University of Ghana http://ugspace.ug.edu.gh

DELINQUENCY AND DEFAULT RISK MODELING OF MICROFINANCE IN GHANA

BY

DENNIS ARKU

(10164162)

A THESIS SUBMITTED TO THE SCHOOL OF GRADUATE STUDIES, UNIVERSITY OF GHANA, IN PARTIAL FULFILMENT OF THE REQUIREMENT FOR THE DEGREE OF MASTER OF PHILOSOPHY IN STATISTICS

JUNE, 2013

University of Ghana http://ugspace.ug.edu.gh

DECLARATION

CANDIDATE'S DECLARATION

This is to certify that this thesis is the result of my own research work and that no part of it has been presented for another degree in this University or elsewhere.

SIGNATURE:…………………………… DATE………………………..….

DENNIS ARKU (10164162)

(CANDIDATE)

SUPERVISOR'S DECLARATION

We hereby certify that this thesis was prepared from the candidate's own work and supervised in accordance with guidelines on supervision of thesis laid down by the University of Ghana.

SIGNATURE:…………………………. DATE………………………..….

DR KWABENA DOKU- AMPONSAH

(PRINCIPAL SUPERVISOR)

SIGNATURE:…………………………. DATE………………………..….

DR ISAAC KWAME BAIDOO

(CO-SUPERVISOR)

University of Ghana http://ugspace.ug.edu.gh

ABSTRACT

The objective of this research was to identify the risk factors that influence loan default of customers in the microfinance sector and to develop model that links these factors to credit default for any customer in the sector. Data from a microfinance institution based in Accra was used. A binomial logistic regression analysis was fitted to a data of 472 customers who were granted credit from January 2011 to December 2012. Based on the Wald criterion, it was realized that among the variables that were considered only six out of the 16 predictor variables significantly influence the probability of loan default. One of the key findings of the study was the fact that the loan officer has a significant effect on loan default risk. Other factors include; Client's age, Assessment, Type of collateral, guarantor, and residential status. The findings suggest that, default rate is higher for trading and manufacturing sectors than for food vendors as well as those in the service sector. Clients in the service industry are relatively less risky. Clients with guarantors or security other than household items perform well in their obligations. The findings indicated that the responsibility of the loan officer have a tremendous impact on loan default. A test of the full model against a baseline model was statistically significant, indicating that the predictors, as a set, reliably distinguished defaulters and non-defaulters. The Receiver Operating Characteristics (ROC) that measures the sensitivity and specificity of the model was significant at 0.05 level. Using the hold out sample, the model is able to classify defaulters and non-defaulters with at least 80% accuracy. This means that for every four out of five clients, the model is able to predict correctly: default or otherwise. The model could serve as tool to manage and improve loan decision and ultimately enterprise profitability.

University of Ghana http://ugspace.ug.edu.gh

ACKNOWLEGDEMENT

Special thanks go to the almighty Yahweh for given me the mental fortitude and wisdom to accomplish this work. I also extend a hand of warm gratitude to Dr. Isaac Kwame Baidoo my supervisor, for his specialized judgment, mentorship, nurturing, patience and willingness to guide my intellectual thoughts on the plains of research. My appreciations also go to my indefatigable supervisor Dr Kwabena Doku-Amponsah for his encouragement and thoughtful comments and suggestions. Most importantly, none of this would have been possible without the love and patience of my family. They have been a constant source of love, concern, support and strength all these years, particularly my brother, Mr. Sampson Arku whose support inspired my academic pursuits up to this point. I will also like to mention my mother, Mrs Felicia Arku who has been a force of encouragement in my pursuit of this degree. My heartfelt appreciation also goes to my lovely wife Comfort Baidoo and kids, Edna and Jehodiah. To the institution that graciously provided the data, I say a big thank to you. And to all and sundry, who contributed in diverse ways in catapulting me onto the summit of this scholarly experience, I am truly grateful and may the Almighty Yewheh reward you in multiple ways.

University of Ghana http://ugspace.ug.edu.gh

TABLE OF CONTENTS

University of Ghana http://ugspace.ug.edu.gh

University of Ghana http://ugspace.ug.edu.gh

LIST OF TABLES

University of Ghana http://ugspace.ug.edu.gh

LIST OF FIGURES

University of Ghana http://ugspace.ug.edu.gh

LIST OF ABBREVIATIONS

AUC ……………………………………………………………...Area under the curve

GHC………………………………………………………………….....Ghana cedis

GAMC…………………………………...Ghana Association of Microfinance Companies

LR……………………………………………………………………...Likelihood Ratio

MFI…………………………………………………………….Microfinance Institution

ROC………………………………………………….Receiver Operating Characteristic

University of Ghana http://ugspace.ug.edu.gh

CHAPTER ONE

INTRODUCTION

1.1 Background of the study

During the past three decades, many developing countries including Ghana have seen their financial system undergo some transformation and innovation due to the emergence of Microfinance Institutions (MFIs). Banks that hitherto did not want to lend to the poor as a result of the poor inability to provide collateral or the high risk associated with lending to the poor are now doing so under the name of microfinance. Microfinance has been defined by Robinson (2001) as the small-scale financial services provided to people into petty businesses in both rural and urban communities. MFIs are primarily established to provide business and consumer credit or bring financial services to poor and vulnerable groups in society with the ultimate goal of improving living standards or eradicating poverty, especially among women. The benefits of Microfinance to poverty alleviation and overall improvement in living standards have been extensively studied by Vukson (as cited in Awotwi, 2011). Vukson showed that, health, education and self-esteem of the borrowers and their families had significantly improved. Khandkar (2005) and Hashemi et al, (1996) in another study also indicated that microfinance had contributed to women‘s empowerment and overall poverty reduction.

Credit risk ranks as the highest risk facing the microfinance industry around the globe. This is due to the nature of clientele base and weak internal control systems that

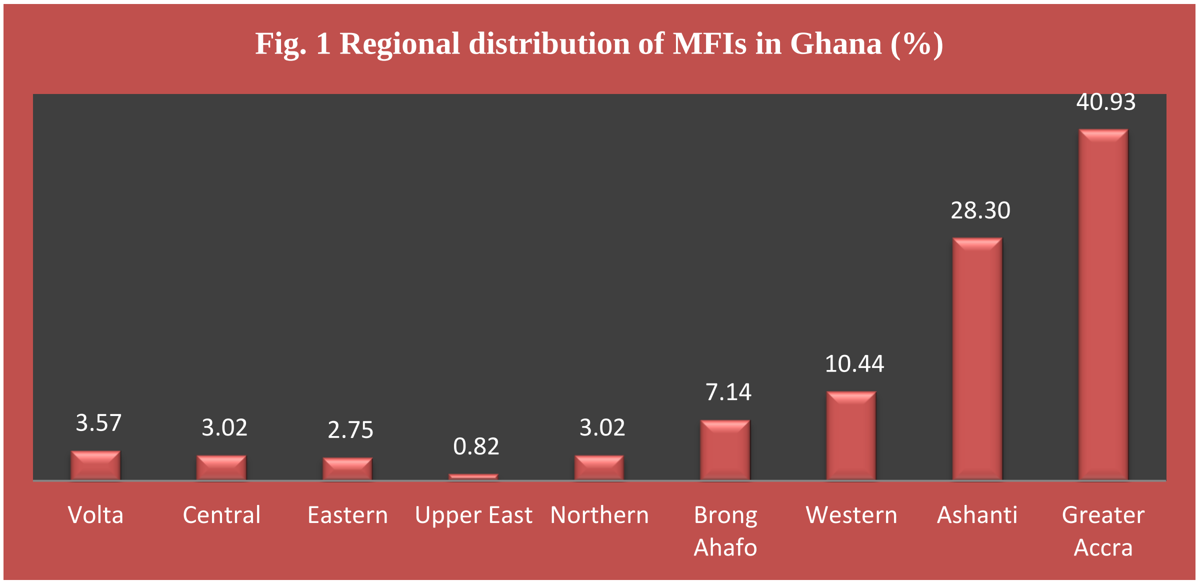
University of Ghana http://ugspace.ug.edu.gh

characterize most MFIs. In Ghana, the same reasons apply. Other issues such as fraud, transaction error etc. related to MFIs are covered.

Greenwood and Jovanovich (1990) maintained that financial deepening, and for that matter, competition promotes innovation and higher return on investment which ultimately lead to long-term growth and survivorship. Rosenberg (2009) also argues that as MFIs compete for clients there is also competition on interest rate, which usually pushes the MFIs to become more efficient. However, McIntosh, de Janvry, and Sadoulet (2004) indicated that increased competition induces a decline in repayment performance and in savings deposited with the MFIs, suggesting multiple loan-taking by clients. Petersen and Rajan (1998), also argued that attrition rates, low portfolio quality, repayment rate, collection methods due to growth in MFI numbers threaten their sustainability.

In Ghana, the number of registered MFIs has grown from 150 in 2008 to 364 in 2011 and continues to surge as the industry appears profitable (GAMC, 2012). For instance the Greater Accra region contributes over 40% of the number of MFIs in Ghana (figure 1).

University of Ghana http://ugspace.ug.edu.gh



Source: Ghana Association of Microfinance Companies report 24th September, 2012.

Figure 1: Regional distribution of MFIs in Ghana(%)

Nonetheless, as the number of MFIs continues to grow and competition becomes intense in the industry, concerns for the survival and long-term sustainability of these institutions also grow. Growth in outreach has closely followed existing high-density areas, leading to these regions becoming more and more saturated, see (figure 1). Multiple borrowing is therefore inevitable in the absence of a fully-functioning centralized database that records clients by their national identity card numbers, and an aggregate total financial obligation across MFIs and other financial institutions.

This growth in competition and the nature of the MFIs calls for innovative ways in assessing the Character, Capacity, and Capital of customers. The character assesses how a person has handled past debt obligations. The character is determined on the basis of customer's background information. The capacity determines how much a borrower can

University of Ghana http://ugspace.ug.edu.gh

comfortably handle based on past records or obligations and cash flow and the capital determines the available assets of the borrower that could be used to repay the debts should payment stalls.

Loan default is one of the major credit risks that Microfinance institutions are faced with. It could come as a result of lack of due diligence on the part of loan officers or the borrowers' personal, demographic and loan characteristics.

This study therefore seeks to ascertain factors that have a major influence on default and develop a function that links client default with these factors.

1.2 Problem statement

The need for micro-credit today is at its highest owing to the structure of the Ghanaian economy where majority (80%) of the workforce works in the informal sector (Boateng and Ampratwum, 2011). The industry is considered inherently risky mostly due to the sector it serves. In most microfinance Institutions (MFIs) the methods of deciding whether to grant a loan to an individual are usually mostly speculative and mostly based on the experience of previous decisions other than objective analysis. Suffice to say that all lending decisions are speculative but some decisions could be data driven.

Usually, the borrower puts in a request for loan and on that basis three things happen:

1. Client makes a deposit (% requested amount)
2. Client's home and business is visited (Assessment)
3. Collateral or personal guarantors are requested

University of Ghana http://ugspace.ug.edu.gh

This is done in order to assess the character, capacity and capital of the client. This exercise is usually carried out by the relationship or loan officer. Objective data is collected in the process but may not be sufficient or credible. The manager thus have little objective basis to relate to in order to make a decision to approve or not. This usually leads to guided but inadequate decisions resulting in increasing rate of default. For instance, with the MFI, it was found that as at December 2012, the company had disbursed GHC 1,739,041.00. Out of this 79% of the loan portfolio has been paid as at 2012 whilst 21% of the portfolio value has been in arrears. In terms of number of clients, the 79% constitute 68% of the clients whilst the 21% arrears constitute 32% of the clients. This result shows poor decision making and the consequence of the outcome may be severe. The question is, can the outcome of the decision to advance a credit facility be successfully predicted with the data available and how can the bad performing customers be isolated or minimized? Better decision mechanism may separate good and bad clients.

Besides, the growing sophistication of lending in the sector requires the development of computational approaches to evaluation of borrower's credit worth.

1.3 Objectives

The overall objective of this research is to develop a model that could be used among other things to assess the credit worthiness of customers in the microfinance sector.

Specifically the research seeks to address two things:

- Identify risk factors that influence loan default among customers in the Microfinance sector

University of Ghana http://ugspace.ug.edu.gh

- Develop a function that links the risk factors to the likelihood of credit default for a customer

1.4 Significance of the study

The study is significant in the three key ways:

1) It will help management to make objective on-the-spot credit decisions.
2) It will help new entrants in the industry to control and manage default risk for a sustainable development
3) To improve portfolio quality of MFIs

1.5 Organization of work

This thesis is structured in five chapters:

Chapter one gives a background introduction on credit risk management and measurement. Chapter two reviews the relevant literature on the subject matter. Chapter three explains the methodology employed to achieve the objectives. The presentation of findings and discussion is captured in chapter four. Chapter five presents the summary of the discussions, conclusions and recommendations.

University of Ghana http://ugspace.ug.edu.gh

CHAPTER TWO

LITERATURE REVIEW

2.0. Introduction

For a complete understanding of the theory and practice behind the study, other related studies were reviewed. This chapter presents a structured review of the related literature. It starts with the concept of loan default and examines the variables commonly used in default models. It finally discusses some default modeling techniques.

2.2. The concept of credit risk and default models

Credit risk is one of the major risks in banking industry. It is the potential for a loss when a borrower cannot make payments as obligated to a lender. It is commonly measured and communicated as the probability of an individual borrower's default.

Gan et al. (2012) did a research on default of mortgage loans in China and observed that a borrower's demographic characteristics impact banks' lending decision process. His results also suggested that the loan amount and interest rate range are positively correlated with the probability of loan default; which means an increase in the loan amount, interest rate level may increase the probability of loan default. Seven variables were identified as having a negative impact on the probability of loan default, good bank rate, occupation such as general manager, general staff, professional employee, whether the borrower resides within the same district with the bank. Amilie and Allen (2006) also

University of Ghana http://ugspace.ug.edu.gh

identified three major financial ratios that significantly influence the probability of default. These are leverage, profitability and liquidity. Length of loan also tested to be highly significant in the default prediction.

