KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY,KUMASI

COLLEGE OF SCIENCE

FACULTY OF PHYSICAL SCIENCES

DEPARTMENT OF MATHEMATICS

Prediction of loan default using logistic regression: A case study of Ahafo Ano Premier Rural Bank

By

FRANKLIN ADU-GYAMFI

A THESIS TO BE SUBMITTED TO THE DEPARTMENT OF MATHEMATICS AND STATISTICS, KNUST IN PARTIAL FULFILMENT OF THE REQUIREMENT FOR THE AWARD OF A MASTER OF SCIENCE DEGREE IN INDUSTRIAL MATHEMATICS

October 16, 2016

Declaration

I hereby declare that this submission is our own work towards the award of a Master of Science in INDUSTRIAL MATHEMATICS and that to the best of my knowledge, it contains no material previously published by another person nor material which has been accepted for the award of any other degree of the University, except where due acknowledgement has been made in the text.

FRANKLIN ADU-GYAMFI	....	....
Student	Signature (PG 8823813)	Date
Certified by:		
Dr. A. O. ADEBANJI	....	....
Supervisor	Signature	Date
Certified by:		
Dr. R. K. AVUGLAH	....	....
Head of Department	Signature	Date

Dedication

This work is dedicated to my parents whose affection, love, encouragement and prayers made me come this far.

Abstract

Advancement of loan facilities among Rural banks to individuals and Small and Medium Enterprises (SMEs) is associated with high risk due to default in repayment of such loan facilities in Ghana.This study seeks to explore the characteristics of customers of the Ahafo Ano Rural Bank that make them more likely to default in loan repayment. Loans form the major part of the assests of banks in Ghana. Loans are the main source of income for these banks which also intends to be very risky to the lender. Rural banks have its constituency in the lending activity of rural folks in Ghana which were set up by Government of Ghana Banking policy.Logistic regression was applied to customer loan application data from Ahafo Ano Premier Rural Bank to determine characteristics of customers who default (dependent variable) with demographic and socio - economic factors as independent variables . A total of 152 customers were considered for the study. Preliminary analysis by the use of test of independence identified, loan type (commercial and susu), marital status (married) from the Chi - Square statistic. Logistic regression obtained from the study had type of loan (commercial), loan repayment period cum number of dependents of customer been statistically significant ($p - values < 0.5$). The model obtained is: $Log(Odds) = 3.863X_1 + 0.088X_2 - 0.234X_3$. In conclusion, the study findings show that customers with more months to pay a loan, married customers, and customers with smaller number of dependents are more likely to default. Results and findings from the study will help the bank and other financial institutions to make informed decisions by identifying low risk and high risk customers when granting loans to their customers.

Acknowledgements

We wish to thank the Almighty God for His love, grace, mercies and protection given to me to complete this work. I also wish to express my profound gratitude to my supervisor for her sage advice, motherly love, insightful criticisms and patience encouragement which aided the writing of this thesis in innumerable ways.

Contents

List of Tables

List of Figures

Chapter 1

INTRODUCTION

Credit conditions continue to get tighter especially for Small and Medium Enterprises (SMEs) which cannot finance themselves from capital markets or cash flows and which lack collateral demand of banks. Lending rates continue to be high while those in the business arena have supported a campaign for lending rates in particular to come down,it has not gone down well with others particularly those in the banking sector who cite low loan recovery and high government borrowing which is more safer among other factors for the high lending rates. They argue that they are not effectively able to differentiate a person who will repay a loan and a person who will not pay hence the decision to increase lending rates. Indeed there is a component of the lending rate computation which is called the default rate. This is charged because they can not differentiate between credit worthy borrowers and loan defaulters. If the default rate is high,the lending rate will be high and vice versa.

Lending forms major part of the assets of bank in an emerging economy like Ghana where capital market is not developed. It is the main source of income on the profit and loss statement of banks but this income comes with enormous risk to the lender. Effective credit management is therefore very crucial to the survival and liquidity of any financial institution. Delinquent loans therefore have adverse effects on the financial status of banks through provision for bad debts which reduces the profitability of the bank (Comptroller,1998). As at December 2012, the total provision for bad and doubtful debts declared by rural and community banks amounted to GHs 36.52 million. This amount will reduce the profitability of the bank since it is an expense which has been incurred to write off these bad debts.

1.1 Background of Study

The Government of Ghana was looking for ways that rural folks can have access to finance. This led to a policy decision which included that universal banks hold more than 20% of their total loan portfolio to the agricultural sector and also with the promulgation of Agricultural Development Bank(ADB)in 1965 with the sole aim of advancing loans to the agricultural sector and its associated businesses in the rural Ghana. These universal banks together with ADB begun operations with outlets in the rural areas to comply with these measures. In spite of these measures, providing financial facilities to the rural Ghana continuously remains very low;since these commercial banks used the deposits from the rural Ghana to lend in the urban areas while these branches in the hinter land made only payments to the cocoa farmers. Initially loans were not extended to these customers in those branches. Later universal banks required high deposit accounts and adequate security for the loan as conditions in order to extend loans to the rural folks. Majority of these fisherman and farmers did not have these kind of accounts and their collateral were not adequate for commercial lending(Andah and Steel 2003). Mensah(1993)and Ranade(1994) opined that the loan portfolio of ADB was very small. Lending to smaller holder farmers made up only 15 percent of their loan portfolio and 27 percent of their branches were found in rural Ghana. Since the policy failed to achieve its intended purpose, Government of Ghana (GoG) decided to provide some form of assistance in relation to promulgation of rural banks in developing communities in Ghana that will have a sole aim of providing financial services in the hinter land.

The BoG was tasked to study the rural banking in Philipines and help to facilitate in the operations of these community - based banks in farming and fishing communities in Ghana. In 1976,the first rural bank was established in Nyakrom in the central region with an initial capital of 60660 old Ghana cedis. Capital contribution was drawn from the inhabitants of the community. The Association of Rural banks was set up to exchange information and ideas and also help improve the performance of banks. The number of rural banks rose sharply from 1980 to 1984 and reached 106. This occured as result of the growing interest from inhabitants of the rural community in setting up their own banks and introduction of Akuafo cheques in 1982 (AjairNair 2010).

A loan that is not income when full repayment of principal and interest has not been settled or the maturity has passed with either the principal or interest in arrears. There is a growing concern in the finance and banking sector in Ghana about the rise in bad loans in the last few decades. The immediate result of this surge in non-performing(bad) loans in the financial system is bank failure.Every bad loan in the banking business is regarded as a result of a poor unbeneficial initiative. It is therefore noted that, attempts to eliminate or minimize loan defaults in the finance sector is a crucial means of improving the sustainability of the banking industry and the economic status.The continuous existence of bad loans in the banking business in the long-run locks scarce resources in non-performing sectors of the economy in general which consequently halts efficiency and growth in the economy. According to Bank for International Settlements(BIS), the standard loan groupings are defined as follows:

1. passed(current): loans that have principal and interest rate paid within the stipulated time. In rural banks, 1 percent is provided for current loans

2. Special mention(olem): loans which are in default for 90 days. A provision of 10 percent is made by the rural banks

3. Substandard: loans in arrears between 90 days and 180 days. Rural banks make 25 percent provision for the loans classified as substandard.

4. Doubtful: loans which have outstanding balance in doubt and there is indication on these accounts that there will be loss. A provision of 50 percent

is provided for these loans

5. loss(unrecoverable): loans which are viewed as uncollectable due to death or bankruptcy of the company. A provision of 100 percent is made for bad debt .

Non-performing loans therefore consists of substandard, doubtful and loss which are further segregated according to their recovery difficulties. Several economies are reducing the time- frame for loans in default to become bad debt in order that lenders are put on alert in the shortest possible time so that precautionary measures are put in place before losses begin to rise. The international Accounting standard 39 revised in 2003 brought measures cum requirements for measurement of financial instruments and subsequently went further to set boundaries for classification of impaired loans. These industry standards were formulated with the aim that lenders won't be caught off guard. Loans and up -front payments are most often the main source of income and assets acquisition for the bank. Weakness in the lending role can demoralize the possibility of the banking sector and the economy as a whole. Due to this reason,the credit unit of banks should be manned and maintained efficiently so as to improve the quality or health in order to: a.minimise risks-loan default and losses b.maximise profitability

In the university of Texas a research was conducted to identify students most likely to default in the repayment of students loan facility. The results showed that, the problem of completion, tenacity and success were significant and strong forecasters of student loan default. Not withstanding the above, sex, enrollment of the university as well as the race or ethnic backgrounds (Herr and et al 2005).

The study was to assess factors affecting probability of default for savings and credit cooperative societies in Tanzania. He found out that age,marital status,loan activity,interest rates,loan duration,value of borrowers' collateral had no effect with the loan default. However years of schooling experience and amount of loan had an influence positively on loan default. Political interference in the activities of savings and credits cooperative accelerated the loan default risk and was observed in particular region where political activities were very high(Magali 2013).

There is therefore the need to investigate into the characteristics of customers to the abundant research carried out in this field to find out whether or not a significant effect of certain variables such as age,gender,geographical location of customers,loan amount,term of loan,type of account,type of loan have adverse effect on loan default.

1.2 Problem Statement

Generally, rural banks in developing countries including Ghana perform very crucial roles in the provision of financial services to a wide range of customers for several economic engagements so as to improve their livelihoods. Financial facilities in the form of loans are core and fundamental functions that rural banks provide to their numerous customers to invest into their respective economic activities to be repaid within a specific time frame.

