Consumer Credit Default Predictors and Risk Modeling: A Case Study

Dr. Geeta Sharma

Reader
Devi Ahilya University
Indore

Ms. Aditi Joshi

Student
Devi Ahilya University
Indore

Sarthak Kumar Jena

Assistant Professor
Srusti Academy of Management
Bhubaneswar

Abstract

In recent years, an increase in consumer spending has resulted in a rise in consumer credit in India to $2,408.3 billion in 2010-11. To maintain adequate profit margins in such a highly competitive and risky business environment, lending institutions must take measures to manage credit risks. Hence banks need to develop automated computer-based credit risk models that can assess the risk of credit default within lesser time and cost. The paper attempts to assess the credit risk of retail NBFC borrowers based on certain consumer-specific characteristics, as well as, loan-specific characteristics of borrowers. The objective of this paper is to find out which are the most predictive variables affecting credit-worthiness of a retail borrower. Data was collected Madhya Pradesh and Chhattisgarh. Logistic Regression and Discriminant analysis were used. Similar results were obtained by using both techniques. Results showed that only marital status of borrower, source of loan, status of the borrower and tenure of loan are the significant factors.

Key Words

Credit Risk Assessment, Personal characteristics, Loan characteristics, Discriminant analysis, Logistic regression

Introduction

Amid dynamic global macroeconomic conditions, credit risk is a dominant source of risk for financial institutions and a subject of strict regulatory oversight and policy framework. Given the increasing interdependencies in the global economy, risk managers of commercial banks and non-banking financial institutions alike, may well be interested in questions like “How credit managers should manage their credit risk so that impact on the credit loss of a given bank (or banks) in a given region if there were large unfavorable shocks to equity prices, GDP or interest rates in that or other regions can be minimized?”

The expression consumer credit may be understood as a form of trade where a person obtains money, goods or services and vouches to pay for this in the future, adding a premium (interest) to the original value. Currently, consumer credit is a large industry operating worldwide. Major retailers spur their sales by supplying credit. Automobile companies, banks and other retail segments utilize consumer credit lines as an additional alternative to make profit. On the other hand, consumer credit injects resources into the economy, permitting production and economic expansion of a country, thereby bringing development to the nation (LEWIS, 1992: 2). However to make credit widely available does not mean to distribute credit at random to all those requesting it; there are factors associated to consumer credit which are crucial in the decision of making credit. The large number of decisions involved in the consumer lending business makes it necessary to rely on models and algorithms rather than human discretion. Credit models are useful to evaluate the risk of consumer loans. The application of the technique with greater precision of a prediction model will provide financial returns to the institution.

Literature Review

Researchers have tried to develop models to predict timely loan repayment usually called as credit risk modeling. In essence, a credit models provide an estimate of a borrower's credit risk – i.e. the likelihood that the borrower will repay the loan as promised – based on a number of quantifiable characteristics. Given its widespread use in developed countries, the benefits of credit risk modeling as well as the methodological issues surrounding credit risk modeling are well established in the academic literature.

ZHU Kong-lai and LI Jing-jing 2010 developed and compared discriminant power and warning effect of models based upon logistic regression and discriminant analysis for assessing credit risk of listed companies.

Oluwarotimi O. Odeh & Allen M. Featherstone, 2006 developed credit risk model for agricultural loans. They examined the performance of logistic regression, artificial neural networks and adaptive neuro-fuzzy inference system in predicting credit default using data from Farm Credit System.

In a paper by Maria Aparecida Gouvêa 2007 a sample set of applicants from a large Brazilian financial institution was focused on in order to develop three models each one based on one of the alternative techniques: Logistic Regression, Neural Networks and Genetic Algorithms. Finally, the quality and performance of these models are evaluated and compared to identify the best one.

Ev•en Koèenda and Martin Vojtek, 2009 developed a specification of the credit scoring model with high discriminatory power to analyze data on loans at the retail banking market. They applied Parametric and non-parametric approaches to produce three models using logistic regression (parametric) and one model using Classification and Regression Trees (CART, nonparametric).