Agarwal et al (2009), assessed the role of individual social capital information characteristics on household default and bankruptcy outcomes. The study used monthly panel data set of more than 170 000 credit card holders over 24 months. With the observations of each borrower's default and bankruptcy filing status, they were able to find distress factors such as riskiness, spending, debt, income, wealth, economic conditions, legal environment and socio-demographical characteristics to significantly affect default risk. Their results showed that a borrower who is married and owns a house of his own has a lower risk of default.

Vasanthi and Raja (2006), also estimated the likelihood of default risk associated with income and other factors with data obtained from Australian Bureau of Statistics. The goal was to establish the relationship between the default risk of homeowners, their socio-economic and housing characteristics. Using a sample of 3431 households, the findings indicated that, repayment rate is substantially high compared to consumer credit amounting to 93.03%.

They also found out that the age of the head of the household is significant: the younger households tend to be adversely affected by the increasing burden of mortgage payments as other consumer loans. According to the paper, income showed to have predictive power: lower income is one of the major contributory factors for default. Another

University of Ghana http://ugspace.ug.edu.gh

important factor was the loan to value ratio indicating that higher loan to value ratio would increase the probability of default. Also, the educational level of the head of household and marital status had significance impact on default. Vasanthi and Raja drew a conclusion that the probability of default is higher with an uneducated, younger and divorced person as head of the family compared to others.

Autio et al. (2009), also conducted a comprehensive study of the use of small instant loans in Finland among 1951 young adults. An open online survey for 18- to 29-year olds included questions about age, gender, financial situation, such as income, employment and occupational status, and family structure. The results showed that the 18- to 23-year-olds use small instant loans more than the 24 to 29-year-olds. The latter group, on the other hand, use consumer credit more, because of their higher income and occupational status, Gender did not seem to have an effect on the number of loans taken, but occupational status, income and household structure did.

Agarwal et al. (2008) studied the determinant of automobile loan default and observed that automobiles are highly visible consumption goods that are often purchased on credit in their article they used a unique proprietary data set of individual automobile loans to assess whether borrower consumption choice reveals information about future loan performance. The result was that an increase in income raises the probability of prepayment, whereas a rise in unemployment increases the probability of default. A decrease in the market rate (the three year Treasury note rate) increases both the probabilities of prepayment and default. They also find that loans on most luxury

University of Ghana http://ugspace.ug.edu.gh

automobiles have a higher probability of prepayment, while loans on most economy automobiles have a lower probability of default. This result is highly correlated to the size of loan. Prepayment is the initial payment the customer makes for the purchase of the automobile.

According to Mario and Claudio (2010), in United States both socio-demographical and behavioral variables have a notable effect on default. The most significant socio-demographical variables are income, time since last moving, age, possession of credit card, education and nationality. Some behavioral variables seemed to have even more predictive power. Those are the amount of scores the customer obtained, loan size and the information if customer has been granted a loan earlier from the same company. Interestingly, the results have variation to some extent when excluding few of the variables outside the model. The predictive power of all three models is adequate and thus can be employed as a reliable credit scoring model for the credit institutions.

According to Lewis (1992), consumer credit has been around for 3000 years since the time of the Babylonians. For the last 750 years of that time there has been an industry in lending to consumers, beginning with the pawn brokers and the insurers of the middle Ages, but the lending to the mass market of consumers in the non-Islamic world is a phenomenon of the last fifty years. In the 1920s, Henry Ford and A.P.Sloan had recognised that it was not enough to produce products, like cars, for the mass market but one also had to develop ways of financing their purchase. This led to the development of finance houses, e.g. GE Capital, GM Finance in the United States.

University of Ghana http://ugspace.ug.edu.gh

Updegrave (1987) found that there were eight variables that affected consumer credit risk: the number of variables, the historic repayment record, bankruptcy history, work and resident duration, income, occupation, age and the state of savings account. Similar results were found by Steenackers and Goovaerts (1989) who collected data on personal loans in Belgian credit company and found out that age, resident and work duration, the number and duration of loans, district, occupation, phone ownership, working in the public sector or not, monthly income and housing ownership have a significant relationship with repayment behaviour.

In Ghana, Awotwi (2011) used a data from an international bank in Ghana, to study the risk of default. The data comprise of 9939 observations of customers of which 14% turned out to be defaulters. According to him, six variables were found to be highly significant in the default. These are Marital Status, Number of months the applicant has been in current employment, interest rate, tenure of loan or repayment period, income level and loan amount.

2.3 Variables commonly used in loan default models

The pragmatism and empiricism of credit scoring implies that any characteristics and environments of the borrower that has obvious connections with default risk should be used in the scoring system (Lewis, 1992). The variables should be sequentially added or deleted to maximize the model's predictive accuracy (Hand and Henley, 1997).

University of Ghana http://ugspace.ug.edu.gh

Lewis suggests that there is no need to justify the case for any variable. If it helps the predictions, it should be used. Dinh and Kleimeier (2007) shared the view that, there are two important standards for variable selection; first, the variables should have significant coefficients and contribute to explanation of the dependent variable's variance. Second, the variables should have close correlation with included variables.

According to (Dinh and Kleimeier, 2007), the major factors commonly used in credit scoring models include the borrowers' income, age, gender, education, occupation, employer type, region, time at present address, residential status, marital status, home phone, collateral value, loan duration, time with bank, number of loans, and current income is a commonly used proxy for the borrower's financial wealth and his or her ability to repay According to the findings, there is a positive relationship between income and the borrowers' default rate; higher income is associated with lower default risk. It is worth noting that, regulatory concerns limit the ability to include every variable in most jurisdictions.

Occupation is a common variable used in credit scoring model and is highly correlated with the borrowers' income level. Education enhances the borrowers' ability to repay. The better educated borrowers are deemed to have more stable and higher income employment and thus a lower default rate. The borrowers' education level distinguished from post-graduate to non-high school graduate. Borrowers with high level of education are more likely to repay their loan since they occupy higher positions and with high

University of Ghana http://ugspace.ug.edu.gh

income levels. Although incidence may be lower, severity may be higher and hence must be studied.

Cook et al., (1992) noted that time with employer measures the number of years that the borrower has been working for the current employer. It reflects the satisfaction of the borrower with the current job. The higher the borrowers' job satisfactions, the more stable their employment will be and the higher their ability to repay their loans
According to Capon (1982), length of time with employer may discriminate against women, since women's average length of employment reduces due to pregnancy and childbearing.

According to Dinh and Kleimeier (2007), the type of company a borrower works in could be a proxy for income level and stability. Missing values of this variable are also very informative as borrowers who do not answer this question show the highest probability of default.

Age was has also been considered as one of the factors used in default risk modeling. Thomas (2000) and Boyle et al. (1992) confirm that older borrowers are more risk adverse, and therefore less likely to default. Hence bank's hesitance to lend to younger borrowers who are more risk embracing.

Arminger et al., (1997) noted that gender in addition to age is one of the most used socio-demographical variables to differentiate the predictive power of gender. Their findings

University of Ghana http://ugspace.ug.edu.gh

suggest clear evidence that women default less frequently on loans possibly because they are more risk averse. According to Coval et al., (2000) gender is a fair discriminatory base on the statistical default rates of men versus women. Region means the area of the country that borrower lives. As people of similar wealth tend to live in the same location, the geographic criterion can indicate a borrower's level of financial wealth. Some suburbs might attract richer residents and this could result increase in housing and property prices. This may also affects the collateral value and probability of default and need to be investigated. It is not clear whether this result is still valid in the present circumstances.

According to Mario and Claudio (2010) the residential variable measures whether borrowers own their home, rent, or live with their parents. This could indicate the borrowers' financial wealth in the case of home ownership. Residential status also indicates financial pressure on borrowers' income, for example rental cost.

According to Crook et al.'s (1992), the default risk drops with an increase in time at present address; it might be a proxy for the borrowers' maturity, stability, or risk aversion. Changing address might be a signal that a borrower's financial wealth is high or improving rapidly. Time addresses the number of years that the borrowers have been living at their current address.

According to Dinh and Kleimeir (2007), marital status affects the borrower's level of responsibility, reliability, or maturity. The probability of default is higher for married than single borrowers. They discover that the marital status is typically related to number

University of Ghana http://ugspace.ug.edu.gh

of dependents which in turn reflects financial pressure on the borrower and borrower's ability to repay a loan.

Collateral is a form of guarantee to support the loan. According to Gup and Kolari (2005), borrowers' collateral can be a sign of default risk. Thus, the probability of default is very low for loans that a house, for example, serves as collateral. This is because of the fear of losing their house. Collateral reduces the bank's risk when it makes a loan. The higher the collateral value the higher the incentive for the borrowers to repay the loan since they do not want to lose their collateral. The collateral value could also be a proxy for the borrowers' financial wealth since it is significantly positive correlated with the borrowers' income (Dinh and Kleimeier, 2007). Loan duration indicates the maturity of loans in months. Loan duration reflects the borrowers' intention, risk aversion, or self-assessment of repayment ability. Time with the bank indicates the borrowers' length banking relationship in years. It can be assumed that the longer a borrower stays with the bank, the more the bank knows about this borrower, and it could lower the probability of default. What is means is that, more knowledge, better decisioning which could lead to lower risk. But this variable should be updated regularly due to adverse and unexpected changes in the borrowers' situation (Dinh and Kleimeier, 2007). It could be seen that more knowledge leads to better decision making and invaribly lower risk of default.

Several studies used loan size as a predictor variable but the overall results are ambiguous and thus no clear expectations can be formed. However, Jacobson and Roszbach (2003),

University of Ghana http://ugspace.ug.edu.gh

showed that loan size has no significant influence on default risk. In the study of Kocenda and Vojtek (2009) small loans appear to be more risky. But the study ignores the severity loan size which is important to the survival of the MFI.

According to Kocenda and Vojtek (2009), borrowers who hold current accounts with their banks have a lower default risk. However, Boyes et al. (1989) recognized that if banks were minimizing default risk, one should find the above variables with positive (negative) effect on the probability of granting a loan and a negative (positive) effect on default risk.

Number of loans or cycle of a loan measures the total number of loans a borrower has received from the bank during the whole relationship with the bank. Today, most borrowers have more than one loan from the same bank. This variable reflects the difficulty for a defaulted borrower to receive further loans from the same bank. This phenomenon is quite severe and if not checked could lead to a total collapse of the MFI system in Ghana.

In Summary, the above reviews suggest and as Lewis (1992) posited that any characteristics and environments of the borrower that has obvious connections with default risk should be used in a credit scoring system.

2.4. Default modelling techniques

There are a number of statistical approaches used to estimate credit scoring models and assessing borrowers' credit worthiness, such as discriminant analysis (Dunn and Frey,

University of Ghana http://ugspace.ug.edu.gh

1976), linear probability models (Turvey, 1991), probit models (Lufburrow et al., 1984) and logit models (Mortensen et al., 1988). The last three methods estimate the default rate based on the historical data on loan performances and the borrowers' characteristics. Linear Probability Model (LPM) assumes there is a linear relationship between the default rate and the factors. The Probit model assumes the probability of default follows the standard cumulative normal distribution function. Discriminant analysis divides borrowers into high and low default-risk classes (Mester, 1997).

In the 1980's a new method for classifying was introduced by Breiman et al. This split data into smaller and smaller pieces. According to the paper, the classification and regression tree approach is an appropriate method for classification of good and bad loans.

Laitinen and Kankaanpää (1999), assessed six alternative methods (LDA, LR, RPA, survival analysis, and HIP) that have been applied to financial failure prediction. The main objective was to study whether the results stemming from the use of alternative methods differ from each other. They used only three financial ratios (total debt to total assets, the ratio of cash to current liabilities and the operating income to total assets) due to methodological issues. The results of 76 randomly selected from Finnish small and medium sized failed firms indicate that no superior method has been found but the predictive power of logistic analysis was best resulting 89.5% versus other methods.

Apilado et al.(1974) applied discriminant analysis to construct their credit scoring models and state that "discriminant analysis firstly distinguishes among group and

University of Ghana http://ugspace.ug.edu.gh

identifies group differences; secondly, it classifies existing and new observation into predetermined groups, and finally it identifies the key variables that contribute the most to the discrimination among groups" Discriminant analysis was used as a credit scoring tool first by Durand (1941) to produce good predictions of credit repayment. Extensive use of discriminant analysis, to build credit scoring models for general banks and credit card sectors have been carried out by Eisenbeis (1983), Martel and Fitts (1981), Grablowsky and Talley (1981), Reichart et al., (1983), Titterington (1992), Desai et al., (1996), Bardos (1990) and Lee et al., (1999).

According to Collins and Green (1982), the linear probability model could present reasonable prediction results compared to discriminant analysis and logit models. However, Pyndick and Rubinfeld (1998), Greene (1997), and Judge et al. (1985) indicated that the linear probability model could predict the default rate, but the predictive value might not necessary lie between zero and one. Moreover, because the variance of the models is generally heteroscedastic, it leads to inconsistent estimation problems and invalid conventional measure of fit such as the R^2.

According to Hand and Henley (1997), the logistic approach is a more appropriate statistical tool than linear regression, when there are two discrete classes (good and bad risks) defined in the model. This gives the logistic approach superior classification rate. The probit model is very similar to the logit model. The logit model is generally preferred to the probit model because of its simplicity (Barney et al., 1999; Novak and LaDue, 1999; Lee and Jung, 1999).