However, repayment of these loan facilities by customers has over the years remains as a major challenge to rural banks and financial institutions alike as most loans granted end up as bad debts due to the inability of management to retrieve both the principal and interest accrued on the loan advanced to customers. Defaults in loam repayment among customers of financial institutions have effects on the banks that affects income levels, development and growth of the bank as well as other customers who will be denied the opportunity to benefit from the various loan facilities provided by financial institutions.

Several studies have investigated and identified the factors or attribute of customers that makes them default in the repayment of loan facilities advanced to them but very little is done from the modelling perspective. This current study therefore seeks to employ a more scientific approach (use of a mathematical model) to identify the determinants of loans defaults among customers of Ahafo Ano Premier rural bank so as to give management the opportunity to be able to discriminant between customers with low and high risk in terms of defaulting from repayment of their respective loans.

1.3 Objectives of the study

The objectives of this study is to

1. Identify factors that affect loan default at Ahafo Ano Premier Rural Bank

2. Predict the probability of loan default and hence classify cases

1.4 Methodology

In order to classify loan applicants as defaulters or non defaulters, the proposed model will based strictly on existing data on expired loans. Logistic regression will used to identify variables that can used as a tool for classification of new observational units especially new respondents into groups in which most probably it belongs. That means each observational unit possesses certain number of characteristics that could be measured and those values vary from one observational unit to another. A sample size of 152 expired customer loans were used. Out of which 100 cases were used for model building and the 52 cases for model evaluation. The dependent variable is default status (default or non-default). The independent variables are age, marital status, income, security, type of loan, type of account, gender

1.5 Significance of the Study

This thesis seeks to classify the loan portfolio of Ahafo Ano Premier Rural Bank. It is anticipated that, the attained model from this study will serve as a vital tool to financial institutions or banks in Ghana when planning and decision making on loan advancements to their respective customers. Otherwise, operating banks will end up becoming bankrupt when they unable to recover all the loans given out to their clients.Hence the need for a more scientific approach to be adopted by banks to ascertain efficient, adequate and effective distribution of funds they have available for loans to ensure and promote the constant growth of these banking institutions.

Findings from the study will provide proposition and recommendations will be given to strength and minimize the weaknesses and challenges associated with loan portfolio allocations.

1.6 Scope of the Study

It will cover the loan portfolio of Ahafo Ano Premier Rural Bank from 2011 to 2014. It will cover clients operating current account, savings account and susu account. The loan types used are commercial loan, susu loan, salary loan.

1.7 Study Organization

This study is structured into five main sections or chapters. The study starts with the first chapter which introduces the work, gives some background on the study area, and presents the problem statement of the work, states the objectives for the entire study through the significance, focus (scope) study organization and main tool (methods) employed. The next chapter, Chapter two is dedicated to reviewing earlier studies carried out in the area of interest which is followed by the fourth chapter that is solely for the description and discussion of methods identified and employed to facilitate the study and also looks at the data cum its characteristics. Chapter four presents results as well as discussions based on the data analysis using the tool identified in three. The last section, chapter five concludes the study by means of giving summary of findings, states the conclusions in relation to the findings and also makes recommendations for financial institution informed by findings.

Chapter 2

LITERATURE REVIEW

2.1 Theoretical Studies

The theory of asymmetric information in banking indicates that it can difficult to distinguish between credit worthy borrowers and loan defaulters (Auronen 2003) which results into hostile selection and moral threat difficulties. Asymmetric statistics in the market implies one entity holds more information than others in terms of decision making and business dealings.

2.2 Information Asymmetry

The existence of information asymmetric characterised by imperfect flow of information between the lenders and borrowers led to collateral lending policies initiated by formal credit institutions. This results in restrictions in the availability of formal credit and as a result informal credit lenders remain the only option for the rural folks. Information asymmetry is one of the causes that impedes information flow from banks to the customers and vice versa. Accurate information is essential for sound economic decisions. Timely information to the rural community by the formal credit lenders to rural folks becomes the power to effective credit delivery (Ekumah, 2003).

2.3 Adverse Selection Theory

Adverse selection is where lenders can not distinguish good borrowers from bad borrowers and this results in a situation where borrowers are charged one price irrespective of their risks. Pagano (1993) shows that information sharing provides information about credit applicants and these information are provided by credit bureaus.Lenders are confronted by a large number of customers and if the lender has no private information about his borrowers, it could result in bad credits being selected. If information asymmetry is greatly reduced through information sharing, good borrowers who were priced out of the credit market because of high rate will now have loans at a reduced rate.Lenders can overcome adverse selection through sharing private information of their borrowers with other lenders. However if banks have monopoly of information about the credit worthiness of their customers, the profit level of borrowers turn to be reduced through the banks charging high interest rates. Information sharing affects market competition,interest rates,volume of lending and social welfare. Information sharing will reduce future interest rates and future profits (Padilla 1997).

Jappelli (2005) posited that if banks share information about borrowers, lenders will have more knowledge about character of the borrower and will prevent borrowers from being a serial defaulter to many bankers.

2.4 Moral hazard

The moral hazard is when a borrower has the incentive to use the funds in risky projects other than purpose specified for the credit. The risk taking party has more informations about its intentions than the party paying the consequences of the risk. If the risk paying party has less information, the borrower is likely to default. Lenders will increase rates in the event of having less information to compensate for the risk (Alary 2001). There has been an upsurge of non-performing loans in banks as a result of contrary choices and moral perils (Bofondi 2003). The continuous being of banking firms in an economy is usually construed in relation to their capacity to overcome the challenges of irregularity of information that eventually results to the declines of NPLs (Uyemura 1993).

2.5 Interest Rate

Interest rate is the margin per unit of duration(time) stated as a percentage of amount loaned out. Interest rate is the profit that accrues from the credit transactions over the loan period. Players in the banking industry in Ghana has a wide perception that there is wide interest rate spread. Ransford (2014) stated that interests rates on financial firms are often influenced by factors such as liquidity of the bank, treasury bill rates, Exchange rates, costs due to overheads, margin of profits, Gross Domestic Product (GDP), prime rates as well as the performance of loans rendered and other underlying factors. Interest rate are related with the maturity of loan that refers to the duration period for which loans are paid back coupled with the fluidity which has to do with the ease of conversion of assests into cash. Different interest rates emerge when liquidity and maturity time relate with other financial instruments (Anyanwu 1990).

Stiglitz (1981) posited that adversative selection facet of interest rate is having diverse debtors with different rates of loan default.Banks use this approach of interest rate as screening tool to separate borrowers. Excessive interest rate will in the long run lead to high performing loans.

Loanable Funds Theory

Loanable funds is the amount of money available for borrowing. The market for loanable funds is where there is a demand and supply of these financial resources. The easy access of funds to be give as loans is influenced by the total revenue in the economy after consumption and government expenditures are settled. The interest rate on loans are therefore determined by the difference between supply and demand in the economy in general(Anyanwu, 1990).

2.6 Credit risk

The loss of a financial reward resulting from the failure on the part of borrowers to settle their loans or in other words meet the contractual commitment is referred to as credit risk. Credit risk arises whenever a borrower is expecting to use future cash flows to pay a current debt. The goal of credit risk management is to maximise bank‘s rate of return and improving bank‘s image. The lender incurs the risk and entails lost interest cum the principal which interferes with cash flows, and as well raise the cost of collection which arise in situations:

1. A consumer does not meet financial obligation on a mortgage loan or other loan
2. A company or an enterprise is having the salaries of its employees in arrears
3. A bond issuer does not pay a coupon or principal when due
4. An insurance company does not honour a contract to a policy holder
5. An illiquid bank folds up and does not pay its clients.

.

Credit risk can be transfered but cannot be completely eliminated. Classification of loans is country specific and different categorisation will lead to different results. Hence the criterion used for the classification of loans is critical for the study of recovery rates (Sauders 1997).

2.6.1 Sources of Credit risk

Most banking crises have arisen either directly or indirectly caused by weakness in credit risk management. Severe credit losses in a banking system usually reflect problems in several areas such as concentration of credit on certain individuals or particular sector of the company and inappropriate due diligence. Some of the major credit risk being discussed are credit concentration,credit appraisal issues and repayment period.

2.6.2 Credit Concentration

Credit concentrations are viewed as pools of transactions that may perform because of a common characteristic or common sensitivity to economic, financial, business development where the losses are large relative to the lending institution safety and soundness. High concentration will give high potential credit losses when there is default (Rich 2003). So the question arises that why do banks allow concentration to develop. In some cases, a pool of loans may represent a concentration of risk that is difficult to avoid. Smaller banks may accumulate credit concentration because of their catchment areas or dynamics of local economies or regulation where banks are required to hold a certain portion of their credit portfolio in a particular sector of the economy. Large banks may develop concentration through acquisition or implementing a strategy to take advantage of investment grade or growing industries which help in asset appreciation of banks (Slywotzky, 2005). However during the recession of 2007-2009, the banking industry experienced significant losses due to exposure to national housing market which suffered declines in asset values. For this reason, bank supervisors need to assess the concentration arising from their credit.

2.7 Credit Appraisal

Most credit problems arise due to weakness in the credit appraisal and monitoring. Competition and loan syndication techniques tend to relegate basic due diligence to the background. Basic due diligence deals with issues relating willigness and ability to pay, the borrowers‘ ability to withstand bad economic conditions, the borrowers‘ credit history,liquidity and marketability of collateral. The credit policy and procedures create parameters for the credit appraisal. The absence of lending procedures and disregard of these policy have led to serious banking problems (Gabriel 2009).