Peter V., and Peter R., 2006, identified income, financial, demographic characteristics, and locational factors as critical determinants of future risk and developed a credit risk model based upon logistic regression technique.

Rationale

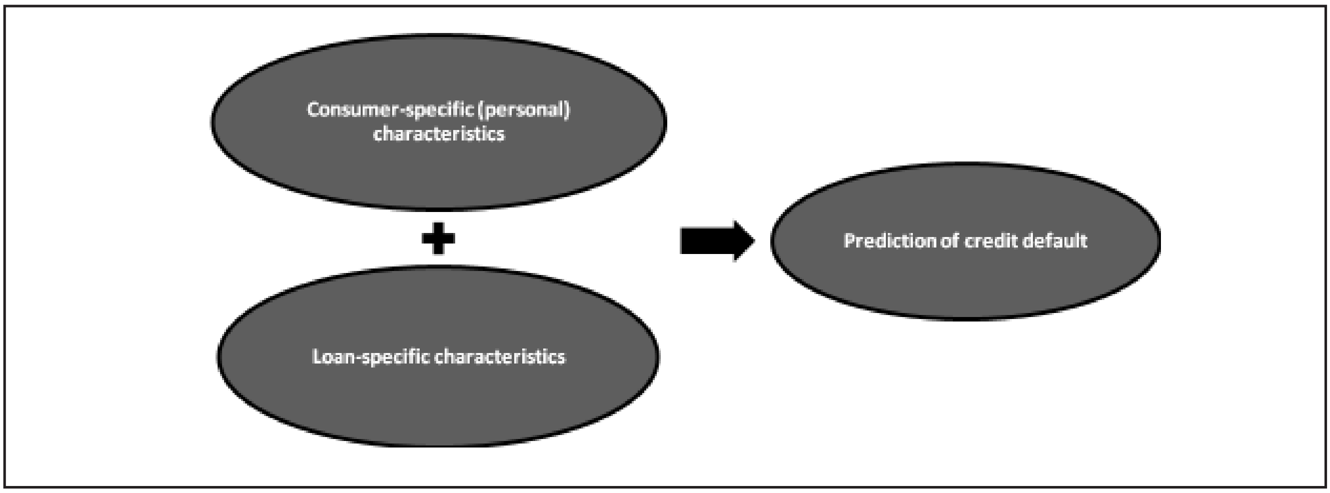
As we went through the literature review we got to know that maximum of the work done on credit risk tends to be focused upon non-retail loans such as corporate lending, housing loans, credit card loans and minimal amount of work has been done upon retail loans and those models which were developed for retail loans were focused upon assessing credit risk for banking institutions. As far as different nature, different set of regulations and different set of risks faced by NBFCs are considered, it becomes obvious that different models would be required for managing credit risk for NBFCs.

The risk management models, which are developed for large business loans, analyze the overall Probability of Default (PD) and Loss Given Default (LGD) of the entire portfolio of retail loans, rather than ascertaining the risk characteristics of a particular borrower (Allen et al., 2004). But Consumer lending business needs a dynamic risk management model because their portfolios are heterogeneous. Thus here we are attempting to develop a credit risk model that can predict credit default by a retail NBFC borrower based upon certain consumer specific and loan specific characteristics.

Objectives

1. To identify the most predictive variables (personal and loan-specific) affecting credit worthiness of retail loan borrowers of NBFC.
2. To develop consumer credit risk models to predict credit default of retail loan borrowers of an NBFC using two statistical techniques: Logistic Regression and Discriminant Analysis.
3. To compare the models developed, in terms of identifying predictive indicators, degree of accurate classification of defaulters, and significance of model.

Conceptualization of the Model



Variables In The Model

Credit Default: If payments are not met for a number of periods (typically three), the lender considers that the borrower has decided to stop payment completely (Quercia, R. G., & Stegman, M. A. (1992)**).** We have converted this qualitative variable into quantitative one on a nominal scale as: Defaulter and Non defaulter.