University of Ghana http://ugspace.ug.edu.gh

Laitinen (2000) continued the work and tested whether Taylor's series expansion can be used to solve the problem associated with the functional form of bankruptcy prediction models. To avoid the problems associated with the normality of variables, the logistic model which describes the insolvency risk was applied. Several financial ratios were employed with estimation sample including 400 firms and the results suggest that the cash to total assets, cash flow to total assets, and shareholder's equity to total assets ratios operationalize the factors affecting the insolvency risk.

Hurdle and Muller (2000) used a semi parametric regression model called generalized partially linear model and showed that performed better than logistic regression. A systematic review of articles that have used logistic regression not only promotes the learning about this method, but also helps suggest new guidelines for principled applications of this versatile technique. Logistic regression, being one special class of regression models, is well suited for the study of categorical outcome variables, such as staying in or dropping out from college. This technique is increasingly applied in educational research.

Clarke, (2005) explained that, logistic modeling approach is commonly used to model bank's lending decision. According to Collins and Green (1982), the logit model can increase the overall classification rate, and substantially reduce the error rate. The logistic approach also gives superior classification compare to discriminant analysis (Wiginto, 1980).

University of Ghana http://ugspace.ug.edu.gh

In summary, based on the literature, there is no best method for estimating risk models. The approach really depends on the data structure and the assumptions related to it. However, the logit models and neural networks have been applied frequently in previous research.

University of Ghana http://ugspace.ug.edu.gh

CHAPTER THREE

RESEARCH METHODOLOGY

3.0 Introduction

This chapter outlines a suitable methodology for modeling the default probability. It outlines how the data was collected, processed and analyzed. The relevant mathematical technique has been presented.

3.1 Research design

The design of the study is strictly quantitative. The study makes use of quantitative data; a historical data that contains the profile of customers of a microfinance institution was collected. Due to company policy, the name of the microfinance company was restricted and will be referred in this study as ‘the MFI’.

3.2 Target population and sample size

The target population was all customers of microfinance institutions in Ghana. However, Out of five institutions contacted, only one made available files and loan profiles for analysis. For many of the MFIs contacted either the data was not adequately captured to fit the purpose of the study or the files were unavailable. The analysis was thus restricted to one data source. The data set comprises of all individual loans disbursed from January 2011 to December 2012.

University of Ghana http://ugspace.ug.edu.gh

3.3 Data collection strategy

Given the sampling frame all the files of loans disbursed under the period of study were grouped into three; on time, not on time, default. On time: clients who were able to discharge their obligation within the time given for payment. Not on time: clients finished payment within one month after contract expired. Default: clients who have neither finished payment nor finished payment one month beyond the contract expiry period.

The profiles of such clients; personal and loan characteristics were then captured with an SPSS database. Only those whose contracts were within the study period were captured.

3.3.1 Predictor variables

In all sixteen (16) variables were considered for this study. The variables were grouped into two main categories: borrower and loan characteristics. Table 1 below presents all variables and its description and the expected influence on default as a whole. It should be noted that, the expected effect of the variables on default risk is based on expert advice of microfinance staff.

Table 1: Overview of variables

Variable name	Expected influencing behavior
Borrower characteristics	
1. Age	Older applicants are expected to have a default risk reducing effect
2. Sex	Women are more risk averse than men
3. Type of business	services and food vending is assumed to be risk reducing while trade is risk increasing

University of Ghana http://ugspace.ug.edu.gh

	effect,
4. Number of years at Residence	More years of stay in community are expected to have a decreasing default risk effect
5. Residential status	House owners have decreasing default risk
6. Marital status	No consensus could be reached
7. Dependents	No consensus could be reached
Loan characteristics	
8. Purpose of loan(sector)	services and food vending is assumed to be risk reducing while trade has risk increasing effect,
9. Amount disbursed	No consensus could be reached
10. Date of disbursement (month)	Raining season month is likely to have increasing default risk
11. Type of collateral	Good collateral lowers risk
12. Guarantor	Guarantors reduces risks of default
13. Assessment	No consensus
14. Officer	No consensus
15. Duration	No consensus
16. Loan cycle	expected decreasing effect on default as a repeat disbursement is seen as a quality stamp

Source: preliminary research findings (2012)

3.3.2 Dependent variable construction and pre-processing

The data preparation step deals with the choice and creation of the desired variables dependent and covariates.

University of Ghana http://ugspace.ug.edu.gh

In this study a binary dependent variable default (Y) was created.

$$Y_i = \begin{cases} 1 \text{ defaulted; if delayed payment is} > 30 \text{ days} \\ 0 \text{ non-default; if delayed payment} \leq 30 \text{ days} \end{cases}$$

This criterion is consistent with general definition of loan default.

Since including all variables will make the model unnecessarily large, the principle of parsimony will justify small model. The researcher employed statistical procedures such as forward and backward selection processes to verify consistency of variables selected in the model.

3.4 Model specification

The study makes use of the logistic regression model. Logistic regression is based on binomial probability theory. It is a mathematical modeling approach used in describing the relationship of several independent variables to a dichotomous dependent variable or a limited dependent variable. The logit function is employed because the dependent variable 'default' is dichotomus, whereas the proposed covariates were mixture of continuous and categorical random variables.

Thus the model was chosen over others due to the data structure and purpose. Also the independent variables need not be interval, nor normally distributed, nor linearly related, nor equal variance within each group (O'Cornell, 2005).

3.4.1. The Odds

The logit model is a derivative of the odds function. The odd of a function is the ratio of the probability of success to that of failure. Thus

$$odds(Y=1)=\frac{P(Y=1/X=x)}{P(Y=0/X=x)} \tag{1}$$

where '$odds(Y=1)$' is the odds of default

$P(Y=1/X)$ is the probability that default occurs given a set of explanatory variables.

$P(Y=0/X)$ is the probability of non-default given set of explanatory variables.

If the odds of default is greater than one it means there is a higher probability of default compared to that of non-default. A value less than one indicate a higher probability of non-default than that of default.

3.4.2 The Logistic model

Given a set of explanatory variables, $X = X_{0'}X_{1'}X_{2'}..., X_n$. The natural log of the $odds(Y=1)$ can be expressed as a linear transform.

$$\ln[odds(Y=1)] = \log it[(Y=1)] = \sum_{i=0}^{n} X_i \beta_i$$

thus

$$\ln[\frac{P(Y=1)}{1-P(Y=1)}] = b_0 + X_1 b_1 + X_2 b_2 + X_3 b_3 ..., + X_n b_n \tag{2}$$

By rearranging (2), the estimated probability of default P(Y=1) is given by

$$P(y=1)=\frac{1}{1+e^{-(b_0+X_1b_1+X_2b_2+X_3b_3...,+X_nb_n+e)}} \quad (3)$$

where $b_0, b_1, b_2, b_3..., b_n$ the coefficients are terms to be estimated and e is the error term.

Proof of equation (3)

To proof that the distribution function of the logit model is given by;

$$P(Y=1/X)=\frac{1}{1+e^{-\sum_{i=0}^{n}X_i\beta_i}}$$

$$from(2)$$

$$\ln[\frac{P(Y=1/X)}{1-P(Y=1/X)}]=\sum_{i=0}^{n}X_i\beta_i$$

$$=\frac{P(Y=1/X)}{1-P(Y=1/X)}=e^{\sum_{i=0}^{n}X_i\beta_i}$$

$$P(Y=1)=e^{\sum_{i=0}^{n}X_i\beta_i}-P(Y=1)e^{\sum_{i=0}^{n}X_i\beta_i}$$

$$P(Y=1)(1+e^{\sum_{i=0}^{n}X_i\beta_i})=e^{\sum_{i=0}^{n}X_i\beta_i}$$

$$P(Y=1)=\frac{e^{\sum_{i=0}^{n}X_i\beta_i}}{1+e^{\sum_{i=0}^{n}X_i\beta_i}}$$

$$hence$$

$$P(Y=1/X,\beta)=F(X,\beta)=\frac{1}{1+e^{-\sum_{i=0}^{n}X_i\beta_i}}$$

From above it follows that;

If $\sum_{i=0}^{n}X_i\beta_i = 0$, then P(Y=1) = 0.5

As $\sum_{i=0}^{n} X_i \beta_i \to \infty$, $P(Y=1) \to 1$

As $\sum_{i=0}^{n} X_i \beta_i \to -\infty$; $P(Y=1) \to 0$

This implies that $0 \leq P(Y=1) \leq 1$

Theorem

Given the binary response variable (default or non-default), the probability distribution of the number of defaults in a given loan portfolio size, for given values of explanatory variables is binomial. Thus the probability that the number of default of a given portfolio size **n** is exactly equal to size x is given by

$$P(X=x) = \frac{n!}{x!(n-x)!} p^x q^{n-x}$$

where q= $P(y=0)$; probability of non-default).

This means that given a portfolio size n and probability of default P(Y=1) from a financial institution, one can use the theorem to perform risk control analysis.

3.4.3 Estimating the parameters of the logit model

The parameters β_i in (3) are estimated using the maximum likelihood estimation. Maximum likelihood estimation involves finding the values of the parameters that give rise to the maximum probability of default. Given that $P(Y_i=1) = F(X_i, \beta_i)$ $P(Y_i=0) = 1 - F(X_i, \beta_i)$, where $i = 1, 2, 3, ..., n$, the likelihood function of the joint distribution is given by

University of Ghana http://ugspace.ug.edu.gh

$$\prod_{i=1}^{n}[F(X_i,\beta_i)^{yi}(1-F(X_i,\beta_i))^{1-yi}] \tag{4}$$

Then the log-likelihood function is

$$\sum_{i=1}^{n}[y_i \ln(F(X_i,\beta_i)(1-y_i)(1-F(X_i,\beta_i)]$$

We note that equation (4) can be rewritten as

$$=\prod_{i=1}^{n}[(\frac{1}{1+e^{-\sum_{i=0}^{n}X_i,\beta_i}})^{yi}(1-\frac{1}{1+e^{-\sum_{i=0}^{n}X_i,\beta_i}})^{1-yi}]$$

And the likelihood function of the joint distribution can be expressed as

$$l=\sum_{i=1}^{n}[y_i \ln(\frac{1}{1+e^{-\sum_{i=0}^{n}X_i,\beta_i}})(1-y_i)\ln(1-\frac{1}{1+e^{-\sum_{i=0}^{n}X_i,\beta_i}})] \tag{5}$$

Note that $(y_i=1)$ or $(y_i=0)$ [default or non-default respectively]

Now, given that the logit model is non-linear, and cannot be made linear by transformation, parameters cannot be estimated with Ordinary Least Squares method. Instead a numerical optimization method such as Newton –Raphson method is employed. Newton Raphson method is used to determine the root of complicated algebraic expressions.

To estimate parameter β we iterate,

$$\beta_{n+1}=\beta_n-[\frac{\partial^2 l}{\partial\beta^2}]^{-1}[\frac{\partial l}{\partial\beta}]_{\beta=\beta} \quad \text{until } |\tilde{\beta}_{n+1}-\tilde{\beta}_n| < \in = 0.1$$

where,

$$\frac{\partial^2 l}{\partial \beta^2} = \sum_{i=1}^{n} \frac{\exp-(X\beta)}{[1+\exp-(X\beta)]^2} X_i X_i$$ [second partial derivative wrt β]

$\tilde{\beta}_n$ is the n-th round estimate and the Hessian and score vectors are evaluated at this estimate.

3.4.3 Estimating the marginal effects of parameters

Given the logistic model; P(Y=1/X) = F (X, β). To obtain the required relationship between changes in say x_2 and P we consider either distribution function when the variable is continuous or the odds ratio when the variable is discrete. Usually, these impacts of incremental changes in an explanatory variable are evaluated by setting each of them to their mean values.

We note that

$$P(Y=1/X) = \frac{1}{1+e^{-\sum_{i=0}^{n} X_i \beta_i}}$$

For continuous covariates the marginal effect is evaluated by considering the density function or the derivative of the logit function.

$$\frac{\delta P}{\delta x_{ij}} = f(X_i;\beta)\beta_j$$, evaluated at $X = \overline{X}$

$$i = 1, 2, 3, \ldots, n; \ j = 1, 2, 3, \ldots, k$$

$$\frac{\delta P}{\delta x_2} = f(X_i;\beta)\beta_2$$; where $f(X_i;\beta)$ is the probability density function of the logit model.

For discrete or categorical covariates the marginal effects will be determined by examining the odds ratio which is the ratio of the odds of the variable to other variable.

University of Ghana http://ugspace.ug.edu.gh

3.5 Assessment of relevance of predictors

After estimating the coefficients, there are several steps involved in assessing the appropriateness, adequacy and usefulness of the model.

In this study, the importance of each of the explanatory variables was assessed by carrying out statistical tests of the significance of the coefficients. The overall goodness of fit of the model is then tested. Additionally, the ability of the model to discriminate between the two groups defined by the response variable is evaluated. Finally, if possible, the model is validated by checking the goodness of fit and discrimination on a different set of data from that which was used to develop the model (hold out sample test).

The significance of the coefficients will be examined using Wald test and likelihood ratio test.

3.5.1 The Wald test

Wald χ^2 statistics are used to test the significance of individual coefficients in the model and are calculated as follows:

$$Wald = \left(\frac{\beta_i}{SE(\beta_i)}\right)^2$$

Each Wald statistic is compared with a χ^2 distribution with 1 degree of freedom. Wald statistics are easy to calculate but their reliability is questionable, particularly for small samples. For data that produce large estimates of the coefficient, the standard error is often inflated, resulting in a lower Wald statistic, and therefore the explanatory variable may be incorrectly assumed to be unimportant in the model. Likelihood ratio tests are generally considered to be superior.