Credit process issues arise out of subjective appraisal of senior management of the bank. This is subjectivity comes as result of a member owing the company or a cronies owing the company or their integrity being compromised. At larger banks, there are separate departments which deals with documentation and appraisal, disbursement and monitoring, collateral appraisal, and independent supervisors with the mandate to check that the processes are aligned with the credit policy and lending standards. At smaller banks, credit officers carry out the appraisal process and the internal and external auditors carry out the checks and balances on these appraisal. Effective credit management helps to avoid bad credits (Sauders 1997). Many banks lend against real collateral. Proper collateral appraisal should be carried out to make assessment of the correlation between the financial condition of the borrower and the market price of the collateral.

2.8 Repayment Period

This is time the borrower is required to service the loan. The time given to borrowers to service the loan is usually too short and the borrower may have used the loan to invest in projects which takes time to help repay in a short time. The longer the term of the loan, the less the chance of defaulting (Adem 2012).

2.8.1 Credit risk and Corporate Governance

Rural banks are usually integrated as limited accountabilities under the code of ethics for companies of 1963 (Act 197) of Ghana and are compulsory to be owned by stakeholders. These shareholders can come from the inhabitants of the catchment area of the bank or any interested individual.Out of these shareholders, board of directors are elected to supervise and put in policies for implementation by the management of the bank. The number of board of directors for rural banks ranges from a minimum of 5 and maximum of 11. Under the board of directors, there are various sub-committee of which are loans committee, finance and audit committee, disciplinary committee. The loans committee meets to approve loans which are above the limits of the General manager. The board of directors have the mandate to set loan approval limits for each supervisor which is always stated in the credit policy of the bank. However,rural banks fail to attract the needed expertise and experienced people to the board because of low renumeration (Ajar, 2010).

In their paper, (Blanchard 2003) proposed that board subcommitee which deals with risk management must be competent and co-opted directors who have no interest in the purchase of the shares even at a later late so that they will be objective in their risk decision. However (Rachdi et al, 2013) proposed that independent directors have no significant effect on credit risk.

Large board of directors, less busy directors and independent directors are associated with lower credit risk levels (Switzer et al 2015) but there was no relationship between corporate governance and credit risk statistically after using CGQ score and Moody‘s credit ratings (Postnova,2012).

The research was undertaken to identify the association between credit risk and governance in Islamic banks. The relationship between credit risk and composition of board of directors,number of board of directors and board subcommittees were statistically insignificant. However there was a positive relationship between credit risk and the bank‘s size (Bourakba et al,2015).

2.9 Empirical Studies

2.9.1 Credit Scoring

The decision to reject or accept a loan facility in the banking industry was on the subjective opinion of the officer in the bank who qualitatively graded the risk after examining the financial statement of the borrower, business plan and also interviewed the customer. Researchers and practitioners therefore found new ways of making the decision objective with the advent of information technology and credit scoring. Credit scoring has been used to control risk, manage losses, evaluating new loans, reducing time for approval of loans, and increasing profitability since there will be reduction of NPLs. Credit scoring is a model used to help the decision maker in the acceptance or rejection of loan based on already defined criteria. Henry Wells was the first person to used credit scoring during the second world war. He used these tools to help inexperienced staff to perform credit appraisal because many of his experienced staff had been sent to war. These models were later employed by the banking industry giving weights to certain characteristics summing the points to a classification score.

Nguta et al (2013) studied the causes of loan default in a District in Kenya using a selected random sample of 400 individuals. The data analysis used were inferential and descriptive statistics. It was found out that the type of business, age of business, number of employees and business profits had a strong relationship with loan default.The study was conducted on 175 farmers of Khorasan-Razavi Province in 2008. It studied factors affecting loan repayment of these farmers who had accessed loans from Agriculture Bank. Among the factors studied were age, farm size (in hectares),customer's of experience, income, interest rate, time laps between loan approval and disbursement, total application cost, whether customer owned farm machinery, loan size, collateral size, number of installments for which loan matures. The logit model was used in the data analysis. He concluded that interest rate was the most important factor (Kohansal et al 2009).

Bichanga and Aseyo (2013) carried out a study in Trans-Nzoia country with a sample size of 100 MFIs loan borrowers and 50 MfIs official loan borrowers drawn from owners of small farms. The study made use of random sampling and questionnaire was the mode of data collection. The data were analysed using qualitative and quantitative techniques and use of frequency tables. The objective of the study was to investigate how non-supervision of borrowers, adverse economic growth and diversion of loan funds leads to default in loan. He found that non-supervision of loan by the MFIs leads to default.

The factors under consideration in that study were age, gender, occupation, loan amount, salary, marital status, term of loans. A total of 15000 loan beneficiaries were analysed of which 1558 were defaulters and 13442 non-defaulters. The results show that age of applicant, gender, occupation, marital status and term of loan influence the ability to repay a loan. The method used in the analysis were parametric estimation and logistic regression (Adem et al 2012).

The study was use two methodologies to evaluate credit risk and compare the two methods which is more reliable. The research used logistic regression and multicriteria decision making. He concluded that logistic regression gave him the probability of loan default whereas multicriteria decision making gave him the classification of clients. He further stated that a small data set and a short repayment history is likely to decrease the quality of credit scoring in logistic regression and in that case multicriteria decision making is more efficient. The study was carried out in a Croatian bank (Sarlija et al 2003)

In their article, (Soureshjani and Kimiagari 2013) tried to find the best model evaluating credit risk using logistic regression and neural network on credit scoring problem. It studied 127 loan applicants of which 21 were defaulters.The variables studied were liquidity ratios, leveraging ratios, activity ratios and profitability ratios. They found out that liquidity ratios, leveraging ratios and activity ratios affect loan default. It also found the best cut off point to minimise overall error of modelling credit risk using the two methods.

The study was conducted in Ogun state with 100 poultry farmers. The variables used were age, education, flock size, household size, home distance from credit source, income, interest rate, loan size, marital status, occupation, financial outlet, preference gender. The probit model shows that flock size of the farmers is the most important factor of loan default though level of farmers‘ education, income, age had significant influence on default (Oni et al 2005).

Hongli and Liwei (2000) carried out a study on real estate industries using logistic regression. The sample used consisted of 37 listed real estate companies. The enterprise's asset scale and profitability were the factors for evaluation of credit risk and logistic regression had a predictive accuracy of 89.2 percent which was good. 18 financial indicators of these 37 listed real estate were analysed.

Dunson and Dadzie (1990) chose sixty rural banks out of one hundred and fourteen rural bnks where an average of 10 questionnaires were given based on their size. The study looked at determinants of loan defaults in rural banks in Ghana. Two methods were used, the first one being discriminant analysis and the second method being logit, ordinary least squares and tobit regressions to analyse secondary data. Gender, collateral type and marital status were significant.

The paper studied factors inhibiting small scale farmers in repayment of their loan facilities in Sene district of Ghana. Descriptive statistics and the probit model were used. The study identified profit margin, age of farmer, years of experience in farming,educational level and number of visits from bank officials as the significant factors. Gender and marital status had no relationship with loan repayment (Wongnaa and Awunyo-Vitor 2013).

A sample of 67 fishermen in Elimina who were randomly selected were used. A standard questionnaire was used to get data from the fishermen. Descriptive statistics and multiple regression analysis were the methods used. The results show that years of education, fishing income, years of fishing experience and loan amount had a positive influence on loan repayment whilst age of the fisherman and investment had a negative relationship with loan repayment(Acquah and Addo 2011).

Mashatola and Darroch (1995) used a logit model to analyse 83 medium scale farmers in Kwazulu-Natal in South Africa. Historical loan farmers were used. The farm size of these farmers had a positive relationship with loan repayment.

A study was conducted in Doma micro-financed farmers in repaying their loans. The loan repayment problem was a function of borrower characteristics, business farm characteristics and micro credit loan characteristics. A logistic regression was used to analyse demographic characteristics of the borrowers. He concluded his study that high farmers' experience leads to high loan repayment(Omini and Imam 2015).

Addisu (2006) studied factors that influence loan repayment in the informal sector. Structured interview,questionnaires and informal discussion was the how data was collected. He concluded that educational level of the borrower was statistically insignificant but small loan amount and diversion of funds increases the probability of loan default.

The aim was to test the superiority of a female in paying her loan. The independent variables used were age, educational level, loan size, interest rate, term of loan and loan cycle. Data was sourced from Opportunity International Savings and Loans, Sinapi Aba trust and Bosomtwe Rural Bank.The sampling used was random sampling of loan clients. The data 484 females and 270 males. The method used for the analysis were descriptive statistics and logistic regression. He concluded that age, loan size, interest rate, term of loan, and loan cycle were predictors of default. Men were better borrowers than females and educational level were found to be statistically insignificant(Afrane and Adusei 2014).

The aim of the research was to compare non-farmers and farmers in the repayment of loans. The scheduled repayment of loans was analysed using tobit model whereas the characteristics of the borrowers was analysed with variable probit model. Farmers have a higher loan repayment performance than non-farmers in Mekong Delta in Vietnam. Group loans was affected by the educational level of farmers and loan to farmers. Independent borrowers was affected by age and gender (Duy 2013).

(Saleem et al 2010) found out the impact of farmer and farmers' characteristics on loan repayment for agriculture. A total of 320 respondents were selected between 2007-2009 in D.I. Khan district using stratified sampling techniques. The results show that all the independent variables (age, education, marital status, farm type, farm size, and farm status) are a significant predictor of loan repayment for agriculture. The t-test and Anova table were used for analysis.

The study was conducted on 350 farmers of Savejbolagh through personal interviews. The data collection was done through stratified random sample. The purpose of their study was to find the impact of physical inputs and human capital factors on the income distribution of farmers in Iran. The logit results show that physical land for production of crops, fruits and animal unit had no effect on income distribution of farmers(Sadeghi and Toodehroosta 2002).