Consumer-Specific Characteristics:

Marital status: Households most likely to default are those headed by a person who is divorced or separated, a single person or household as suggested by Burrows' (1998), Vandell and Thibodean's (1985). Marital status is again a nominal variable with only two categories: Married and Single

Age: Avery, R. B., Brevoort, K. P., & Canner, G, (2004), also considered age to be an important demographic factor for predicting default. We have classified our sample into four classes on basis of age as:20-30 years, 30-40 years, 40-50 years and 50-60 years.

Existing or New Borrower: The fact that the customer is new or exisiting borrower of the instituion is a very important factor in predicting loan default by a retail customer because institutions extend loan further to only those customers who have good track record of performance of repayment in their previous interactions with the institute. Thi huyen and Kleimeier S., 2007 in their study also considered credit history of borrower with the institution an important variable in developing credit scoring model.

Occupation: Occupation appears to be important in determining whether a borrower will default or not, with office workers perhaps being the "safest" (Moffatt, P. G. (2005)**.** This being a qualitative variable, is measured on a nominal scale as: salaried class and selfemployeed class.

Loan specific characteristics:

Loan Tenure: We see as the tenure of loan increases, probabilty of credit default also increases because the financial and attitudnal characterstics of the borrower that determine ability and willingness to repay loan can change over the period of time(John M. Chapman). On the basis of our data sample we have classified sample as: Medium term loans: upto three years and Long term loans: more than 3 years.

Loan amount: Loan amount can be a factor affecting credit default by a retail cuatomer, because it is easier to repay samller loan amount than the larger one (John M. Chapman) We have divided our sample into 2 classes on basis of loan amount in terms of rupees as:10000 to 75000 and 75001 to 150000.

Loan source: In this classification, internal source of loan refers to loan applications when brought to the organization by its own on roll employees where as alternate source of loan refers to when loan applications are brought to the organization through alternate channels that are external to organization i.e through outsourcing. Such chhanels includes web based chhanels or other business chains.

Research Methodology

Design of the study: this is more of an exploratory study, also a little descriptive in nature as it aims at determining whether certain variables under study can be successfully used to predict credit default in retail loans and then to construct a model based upon significant variables for predicting the probability of credit default by a retail NBFC borrower.

Data collection method: For this research we collected primary data from a leading international NBFC's various branches located at Madhya Pradesh and Chhttisgarh.

Sample Design: The sampling frame comprised of customers from Madhya Pradesh and Chhattisgarh region to whom loan was disbursed with in a period of six months ranging from January-to-August 2012 by a leading National NBFC. Total number of such customers was 5000. Quota sampling technique was used to generate a final sample of 196 customers out of a sampling frame of 5000.

Data descriptive statistics: Out of 197 customers 90 are defaulters where as 107 customers are non defaulters. Out of total sample 68% are salaried employees, 43% are existing borrowers, 88% are males. 136 of total 197 i.e. 69% customers are married. Observing the loan characteristics we find that equal number of applications has been sourced from internal and alternate channels, 52% cases belong to a loan amount up to Rs.75000 and only 27% cases are from medium term loan.

Statistical tool applied: Based upon the findings and techniques used by Simangaliso Biza-Khupe (2011), Fennee Chong, Fareiny Morni, Rosita Suhaimi (2010), Lee J., (2009), Maria Aparecida Gouvêa (2007), Srinvas Gumparthi & Dr. Manickavasagam V.(2010) and ZHU Kong-lai & LI Jing-jing(2010) we opt for using two techniques for designing credit risk model, that are logistic regression and discriminant analysis.

Results and Discussions

The section provides a summary of results obtained using the two models. A comparison of the two models has also been included below.

Logistic Regression Model

Y = In (P/1-P) = á + b1X1 + b2X2 + b3X3 + b4X4 + b5X5+ ………. bkXk+ ì

Where, Y= binary dependent variable

p = probability of success, Xk= Independent variable k, bk = parameter to be estimated

Hypothesis:

Null hypothesis: There is no significant effect of age, gender, income, occupation, existing borrower new borrower status, loan amount, and loan tenure and loan source in predicting probability of credit default by a borrower. Mathematically represented as: H0:b1=b2=b3=b4=b5=b6=b7=b8=0

Alternate hypothesis: There is significant effect of age, gender, income, occupation, existing borrower new borrower status, loan amount, and loan tenure and loan source on probability of credit default by a borrower.