3.5.2 Likelihood ratio test

The likelihood ratio test for a particular parameter compares the likelihood of obtaining the data when the parameter is zero (L_0) with the likelihood (L_1) of obtaining the data evaluated at the MLE of the parameter. The test statistic is calculated as follows:

$$-2\ln(LR) = -2\ln\frac{L_0}{L_1} = -2[\ln L_0 - \ln L_1]$$

It is compared with a χ^2 distribution with 1 degree of freedom.

The variables selected by forward or backward stepwise should all have significant changes in -2log-liklihood. Thus a predicted variable is added if significant value is less than 0.05.

The change in -2log-likelihood is generally more reliable than the Wald statistic. If the two disagree as to whether the predictor is useful to the model trust the change in -2log-likelihood.

3.6 Goodness of fit of model

The goodness of fit of a model measures how well the model describes the response variable. Assessing goodness of fit involves investigating how close values predicted by the model are to the observed values. For this study, the researcher examines the

University of Ghana http://ugspace.ug.edu.gh

Hosmer–Lemeshow test as a commonly used test for assessing the goodness of fit of a model.

3.6.1 The Hosmer–Lemeshow test

The Hosmer–Lemeshow test is a commonly used test for assessing the goodness of fit of a model and allows for any number of explanatory variables, which may be continuous or categorical. The test assesses whether or not the observed event rates match expected event rates in subgroups of the model population. The Hosmer–Lemeshow test specifically identifies subgroups as the deciles of fitted risk values.

The Hosmer–Lemeshow test statistic is given by:

$$HL = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{N_i p_i (1 - p_i)}$$

Here O_i, E_i, N_i, and P_i denote the observed events, expected events, observations, predicted risk for the i^{th} risk decile group, and n is the number of groups. It can be shown that the test statistic asymptotically follows a χ^2 distribution with n-2 degrees of freedom. The number of risk groups may be adjusted depending on how many fitted risks are determined by the model. This helps to avoid singular deciles groups. The Hosmer-Lemeshow statistic indicates a poor fit if the significant value is less than 0.05.

3.6.2 Pseudo R^2

Pseudo R-square provides further statistics that may be used to measure the usefulness of the model similar to the coefficient of determination (R^2) in linear regression. The Cox &

University of Ghana http://ugspace.ug.edu.gh

Snell and the Nagelkerke R^2 are two such statistics. The Nagelkerke R^2 is an adjusted version of the Cox & Snell R^2 and covers the full range from 0 to 1, and therefore it is often preferred. The R^2 statistics do not measure the goodness of fit of the model but indicate how useful the explanatory variables are in predicting the response variable and can be referred to as measures of effect size. They are mostly useful when comparing models.

3.6.3 Reciever Operating Characteristic (ROC) curve and AUC

The ROC curve is a measure of model fit. The curve basically tells how the model trade-off between sensitivity and specificity. Any point on the line indicates how the probability of correctly predicting a 1 is traded off against the probability of correctly predicting a 0. The statistic that conveys information about the ROC is AUC. When the area under the curve (AUC) is 1, there is no trade-off between 1s and 0s and the model is predicting everything. It falls as the model becomes worse.

3.7 Model validation

For the study, the validity of the model was assessed by carrying out tests of goodness of fit and discrimination from a randomly selected data set from the original one.

The data set was divided into two 70% for the model formation and 30% for validation to examine the discriminatory power of the model

3.8 Data Processing and Analysis

The data was processed and analysed using SPSS.

University of Ghana http://ugspace.ug.edu.gh

CHAPTER FOUR

RESULTS AND DISCUSSION

4.0 Introduction

This chapter presents the analysis of the data. It starts with data description and explains the dependent and independent variable coding used for the model; its effects and relevance. It finally presents the findings in relation to the research objectives.

4.1 Data description

The study used a dataset from a microfinance company in Accra, Ghana. The company provides microfinance services to SMEs in Ghana. For the purpose of the study the microfinance company will be called 'the MFI'. The main clientele base include: Petty traders, Food vendors, Restaurant or Bar operators, Seamstress, Taylor etc.

Table 2: Gender analysis of clients

Gender	Clients	Clients per sex	OLB
F	329	69.70%	138,025.29
M	143	30.30%	183,643.17
TOTAL	**472**	**100.00%**	**321,668.46**

In all there are 472 clients within the period under study, 69.7% are females and 30.3% are males. The data also includes other variables such as Age, Sex, Marital status, Sector, residential status etc.

In all as at December 2012, the company had disbursed GHC 1,739,041.00. Out of this 79% had been paid as at end December, 2012 whilst 21% of portfolio value was in

University of Ghana http://ugspace.ug.edu.gh

arrears. Out of the 472 clients disbursed within the period about 68% performed well whilst 32% did not perform well.

4.2 Independent Variable coding

The codes for the levels of the variables used in developing the model have been presented in Table 3.

Table 3: Definition of study variables

No.	CATEGORICAL VARIABLE	CODE
1	OFFICER	1-A
		2-B
		3-C
		4-D
		5-E
		6-F
2	SEX	1 – Male
		2 –Female
3	SECTOR	1-Food industry
		2-Manufacturing
		3-Trade industry
		4-Service
4	TYPE OF RESIDENCE	1-Rented
		2-Own house
		3-Family house
5	DISTANCE(15km)	1-Yes
		2-No
6	MARRIAGE	1-Married
		2-Single
		3-Divorced
		4-Widow/Widower
7	AGE	Age of clients
8	GUARANTOR	1-Yes
		2-No
9	ASSESSMENT	1-Yes
		2-No
10	COLLATERAL	1-Stock
		2-Household items

University of Ghana http://ugspace.ug.edu.gh

		3-Car/Vehicle
		4-Cheque
11	MONTH	1-1st Quarter
		2-2nd Quarter
		3-3rd Quarter
		4-4th Quarter
12	AMOUNT	Amount disbursed
13	DURATION	Number of months
14	RESIDENCE	Years in residence
15	CYCLE	Number of loan cycle
16	DEPENDENTS	Number of dependents

University of Ghana http://ugspace.ug.edu.gh

4.3 Dependent variable coding

The dependent variable is coded 1 if client defaulted and 0 if the client did not default. The criteria for default have been clarified in the methodology section.

4.4 Descriptive statistics of variables used in the model

The data consist of sixteen variables including demographic and loan characteristics. These include Age, Sex, Marital status, Sector, Residential status, Cycle, Guarantor, number of dependents, years in community, amount disbursed, type of collateral, distance from office to clients location, duration and loan officer. The descriptive statistics for the scale variables have been presented in Table 4 below.

Table 4: Descriptive statistics of scale variables

	N	Minimum	Maximum	Mean	Std. Deviation	Variance
Age	472	22	68	39.24	7.96	63.29
Amount disbursed	472	200	10000	1672.70	1432.87	2.05
Cycle	472	1	5	1.71	0.89	0.79
Number of Dependents	472	0	10	2.16	1.30	1.69
Duration	472	1	6	5.29	1.10	1.20
Years in community	472	0.5	40	6.44	5.68	32.22

Source: research findings (2013)

University of Ghana http://ugspace.ug.edu.gh

From Table 4 the average age of customers was 39years. The minimum age is 22 and the maximum age is 68.

The minimum amount disbursed according to the data is 200 whilst the maximum is 10,000 cedi. The average disbursed amount is 1,432.00.

The average number of dependents is 2, whilst the highest is 10. The minimum year in residence is 6 months whilst the maximum is 40 years. The average year in residence is 6 years.

Also the average month for a loan to mature is 5 months. The minimum loan duration is 1 month whilst the maximum is 6 months.

4.5 Explanation of data processing

There are sixteen (16) independent variables with 472 applicants previously given loans. Out of the 472 clients 332 (70%) was randomly selected for building the model, whilst 140 (30%) was used to validate the model. The dependent variable (default) was created by the researcher. Thus it was categorised 1 when default occurs and 0 if default did not as shown in Table 5 below.

Table 5: Dependent variable coding

Original Value	Internal Value
Non-default	0
Default	1

Source: research findings (2013)

University of Ghana http://ugspace.ug.edu.gh

Table 6 below presents a summary of categorical variable codings. For each variable the last attribute is the reference variable. Thus it is use as a reference to determine the importance of the independent variables relative to the reference variable. Thus for each loan officer, the statistics generated will be analyzed relative to loan officer designated F.

Table 6 : Categorical variable coding

Variable	**Description**	**Frequency**	**Parameter coding**				
			(1)	**(2)**	**(3)**	**(4)**	**(5)**
Officer	A	100	1.000	.000	.000	.000	.00
	B	122	.000	1.000	.000	.000	.00
	C	62	.000	.000	1.000	.000	.00
	D	49	.000	.000	.000	1.000	.00
	E	84	.000	.000	.000	.000	1.00
	F	55	.000	.000	.000	.000	.00
Amount disbursed (grp)	Below 2000	332	1.000	.000	.000	.000	
	2000-2900	98	.000	1.000	.000	.000	
	3000-3900	10	.000	.000	1.000	.000	
	4000-4900	17	.000	.000	.000	1.000	
	Above 5000	15	.000	.000	.000	.000	
Type of collateral	Stock	198	1.000	.000	.000		
	Household items	167	.000	1.000	.000		
	Car	66	.000	.000	1.000		
	Cheque	41	.000	.000	.000		
Sector	Food industry	83	1.000	.000	.000		
	Manufacturing	27	.000	1.000	.000		

University of Ghana http://ugspace.ug.edu.gh

	Trade industry	206	.000	.000	1.000
	Service industry	156	.000	.000	.000
Marital status	Married	352	1.000	.000	.000
	Single	103	.000	1.000	.000
	Divorced	6	.000	.000	1.000
	Widow/widower	11	.000	.000	.000
Quarter disbursed	1st quarter	100	1.000	.000	.000
	2nd quarter	170	.000	1.000	.000
	3rd quarter	126	.000	.000	1.000
	4th quarter	76	.000	.000	.000
Age grp	Below 30yrs	47	1.000	.000	.000
	30-39yrs	248	.000	1.000	.000
	40-49yrs	135	.000	.000	1.000
	50yrs +	42	.000	.000	.000
Type of residence	Rented	316	1.000	.000	
	Own house	117	.000	1.000	
	Family House	39	.000	.000	
Sex	Male	209	1.000		
	Female	263	.000		
Assessment	Yes	412	1.000		
	No	60	.000		
Guarantor	Yes	135	1.000		
	No	337	.000		
Distance within 15km	Yes	431	1.000		
	No	41	.000		

Source: research data (2013)

University of Ghana http://ugspace.ug.edu.gh

4.6 Presentation of findings

The objective of this research was to iidentify risk factors that influence loan default among customers in the Microfinance sector and finally to come out with a function that links the risk factors to credit default for a customer. To this end, several model considerations were made and based on the data considerations a logit model was considered. First the researcher presents a case where no attributes of the borrower was considered and explored the case where the covariates were considered.

4.6.1 The baseline model

Table 7, presents the baseline default risk model, thus, the model with only the constant included before any coefficients were added or entered into the equation. The researcher compares this model with a model including all the predictors to determine whether the latter model is more appropriate. In the table 'Selected Cases' were the number of data sets selected for the model whilst 'Unselected Cases' were those selected for validation. The information in the table suggests that if one knew nothing about the client requesting for a loan and guessed that a person will default, one will be correct 65.4% of the time based on the used cases, and 73.6% based on the out of sample files. This means that the baseline model has some predictive power but only 69.5% on average.

University of Ghana http://ugspace.ug.edu.gh

Table 7: Classification Table for the baseline model

Observed			Predicted: Selected Cases: Defaulted: Non-default	Selected Cases: Defaulted: Default	Selected Cases: Percentage Correct	Unselected Cases: Defaulted: Non-default	Unselected Cases: Defaulted: Default	Unselected Cases: Percentage Correct
Step 0	Defaulted	Non-default	217	0	100.0	103	0	100.0
		Default	115	0	.0	37	0	.0
	Overall Percentage				65.4			73.6

Source: research findings (2013)

4.7 The model with explanatory variables

The researcher then includes explanatory variables into the model. The researcher uses the likelihood ratio method (Forward stepwise). With this method, the predictor with the largest score statistic whose significance value is less than a specified value (by default 0.05) is added to the model. The variables left out of the analysis at the last step all have significance values larger than 0.05, so no more are added. The variables chosen by the forward stepwise method should all have significant changes in -2 log-likelihood.

4.7 .1 Parameter estimates

The estimates in Table 8 summarize the model coefficients and the effect of each predictor variable. In table 8 below, column 'B' represents the coefficient of the predictor and SE is the standard error. The ratio of the coefficient to its standard error, squared, equals the Wald statistic. Exp (B) represents the ratio-change in the odds of default for a one-unit change in the predictor or the odds ratio.