Addae-Korankye (2014) concluded by stating that poor loan appraisal, interest rate, poor monitoring, insufficient loan sizes and improper customer selection are the causes of loan default in microfinance institutions. He recommended that credit policies should be adhered and periodically reviewed.

Oke et al (2007) used multi-stage stratified random sampling technique for his data collection whereas multiple linear regression was used to analyse the data. The objective of the study was to find variables for micro credit repayment. He concluded his research with poverty hampering loan repayment. Income and loan repayment were found to statistically insignificant.

The study was to find the relationship between factors of the borrower and transactions cost with the view to identify specific component of the transaction cost that turn to impede access to credit. Regression analysis identified loan size and borrowing experience positively influence transaction cost of a credit (Masuko and Marufu 2003).

In a study by Igwe et al (2013) to find out determinants of transaction cost for borrowers among farmers, multiple regression analysis was used to analyse the research. Among the factors found to be statistically significant are distance to the bank, age, interest rate and loan amount.

In the study to find out the policy variables from Anambra state, discriminant analysis was used. Rural folks were surveyed to determine which factors influence their use of banks. Household income, years of formal education, gender, awareness of rural bank branch were significant(Okorie1992).

49 MFIs was used for the analysis. The study was to determine the effect of MFIs specific factors, self help groups specific factors and external factors on loan delinquency in MFIs in Kenya. The research first run individual factors on loan delinquency. Under external factors, economic downtown and weather conditions was statistically significant to defaults in MFI. The results show that a positive significant relationship between SHG's good governance and management on loan portfolio. Multiple regression was used in the data analysis.

The model had around predictive accuracy of 85 percent which indicates that independent variables explained the dependent variables well. Logistic regression was used to study the time for students to graduate and the factors which tend to classify these different types of students. Kidanekal and Assefa (2011) concluded that women have better chance to complete than men.

Kabhori et al (2010) studied failing tuberclosis completion course as dependent variable where sex, nationality, being in prison, type of TB,age and weight. A sample of 9672 were used where logistic regression was used for the analysis. All the variables studied were found to statistically explained the failing tuberclosis completion course. The logistic regression had 82.05 percent classification revealing how good the model was.

Manuel et al (1999) compared three methods(discriminant analysis,neural networks and logistic regression)for predicting distributions of species. He studied 32 variables which explains specie distribution. He concluded from this findings that in all the entire data, neural networks classified more cases than logistic regression and discriminant analysis though logistic regression had a predictive power of 83 percent. However when some of the independent variables were left out of the model, the logistic regression had 82 percent prediction which was higher than neural networks and discriminant analysis.

Memic (2015) classified companies into good borrowers and bad borrowers. He investigated factors that led to loan default. Logistic regression and multiple discriminant analysis were used and their results compared. Return on Assets was found to statistically explained loan default. Multiple discriminant analysis had a less predictive power than logistic regression.

Chapter 3

METHODOLOGY

3.1 Introduction

In this chapter, the main scientific methods employed in data collection required for the study are presented. It also gives description of data analysis to achieve the objectives of the study. The model adopted and it parameters estimation procedure are further discussed in this chapter.

3.1.1 Data Collection

Data to be employed for the study is a secondary data on random sample of 150 customers who have taken various loan facilities is from the Ahafo Ano Premier Rural Bank in the Ashanti region of Ghana. The data contains variables such as; age of customer, sex, type of loan, type of account, amount of loan, number of dependent, income and duration of loan repayment in months and the default status (yes for default customers and no for non-default customers).

For the purpose of this study, the default status (yes or no) will be the response of the dependent binary variable and the all other variables stated above are considered independent variables that would be used to predict the chance of a particular customer of the bank to default in his or her loan facility repayment from the bank.

3.1.2 Logistic Regression Models

These are category of statistical models belonging to the class of models referred to in general as Generalized Linear Models (GLM). This collection of models entails several models that are often used in statistical analysis with some assumption and conditions that must be satisfied. Some of the commonly employed and applied GLMs are regression models, Analysis of Variance (ANOVA) and several other essential multivariate models like ANCOVA etc. GLMs are well treated and presented by (Agresti, 1996).

Logistic regression is mostly used for prediction of outcomes that are discretized in nature for instance membership of groups, from a combination of predictor or explanatory variables that may be binary, categorical, metric or in some cases a blend of the above. The response or outcome variable is dichotomous examples of which include absent or present of a condition, dead or alive and many others. Unlike other generalized linear models that can not be employed in certain analysis where independent variables must be metric. Hence, in situations where predictors are a blend of continuous and categorical in nature, then the logistic model is the most desired among other linear models.

It is important to note however, that in case of multiple linear regression, the underlying idea is to obtain the least squares best fit line around which the Y values (response or dependent variable) are distributed. In contrast to that, with logistic regression model,the main aim is focused on determining the likelihood of an individual falling into one category of the response or the other. The idea of probability helps to interpret the coefficients in the logistic regression model in a meaningful manner similar to that the ordinary linear regression model.

3.1.3 The Logistic Model

logistic regression models are required in situations where the response variable is dichotomous or binary ,implying that, the response variable could assume a value of 1 with likelihood of success p, or the value 0 with likelihood of failure 1-p. This type of variable is called a Bernoulli (or binary) variable. The extended form of the binary logistic regression called the Multinomial logistic regression can be applied in instance whereby the outcome variable is more than two groups (Tabachnick and Fidell, 1996).However, the less emphasis is often given to the this subject matter compared to the binary logistic model.

As already mentioned earlier, on the characteristics of predictors of the logistic regression models, they can be categorical or metric in nature and thus do not require stringent assumptions as in the case of linear regression models where the assumption of normality is a very important requirement. Moreover, there is no linear relationship between the outcome and predictors, but rather, the logit transformation of P is used.

In the simplest case of one predictor X and one binary or dichotomous outcome variable Y, the logistic regression model predicts the logit of Y from X. The logit is the natural logarithm of odds of Y= 1 (the outcome of interest). The simple logistic model has the form:

$$\log(p/1-p) = \log(odds) = \beta_0 + \beta_1 x \tag{3.1}$$

Hence,

$$Probability(Y = outcome \quad of \quad interest|X = x) = \frac{e^{(\beta_0+\beta_1 x)}}{1+e^{(\beta_0+\beta_1 x)}} \tag{3.2}$$

Where P is the probability of the outcome of interest (or the "event") under variable Y, β_0 is the Y intercept, and β_1 is the slope parameter. Both the Y intercept and the slope parameter are estimated by the method of Maximum Likelihood (ML). The ML method is designed to maximize the likelihood of obtaining the data given its parameter estimates. As equation (3.2) illustrates, the relationship is nonlinear between parameters (β_0,β_1) and the probability of observing a particular outcome in an observation (a customer defaulting in loan facility repayment). Within the inferential framework, the null hypothesis states that P equals zero in the population. Rejecting such a null hypothesis implies that a relationship exists between X and Y. If the predictor is binary, such as gender, the exponentiated e^{β_1} is simply the odds ratio, or the ratio of two odds. The logistic function, that is, the g(x) in Equation 3.2, has the following unique characteristics as further illustrated in the figure above:

1. Unless P = 0, the binary logistic regression maps the regression line onto the interval (0,1) which is compatible with the logical range of probabilities.

2. The regression line is monotonically increasing if $\beta_1 > 0$, and monotonically decreasing if $\beta_1 < 0$.

3. The function takes the value of 0.5 at $x = \frac{-\beta_0}{\beta_1}$ and is symmetric to the point of $(\frac{-\beta_0}{\beta_1}, 0.5)$.

As these properties above demonstrate, the logistic regression model guarantees that (a) the predicted probabilities P will fall within the range of 0 to 1; (b) the slope parameter has the same meaning as the slope parameter in least squares regression models; and (c) the logistic function has a point of inflection corresponding exactly to 50% probability. Employing the same logic as that underlying the simple logistic regression, a complex model can be constructed to improve the prediction of the logit by including several predictors (explanatory variables). The complex logistic model is in the same form as multiple regression equations. It is given by:

$$\log(p/1-p) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k \tag{3.3}$$

Therefore it follows that Probability($Y = outcome \quad of \quad interest | X_1 = x_1, X_2 = x_2, ..., X_k = x_k$)

$$P(x) = \frac{e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_K X_K)}}{1 + e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_K X_K)}} \tag{3.4}$$

Where P is once again the probability of the "event" under the outcome variable Y, β_0 is the Y intercept parameter, β_s are slope parameters, and X_s are a set of predictors. Once again, β_0 and β_s are estimated by the ML method. The interpretation of β_s is rendered using either the odds ratio (for categorical predictors) or the delta - P (for continuous predictors). The null hypothesis states that all

β_s equal to zero. A rejection of the null hypothesis means that at least one of the co-efficients is significantly different from zero.

3.1.4 Assumptions of the Logistic Model

I The model do not rely on linearity assumption between the outcome and predictors variables.

II The outcome variable is binary in nature (2 categories)

III Predictors can be of any scale of measurement, no normality and variance assumption of predictors

IV Groups are non-overlapping and independent.

V For good, stable and reliable maximum likelihood estimates, large samples are often required . A recommendation of a predictor to 20 cases (observations)

3.1.5 Odds and Odds Ratio

The odds of an event happening is the probability that the event will happen divided by the probability that the event will not happen. Mathematically, if there is a P probability of an event happening, then the odds can be considered as the number of successes you expect to get for every failure on average. The odds of an event is given as

$$odds(event) = \frac{probability \quad of \quad success}{probability \quad of \quad failure} = \frac{p}{1-p} \tag{3.5}$$

The odds has a range of 0 to ∞ with values greater than 1 associated with an event being more likely to occur than to not occur and values less than 1 associated with an event that is less likely to occur than not occur.

