created the phenomenon of credit risk and credit worthiness. The risks involved in lending of money are related to human history since long. The concept of lending and borrowing is related to human history.

[*]Assistant Professor, Abdul Wali Khan University, Mardan. arifhussain@awkum.edu.pk
[†]Lecturer, Abdul Wali Khan University, Mardan. muhammadkhan@awkum.edu.pk
[‡]Assistant Professor, Northern University, Nowshera. shams.mahsud@gmail.com
[§] Lecturer, Northern University Nowshera, asiya@northrn.edu.pk

Various methods of credit scoring are applied to examine worthiness of a credit customer. Information provided by a loan customer is applied to prepare credit score for every customer which are used to identify good loans and bad loans. Credit scores are usually divided in statistical credit scoring and judgmental credit scoring by any lending institution (Caire, 2004). Credit scoring model determines a customer credit worthiness by grouping certain important attributes of a loan applicant.

Various empirical evidence shows that the major loss of any financial institution is due to inefficient and ineffective process of credit risk management which is a major cause of bankruptcy and insolvency with in a financial institution. Therefore, all the credit institutions should gather complete information on all aspects of their credit customers effecting their credit risk and also make regular monitoring of the customers to whom credit has been granted. It is therefore deducted that strategies for credit risk and continuous managerial supervision has an effect on risk and return of credit portfolio in any financial institution n order to remain competitive in terms of increasing profits, cost minimization and to remain solvent. Risk managers in any credit institution should devise such credit instruments which can efficiently assess creditworthiness, should minimize turnaround time for the assessment of credit customer and also to maintain a minimum unit processing cost. Preparing scoring model for emerging markets is a bit complex because of accessibility, availability and interpretation techniques for data are usually scares and makes the whole process of credit risk assessment a challenging activity. Problems usually faced by small banks and banks in emerging markets regarding credit scoring estimations include complexity in data mining applications and tools, lack of databases for internal decisions, shortages of software applications and inexperience staff. Managers in such financial institutions do not show their keenness and motivation to adopt modern software and making databases for credit scoring procedures. They usually refer these efforts as uneconomical and costly. They mostly depend upon traditional and routine judgmental techniques for the assessment of default risk of their credit applicants

Literature Review

Walter (2012) contended that applying credit scores minimizes the cost of loan appraisal process and also minimizes the role of credit officer in this respect. The objective of using statistical scoring methods are to reduce errors, risk based pricing implementation, differentiated credit products development and to improve lending efficiency. The evaluation of accuracy prediction is therefore important in the whole process. The commonly applied scorecards include behavioral and application scorecards. The application scorecards make default prediction on the basis of loan characteristics, financial data, demographics, availability of data from credit bureau and nature and type of business. Behavioral scoring models predict the probability of default on the basis of repayment schedule of an existing customer. Japeli and Pagano (2012) further argued that scoring models are adopted to reduce errors in the loan granting procedure. Countries where credit information about bank borrowers are shared have lower credit risk and higher rate of lending. Luo et.al. (2004) contended that financial institutions adopt subjective scoring for the analysis of credit risk and business risk. In a subjective credit scoring procedure various qualitative indicators are used as decision criterion. Subjective scoring mechanism is time consuming for loan officers and need a lot of training and time. On the other hand, statistical scoring procedure use database to quantify and forecast credit risk on the basis of various characteristics entered in a database.

The first ever credit scoring model was designed in 1958 for FICO company of USA. Montgomer Ward in 1960 introduced for the first time scoring model for credit card operations. The American Bank and Trust started in 1970. Wall (1928) introduced Credit-Men model to estimate a company's position relative to other companies in similar industry. To date several credit scoring models are applied successfully for the assessment and approval of credit decisions. Bovilla et.al. (2003) in Spain made a comparison between non parametric models i.e. neural networks and decision trees and parametric techniques i.e. logit and discriminant analysis. They applied these models for assessment of credit card risk and concluded that the predictive ability of parametric models is more than non-parametric models with greater accuracy.

Billoti and Crook (2009) introduced macroeconomic variables like unemployment rate and interest arte in their credit scoring model using survival analysis. For prediction purposes, survival analysis is more powerful then logistic regression for preparing credit scoring model. Azam et.al. (2013) analyzed the socioeconomic characteristics of credit customers on loan decisions on by using logistic regression and descriptive statistics. The study concluded that only three independent variables i.e. years with existing employer, region and status of residence have impact on customer loan decisions.