Table 8: Model parameter estimates

	Variable	B	S.E.	Wald	Df	Sig.	Exp(B)
Step 1	collatype1			61.324	3	.000	
	collatype1(1)	19.758	8038.602	.000	1	.998	3.810E8
	collatype1(2)	21.803	8038.602	.000	1	.998	2.944E9
	collatype1(3)	19.569	8038.602	.000	1	.998	3.152E8
	Constant	-21.203	8038.602	.000	1	.998	.000
Step 2	Officer			35.045	5	.000	
	officer(1)	2.701	.669	16.322	1	.000	14.894
	officer(2)	.820	.640	1.642	1	.200	2.270
	officer(3)	1.240	.714	3.014	1	.083	3.457
	officer(4)	1.152	.788	2.139	1	.144	3.164
	officer(5)	2.786	.688	16.383	1	.000	16.215
	collatype1			60.090	3	.000	
	collatype1(1)	19.596	7586.761	.000	1	.998	3.239E8
	collatype1(2)	21.967	7586.761	.000	1	.998	3.469E9
	collatype1(3)	18.911	7586.761	.000	1	.998	1.633E8
	Constant	-22.813	7586.761	.000	1	.998	.000
Step 3	Officer			44.468	5	.000	
	officer(1)	3.984	.773	26.576	1	.000	53.710
	officer(2)	.803	.638	1.584	1	.208	2.231
	officer(3)	1.244	.712	3.050	1	.081	3.470
	officer(4)	1.797	.793	5.138	1	.023	6.033
	officer(5)	3.532	.738	22.939	1	.000	34.204
	quarantor1(1)	-2.571	.492	27.313	1	.000	.076
	collatype1			43.143	3	.000	
	collatype1(1)	19.341	6885.880	.000	1	.998	2.511E8
	collatype1(2)	21.407	6885.880	.000	1	.998	1.981E9
	collatype1(3)	18.217	6885.880	.000	1	.998	8.155E7
	Constant	-22.213	6885.880	.000	1	.997	.000
Step 4	Officer			42.640	5	.000	
	officer(1)	4.041	.799	25.565	1	.000	56.860

	officer(2)	.987	.662	2.226	1	.136	2.684
	officer(3)	1.624	.745	4.753	1	.029	5.072
	officer(4)	1.683	.819	4.222	1	.040	5.384
	officer(5)	3.795	.772	24.176	1	.000	44.462
	quarantor1(1)	-2.430	.495	24.065	1	.000	.088
	collatype1			39.807	3	.000	
	collatype1(1)	19.015	6911.584	.000	1	.998	1.812E8
	collatype1(2)	21.101	6911.584	.000	1	.998	1.459E9
	collatype1(3)	18.050	6911.584	.000	1	.998	6.902E7
	agw1	.078	.022	12.336	1	.000	1.081
	Constant	-25.206	6911.584	.000	1	.997	.000
	Officer			43.370	5	.000	
Step 5	officer(1)	4.504	.855	27.758	1	.000	90.415
	officer(2)	1.146	.682	2.822	1	.093	3.144
	officer(3)	1.765	.764	5.344	1	.021	5.842
	officer(4)	1.838	.830	4.905	1	.027	6.284
	officer(5)	4.131	.820	25.388	1	.000	62.249
	RES			9.506	2	.009	
	RES(1)	1.028	.752	1.871	1	.171	2.796
	RES(2)	-.463	.822	.318	1	.573	.629
	quarantor1(1)	-1.712	.524	10.676	1	.001	.181
	collatype1			39.822	3	.000	
	collatype1(1)	19.363	7068.589	.000	1	.998	2.567E8
	collatype1(2)	21.476	7068.589	.000	1	.998	2.123E9
	collatype1(3)	18.093	7068.589	.000	1	.998	7.206E7
	agw1	.082	.023	12.866	1	.000	1.085
	Constant	-26.665	7068.589	.000	1	.997	.000
Step 6	Officer			41.569	5	.000	
	officer(1)	4.511	.877	26.446	1	.000	90.996
	officer(2)	1.081	.673	2.574	1	.109	2.947
	officer(3)	1.617	.758	4.556	1	.033	5.038
	officer(4)	1.606	.831	3.734	1	.053	4.981
	officer(5)	3.957	.815	23.575	1	.000	52.303
	RES			7.755	2	.021	

University of Ghana http://ugspace.ug.edu.gh

RES(1)	.988	.789	1.568	1	.210	2.686
RES(2)	-.363	.865	.176	1	.675	.696
quarantor1(1)	-2.212	.615	12.911	1	.000	.110
Asse1(1)	-1.761	.891	3.905	1	.048	.172
collatype1			33.147	3	.000	
collatype1(1)	19.570	7307.668	.000	1	.998	3.155E8
collatype1(2)	21.514	7307.668	.000	1	.998	2.205E9
collatype1(3)	18.313	7307.668	.000	1	.998	8.975E7
agw1	.084	.023	12.990	1	.000	1.088
Constant	-25.028	7307.668	.000	1	.997	.000

Source: research findings (2013)

The predictors and coefficient values shown in the last step (Step 6) are considered. This is because the procedure used six steps to arrive at significant predictive variables. The significant variables include Loan officer, Residential status, Guarantor, Type of collateral, Assessment and Age, whilst the non-significant variables include; amount disbursed, sex, years in residence, cycle, term of loan, number of dependents, distance from office to clients location, marital status, sector and date of disbursement. The above results have presented below in Table 9.

Table 9: Predictive variable classification

Significant Predictive variables	Non-significant variables
Age of client	Amount disbursed
Assessment	Sex
Guarantor	Years in residence
Type of collateral	Cycle
Loan officer in charge	Term of loan
Type of residence	Distance
	No. of dependents
	Marital status
	Sector
	Date of disbursement

Source: research findings (2013)

University of Ghana http://ugspace.ug.edu.gh

The Table 10 below presents the default rate for the selected variables commonly examined in default analysis.

Table 10: Default rate classification

Variable	Level	Frequency Non-Default	Default	Total	Default rate
SECTOR	Food industry	72	11	83	13%
	Manufacturing	19	8	27	30%
	Trade industry	109	97	206	47%
	Service industry	120	36	156	23%
RESIDENCE	Rented	197	119	316	38%
	Own house	91	26	117	22%
	Family house	32	7	39	18%
AMOUNT DISBURSED	Below 2000	224	108	332	33%
	2000-2900	73	25	98	26%
	3000-3900	8	2	10	20%
	4000-4900	10	7	17	41%
	5000+	5	10	15	67%
AGE	<30 Years	43	4	47	9%
	30-39 Years	179	69	248	28%
	40-49 Years	78	57	135	42%
	50+ Years	20	22	42	52%
MARITAL STATUS	Married	232	120	352	34%
	Single	80	23	103	22%
	Divorced	1	5	6	83%
	Widow/Widower	7	4	11	36%
SEX	Male	152	57	209	27%
	Female	168	95	263	36%
COLLATERAL	Stock	162	36	198	18%
	Household Items	65	102	167	61%
	Vehicle	54	12	66	18%
	Cheque	39	2	41	5%
GUARANTOR	Yes	119	16	135	12%
	No	201	136	337	40%

Source: research findings (2013)

The findings in Table 10 above revealed that about 13% default rate for those in the food sector, 30% for those in the manufacturing sector, 47% for those in the trade sector and

University of Ghana http://ugspace.ug.edu.gh

23% for those in the service sector. These results suggest that, those in the food vending are less risky, than the rest of the sectors. This is followed by those in the service sector, the manufacturing and trade sector. Thus according to the data, customers in the trade sector are highly risky. They have the highest default rate. However, the model did not found the sector variable as a significant predictor variable for default.

4.8 Discussion of results

4.8.1 Age of client

The variable age represents the age of client at the time of applying for the loan. During the period in question, the minimum age of applicant was 22 years and the maximum age was 68 years. The average age recorded was 39 years. For analytical purposes, the age was categorized and the default rate for the various categories was computed. The result is present in Table 10 above. The result indicates that those who are less than 30 years have lower rate of default (9%), whilst those above 50 years recorded higher rate of default. This is probability because younger once are more aggressive and able to turn around the loan in a more productive way to pay back the loan other than the older once.

In Table 8, the parameter estimate for age is 0.084 with an odds ratio of 1.09. This indicate that the all things being equal, the odds of default for a 41 year old customer is 1.09 times the odds of default for a customer who is 40 years old. It means a unit change in age affects the probability of default by 9% positively. Unlike, Thomas (2000) and Boyle et al. (1992) who found that older borrowers are more risk adverse, and therefore less likely to default, in this study older client are more likely to default.

University of Ghana http://ugspace.ug.edu.gh

4.8.2 Type of residence

It was observed that the kind of apartment a customer lives affects the rate of default. Thus clients who live in rented apartment increases the risk of default whilst those in own house reduces the rate of default. From the results, clients who live in rented apartment are 2.68 (odds ratio) times more likely to default than those in family house. However those who live in their own house are 30% less likely to default than those in family house. The study showed 38% default rate for individuals living in rented houses, 22% for own house and 18% for those living in family houses. This shows that default risk for those in family house is lower than that living in own house. The trend is consistent with what other researchers have come out with except that in this case those living in family house performed better than those living in own house. Similarly, the residential variable measures whether borrowers own their home, rent, or live with their parents. According to Mario (2010), the type of residence indicates the borrowers' financial wealth in the case of home ownership or financial pressure on borrowers' income, for example rental cost.

4.8.3 Type of collateral

It was equally observed that the type of collateral used affects the chances of default. According to the findings, clients who use household items as collateral were more likely to default than those who use stock of goods and those who use vehicles as collateral. Borrowers' collateral can be a yardstick for default performance assessment for a client. For instance, the loans that the house serves as collateral will have a low probability of default. This is probably because borrowers are risk adverse and fear of losing their

University of Ghana http://ugspace.ug.edu.gh

house. Collateral reduces the bank's risk when it makes a loan. The higher the collateral value the higher the incentive for the borrowers to repay the loan since they do not want to lose their collateral. The collateral value could also be a proxy for the borrowers' financial wealth since it is significantly positive correlated with the borrowers' income (Dinh and Kleimeier, 2007). It was noticed in the study that, the variable contributes significantly to the model. In terms of default rate, the findings suggest that, those who used the stock of goods in shop or store as collateral have a default rate of 18.2%, those with household items have a default rate of 61.2%, those who used vehicles were also 18.2% and those who guarantee with cheques usually have a default rate of 4.9%. This means that those who guarantee with household items have the highest rate of default.

4.8.4 Loan officer

The loan officer variable was used as proxy to measure the performance of loan officers and the extent of influence on default risk. In the researchers view, the ability of loan officers to assess clients affects the performance of loan. This is absolutely necessary because the knowledge and attitude of credit officers to very large extent affect the performance of the client. The position is consistent with what Lewis (1992) indicated that any characteristic and environments of the borrower that has obvious connections with default risk should be used in credit analysis. From Table 8, the odds ratio for officer (1), officer (2), officer (3), officer(4) and officer(5) were 90.996, 2.947, 5.038, 4.981, and 52.303 respectively. There were six officers but that officer was used as a reference value for comparative analytical purposes. For instance officer (2) is 194.7% more likely to default than officer six, whilst officer (3) is 403.8% more likely to default than officer six all things being equal. This means that that officer (2) is relatively better

University of Ghana http://ugspace.ug.edu.gh

than officer (3). The parameter estimates were all positive indicating that, there is direct relationship between probability of default and the officers.

4.8.5 Assessment

From the findings a client duly assessed by verifying residence and business reduces the rate of default by at least 82.8%. This means that a customer who is assessed is 84.2% less likely to default than if not assessed. In other words default rate is higher when customers were not assessed.

4.8.6 Personal guarantor

A guarantor assures and usually agrees to be responsible for another person's liability. It is usually requested when a customer does not usually provide adequate security against the loan. The findings show that guaranteed loans are less likely to default than those without guaranteed by about 89%. It can also be confirmed in Table 9, that the default rate for guaranteed loans were 11.9% and the rate for unguaranteed was 40.4%.

4.8.7 Years in residence

Crook et al.'s (1992) also pointed out in his study that, default risk drops with an increase in time at present address; it might be a proxy for the borrowers' maturity, stability, or risk aversion. Changing address might be a signal that a borrower's financial wealth is high or improving rapidly. Time addresses the number of years that the borrowers have been living at their current address. However, this study did not find out years in residence as being significant to the model. This may be due to the nature of the micro credit system. Most of these clients are petty traders or engaged in small scale businesses. A client might have stayed in a particular locality for a while, but as soon as a facility is granted, the person may run to home town and might never be traced. The study found this variable less relevant. This may be because the customers themselves are not able to provide accurate data on such variables as they will always try to impress.

4.8.8 Marital status

Dinh and Kleimeir (2007) intimated that, marital status affects the borrower's level of responsibility, reliability, or maturity. The probability of default is higher for married than single borrowers. They discover that the marital status is typically related to number of dependents which in turn reflects financial pressure on the borrower and borrower's ability to repay a loan. This study relates very well with the findings made in the study. The study showed that, default risk for married clients is 34%, not married before (single) is 22%, divorced is 83% whilst widow or widower had a default rate of 36%. Nonetheless, the variable was not significant to the model, hence excluded from the model.

4.8.9 Duration

Loan duration or term of loan was equally considered as an important factor. Loan duration reflects the borrowers' intention, risk aversion, or self-assessment of repayment ability. It can be assumed that the longer the loans duration, the lower the probability of default. The clientele base and the nature of business of microfinance institution is usually seasonal. For this reason, most of the client's term of payment is usually limited to 6 months. The average payment term according to the study is 5 months. The study noted a marginal contribution of 7% to the model default risk. It also found to be inversely related to default risk. What it means is that the higher the loan terms the less risky the loan. This is probably because it makes them less stressful and able to pay it in

University of Ghana http://ugspace.ug.edu.gh

bit. This finding is consistent with the findings of Steenackers and Goovaerts. The procedure in Table 8 did not find the variable too predictive for default.

4.9.0 Loan Cycle

Number of loans the clients have borrowed over time can be used as measure of relationship with the bank. Of late, most borrowers have more than one loan from the same bank or other banks. This variable reflects the difficulty for a defaulted borrower to receive further loans from the same bank. In this study the variable was not found to be significant to the model. This may be because business circumstances change overtime so the fact that the person performed well does not necessarily mean that it will perform the next time. Hence, its impact was essentially insignificant to the model. It can be concluded that the impact is ambiguous.

4.9.1 Loan size

The loan size is the amount of loan being requested. Sometimes the same amount is granted or reduced depending on the circumstances of the client. Several studies use loan size as a predictor variable but the overall results are ambiguous and thus no clear expectations can be formed. The study did not find it useful either. Nonetheless, from Table 9, the highest rate of default for that category was those granted above Gh5000 (67%), the lowest was recorded for those who were granted between Gh3000-3900 (20%). Jacobson and Roszbach (2003) show that loan size has no significant influence on default risk. However, in the study of Kocenda and Vojtek (2009) small loans appear to be more risky.