The odds ratio is one of a range of statistics used to assess the risk of a particular outcome (or disease) if a certain factor (or exposure) is present. The odds ratio is a relative measure of risk, telling us how much more likely it is that an individual who is exposed to a certain factor under study will develop the outcome as compared to someone who is not exposed to the said factor (Westergrenet al,2001). The odds ratio which is the ratio of two odds can be expressed as

$$OR = \frac{odds(1)}{odds(2)} = \frac{(p_1/1 - p_1)}{(p_2/1 - p_2)} = \frac{P_1(1 - P_2)}{P_2(1 - P_1)} \tag{3.6}$$

3.1.6 Method of Estimation

Under this section we discuss the main estimation method that will be employed to estimate the (co-efficients) parameters of the logistic regression model. However, unlike the classical regression model where the ordinary least squares method is used, the Maximum likelihood estimation method or technique will be employed here. The method of maximum likelihood estimation (MLE) is a comparatively simple technique for estimation of the unknown parameter of interest θ. The MLE approach of model parameter estimation was identified and introduced in 1912 by one of the renowned English statistician called Fisher who is well recognized for his great contribution to the study of statistics. The method of maximum likelihood is intuitively interesting and attractive, because our main focus is to obtain parameter estimates of the true parameters that have characteristics or attributes of the real observed data. The Maximum Likelihood method performance better and efficiently with large data.

The maximum likelihood method basically implies choosing the parameter that makes the likelihood of having the observed data at hand maximum or optimal. For the case of discrete distributions, the likelihood is the same as the probability. According to Walpole et al, (2007), parameters are usually selected to ensure that, they at least maximize the probability density from which the observed data originates from

From the theoretical framework, the MLE method is employed in the absence of real data on random variables to yield the what is called the maximum likelihood estimate denoted by $\hat{\theta}$.In the presence of real (actual) data, the ML estimates assumes specific (real) values that seeks to maximize the estimator.

MLE demands us to maximize the likelihood function $L(\theta)$with respect to the unknown parameter$\theta.L(\theta)$ is defined as a product of n terms, which is a bit challenging to maximize. However,since the maximization of $L(\theta)$is the same as maximizations of the log $L(\theta)$ which is made possible due to the fact that, the log is a monotonic increasing function. We define log $L(\theta)$as log likelihood function, we denote it as $l(\theta)$ i.e

$$l(\theta) = \log L(\theta) = \log \prod_{i=1}^{n} f(X_i|\theta) = \sum_{i=1}^{n} logf(X_i|\theta) \tag{3.7}$$

Maximizing the $l(\theta)$ with respect to θ gives rise to the Maximum Likelihood Estimation MLE (Fisher, 1912).

3.1.7 Properties of MLE Estimators

The maximum likelihood estimates are most often referred over other estimators due to the under-listed properties:

i Estimates are asymptotically consistent.

ii The ML estimator is asymptotically efficient, implying that the variance of estimates are smallest relative to all other consistent estimators.

iii The ML estimator is asymptotically normally distributed, which justifies various statistical tests.

3.1.8 MLE for model parameters

The parameter estimation process begins first and foremost with identifying class GLM in which the response variables are measured on a binary scale. Example of the responses may be alive or dead, or present or absent. “Success ”and “failure ”are used as generic terms of the two categories. Next, we define the binary random variable

$$Y = \begin{cases} 1, & if \text{ the outcome is a } success \\ 0, & if \text{ the outcome is a } failure \end{cases} \quad (3.8)$$

with probabilities $P(Y = 1) = \pi$ and $P(Y = 0) = 1 - \pi$ which follows logically that, if there are n such random variables $Y_1, Y_2, ..., Y_n$ which are independent with $P(Y_j) = \pi_j$ then their joint probability is

$$\prod_{j=1}^{n} \pi_j^{y_i}(1-\pi_j)^{1-y_j} = \exp[\sum_{j=1}^{n} y_j \log \frac{\pi_j}{1-\pi_j} + \sum_{j=1}^{n} log(1-\pi_j)] \quad (3.9)$$

and this is a member of the exponential family. Next,for the case where the π_j's are all equal, we can define

$$Z = \sum_{j=1}^{n} Y_j \quad (3.10)$$

so that Z is the number of successes in n trials. The random variable Z has the distribution binomial (n, π):

$$P(Z = z) = \begin{pmatrix} n \\ y \end{pmatrix} \pi^y (1-\pi)^{n-y}, y = 0, 1, 2, 3, ...n \quad (3.11)$$

Finally,we consider the general case of N independent random variables $Y_1, Y_2, ..., Y_n$ corresponding to the number of successes in N different subgroups or strata. If $Y_i \sim binomial(n_i, \pi_i)$ the log-likelihood function,

$$l(\pi_1, \pi_2, ..., \pi_n; Y_1, Y_2, ..., Y_n) = \exp[\sum_{i=1}^{N} y_i \log \frac{\pi_i}{1-\pi_i} + n_i \log(1-\pi_i) + log \begin{pmatrix} n_i \\ y_i \end{pmatrix}] \quad (3.12)$$

and from Equation 3.4, we can express

$$\log \frac{\pi_1}{1-\pi_i} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k \quad (3.13)$$

and

$$\log(1-\pi_i) = -\log(1+e^{\beta_0+\beta_1 X_1+\beta_2 X_2+\ldots+\beta_k X_k}) \tag{3.14}$$

and the log-likelihood function becomes

$$l = exp[\sum_{i=1}^{N} y_i(\beta_0+\beta_1 X_1+\ldots+\beta_k X_k) - n_i \log(1+e^{\beta_0+\beta_1 X_1+\beta_2 X_2+\ldots+\beta_k X_k}) + \log \begin{pmatrix} n \\ y \end{pmatrix}] \tag{3.15}$$

Then taking the partial derivatives of the log-likelihood above with respect to the β_i's, and illustrating for the simple case where we have one independent variable(Rao,1973).

$$\frac{\partial l}{\partial \beta_0} = \sum_{i=1}^{n} y_i - n_i[\frac{e^{\beta_0+\beta_1 X_i}}{1+e^{\beta_0+\beta_1 X_i}}] = \sum (y_i - n_i\pi_i) \tag{3.16}$$

and

$$\frac{\partial l}{\partial \beta_1} = \sum_{i=1}^{n} y_i x_i - n_i x_i[\frac{e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}}] = \sum X_i(y_i - n_i\pi_i) \tag{3.17}$$

From which we proceed to obtain the second partial derivatives to enable as obtain the information matrix,

$$\begin{bmatrix} \sum n_i\pi_i(1-\pi_i) & \sum n_i x_i \pi_i(1-\pi_i) \\ \sum n_i x_i \pi_i(1-\pi_i) & \sum n_i x_i^2(1-\pi_i) \end{bmatrix}$$

Maximum likelihood estimates are then obtained by numerical iterative procedures called the Newton Raphson method.

3.1.9 Minimum Observation to Predictor Ratio

Just like other linear statistical models, a logistic regression model fitted from a sample of data is subject to some sampling errors. Consequently, estimates for the regression coefficients become unstable for small samples. So, the question often asked, what is the recommended minimum for the observation per predictor ratio? This question cannot be answered directly since most literature have not given clear or specific rules applicable to logistic regression models. Most rules we found were presented within the context of the ML method, for which some of the important desired properties are; assumption of normality, asymptotic efficiency and consistency. Furthermore, in terms of statistical test of significance, the greater the sample size, the better the X^2 approximation to the sampling distribution of ML estimators. Since ML is the method of choice for estimating logistic regression coefficients, we most at times rely on these rules to assess the adequacy of sample sizes used .

As Long (1997) pointed out in his study that, "Generally, it is dangerous to employ the method of maximum likelihood estimation in scenarios where observations are less than 100 but samples larger than 500 seems appropriate ". He further stated: "a rule of a minimum of one parameter to 20 observations is deemed adequate for logistic models ". In a related study, Lawley and Maxwell (1971) suggested that the significance test of the ML factor analysis solutions is appropriate if the sample contains at least 51 more cases than the number of variables under consideration. That is, $50 < N - k - 1$, where N is the sample size and k is the number of variables (or predictors in the case of logistic regression). This is only a general rule of thumb, as Kim and Mueller (1978) correctly noted. Even though the required minimum number of cases per response variable ratio to accomplish steadiness of parameter estimates differs from one author to another, numerous researcher do recommend the least ratio of 10 to 1 with a minimum sample size of 100 or 50 plus a variable number that is a function of the number of predictors.

It is well noted that, researchers in general agreed that (a) the ML estimators of logistic regression coefficients would be stabilized with large samples, (b) certain regression models and/or data structures seem to require even larger samples, and (c) a conservative significance level should be adopted as evidence against the null hypothesis in small samples (Allison, 1995; Long, 1997).

3.1.10 Evaluations of Logistic Regression Models

Once a logistic model is formulated and fitted to data set under study, its adequacy is evaluated by a variety of statistical tests and indexes proposed by pioneers in this field. These include among others:

(a) Tests of individual parameter estimates

(b) Tests of the overall model,

(c) Validation of predicted probabilities, and

(d) Goodness-of-fit statistic.

Individual parameter estimates are tested by the likelihood ratio test, the Wald statistic, or the Score test. According to Jennings (1986), Long (1997), and Tabachnick and Fidell (1996), the likelihood ratio test is more powerful than the Wald test, while the Score test is a normal approximation to the likelihood ratio test. A logistic model is said to be a good-fit to the data if it shows some level improvement beyond the intercept or constant - only model (also called the null model). Such an improvement is scrutinized by theses tests: the likelihood ratio, Score, and Wald tests.