H1: b1 $\neq$ b2 b3 b4 b5 b6 b7 b8 0

Logistic Regression Results Summary

Logit (p)=1.841 + 1.815 (EBNB status) - 2.845 (loan tenure) - 1.07 (Loan Source) - 2.027 (marital status)

A test of the full model against a constant only model was statistically significant, indicating that the predictors as a set reliably distinguished between defaulters and non defaulters (chi square = 87.582, p < .000 with df = 8). Thus null hypothesis was rejected. Nagelkerke's R2 of .525 indicated a moderate relationship between prediction and grouping. Prediction success overall was 80.9%. (Annexure 1) The Wald criterion (Annexure 2) demonstrated that EBNB Status (p=.000), Loan Tenure (p=.000), Marital Status (p=.003) and Loan source (p=.015) made a significant contribution to prediction. (Annexure 2) Age, Gender, Occupation and loan amount were not significant predictors. EXP(B) depicts that a new borrower are 6 times more likely to make a credit default than existing borrowers, increase in loan tenure by one month increases probability of credit default by .058 times, customer whose loan application has been brought by an internal source is .341 times more likely to make a credit default than that brought by alternate source and those who are single are having .132 times less probability of credit default.

Discriminant Analysis:

DA involves the determination of a linear equation like regression that will predict which Group the case belongs to. The form of the equation or function is:

D= a + v1X1 + v2X2 + v3X3…. + viXi

Where D=discriminate function

v =the discriminant coefficient or weight for that variable

X =respondent's score for that variable

a =a constant

i =the number of predictor variables

Discriminant Analysis results summary:

D = -1.467-(1.164 *ebnbstatus*) + (1.921 *tenure*) + (.594 loansource) + (1.018 *marital status*)

Test of equality of group means table provides strong statistical evidence of significant differences between means of non defaulters and defaulters for existing borrower new borrower status, tenure, loan source, loan amount and marital status at p<.0001 level of significance. The discriminate function revealed a significant association between groups and existing borrower new borrower status, marital status, loan source and loan tenure and closer analysis of the structure matrix (correlation of predictor variables with discriminant function) (Annexure 4) revealed the same as existing borrower new borrower status (-.557), marital status (.264), loan source (.189) and loan tenure (.728), while age (-.136), loan amount (.246), gender (.014) and occupation (-.035) were proved to be poor predictors. The standardized discriminant coefficients revealed that order of factors in magnitude of their impact upon credit default is loan tenure (.728) having highest impact followed by existing borrower new borrower status (-.530), marital status (.424) and at last loan source (.295). The classification showed that overall 79.8% were correctly classified (Annexure 3).

Comparison of both the models:

Both the models can be majorly compared upon these four points:

1. **Significant predictor variables:** Both the models indicated that four out of eight predictor variables can significantly predict credit default by a retail NBFC borrower.

2. Interestingly, both the models indicated existing borrower and new borrower status, loan tenure, loan source and marital status can significantly predict credit default.

3. **Direction of impact of all four factors:** Both the models indicated the same movements of credit risk in same direction in response to variation the predictor variable.

   Existing borrower new borrower status: New Borrowers are more likely to be defaulter than existing borrower.

   Loan tenure: loans with longer duration are less risky as customers having medium loan duration are having greater probability of making a credit default.

   Loan source: internal loan source is riskier than alternate, i.e. customer whose loan application has been sourced from internal sources has lesser probability of making a credit default.

   Marital status: married customers have a greater probability of credit default than those who are single.

4. **Percentage of accurate classification as shown by classification matrix:** Both the models

successfully classified defaulters as defaulters and non defaulters as non defaulters up to an extent of 80.9% (logit model) and 79.8% (discriminant model).