4.9.2 Sector

With regard to sectors, the MFI categorized customers into four namely; Food sector, Manufacturing, Trade sector, Service sector. Customers in the food sector basically sell

University of Ghana http://ugspace.ug.edu.gh

foodstuffs and cooked food. These include chop bar operators, koko sellers, waakye, rice, vegetables, fruits, butchers, bakers etc.

The trade sector basically involves general trading; provisions, stores with all kinds of items, hawkers etc.

The manufacturing sector includes those in creative arts such as sculptures, shoe makers etc.

The service sector comprise of seamstress and tailors, carpenters, masons, drivers etc. and formal sector workers.

4.10 The model formation

After determining the factors that affect default risk, the researcher also intended to come up with a function that can adequately describe a relationship between the risk factors and the probability of default.

The Table 11 below summarizes the model variables, their estimates and odds ratio.

Table 11: Summary of model variables

Variable	Description	Model variable	B: (Coefficients)	Exp(B): Odds ratio
Constant	Constant	X_o	-25.03	0.00
Officer(1)-A	1 if client was handled by A and 0 otherwise	X_1	4.51	90.99
Officer(2)-B	1 if client was handled by B and 0 otherwise	X_2	1.08	2.95
Officer(3)-C	1 if client was handled by C and 0 otherwise	X_3	1.62	5.04
Officer(4)-D	1 if client was handled by D and 0 otherwise	X_4	1.61	5.00

Officer(5)-E	1 if client was handled by E and 0 otherwise	X_5	3.96	52.30
RES(1)	1 if client lives in rent and 0 otherwise	X_6	1.00	2.69
RES(2)	1 if client lives in own house and 0 otherwise	X_7	-0.36	0.70
Guarantor	1 if client has a guarantor and 0 otherwise	X_8	-2.21	0.11
Assessment	1 if client was assessed and 0 otherwise	X_9	-1.76	0.17
Collatype(1)	1 if client used stock and 0 otherwise	X_{10}	19.57	3.16E8
Collatype(2)	1 if client used Household items and 0 otherwise	X_{11}	21.51	2.21E9
Collatype(3)	1 if client used Vehicle and 0 otherwise	X_{12}	18.31	8.98E7
Age	Age of client	X_{13}	0.08	1.09

Source: research findings (2013)

After careful test and retest of various models and structural forms the model below is presented;

$$P(Y=1)=\frac{1}{1+e^{-(-25.03+4.51X_1+1.08X_2+1.62X_3+1.61X_4+3.96X_5+0.99X_6-0.36X_7-2.21X_8-1.76X_9+19.57X_{10}+21.51X_{11}+18.31X_{12}+0.08X_{13})}}$$

Model interpretation

The 'sign'(- or +) associated with the coefficient indicates the direction of relationship of the variable to the dependent variable.

The value of the odds ratio also gives the direction of relationship to probability of default. If the value is less than 1 the relationship is negative and positive otherwise. For categorical variables the odds ratio indicates the strength of the variable to default probability relative to a reference variable. For instance the odds ratio for officer B is 2.95. What it means is that, a client being supervised by officer B is 2.95 times more likely to default compared with officer F. The reference variable is F, since from the data officer F has the lowest default rate.

University of Ghana http://ugspace.ug.edu.gh

From the figures shown in the table 10 it could be seen that loan officer B is relatively better in terms of client handling, followed by D and C. The model suggests that the role of the loan officer is important and influences default or the performance of the customers.

Similarly the odds ratio for Age is 0.08. This is a continuous variable. The positive sign suggests that age is directly related to risk of default.

4.11 Model diagnostics

According to, Field (2000) 'running a logistic regression without checking how well the model fits the data is like buying a new pair of trousers without trying them on'.

The researcher therefore examines how good the model fit the data using the Hosmer Lame show's test, ROC curve and the classification table.

Firstly the researcher looks at the proportion of risk of default that could be explained by the covariates (see Tables 12 and 13 and figure 2).

Table 12: Model summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	336.445[a]	.242	.334
2	292.899[a]	.335	.462
3	257.291[a]	.403	.556
4	244.076[a]	.426	.588
5	233.597[a]	.444	.612
6	229.335[a]	.451	.622

Source: research findings (2013)

University of Ghana http://ugspace.ug.edu.gh

Consider the value of Nagelkerke R square statistic (0.622) in the last column of the last row. The statistic suggests that about 62% of the variability or risk of default could be explained by only six (6) covariates.

Table 13: Hosmer and Lemeshow test

Step	Chi-square	Df	Sig.
1	.000	2	1.000
2	16.423	8	.037
3	31.715	8	.000
4	3.735	8	.880
5	17.886	8	.022
6	13.563	8	.094

Source: research findings (2013)

The Hosmer Lameshow test also tests whether or not the model is an adequate fit to the data. The null hypothesis is that the model is a good fit to the data and will only reject this if there is sufficiently strong grounds to do so; conventionally if p- value <0.05. From Table 13, step 6, the p-value is 0.094. The researcher refused to reject the null hypothesis. Hence the model is a good fit to the data.

To validate the above results, ROC and AUC were also examined as shown in figure 2 and Table 14 respectively.

University of Ghana http://ugspace.ug.edu.gh

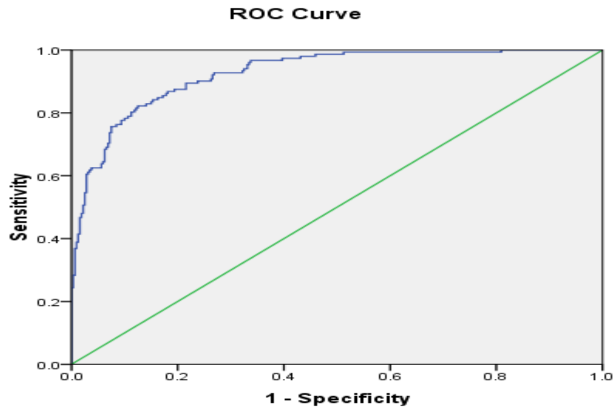


Figure 2: Reciever Operating Characteristic (ROC) Curve

The ROC measures both the sensitivity and specificity pairs for possible cut off points. The curve shows the 45 degree line and the ROC curve. The further the ROC is away from the 45 degree line the better the model predicts both 1s and 0s.

Similarly, the area under the curve as shown in Table 14 is 0.926. According to the findings the Area under Curve is within (0.902, 0.95) with 95% level of confidence. From the Table, since P-value (0.00) is significant (less than 0.05), it means that the model classifies applicants default or non-default better than by chance.

Table 14: Area under curve (AUC)

Area	Std. Error	Asymptotic Sig.	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.926	.012	.000	.902	.950

More importantly, how the model is able to classify defaulter's non and-defaulters or how the model is able to isolate defaulters and replicate non-defaulters have been displayed in Table 15 below. A good model should be able classify a higher proportion of the two

University of Ghana http://ugspace.ug.edu.gh

categories. Table 15 below summarizes how the customers were classified statistically based on the model. It classifies the customers using the in-sample data indicated as 'Selected Cases' and the hold out sample data indicated as 'Unselected cases'.

Table 15: Model's Classification Table

			Predicted					
			Selected Cases			Unselected Cases		
			Defaulted		Percentage	Defaulted		Percentage
	Observed		Non-default	Default	Correct	Non-default	Default	Correct
Step 1	Defaulted	Non-default	172	45	79.3	83	20	80.6
		Default	33	82	71.3	17	20	54.1
	Overall Percentage				76.5			73.6
Step 2	Defaulted	Non-default	204	13	94.0	98	5	95.1
		Default	54	61	53.0	22	15	40.5
	Overall Percentage				79.8			80.7
Step 3	Defaulted	Non-default	204	13	94.0	96	7	93.2
		Default	46	69	60.0	16	21	56.8
	Overall Percentage				82.2			83.6
Step 4	Defaulted	Non-default	197	20	90.8	92	11	89.3
		Default	35	80	69.6	15	22	59.5
	Overall Percentage				83.4			81.4
Step 5	Defaulted	Non-default	198	19	91.2	94	9	91.3
		Default	32	83	72.2	17	20	54.1
	Overall Percentage				84.6			81.4
Step 6	Defaulted	Non-default	199	18	91.7	93	10	90.3
		Default	32	83	72.2	13	24	64.9
	Overall Percentage				84.9			83.6

Source: research findings (2013)

University of Ghana http://ugspace.ug.edu.gh

The classification table in Table 15 suggests that the variables were entered stepwise, and the classification results indicated at each stage shows an improvement in classification as additional variables were added. At the sixth step no more variables were added because the remaining variables have significant values larger than 0.05. Thus the variables chosen by forward stepwise approach should have significant changes in -2 log-likelihood.

The information in the last step (Step 6) shows the practical results of using the logistic regression model. Cells on the diagonal are correct predictions and cells off the diagonal are incorrect predictions. Table 15 suggests that 199 out of the 217 (91.7%) applicants who did not defaulted are classified correctly. Also 83 out of the 115 (72.2%) defaulters are classified correctly. Overall, 84.9% of the applicants were classified correctly. Subset classification was obtained by classifying customers who were not used to create the model. These results are shown in the 'Unselected Cases' section of the table. Thus 83.6% of these cases were correctly classified by the model. This suggests that, overall; the model is in fact correct about four out of five times.

4.12 Marginal effects of covariates

One other objective of the study is to examine the dynamics of the factors that affect default. To achieve this objective, the researcher examines the marginal effects of the variables and its relationship with the risk of default.

With the marginal effects, the researcher examines the cumulative distribution function and the probability density function of the model.

$$F(X,\beta) = \frac{1}{1+\exp-\sum_{i=0}^{13}(X_i,\beta)} \text{ [distribution function]} \quad (6)$$

where $\beta_0, \beta_1, \beta_2, ..., \beta_{13}$

$$f(X,\beta) = \frac{\exp-\sum_{i=0}^{13}(X_i,\beta)}{[1+\exp-\sum_{i=0}^{13}(X_i,\beta)]^2} \text{ [Probability density function]} \quad (7)$$

Thus the marginal effect for the variable X can be expressed as

Marginal Effect = $f(X,\beta)\beta$ (8)

evaluated at $X = \bar{X}$;

Now, from the data $f(X,\beta)$ and $F(X,\beta)$ can be expressed as follows;

$$P(Y=1) = F(X,\beta) = \frac{1}{1+e^{-(-25.03+4.51X_1+1.08X_2+1.62X_3+1.61X_4+3.96X_5+0.99X_6-0.36X_7-2.21X_8-1.76X_9+19.57X_{10}+21.51X_{11}+18.31X_{12}+0.08X_{13})}} \text{..(9)}$$

$$f(X,\beta) = \frac{e^{-(-25.03+4.51X_1+1.08X_2+1.62X_3+1.61X_4+3.96X_5+0.99X_6-0.36X_7-2.21X_8-1.76X_9+19.57X_{10}+21.51X_{11}+18.31X_{12}+0.08X_{13})}}{[1+e^{-(-25.03+4.51X_1+1.08X_2+1.62X_3+1.61X_4+3.96X_5+0.99X_6-0.36X_7-2.21X_8-1.76X_9+19.57X_{10}+21.51X_{11}+18.31X_{12}+0.08X_{13})}]^2}$$

Hence to find the marginal effect for age of client (X_{13}) or how age alone could influence default rate, we have marginal effect = $f(X,\beta)\beta_{13}$

where $X_{1,} ..., X_{13}$ and **β** have been described in Table 16 below

University of Ghana http://ugspace.ug.edu.gh

Table 16: Parameter estimates summary

Variable	Description	Model variable	B: (Coefficients)	Exp(B): Odds ratio
Constant	Constant	X_o	-25.03	0.00
Officer(1)-A	1 if client was handled by A and 0 otherwise	X_1	4.51	90.99
Officer(2)-B	1 if client was handled by B and 0 otherwise	X_2	1.08	2.95
Officer(3)-C	1 if client was handled by C and 0 otherwise	X_3	1.62	5.04
Officer(4)-D	1 if client was handled by D and 0 otherwise	X_4	1.61	5.00
Officer(5)-E	1 if client was handled by E and 0 otherwise	X_5	3.96	52.30
RES(1)	1 if client lives in rent and 0 otherwise	X_6	1.00	2.69
RES(2)	1 if client lives in own house and 0 otherwise	X_7	-0.36	0.70
Guarantor	1 if client has a guarantor and 0 otherwise	X_8	-2.21	0.11
Assessment	1 if client was assessed and 0 otherwise	X_9	-1.76	0.17
Collatype(1)	1 if client used stock and 0 otherwise	X_{10}	19.57	3.16E8
Collatype(2)	1 if client used Household items and 0 otherwise	X_{11}	21.51	2.21E9
Collatype(3)	1 if client used Vehicle and 0 otherwise	X_{12}	18.31	8.98E7
Age	Age of client	X_{13}	0.08	1.09

4.13 Model Validation

About 140 customers that were not part of the model data was used to examine the validity of the model. The summary has been presented in Tables 17 and 18. Table 17 presents the summary of the predicted results using the model formulated in equation (9) against what was actually observed.

University of Ghana http://ugspace.ug.edu.gh

Table 17: Predicted group versus Observed group

Validate			Observed group: Non-default	Observed group: Default	Total
0	Predicted group	Non-default	93	13	106
		Default	10	24	34
	Total		103	37	140
1	Predicted group	Non-default	199	32	231
		Default	18	83	101
	Total		217	115	332

The findings in Table 17 indicate that the model is able to predict Non-defaulters correctly 93 out of 106 (87.74%) and defaulters correctly 24 out of 34 (70.59%). Overall the model could predict correctly 85% of the 140 out of sample customers.