Several studies have opined that, validation of predicted probabilities is usually presented in terms of percentages of correct classifications, Somers' D statistic, sensitivity, specificity, false positive, false negative, or concordance pairs. The goodness-of-fit statistic is reported either as chi-square tests or the R^2 type of indexes. The chi-square test is based on Hosmer and Lemeshow statistic also referred to as the deviance. The Hosmer and Lemeshow statistic is both conservative and sensitive to the way in which predicted probabilities are grouped (Hosmer and Lemeshow, 1989; Ryan, 1997). Another school of thought however suggests that, Deviance - based chi-square tests are correct only if they are based on the difference between two deviances rather than the deviance a model alone (McCullagh and Nelder, 1989).

The methods adopted to ascertain whether to include or not to include a parameter estimate based on their relevance or significance in the model are presented and discussed below.

3.1.11 The Wald Statistic

This test is used to assess the statistical significance of each coefficient (β) in the model. A Wald test calculates a Z statistic, which is:

$$Z = \frac{\hat{\beta}}{SE(\hat{\beta})} \tag{3.18}$$

This Z value is when squared results in a Wald statistic with a chi-square distribution. However, several authors have noticed some problems with the use of the Wald statistic.For example, Menard (1995) warned that for large coefficients, standard errors are often is inflated in some way, there by lowering the Wald statistic value. Agresti (1996) also stated that, the likelihood-ratio test is more reliable than the Wald test when dealing with small sample sizes of data.

3.1.12 Likelihood Ratio Test

This test uses the ratio of the maximized value of the likelihood function for the final model (L_1) over the maximized value of the likelihood function for the simpler model (L_0). The likelihood-ratio test statistic equals

$$-2\log(\frac{L_0}{L_1}) = 2(\log L_1 - \log L_0) \tag{3.19}$$

The log transformation of the likelihood functions results into a chi-squared statistic. This test procedure is mostly preferred over other statistics when the model is obtained by the backward stepwise elimination approach.

3.1.13 Hosmer and Lemeshow Test

The Hosmer-Lemshow test is mostly used to assess the appropriateness of estimated model to the data by constructing ten ordered sets of subjects and then relates the observed and predicted frequencies in the each of the sets formed. Hence indicating that, the test statistic depicts a chi-square statistic with a desired non-significance conclusion result to illustrate the prediction adequacy of the model since the observed and predicted values of groups are similar. The hypotheses involved are stated below:

H_0 : The Model fits the data well

versus

H_1: The Model does not fit the data

Hence, the rejection of the null hyothesis, H_0 implies that logistic model to the data is not a good fit to the data and the reverse indicates a good fit to the data.

Chapter 4

Data Analysis and Results

4.1 Introduction

This chapter discusses the results of the data used for the study of loan default customers of the Ahafo Ano Premier bank for this study. An extract of the summary of some characteristics of customers and their associated demographics are presented and discussed below in the empirical Sessions and more detailed results in further analysis session.

4.2 Empirical Data Analysis

The study randomly obtained information on 152 customers of the Ahafo Ano Premier Rural bank to explore their characteristics. The results presented exclude all missing observations under any other of the variables in this study.

Table 4.1: Characteristics of customers

Variable	Response	Frequency	Percentages (%)
Default Status	no	78	52.0
	yes	72	42.0
sex	Female	83	55.0
	male	68	45.0
marital Status	single	84	55.6
	married	67	44.4
Location	urban	62	41.1
	rural	89	58.9
Type of account	susu	64	42.4
	savings	4	2.6
	current	83	55.0
Type of loan	salary	116	77.8
	susu loan	11	7.4
	commercial	22	14.8

From the total valid customers of 150, 48 percent of the customers had defaulted on their loan repayment and the other 52 percent did not. These customers were made of a slightly more females who constitute 55 percent and 45.0 percent representing males. On marital status of these customers, the proportions are 44.4 and 55.6 percent respectively of the married and single group. The bank served 58.9 percent of rural residents and 41.1 % of customers residing in the urban areas. The results also indicates that, customers of the bank operates various accounts most commonly are susu, savings and current accounts representing 42.4 percent , 2.6 percent and 55 percent respectively. It also shows most of the loans offered by the bank were salary loans accounting for 77.8 percent, commercial accounts were 14.8 percent and susu loan recorded just 7.4 percent. On the age distribution

Table 4.2: Age Statistics of Customers

	Statistic	S.E
Mean	45.52	.953
95% C.I Lower Bound	43.63	
95% C.I Upper Bound	47.40	
Median	46.00	
Variance	137.065	
Std. Deviation	11.707	
Minimum	24	
Maximum	72	
Skewness	-.010	0.197
Kurtosis	-1.097	0.392

of the cherished customers of the bank, the results shows the minimum age of these customers is 24 years and a maximum age of 72 years. The mean age of the customers recorded a value of 45.52 years and a median age of 46 years. The mean and median age of the customers indicates that, the age distribution of the customers of the bank is slightly skewed to the left with a corresponding skewness value of -0.010. And the kurtosis value of -1.097 which measures the degree of peakness of age further suggests that, the ages are flattened than that of the normal distribution.

The income distribution of customers from the table above shows a minimum income earned among the customers is 50.0 cedis and the maximum income of 100,000 cedis. The mean and median incomes of 4,753.32 and 540.50cedis respec-

Table 4.3: Income Statistics of Customers

	Statistic	S.E
Mean	4,753.32	1070.417
95 % C.I Lower Bound	2,638.16	
95 % C.I Upper Bound	6,868.48	
5% Trimmed Mean	2,442.72	
Median	540.50	
Variance	171,869,032.454	
Std. Deviation	13,109.883	
Minimum	50	
Maximum	100,000	
Skewness	5.164	0.198
Kurtosis	30.039	0.394
Kurtosis	-1.097	0.392

tively which are far apart suggests, the incomes of these customers are skewed to right with a skewness value of 5.164 which is very significantly different from zero for Symmetric distributions. Security and collateral for loan facilities are very im-

Table 4.4: Security of Loans

Type of security	Frequency	Percentage(%)
	1	0.7
Building	15	9.9
Car	7	4.6
Guarantors	104	68.4
House	3	2.0
Lien	15	9.9
Lien and Guarantors	1	0.7
Salary	1	0.7
Salary and Guarantors	5	3.3

portant to the financial institutions and banks alike, the securities or collaterals provided by customers of the bank in this study included among others; Buildings (9.9%), Cars (4.6 percent), Guarantors (68.4 percent) salary (0.7 percent) and salary and Guarantors (3.3 percent).

4.3 Further Analysis

This section presents results on some key factors that can contribute to a customer defaulting in a loan facility repayment when such periods are due. Chi - square test statistics are first presented with the cross - tabulations of such factors. From the cross - tabulations of Sex of customer and default status as shown below in Table 4.5. The results show that, from the 82 females, 36 presenting 46.9 percent defaulted from their loan repayments whiles, remaining 53.1 percent did not default. Contrary to that of the males where from a total of 68, 36 representing 52.9 percent and 47.1 percent however, did not default from their loan commitment. This shows that, males are more likely to default than their females counter parts.

Table 4.5: Cross - tabulation of Sex and Default Status

Sex/Default Status	Yes	No	Percentage Correct
female	46 (56.1%)	36(43.9%)	82
male	32(47.1%)	36 (52.9%)	68
Total	78 (52.0%)	72 (48.0%)	150

Though there exist some differences in the proportions of males and females who default in their loan repayments however, the Chi - square test of independence shows that, the difference are not significant between males and females with regards to defaulting in loan repayment. Thus, with a Chi - Square value of 1.217 and p - value of 0.270, it implies defaulting in a loan repayment does not depend on the sex of the customer. On the cross-tabulation of loan type

Table 4.6: Chi-sqare Statistics

	Value	df	Asynp.Sig.(2-sided)
Pearson Chi-square	1.217	1	0.270
Continuity Correction	0.882	1	0.348
Likelihood Ratio	1.218	1	0.270
N of Valid Cases	150		

and default Status, the results shows that, for customers with salary loans, 37.1 percent of them defaulted as compared with 85.7 and 100 percent of commercial and susu loan customers who defaulted in repayment of their loans. The results also suggest that, there is a high default rate associated with commercial and susu loan seekers from the bank.

The Chi - square test with a statistic value of 29.386 with degrees of freedom

Table 4.7: Cross-tabulation of Loan type and Default Status

Loan Type/Default Status	Yes	No	Total
salary	73(62.9%)	43(37.1%)	116
susu loan	0(0.0%)	11(100.0%)	11
commercial	3(14.3%)	18(85.7%)	21
Total	76(51.4%)	72 (48.6%)	148

2 and a corresponding $p - value < 0.001$ shows that, there exist an association between loan type and defaulting in a loan repayment. Hence, customers who seek for susu and commercial loans from the financial institution are more likely to default in payment than customers on salary loans. With the identified as-

Table 4.8: Chi-square Statistics

	Value	df	Asymp. Sig.(2-sided)
Pearson Chi-square	29.386	2	0.00
Likelihood Ratio	34.876	2	0.00
Linear-by-Linear Association	23.542	1	0.00
N of Valid Cases	148		

sociation in loan type and default status of the customer, it worth assessing the strength of association that exists between them. The Cramer's values of 0.446 with a p- value of 0.000 indicates that, the association between loan type and default status is relatively a strong association(Rea and Parker, 1992).

Going further, the study will also explore the association between marital status of customers and default status. The cross - tabulation below indicates that, out of the 84 customers who are single, 34 (40.5 percent) did default in repayment of the loan facilities they went in for whiles, with the married customers 57.6 percent defaulted in repayment. This suggests that, married customers are on the whole risky in terms of defaults as compared to single customers.