5. **Moreover significance of logistic regression model was better than that of discriminant model.**

Applicability of The Findings

The results of the research can be employed by NBFCs in balancing their retail credit portfolios among different classes of borrowers based upon the characteristics found to significantly affect credit default.

Limitations of The Research

1. The findings can be applied only to NBFCs personal loan segment.

2. Some of the personal characteristics of borrowers such income of borrower, number of dependents, distance of home from NBFC office, availability of permanent residence identified as important for analyzing credit risk could not be included in the research because of data availability constraint, since the actual corporate data was collected, some of the characteristics of borrowers couldn't be disclosed by them due to policy issues.

Further Scope of Study

1. Credit risk model that we have tried to develop can be further refined by excluding insignificant variables and including more factors that can be significant predictors of credit default.

2. Similar model can be developed for credit card, mortgage, vehicle loans and other sub segments of personal loans.

Conclusion

In this research we identified the factors that can be significant in predicting credit default by a retail borrower of NBFC using logistic regression and discriminant analysis. Both the models gave similar results. Our results showed that NBFCs should rely more upon alternate sources of loan rather than increasing internal sources, they should focus their resources upon retaining existing borrowers rather than exploring new borrowers, increase proportion of medium term loans in their portfolio and also percentage of single customers as compared to married ones.

References

Alan Agresti, An introduction to categorical data analysis.

Avery, R.B., Calem, P.S., Canner, G.B., 2004. "Consumer credit scoring: Do situational circumstances matter?" *Journal of Banking and Finance* 28, 835–856.

Burrows R. 1998, "Mortgage indebtedness in England: an 'epidemiology", *Housing Studies*, 13(1): 5-22.

Chapman J.M.,1940 "Factors Affecting Credit Risk in Personal Lending". National Bureau of economic research.

Chong, F., Morni, F. and Suhaimi, R. (2010). "Demographic Factors and Repayment Performance of NBFI Customers in Kuching"

David, J., (2001). "Modeling consumer credit risk", *IMA journal of management mathematics* (2001) 12, 139-155

Fennee .C, Fareiny .M, Suhaimi R., 2010 "Demographic factors and repayment performance of NBFI Customers in Kuching"

Gouvêa M. A., (2007). "Credit risk analysis applying logistic regression, neural networks and genetic algorithms models." POMS 18th Annual Conference May 4 to May 7, 2007, Dallas, Texas, U.S.A.,

Gumparthi S.,(2010), "Risk assessment model for assessing NBFCs' (asset financing) customers" *International journal of Trade, Economics and Finance*, vol. 1, no. 1, june, 2010 2010-023

Lee J., (2009), "Racial/Ethnic disparities in household debt repayment", The Ohio State University

Lewis, E. M. (1992) *An Introduction to Credit Scoring*. San Rafael: Fair Isaac and Co., Inc.

Koèenda E., and Vojtek M., (2009). "Default Predictors and Credit Scoring Models for Retail Banking", CESIFO working paper no. 2862, category 12: empirical and theoretical methods.

Moffat, P. G. (2005) "Hurdle Models of Loan Default", *Journal of the Operational Research Society*

Odeh O. O., Featherstone A. M., and Das Sanjoy, 2006, "Predicting credit default in an agricultural bank: methods and mssues" Selected Paper prepared for presentation at the Southern Agricultural Economics Association Annual Meetings Orlando, Florida, February 5-8, 2006

Simangaliso B.K., 2011. "The significance of demographic factors on key-indicator variables of consumer credit decisions" *Far East Journal of Psychology and Business* Vol 3 No 2.

Vandell K & Thibodean T. 1985, "Estimation of mortgage defaults using disaggregate loan history data", *AREUEA Journal*, 13(3): 292-316.

ZHU K., and LI Jing-jing, (2010), "Studies of Discriminant Analysis and LogisticRegression Model Application in Credit Risk for China's Listed Companies", *Management Science and Engineering* Vol. 4, No. 4, 2010, pp. 24-32 ,www.cscanada.org