4.14 Practical Illustration of results

To demonstrate the practical usefulness of the model a random sample of 40 customers were randomly drawn from the hold out sample and probability of default computed. The results were then compared with what was actually observed (after the fact). In Table 18 below, the five predictive variables have been indicated from columns 2-7. Column 9 is what was actually observed from the customer data, column 8 indicate the predicted probability using the model and the last column indicate the predicted group based on the predicted probability. If the predicted probability is greater than 50% it is classified as default, if it is less than 50% it is classified as non-default. If it is exactly 50% it is indecision.

Table 18: Real data validation illustration

	Officer	Residence	Age	Guarantor	Assessment	Collateral	Predicted Prob.	Observed	Predicted group
1	B	Own house	49	No	Yes	Household items	0.28	Non-default	Non-default
2	B	Rented	40	No	Yes	Stock	0.10	Non-default	Non-default
3	B	Own house	36	No	Yes	Household items	0.17	Non-default	Non-default
4	C	Rented	35	Yes	Yes	Car	0.01	Non-default	Non-default
5	C	Rented	45	No	Yes	Cheque	0.00	Non-default	Non-default
6	C	Rented	32	Yes	No	Cheque	0.00	Non-default	Non-default
7	C	Rented	39	No	Yes	Cheque	0.00	Non-default	Non-default
8	C	Rented	26	No	Yes	Stock	0.12	Non-default	Non-default
9	C	Rented	38	No	No	Cheque	0.00	Non-default	Non-default
10	B	Rented	42	No	Yes	Household items	0.54	Non-default	Default
11	B	Rented	36	No	No	Household items	0.48	Non-default	Non-default
12	C	Own house	32	Yes	No	Cheque	0.00	Non-default	Non-default
13	C	Own house	37	Yes	No	Cheque	0.00	Non-default	Non-default
14	C	Rented	32	No	Yes	Cheque	0.00	Non-default	Non-default
15	C	Own house	22	No	Yes	Stock	0.02	Non-default	Non-default
16	C	Rented	33	Yes	No	Cheque	0.00	Non-default	Non-default

17	C	Rented	47	No	Yes	Stock	0.40	Non-default	Non-default
18	C	Rented	35	Yes	No	Cheque	0.00	Non-default	Non-default
19	C	Rented	36	No	Yes	Household items	0.60	Non-default	Default
20	B	Own house	38	No	Yes	Household items	0.16	Non-default	Non-default
21	B	Rented	35	No	Yes	Stock	0.11	Non-default	Non-default
22	B	Rented	35	No	Yes	Stock	0.11	Non-default	Non-default
23	B	Rented	34	No	Yes	Stock	0.11	Non-default	Non-default
24	B	Rented	48	No	Yes	Household items	0.67	Non-default	Default
25	B	Rented	28	No	Yes	Stock	0.08	Non-default	Non-default
26	B	Own house	42	No	Yes	Household items	0.22	Non-default	Non-default
27	B	Rented	32	No	Yes	Household items	0.39	Non-default	Non-default
28	B	Rented	44	No	Yes	Stock	0.19	Non-default	Non-default
29	B	Family House	44	No	Yes	Stock	0.06	Non-default	Non-default
30	B	Rented	36	No	Yes	Household items	0.48	Non-default	Non-default
31	B	Rented	47	No	Yes	Stock	0.25	Non-default	Non-default
32	B	Own house	33	No	Yes	Household items	0.12	Non-default	Non-default
33	B	Own house	37	No	Yes	Stock	0.02	Non-default	Non-default
34	B	Rented	33	No	Yes	Car	0.01	Non-default	Non-default
35	B	Own house	36	No	Yes	Stock	0.02	Non-default	Non-

University of Ghana http://ugspace.ug.edu.gh

									default
36	B	Rented	55	No	Yes	Stock	0.32	Non-default	Non-default
37	B	Rented	42	No	Yes	Stock	0.23	Non-default	Non-default
38	B	Rented	49	No	Yes	Stock	0.21	Non-default	Non-default
39	B	Rented	42	No	Yes	Stock	0.12	Non-default	Non-default
40	B	Rented	46	No	Yes	Car	0.03	Non-default	Non-default

It could be observed from the table above that whilst majority was correctly predicted others were not. For instance for client numbered 10 and 24, it was predicted to default because the probability was more than 50%. In actual sense, the empirical results show that the client did not default. This may probably be due to other unforeseen concourse of circumstances, such as ill health, accident, robbery, and negligence or otherwise of duty by the loan officer etc. which the model does not consider.

University of Ghana http://ugspace.ug.edu.gh

CHAPTER FIVE

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

5.0 Introduction

This chapter summarizes the results and discussion and makes conclusions that reflect the research objectives.

5.1 Summary

One of the objectives was to identify the risk factors that can influence loan defaults among customers in the microfinance sector.

From the discussions, the major determinants include the loan officer in charge of the customer, type of residence, age of client, presence of personal guarantor and assessment done to verify client background.

It could be noticed from the findings that the responsibility of the loan officer cannot be underestimated, it affect the probability of default directly. Thus knowledge and attitude of the loan officer affect the performance of the client.

It was also discovered that customers in the food vending and services sector have a lower risk compared to manufacturing and trade. Thus the Trade sector has relatively higher risk of default

The model also suggests that those who reside in family house and own their houses have relatively less risk than those in a rented apartment.

University of Ghana http://ugspace.ug.edu.gh

Term of loan is inversely related with default risk, the longer the term the lower the risk but was not significant to the model.

Age of client affects the level of risk of loan. The age of clients is directly related to risk of default. Thus if applicant age is less than 30yrs its probability of default is relatively higher than those whose age is above 32years.

Applicants who use guarantor have a relatively lower risk than those without guarantors

The type of collateral used correlate positively with risk of default.

The finding suggests that clients who used household items to support their loan are relatively more risky than those with stocks. However, those who use post-dated cheques are less risky than those who use stock of goods and household items as collateral. Those who use vehicle as collateral have relatively lower probability of default.

5.2 Conclusions

The findings indicated that a fairly reasonable way of accessing client (Y) giving a set of characteristics could be expressed as;

$$P(Y=1/X_i)=\frac{1}{1+e^{-(-25.03+4.51X_1+1.08X_2+1.62X_3+1.61X_4+3.96X_5+0.99X_6-0.36X_7-2.21X_8-1.76X_9+19.57X_{10}+21.51X_{11}+18.31X_{12}+0.08X_{13})}}$$

where, $X_1, X_2, X_3, ..., X_{13}$, are the explanatory variables (see Table 19).

University of Ghana http://ugspace.ug.edu.gh

Table 19: Predictor variables and description

Variable	Description	Model variable
Constant	Constant	X_o
Officer(1)-A	1 if client was handled by A and 0 otherwise	X_1
Officer(2)-B	1 if client was handled by B and 0 otherwise	X_2
Officer(3)-C	1 if client was handled by C and 0 otherwise	X_3
Officer(4)-D	1 if client was handled by D and 0 otherwise	X_4
Officer(5)-E	1 if client was handled by E and 0 otherwise	X_5
RES(1)	1 if client lives in own house and 0 otherwise	X_6
RES(2)	1 if client lives in rented house and 0 otherwise	X_7
Guarantor	1 if client has a guarantor and 0 otherwise	X_8
Assessment	1 if client was assessed and 0 otherwise	X_9
Collatype(1)	1 if client used stock and 0 otherwise	X_{10}
Collatype(2)	1 if client used Household items and 0 otherwise	X_{11}
Collatype(3)	1 if client used Vehicle and 0 otherwise	X_{12}
Age	Age of client	X_{13}

Source: research findings (2013)

5.3 The usefulness of the model

The researcher considers a practical way of utilizing the model. First a template for the model has been developed. The template captures all the significant variables and automatically assesses a client's risk of default given the set of variables are plugged in. The researcher illustrates the usefulness of the model as below. Consider three customers who applied for a loan in a microfinance company with characteristics including the attributes in Table 19.

University of Ghana http://ugspace.ug.edu.gh

The objective will be to identify the risk for advancing credit to such individuals using the predictors identified in Table 19.

Key = plug in 1 if event occurred or plug in 0 if otherwise

By plugging in the inputs 1 or 0 or as appropriate as the customer data apply into the model

$$P(Y=1)=\frac{1}{1+e^{-(-25.03+4.51X_1+1.08X_2+1.62X_3+1.61X_4+3.96X_5+0.99X_6-0.36X_7-2.21X_8-1.76X_9+19.57X_{10}+21.51X_{11}+18.31X_{12}+0.08X_{13})}}$$

We have the result shown in Table 20 (last column);

Table 20: Template demonstration of the usefulness of the model

	Officer	Residence	Age	Guarantor	Assessment	Collateral	Predicted Prob.	Predicted group
1	A	Rent	39	Yes	Yes	Household items	0.281696	Non-default
2	B	Rented	40	No	Yes	Stock	0.096157	Non-default
3	B	Own house	36	No	Yes	Household items	0.174149	Non-default

If the predicted probability is less than 50%, the client is classified as likely to be a good client and if more than 50%, the client is likely to be a bad client and need further controls to mitigate the impact in case of default.

University of Ghana http://ugspace.ug.edu.gh

5.4 Recommendations

5.4.1 Policy consideration

Based on the research findings, three key observations are recommended for policy consideration:

- Microfinance Institutions should adopt among others the default risk model to ascertain the level of risk. Since it's relatively efficient and cost effective.
- There should be adequate training for Loan officers in the MFI in order to improve on their skills and methodology.
- Bank of Ghana should consider enacting a law that will ensure that all MFIs in Ghana are put on a centralized database to check multiple borrowing and also serve as an internal control measure for sustainable MFIs in Ghana.

5.4.2 Recommendation for future research

It is recommended that future studies consider how rate of interest charged by the MFIs influence default risk.

Future studies should focus on developing random effects logit model to capture discrepancies of loan assessment processes.

University of Ghana http://ugspace.ug.edu.gh

BIBLIOGRAPHY

Adel, L., Ainon, R.N. and Wah, T.Y (2008). Credit scoring models using soft computing methods. *The International Arab Journal of Information Technology*, 7(2); Faculty of Computer Science and Information Technology, Malaya.

Agarwal, S., Chomsisengphet, S. and Liu, C. (2008). Consumer Bankruptcy and Default: The Role of Individual Social Capital. *Working Paper*. Available at SSRN: http://ssrn.com/abstract=1408757.

Allen, L., DeLong, G. & Saunders, A. (2004). Issues in the Credit Risk Modelling of Retail Markets. *Journal of Banking and Finance*, Vol. 27, Issue 4, p. 221-342.

Allen, M.F and Amelie, J. (2006), Determining the probability of default of Agricultural Loans in French Bank. *Journal of Agricultural Economics.* 67:513-20.

Altman, E. I. (1968). Financial Ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance*, Vol. 23, Issue 4, p. 589-609.

Altman, E. I., and Saunders, A. (1997). Credit risk measurement: developments over the last 20 years. *Journal of Banking and Finance*, Vol. 21, Issue 11-12, p. 1721-1742.

Altman, E. I., Eom, Y. H. and Kim, D. W. (2007). Failure prediction: evidence from Korea. *Journal of International Financial Management and Accounting*, Vol. 6, Issue 3, p. 230-249.

Altman E.I., Marco G., Varetto F., (1994). Corporate distress diagnosis; comparisons using linear discriminant analysis and neural networks (the Italian experience). *J. Banking and Finance.* 18, 505- 529.

Agresti, A. (1996). *An Introduction to Categorical Data Analysis*. John Wiley and Sons.

Apilado, V.P., Warner, D.C. and Dauten, J.J.(1974). Evaluative technologies in consumer finance-experimental results and policy implications. *J. Finance Quant.Anal, Mar*., 275-283.

Autio, M., Wilska, T-A., Kaartinen, R. and Lähteenmaa, J. (2009). The Use of Small Instant Loans Among Young Adults – a Gateway to a Consumer Insolvency. *International Journal of Consumer Studies*, Vol. 33, Issue 4, 407-415.

University of Ghana http://ugspace.ug.edu.gh

Arminger, G., Enache, D. and Bonne, T. (1997). Analyzing credit risk data: A comparison of logistic discrimination, classification tree analysis, and feed forward network. *Computational Statistics*, Vol. 12, Issue 2, p. 293-310.

Awotwi, E.(2011). Estimation of probability of default of consumer credit in Ghana: A case study of an international bank. Kumasi:KNUST.

Banasik, J., Crook, J. and Thomas, L. (2003). Sample selection bias in credit scoring models. *Journal of the Operational Research Society*, Vol. 54, Issue 8, p. 822-832.

Bank of Ghana, BOG (2013). Microfinance in Ghana. retrieved from http//:www.modernghana.com/news/317528/1/loan-default on 15th May 2013.

Barney, D.K., Graves,O.F. and Johnson, J.D (1999). The farmers home administration and farm debt failure prediction. *Journal of Accounting and Public Policy*. 18,99-139.

Boateng, C.O and Ampratwum, E. (2011). The Informal Sector in Ghana. Friedrich Ebert Stiftung. Ghana Office.

Bofondi, M. and Lotti, F. (2006). Innovation in the Retail Banking Industry: The Diffusion of Credit Scoring. *Review of Industrial Organization*. Vol. 28, Issue 1, p. 343-358.

Boyes, W. J., Hoffman, D. L. and Low, S. A. (2002). An econometric analysis of the Bank credit scoring problem. *Journal of Econometrics*, Vol. 40, Issue 1, p. 3-14.

Brown, S., Taylor, K. and Price S. W. (2005). Debt and distress: evaluating the psychological cost of credit. *Journal of Economic Psychology*, Vol. 26, Issue 5, p. 642-663.

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. Oxford University Press: Oxford, pp 75-90.

Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J. (1984). Classification and regression Trees. Wadsworth Belmont: California.