The difference in percentages between married and single customers of the bank

Table 4.9: Cross -tabulation of Marital Status and Default Status

Marital Status/Default Status	Yes	No	Total
Single	50(59.5%)	34 (40.5%)	84
Married	28(42.4%)	38(57.6%)	66
Total	78 (52.0%)	72(48.0%)	150

respective of their default status on loans, the test statistics from the Chi - square test in Table 4.11 with a value of 4.330 with a p- value of 0.037 implies that, there is an association or dependence between marital status and loan default. Thus, indicating that married customers are more likely to default than customers who are single. The symmetric measures tables below with a significant Phi co-efficient

Table 4.10: Chi-sqare Test Statistics

	Value	df	Asymp. Sig.(2-sided)
Pearson Chi-square	4.330	1	0.037
Continuity Connection	3.672	1	0.055
Likelihood Ratio	4.347	1	0.037
N of Valid Cases	150		

value of 0.170 however, suggests that, the association between marital status and loan default is a weak association as opined expressed by (Rea and Parker, 1992).

4.3.1 Logistic Regression analysis of loan default

In this section of the study, logistic regression model is used to model the default of customers with some risk factors that contribute to the likelihood of customers defaulting in repayment. The model is obtained by using 100 cases and the remaining cases are used to validate the model to ascertain its goodness - of - fit on the data.

Table 4.11: Dependent Variable Encoding

Original Value	Internal Value
No(Non default)	0
Yes(Default)	1

Table 4.12: Initial Classification Table

Ob./Predicted Status	Yes	No	Percentage Correct
Yes	0	48	0.0
No	0	49	100.0
Overall Percent			50.50

From the beginning block, the proportion by chance accuracy is $0.495^2 + 0.505^2 = 0.5001$ which is 50.01% of grouping and prediction.

Table 4.13: Omnibus Tests of Model Coefficients

	Chi-square	Df	Sig.
Model	52.658	6	0.000

Statistical significance level of the overall model

The Omnibus Tests of Model Coefficients for the independent variables entered in block 1 above has a chi-square value (53.559) with a corresponding significance level (0.000) less than 0.05. Hence the existence of a significant relationship between the response and independent variables and the null hypothesis that there is no difference between the model with only the constant and the model with the independent variable is rejected. Thus the hypothesis tested stated as follows;

$H_0 : \beta_i = 0$

Vrs

$H_0 : \beta_i \neq 0$

Table 4.14: Model Summary

-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
81.802	0.419	0.559

4.3.2 Hosmer and Lemeshow Test

This is a reliable goodness of fit test of the model. The model is a good $-$ fit of the data when the significance value is greater than 0.05.

Hypothesis testing

H_0: The model fits the data

H_a: The model does not fit the data

From the table above, the Hosmer and Lemeshow $p\!-\!value$ of 0.973 is greater than 0.05. Therefore the null H_0 is not rejected and we conclude that, the observed number of customers who default in loan repayment are not significantly different

Table 4.15: Hosmer and Lemeshow Test

Chi-square	Df	Sig.
2.633	8	0.955

from those predicted by the model and hence the overall model is a good fit of the data.

From the beginning block, the proportion by chance was estimated at 47.05

Table 4.16: Final Classification Table

Ob./Predicted Status	Yes	No	Percentage Correct
Yes	41	7	85.4
No	12	37	75.5
Overall Percent			50.50

percent and for the overall accuracy rate of the final model classification should be at least 25 percent more than the proportion by chance. The final model overall accuracy of 79.4 percent is more than $(1.25 \times 0.4705) = 0.5881$ which is 58.81 percent. Thus the final classification is accurate for the expected and observed cell frequencies.

The variables in the logistic model are selected based on their significance levels which are less than 0.05. From the standard errors of the independent variables, the problem of multi - collinearity to the model is checked. And since the standard errors are less than 2, there is no problem with multi - collinearity in the model. From the table below with the variables in the equation, the final model for the prediction of defaulting status (default)of the customers is given as;

$$Log(Odds) = 3.863X_1 + 0.088X_2 - 0.234X_3 \tag{4.1}$$

And for predicting the probability of a particular case we can use

$$\pi = \frac{e^{3.863X_1+0.088X_2-0.234X_3}}{1+e^{3.863X_1+0.088X_2-0.234X_3}} \tag{4.2}$$

Where;

X_1 = type of loan (commercial)

X_2 = Loan repayment period (duration) in months

X_3 = Number of Dependents of customer

Table 4.17: Test of Significance of Coefficients

Variable	β	S.E	Wald	Df	Sig.	Exp(β)	95% L.B	95% U.B
Type of Loan			15.936	1	0.000			
Loan(Commercial)	3.863	0.968	15.936	1	0.000	47.632	7.147	317.465
Payment period	0.088	0.032	7.731	1	0.005	1.092	1.026	1.162
Number of Dependents	-0.234	0.117	3.985	1	0.046	0.791	0.629	0.996
Marital Status(1)	-1.685	0.892	3.567	1	0.059	0.185	0.032	1.066
Age	-0.034	0.031	1.182	1	0.277	0.967	0.909	1.028
Constant	0.344	1.633	.044	1	0.833	1.411		

The logistic regression model obtained above was used to classify the left out sample of 51 customers of the bank for the purpose of validating and assessing the performance of the fitted model. From the validation results presented in Table 4.20, it is observed that, 67.86 percent of non- defaulting customers were correctly classified and 69.57 percent of loan defaulting customers were also correctly classified. The overall percentage correct classification of the fitted model is 68.63 which 35 out of the 51 left out sample.

Table 4.18: Validation of Model

Ob./Predicted Status	Yes	No	Percentage Correct
Yes	19	9	67.86
No	7	16	69.57
Overall Percent			68.63

4.3.3 Sumary of the model

A logistic regression analysis to predict the default status (default or non- default) of customers of a bank involving 100 customers using variables such as amount of the loan, age of the client, type of loan facility, marital status, number of dependents, repayment period and many others as predictors for each customer. A test of the full model against the constant only term is statistically significant ($0.000 < 0.05$), indicating that, the predictor variables significantly distinguished between defaulting customers a loan and non- defaulting customers of the bank under study ($Chi - square = 52.685$, $p < 0.000$ with $df = 7$).

The Nagelkerke's R - square of 0.559 indicates a moderately strong relationship between prediction and grouping. The overall success of prediction is observed to be 80.41 percent (85.4 percent for non- defaulting customers and 75.5 percent for defaulting customers).The Wald criterion for including variable in the final model shows that number of dependents, repayment period, loan type (commercial)made significant contributions to the model with significance level less than 0.05 each as in the Table 4.19.

Chapter 5

Summary of Findings, Conclusion and Recommendations

5.1 Introduction

As a concluding chapter to this research work, this section of the report provides the summary of findings from the study, conclusions drawn from the analysis of data collected on customers of Ahafo Ano Premier Rural Bank also suggests some recommendations with respect to the major findings and conclusion from the study.

5.2 Summary of Findings of the Study

From a total 152 customers of the Ahafo Ano Premier Rural Bank who were randomly selected for the purpose of this study in line with the intended objectives mentioned earlier, an exploratory data analysis was carried out on the characteristics of both defaulting and non - defaulting customers of the bank under study. Results from the analyses showed that 48 percent of the customers had defaulted on their loan repayment and the other 52 percent did repay their loan facilities when the period for repayment was due. The results of this study also show that of customers of the bank were both rural urban folks with the rural customers representing 59.2 percent and 40.8 percent were urban dwellers.

With regards to factors that influence the likelihood of a customer of the bank to default in a loan facility repayments, the Chi - Square test of independence of association between the default status of customers and factors or characteristics of the customers such as; sex, type of loan taken, duration of the loan, marital status, number of dependent of customers among other were considered and tested.

The results show that though males were slightly more likely to default in a loan repayment with a probability of 0.529 as compared females with a probability of 0.469 but the Chi - square test statistics showed no difference in proportion with respect to defaulting was not significant with a $p-value$ of 0.270.suggesting that sex of customer and defaulting in a loan repayment are independent.

With a Chi-square test value of value of 29.386 and a corresponding $p\hat{a}value < 0.001$, there exist an association between loan type and defaulting in a loan repayment where 85.7 and 100 percent of commercial and susu loan customers defaulted respectively in repayment of their loans as compared 37.1 % of salary loan customers. Hence, customers who seek for susu and commercial loans from the financial institution are more likely to default in payment than customers on salary and the association classified as relatively strong from the Cramer's value of 0.446 (Rea & Parker, 1992).

Defaulting status was identified to be significantly associated with marital status of customers with married customers been more likely to default than their single counterparts but, the association between marital status and loan default showed a weak association with a Phi co- efficient value of 0.170 as expressed by (Rea & Parker, 1992).

The binary logistic regression model fitted to the data to predict the status of customers based on their characteristics, significantly distinguished between defaulting customers and non - defaulting customers of the bank under study ($Chi-square = 52.658, p < 0.000 with df = 7$). The significant variables in the model were selected based on their p - values and the Wald test criterion and included among others; number of dependents, repayment period, and loan type (susu). The Nagelkerke's R-square of 0.566 indicates a moderately strong relationship between prediction and grouping. Hence, overall success of prediction of the model is observed is 80.41% (85.4% for non- defaulting customers and 75.5%

for defaulting customers).

And for predicting the probability of a particular case we can use

$$\pi = \frac{e^{3.863X_1+0.088X_2-0.234X_3}}{1+e^{3.863X_1+0.088X_2-0.234X_3}} \tag{5.1}$$

Where;

X_1 = type of loan (commercial)

X_2 = Loan repayment period (duration) in months

X_3 = Number of Dependents of customer

5.3 Conclusion

Defaulting status was identified to be significantly associated with marital status of customers with married customers been more likely to default than their single customers.