Research Methodology

Data Collection and Variables

The credit scoring model in this study is developed on the basis of data provided by Askari Bank Limited. Askari Bank limited formerly known as Askari Commercial Bank was found in 1991 and listed on Pakistan Stock Exchange. It has 392 branches all over Pakistan. Askari Bank Limited is operating one of the biggest consumer loans division among all banks listed in Pakistan Stock Exchange. This credit scoring model was prepared using 800 personal loans of Askari Bank. It includes 600 good loans in all respect (75%) and 200 bad loans (25%). The credit scoring for each credit customer is based on 20 independent variables as given in Table 1. The loan quality which is dependent variable of the study is characterized with two categorical values i.e. paid/good loan as 1 and defaulter/bad loan as 0.

The proposed variables over which the credit scoring model is based upon are given below.

Table: 2Credit Scoring Model

S/No	Description
1	Nature of job
2	Applicant age
3	Marital status
4	Education status
5	Gender
6	Occupation Spouse Education
7	Home ownership
8	Number of dependents
9	Loan purpose
10	Level of income
11	Time spent at existing address
12	Employment status
13	Total period of employment
14	Monthly income
15	Place of living
16	Type of loan
17	Loan from other banks
18	Loan default declaration
19	Home telephone
20	Loan duration

Credit Scoring Models

Logistic Regression

Logistic regression is used when probability of a dichotomous variable (one or zero) depends upon various potential independent variables in the following statistical form:

$$Log\ (p/(1-p)) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \ldots\ldots\ldots\ldots + \beta_n X_n$$

Where P = Probability, B = Respective coefficients, Xi_{-n} = Respective independent variables, Log (p/(1-p)) = log of ratios of probabilities of outcomes

Discriminant Analysis

Discriminant analysis is a classification problem, where two or more groups or clusters or populations are known *a priori* and one or more new observations are classified into one of the known populations based on the measured characteristics. Discriminant analysis needs the data to be normally distributed and independent. Discriminant analysis statistically can be presented as:

$Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \ldots\ldots\ldots\ldots + \beta_n X_n$, Where, β_0 = Constant term, β_i = Respective coefficients, Xi_{-n} = Respective independent variable from i = 1 to n

Probit Analysis

PA is a technique that finds coefficient values, such that this is a probability of a unit value of a binary coefficient. As such Probit means "probability unit". Under a probit model, a linear combination of the independent variables is transformed into its cumulative probability value from a normal distribution. The method requires finding value for the coefficients in this linear combination, such that this cumulative probability equals the actual probability that the binary outcome is one, therefore:

$$Prob\ (y = 1) = \phi\ (\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \ldots\ldots\ldots\ldots + \beta_n X_n)$$

Where *y* is the zero-one binary outcome for a given set of value. ϕ is the value from the cumulative normal distribution function. β_0 is the intercept term, and β_i represents the respective coefficient in the linear combination of explanatory variables, *Xi,* for *i* = 1 to n.

Data Analysis

The three credit scoring techniques including the logistic regression, discriminant analysis and probit analysis were applied for modeling purposes.

Logistic Regression

Using the predicted variables for designing logistic regression credit scoring model, the results are given in table 3. The table shows that with a fifty percent cut off point the correct classification rate in LR model is 81%. There is a significant relationship among the variables as the probability value for the model is lower than 0.01 with 99% confidence level. Also the probability value of the residuals is higher or equal to 0.10, it shows that the model best fits the data with 90% confidence interval. The correct highest classification rate was 88.91% using a 0.60 cut off value.

Table: 3 Classification of results using logistic regression (LR)

Group	Predicted group	
	Good	Bad
Good	558	42
Bad	62	138

Correct classification percentage = 81%, Cut off rate = 50%

Discriminant Analysis

In order to develop set of discriminant functions, the discriminant credit scoring model are applied as it helps to predict the dependent variable. A one discriminating function at 95% confidence interval was significant with a probability value of 0.000. The result for discriminant analysis shows that the correct classification rate with 0.50 cut off point is 85.05%.

Table: 4 Classification of results using discriminant analysis (DA)

Group	Predicted group	
	Good	Bad
Good	510	90
Bad	29	171

Correct classification percentage = 85.05%, Cut off rate = 50%

Probit Analysis

The probit analysis technique was formulated to estimate the relationship between independent variables of credit scoring and the dependent variable of loan quality. As the probability values for the model are less than 0.01 therefore it is concluded that there is a significant relationship between variables at 99% confidence level. Probability values for the residuals are equal or greater than 0.10, it shows that the data best fits the model at 90% confidence level. The probit credit scoring model shows 85.1 correct classification rate for the probit model with 0.50 cut off level.