Buckley, J and James, I (1979). Linear regression with censored data. Biometrika 66, 429- 436.

University of Ghana http://ugspace.ug.edu.gh

Chen, M. C. and Huang, S. H. (2003). Credit scoring and rejected instances reassigning through evolutionary computation techniques. Expert Systems with Applications, Vol. 24, Issue 4, p. 433-441.

Claessens, S., Krahnen, J. and Lang, W. W. (2005) The Basel II reform and retail credit markets. *Journal of Financial Services Research*, Vol. 28, Issue 1-3, p. 5-13.

Capon, N.(1982). Credit scoring systems: a critical analysis. J. Marketing, 46, 82-91.

Chandler, G.G and Ewert, D.C. (1976). Discrimination on basis of sex and the equal credit opportunity Act. Credit Research Centre: Purdue University, Indiana

Chatterjee S., Barcun S., (1970). A nonparametric approach to credit screening. *J. American Statistical Assoc.* 65, 150-154.

Cheng B., Titterington D.M., (1994). Neural Networks: A review from a Statistical Perspective, Statistical Science 9, 2-30.

Choi S.C. (1986). Statistical Methods of discrimination and classification. Pergamon Press: New York.

Clarke, J.A.(2005). Implications of stratified sampling for fair lenders binary logit models. *Journal of Real Estate Finance and Economics*, 30(1), Pg 5.

Coffman J.Y., (1986). The proper role of tree analysis in forecasting the risk behaviour of Borrowers. MDS Reports, Management Decision Systems, Atlanta, 3,4,7 and 9.

Collins, R. A. and Green, R.D.(1982). Statistical Methods for Bankruptcy Forecasting. *Journal of Economics and Business*, 34, 349-354.

Coval, J. D., Jurek, J. W., & Stafford, E. (1992). Economic catastrophe bonds. *The American Economic Review*, 628-666.

Coval, J. and Shumway, T. (2000). Do behavioral biases affect prices? University of Michigan, *Working Paper.*

Crook, J. N., Hamilton, R. and Thomas, L. C. (1992). A Comparison of a credit Scoring model with a credit performance Model. *The Service Industries Journal*, Volume 12, Issue 4, p. 558-579.

Desai, V.S., Crook, J.N. and Overstreet, G.A. (1996). A comparison of neural networks

University of Ghana http://ugspace.ug.edu.gh

and linear scoring models in the credit environment. *European J. Operational Res*.95, 24-37.

Dirickx Y.M.I. and Wakeman L. (1976). An extension of the Bierman-Hausman Model for credit granting. Management Science 22, 1229-1237.

Dinh, T. H. and Kleimeier, S. (2007). A Credit Scoring Model for Vietnam's Retail Banking Market. *International Review of Financial Analysis*, Vol. 16, Issue 5, p. 571-495.

Dunn, L. F. and Kim, T.(1999). An Empirical Investigation of Credit Card Default.

Working paper, Ohio State University, Department of Economics, 99-13.

Durand, D. (1941). Risk elements in consumer instalment financing. National Bureau of Economic Research: New York.

Desai, V. S., Crook, J. N. and Overstreet, G. A. (1996).A comparison of neural networks and linear scoring models in the credit union environment. *European Journal of Operational Research*, Vol. 95, Issue 1, p. 23-37.

Edelman, D.B. (1997). Credit scoring for lending to small businesses, Proceedings of Credit Scoring and Credit Control (V). Credit Research Centre: University of Edinburgh.

Eisenbeis, R.A. (1977). Pitfalls in the application of discriminant analysis in business, finance and economics. *Journal of Finance* 32, 875-900.

Eisenbeis, R.A. (1978). Problems in applying discriminant analysis in credit scoring models. *Journal Banking and Finance* 2, 205-219.

Freed N., Glover F., (1981b). Simple but powerful goal programming formulations for the discriminant problem. *European Journal of Operational Research* 7, 44-60.

Frydman H., Kallberg J.G., Kao D-L, (1985). Testing the adequacy of Markov chains and Mover-Stayer models as representations of credit behavior. Operations Research 33, 1203-1214.

Fung, R., Lucas, A., Oliver R. and Shikaloff N. (1997). Bayesian Networks applied to credit scoring, Proceedings of Credit Scoring and Credit Control V. Credit Research Centre: University of Edinburgh.

University of Ghana http://ugspace.ug.edu.gh

Gan, C., Li, Z., Wang, W., & Kao, B. (2012). Credit scoring in mortgage lending: evidence from China. *International Journal of Housing Markets and Analysis*, *5*(4), 334-350.

Ghana Association of Microfinance Companies, GAMC . Retrieved from http://ghana-microfinance.com, on 20th March.2012.

Grablowsky,B.J. and Talley, W.W (1981). Probit and Significant functions for classifying credit applicants; a comparison. *J. Economics and Business*, 33,254-261.

Greene, W.(1997). Economic Analysis, (3rd ed): Prentice Hall

Greenwood, J. and Jovanovic, B. (1990). Financial Development, Growth and the Distribution of Income, University of Western Ontario, The center for study of International Economic Relations working papers 9002.

Gup, B. E., & Kolari, J. W. (2005). Commercial Banking: The Management of Risk (3rd ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Hashemi, S. M. (1990). Rural Credit program and Women Empowerment; World Development 24(4):635-53

Hand, D. J. and Henley, W. E. (1997). Statistical classification methods in consumer credit scoring: a review. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *160*(3), 523-541.

Ignizio J.P. and Soltys J.R. (1996). An ontogenic neural network bankruptcy classification tool, IMA *Journal of Mathematics applied in Business and Industry*. 7, 313-326.

Jacobson, T. and Roszbach, K. (2003). Bank lending policy, credit scoring and value-at-risk. *Journal of Banking and Finance*, Vol. 27, Issue 4, p. 615-633.

Judge G.G., Griffiths, W.E., Hill, R.C., Helmut, L. and Lee,T.C (1985). The Theory and Practice of Econometrics. 2nd ed, New York,NY: John Wiley and Sons.

Laitinen, E. and Laitinen, T. (2000). Bankruptcy prediction: Application of the Taylor's Expansion in logistic regression. International Review of Financial Analysis, Vol. 9, p. 327-349.

Laitinen, T. and Kankaanpää, M. (1999). Comparative analysis of failure prediction methods: The Finnish case'. European Accounting Review, Vol. 8, Issue 1, p. 67-92.

University of Ghana http://ugspace.ug.edu.gh

Lawrence, E. C. and Arshadi, N. (1995). A multinomial logit analysis of problem loan resolution choices in banking. Journal of Money, Credit and Banking, Vol. 27.

Lee, T. S., Chiu, C. C., Lu, C. J. and Chen, I. F. (1982). Credit scoring using the hybrid neural discriminant technique. Expert Systems with Applications, Vol. 23, Issue 3, p. 245-254.

Lewis, E.M. (1992). *An introduction to credit scoring*. Athena Press, San Rafael, California.

Lieli, R. P. and White, H. (2008). The Construction of Empirical Credit Scoring Models Based on Maximization Principles. *Journal of Banking and Commerce*, Vol. 27, Issue 4, p. 615-633

Lufburrow, J., Barry, P.J. and Dixon, B.L.(1984). Credit Scoring for Farm Loan Pricing. *Agricultural Finance Review* 44 : 8-14.

Khandker, S.(2003). Microfinance and Poverty: Evidence using panel data from Bangladesh; World Bank Policy Research Working Paper 2945.

Kočenda, E. and Vojtek, M. (2009). Default Predictors and Credit Scoring Models for Retail Banking. CESifo Working Paper No. 2862.

Luo, J-h. and Lei, H. Y. (2008). Empirical study of corporate credit default probability based on Logit model. *Journal of Banking and Finance.* Vol. 27, Issue 4, p. 615-633.

Mario,J.M. and Claudio, G.(2010). Systemic Risk, Index Insurance, and Optimal Management of Agricultural Loan Portfolios in Developing Countries. *American J. of Agricultural Economics.* VOL.93,Issue 2, P. 399-406

Martell, T.F. and Fitts, R.L. (1981). A quadratic discriminant analysis of bank credit card user characteristics, *J. of Economics and Business* 33, 153-159

Martin, D. (1977). Early Warning of Bank Failure: A Logit Regression approach. *Journal of Banking and Finance,* Vol. X, Issue X, p. 249-276.

McIntosh, C. De Janvry, A, and Sadoulet(2005). How rising competition Among Microfinance Institutions Affects Incumbent Lenders. *The Economic Journal* V.115,issue 586, 987-1004

Mester, L. (1997). What's the point of credit scoring? Federal Reserve Bank of Philadelphia Business Review, September/October, p. 3-16.

University of Ghana http://ugspace.ug.edu.gh

Musto, D. K. and Souleles, N. S. (2006). A Portfolio View of Consumer Credit. *Journal of Monetary Economics, Vol. 53, Issue 1, p. 59-84.*

Merton, R.C. (1974). On the pricing of corporate debt: the risk structure of interest rates, *Journal of Finance 29, 449-470.*

Mortensen, T., Watt, D.L. and Leistritz, F.L.(1988). Predicting Probability of Loan Default. *Agricultural Finance Review* 48 : 8-14

Narain, B. (1992). Survival Analysis and the credit granting decision, in Credit Scoring and Credit Control. Oxford: Oxford University Press, pp 109-122

Neophytou, E. and Charitou, A. (2000). Predicting Corporate Failure: Empirical Evidence for the UK. Working paper, University of Southampton, Department of Accounting and Management Science, No. 01-173.

Novak, M.P. and LaDue E.L.(1999). Application of Recursive Partitioning to Agricultural Credit Scoring*. J. Agric and Appl. Econ*. 31,1:109-122.

O'Connell, A. A.(2005). Logistic Regression Models for Ordinal Response Variables. Quantitative applications in the social sciences, Thousand Oaks, CA: Sage Publications, Volume 146.

Özdemir, Ö. and Boran, L. (2004). An Empirical Investigation on Consumer Credit Default Risk. *Turkish Economic Association Working* Paper 2004 / 20.

Petersen, M. and Rajan, R.(1995). The Effect of Credit Market Competition on Lending Relationships. *Quarterly Journal of Economics*, 110,407-443

Peltoniemi, J. (2004). The value of relationship banking: empirical evidence on small business financing in finnish credit markets. *MSc thesis*, University of Oulu, Oulu.

Pyndick, R.S. and Rubinfeld, D.L.(1998). *Econometric Models and Economic Forecast*. 4th ed, McGraw Hill.

Reichert, A.K.,Cho, C.C. and Wagner, G.M.(1983). An examination of the comceptual issues involved in developing credit scoring models. *Journal of Business and Economic Statistics*, 1(10),101-114

Robinson, M. (2001). *The microfinance revolution*. Washington DC, World Bank Publications.

University of Ghana http://ugspace.ug.edu.gh

Roszbach, K. (2004). Bank Lending Policy, Credit Scoring and the Survival of Loans. *Review of Economics and Statistics*, Vol. 86, Issue 4, p. 946-958.

Rosenberg, E. and Gleit, A. (1994). Quantitative methods in credit management: a survey, Operations Research 42, 589-613.

Steenackers, A. and Goovaerts, M. J. (1989). A Credit Scoring Model for Personal Loans. Insurance: Mathematics and Economics. Vol. 8, Issue 1, p. 31-34.

Stiglitz, J. E. and Weiss, A. M. (1981). Credit Rationing in Markets with Imperfect Information. *American Economic Review*, Vol. 71, Issue 3, p. 393-410.

Straka, J. W. (2000). A Shift in the Mortgage Landscape: The 1990s Move to Automated Credit Evaluations. *Journal of Housing Research*, Vol. 11, Issue 2, p. 207-232.

Thomas L.C., (1992). *Financial risk management models, in risk analysis, assessment and management*. Wiley: Chichester.

Thomas, L. C. (2000). A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers. International Journal of Forecasting, Vol. 16, Issue 2, p. 149-172.

Titterington, D.M. (1992). Discriminant Analysis and related topics, in Credit Scoring and Credit Control. pp 53-73, Oxford University Press, Oxford.

Tsai, M. C., Lin, S. P., Cheng, C. C. and Lin, Y. P. (2009). The consumer loan default predicting model – An application of DEA-DA and neural network. Expert Systems with Applications, Vol. 36, Issue 9, p- 11682 11690.

Turvey, C.G.(1991). Credit Scoring for Agricultural Loans: A Review with Applications. *Agricultural Finance Review* 51 : 43-54.

Updegrave, W.L.(1987). How Lender Size You Up. *Money*. Pp. 23-40

Van Order, R. and Zorn, P. M. (2000). Income, Location and Default: Some implications for community lending. Real Estate Economics, Vol. 28, Issue 3, p. 385-404.

Vasanthi, P. and Raja, P. (2006). Risk Management Model: an empirical assessment of the risk of default. *International Research Journal of Finance and Economics*, Vol. 1, Issue 1.

Weizhuo, W. (2010). The Probability of Chinese Mortgage Loan Default and Credit

University of Ghana http://ugspace.ug.edu.gh

Scoring. Retrieved from http://hdl.handle.net/10182/2346

Wiginton, J. C. (1980). A note on the Comparison of Logit and Discriminant Models of Consumer Credit Behaviour, *Journal of Financial and Quantitative Analysis* 15, 757-770

Wilson, N., Summers, B. and Hope, R. (2000). Using Payment Behaviour Data for Credit Risk Modelling. *International Journal of the Economics of Business*, Vol. 7, Issue 3, p. 333-346.

Yang, Y., Nie, G. and Zhang, L. (2009). Retail Exposures Credit Scoring Models for Chinese Commercial Banks. Computer Science, Vol. 5545, Issue 1, p. 633-642.

Yeh, I., and Lien, C. H. (2009). The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. *Expert Systems with Applications*, *36*(2), 2473-2480.