In the conclusion, the study on default status and the contributing factors for a customer to default in a loan repayment identified commercial loans as high risk loans that could not be repaid on time. Customers who had a smaller number of dependents also had high probability of defaulting as compared with customers with more dependents since customers with more dependents and odds of 0.7913. Customers with more months (duration) to repay their loan facilities were also observed as potential defaulters and finally the results show that married customers are more likely to default than single customers.

The model corrected predicted and classified 68.6% of the testing sample (cases)

5.4 Recommendations

Based on the conclusions arrived at in this study, the following recommendations are made to management of Ahafo Ano Premier Rural Bank, other financial institutions, stakeholders and customers as well in the financial and business environment;

1. The Ahafo Ano Premier Rural Bank should put in place measures to reduce the occurrence loan default for commercial and susu loan facilities more specifically.

2. The Ahafo Ano Premier Rural Bank should advance more loan facilities to salary workers since they are less likely to default.

3. Customers of the bank should endeavor to repay their loan facilities so as to keep the bank in business for the benefit of other customers.

4. Further research in the future should consider the comparison approach of classification methods.

REFERENCES

Acquah,H and Addo,J(2011). Determinants of loan repayment performance of fisherman; empirical evidence from Ghana, XLIV

Adem ,O and Crichuhi , A and Otieno, R(2012). Parametic modeling of probability of bank loan default in Kenya JAGST, 2012,14

Addisu, M(2006). Microfinance repayment problems in the informal sector in Addis Ababa, Ethopian Journal of business and development.

Afrane, S.K and Adusei, M (2016). The female superiority in loan repayment Hypothesis is it valid in Ghana? Advances in management and applied economics, 2016 Pg 4

Agresti, A. (1996). *An introduction to categorical data analysis* (Vol. 135). New York: Wiley.

Ajair Nair, A.F (2010). Rural banking: The case of rural and community banks in Ghana, World Bank Agriculture and rural development discussion paper.

Alary. D and C.Goller (2001). Strategic default and penalties on the credit market with potential judgment errors, EUI working paper,

Alex Addae-Korankye (2014). Cause and control of loan default/delinquency in microfinance institutions in Ghana, American international journal of contemporary research Vol.4 No.12 December 2014

Allison, P. D. (1995). *Survival Analysis Using the SAf System*: A Practical Guide. Cary,NC: SAS Institute Inc.

Anyanwu J.C (1990) Monetary Economics Theory policy and institution. Hybrid publishers ltd PP 247-274

Auronem, L(2003) Asymmetric information: Theory and Applications Helsinki University of Technology.

Bicharga,D.W.O and Aseyo,M.L (2013)– causes of loan default within microfinance institutions in Kenya interdisciplinary journal of contemporary research in Business,Pg 4

Blanchard D. Dionme G.(2003). Risk management and corporate governance working paper, Pg 3-4.

Bofondi, M and Gobbi, G (2003). Bad loans and entry in local credit markets, Bank of Italy research.

Bourakba,C and Zerargui, H(2015). The relationship between credit risk and corporate Governance in Islamic banking. Issues in business management and economics, Vol. 3.

Dunso,B.H and Dadzie,K.Q (1990). The determinants of loan default and delinquency in rural credit program in Ghana, USAID.

Duy, V.q (2013). Is the repayment performance of farmers better than that of Non-Framers? A case study of borrowers of formal Bank credit in the Mekong Delta, Vietnam, CAS Discussion paper N8,88

Evelyn Richard (2011). Factors that cause Non-Performing loans in commercial banks in Tanzania and strategies to resolve them journal of management policy and practice vol. 1/2/7, 2011

Ekumah, E.K and Essel, T.T(2003) Information is power , the problem with credit accessibility in rural banks in Ghana, International Labour Organization.

Gabriel, A. (2009). Lesson from the current Financial crisis; A risk management Approach, Review of Economic and Business studies, Pg 3

Hongli, L and Liwei,S (2005). Assesment of credit risk of real estate based on logistic regression model, Atlantic press.

Hosmer, D. W., Jr., and Lemeshow, S. (1989). *Applied Logistic Regression*. New York:John Wiley and Sons, Inc.

Jennings, D. E. (1986).*Judging inference adequacy in logistic regression*. Journal of the American Statistical Association 81: 471-476.

Jappelli, T and Marco (2005). Role and effects of credit information sharing, centre for studies in economics and finance.

Igwe, K and Egbuson, O (2013). *Determinants of transaction cost for borrowers among farmers in Ikwuano local Government Area, Abia State Nigeria*. American journal of rural Development, Vol 1

Kalhori, S.R.N, Nashi, M and Zeng, X.J(2010). *A logistic regression model to predict high risk patients to fail in tuberculosis treatment course completion*, LAENG International Journal of Applied mathematics

Kohansal, M.R And Mansoori, T (2009). Factors affecting of loan repayment performance of farmers in Khorasan-Razari province of Iran, Conference on international research and food security, 2009

Kidanekal, H and Assefa, E (2001). Degree prediction using logistic regression, centre of mathematical science

Kim, J. O., and Mueller, C. W. (1978).*Factor analysis:* Statistical methods and practical issues. Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 07-014. Beverly Hills and London: Sage Publications**.**

Lawley, D. N., and Maxwell, A. E. (1971).*Factor Analysis as a Statistical Method*.London: Butterworth and Co.

Long, J. S. (1997). *Regression Models for Categorical and Limited Dependent Variables*.Thousand Oaks, CA: Sage Publications, Inc.

Masuko,L and Manfu, D (2003). The determinants transactions cost and access to credit by small and medium (SMEs) and poor farmers in Zimbabwe, International labour Organisation,

Mannel, S., Dias, J.M, And Omerod, S.J (1999). Comparing discriminent analysis, neural networks and logistic regression for predicting specie distribution, Ecological modeling, Pg 120

Mashatola, M. and Darroch, M (2002). Factors affecting the loan status of sugar cane farmers using a graduated mortgage loan repayment scheme in Kwazulu – Natal,iran

Memic,D (2015). Assessing credit default using logistic regression and multiple discriminant analysis ; Empirical evidence from Bosnia and Herzegovinia, Interdisciplinary description of complex systems, 2015,

Nguta, M. H. And Huka,G.S (2013). *Factors influencing loan repayment default microfinance*: The experience of Imenti North district, Kenya, international journal of applied science and technology, 2013 Pg 3.

Oni, O, Oladele,O and Oyewole,I.K(2005). Analysis of factors influencing loan default among poultry farmers in Ogun state Nigeria, journal of central European Agricultural, Pg 6

Quaryme C. R, And Ababio, O. (2014). *The determinants of Bank interest rates spread in Ghana*, International Journal of economic behavior and Organisation

Oke, J, Adeyemo,R And Agbonlahor,M(2007). *An empirical analysis of micro credit in Southwestern Nigeria*, Humanity and social science Journal,Vol 2.

Okorie, A (1992). Rural banking in Nigeria: Empirical evidence of indicative policy variables from Anambra state, Agricultural economics.

Pagano, M AND Jappeli, T.(1993). Information sharing in credit markets, Journal of finance in credit markets.

Padilla, A and Marco (1997). Endo Communication among lenders and entrepreneurial incentives. The review of financial studies

Rich, J and Tange, C.(2003). Credit Risk measurement, A portfolio Review

Stiglits, J.E and Weiss,A(1981). Credit rationing in markets with imperfect information, The American Economic review, Pg 71

Slynotzky, A.J and Drzik, J(2005)countering the biggest risk of all, Harvard Business Review, 2005 Pg 82

Sanders A (1997). Financial institutions management: A modern perspective Irwin inc.

Sadeghi,J.M and Toodehroosta, M (2002) Physical and human capital factors affecting income distribution among farmers of sarejbolagh township, Iran, METU/VI conference, 2002.

Sarlija,N, Soric,K and Vlah S(2012). logistic regression and multi - criteria decision making in credit scoring, University of Zagreb, Faculty of Economics, 2012

Soureslijahni,M.H and Kimiagari,A.M (2013). Calculating the best cut off point using logistic regression and neural network on credit scoring – A case study of a commercial bank, 2013, Pg 7

Uyemura, D and Deventer, D(1993). Risk management in Banking: Theory and Application of Assets and Liabilities management, Banking Publication

Wameru,N.M and Kalani,V.M (2009). *Commercial banking crises in Kenya: causes and remedies* Global journal of finance and banking issues, vol (3)

Wongnaa,C and Awinyo – Vitor D (2013). *Factors affecting loan repayment performance among yam farmers in the sene district, Ghana*. Agric on-line papers in economics and informatics, vol(5)

Postnora, A (2012). Does Good corporate Governance reduce credit risk? Hanken school of economics,

Switzer L,N and Sachnt, J,M(2015). Bank credit risk and corporate Governance Autorite des Marches Financiers,

Tabachnick, B. G., and Fidell, L. S. (1996).*Using Multivariate Statistics, 3rd ed*. NewYork: Harper Collins.

Walpole, R.E, Myers, R.H, Myers, S.L and Ye, K.(2007). *Probability & Statistics for Engineers & Scientists*, 5th ed. Upper Saddle River, N.J: Prentice Hall.

Westergren A., Karlsson S., Andersson P., Ohlsson O., Hallberg I.R. (2001).*Eating difficulties, need for assisted eating, nutritional status and pressure ulcers in patients admitted for stroke rehabilitation*. Journal of Clinical Nursing 10, 257-269