Table:5 Classification of results using Probit analysis (PA)

Group	Predicted group	
	Good	Bad
Good	523	67
Bad	34	166

Correct classification percentage = 85.1%, Cut off rate = 50%

Result comparison for credit scoring models

It is necessary to compare the results of credit scoring models used in this study. For this purpose, the average classification rate is an important criterion to evaluate the credit scoring models. Table 6 shows the evaluation of credit scoring techniques. The table shows that the probit analysis credit scoring model shows the highest average classification rate. Therefore, the probit analysis best predicts the probability of a good loan application.

Table 6 Result comparison for credit scoring models

Credit Scoring Technique	Credit Scoring Models Result		
	(1-1)	(0-0)	Average
Logistic Regression (LR)	(558/600) 93%	(138/200) 69%	81%
Discriminant Analysis (DA)	(510/600) 85%	(171/200) 85.5%	85.05%
Probit Analysis (PA)	(523/600) 87.2%	(166/200) 83%	85.1%

Conclusion

In today's competitive environment of financial institutions with the focus on commercial banking sector credit risk evaluation is the main stream activity for commercial banks in particular and financial institutions in general. Credit risk evaluation is getting importance due to competition and also the Basel requirements. This study is based on the evaluation of credit scoring models for retail bank loan customers of Askari Bank Limited in Pakistan. Credit risk models further brings efficiency in the evaluation of a loan application which ultimately reduces the overall risk of bank and also reduces the probability of default on loans. This study has used three credit scoring models of logistic regression, discriminant analysis and probit analysis for the evaluation of credit applicants. The ranking criteria for ranking among the various models is based on the average correct classification rate. Therefore, the ranking shows that Probit analysis techniques has the highest average correct classification rate while Logistic regression technique has the lowest average correct classification rate. These statistical credit scoring approach further augments the judgmental approach of credit application evaluation procedure. Further studies can be carried out on an extended data sample of retail loans. Other important variables can be included in the model to further improve the validity of the credit evaluation procedure according to the regulatory requirements. The statistical procedures of decision trees and neural networks can also be applied in credit scoring procedures to further make it more authentic and reliable.

References

Caire, D. (2004). Building credit scorecards for small business lending in developing markets. London: Annock Consulting,

Gool, J. V., Verbeke, W., Sercu, P. and Baesens, B. (2012). Credit scoring for microfinance: Is it worth it? *International Journal of Finance and Economics.17.* 103–123.

Banasik, J., and Crook, J. (2010). Reject inference in survival analysis by augmentation. *Journal of Operational Research Society 61(3)*. 473–458.

Sur, K. (2008). Credit risk analytics: A cornerstone for effective risk management, an Oracle White Paper, Oracle financial services, October 2008.

Walter, O. (2012). The predictive accuracy of credit ratings: Measurement and statistical inference. *International Journal of Forecasting, 28(1),* 288–296

Japelli, T. and Pagano, M. (2000). Information sharing, lending and defaults: Cross-country evidence. *Journal of Economic Literature.*

Luoto, J., McIntosh, C. and Wydick, B. (2004). Credit information systems in less-developed countries: Recent history and a test. *Journal of Economic Literature*.

Schreiner, M. (2003). Scoring: The next breakthrough in micro credit?. Washington. *CGAP. Occasional Paper No.7,* 1-60.

Wall, A. (1928). Ratio method and statement analysis. Nueva York: Harper.

Bonilla, M, Olmeda, I., & Puertas, R. (2003). Modelos paramétricos paramétricos en problemas de Credit Scoring. *Revista Española de Financiación y Contabilidad*, 118(32), 833-869.

Hand, D. J., Jacka, S. D. (1998). Statistics in finance, *Arnold Applications of Statistics: London.*

Lee, T., Chen, I. (2005). A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression. *Expert Systems with Applications* 28 (4): 743-752.

Boyle, M., Crook, J. N., Hamilton, R., Thomas, L. C. (1992). Methods for credit scoring applied to slow payers. In Credit scoring and credit control, Thomas, L. C., Crook, J. N., Edelman, D. B., eds., *Oxford University Press, Oxford,* 75-90.