Strathmore UNIVERSITY

Measuring the effect of characteristics of loan obligors on default risk in a portfolio of individual loans using Generalized Linear Models

Masimane Mercy Njeri 070851

Submitted in partial fulfilment of the requirements for the Degree of BBS Financial Economics at Strathmore University

School of Finance and Applied Economics
Strathmore University

Nairobi, Kenya

November, 2015

Strathmore UNIVERSITY

SCHOOL OF FINANCE AND APPLIED ECONOMICS

RESEARCH PROJECT SUBMISSION FORM

This form must be completed and handed to the School at the time of submission of the Three (3) spiral-bound copies of the project.

A. To be Completed by the Student

Student ID and Name: 070851 MERCY NJERI MASIMANE	
Programme of Study: BBS FINANCIAL ECONOMICS	
Title of Project/Dissertation: MEASURING THE EFFECT OF CHARACTERISTICS OF LOAN OBLIGORS ON DEFAULT RISK IN A PORTFOLIO OF INDIVIDUAL LOANS USING A GENERALIZED LINEAR MODEL	
Word Count: 7,647	
Signature:	Date: 17/11/2015

B: To be Completed by the Supervisor(s)

Name(s): DR. JOHN OLUKURU.	
Comment(s) on the Supervision Seen and approved for submission.	
Signature:	Date: 17/11/2015.

C: For Office Use Only

This is to certify that Mr/Ms/Mrs .... has submitted .... spiral-bound copies of project on .... at the School of Finance and applied Economics

....

Officer's Name **Officer's Signature**

DECLARATION

I declare that this work has not been previously submitted and approved for the award of a degree by this or any other University. To the best of my knowledge and belief, the Research Proposal contains no material previously published or written by another person except where due reference is made in the Research Proposal itself.

...MERCY NJERI MASIMANE... [*Name of Candidate*]

.... [*Signature*]

..........17\11\2015.............................. [*Date*]

This Research Proposal has been submitted for examination with my approval as the Supervisor.

..........JOHN OLUKURU............ [*Name of Supervisor*]

.... [*Signature*]

..............................17/11/15...... [*Date*]

School of Finance and Applied Economics
Strathmore University

Table of Contents

Abstract

A large part of bank profits is earned from the interest charged on the different loans they advance to their customers; therefore proper management of credit risk is crucial for commercial banks. This paper studies the characteristics of loan obligors in a portfolio of individual loans in a German Bank and how these characteristics contribute to the individual's probability of default. The study mainly seeks to determine the cross sectional dependence of default rates on various characteristics of loan obligors. The findings of this paper will enable banks to build credit scoring models based on the characteristics of loan obligors in order to better manage credit risk. Eighteen characteristics are considered in this study but only the five most significant attributes are analysed. The research found that loan obligors with the highest probability of default have the following set of characteristics: male, have other payment plans, have been employed for a longer period, have low loan repayment duration and have a low checking status.

Key words: generalised linear model, cross sectional dependence, credit risk, probability of default.

CHAPTER 1

1.0 Introduction

1.1 Background

Credit risk is the risk of an economic loss from the failure of a counterparty to fulfill its contractual obligations (Jorion, 2009). Credit risk involves the possibility of nonpayment, either on a future obligation or during a transaction.

The goal of credit risk management is to maximise a bank's risk-adjusted rate of return by maintaining credit risk exposure within acceptable parameters. Banks need to manage the credit risk within the entire portfolio as well as the risk in individual credits or transactions. The effective management of credit risk is a critical component of a comprehensive approach to risk management and essential to the long-term success of any banking organisation (Basel Committee on Banking Supervision, 2004).

The importance of credit risk, as part of financial risk analysis, comes from the New Basel Capital Accord - Basel II, published in 1999 and revised in 2004 by the Basel Committee for Banking Supervision.In June 1999, the Basel Committee on Banking Supervision released a proposal to replace the 1988 Basel Capital Accord with a more risk-sensitive framework.

The New Basel Capital Accord (2004) proposed new regulatory rules for banks' capital adequacy evaluations. The main innovations relate to credit and operational risk. In terms of credit risk, the NBCA revised the 1988 Accord by proposing a more risk-sensitive methodology for assessing the default risk of banks' clients. The risk inputs entering the final capital adequacy computations were closely related to the risk characteristics of individual bank clients. In this sense, the proposed methodology opted for the adoption of ratings, developed by external agencies or by banks themselves, in quantifying and signalling to the bank the default risk of individual borrowers.

During recent years, financial institutions have devoted important resources to build statistical models to measure the potential losses in their loans portfolios (Dietsch & Petey, 2002). Banks and also other financial institutions must make successful loans that are paid back in full in order to earn high profits. Adverse selection and moral hazard increase the likelihood of default risk. Adverse selection is where borrowers that are more likely to take part in risky projects are the ones who seek credit. Moral hazard is where borrowers take up risky projects at the expense of the bank, which increases the likelihood of default.

Credit risk can be managed in various ways. For example, lenders collect reliable information from prospective borrowers or specialize in lending to local firms or to firms in particular industries or enter loan contracts that restrict borrowers from engaging in risky activities. Lenders can also emphasize on collateral requirements for loans, credit rationing, among other ways of screening for credit risk.

Large banks, whose commercial borrowers may number in the tens of thousands, can use internal ratings which are essential for effective credit risk management (William & Carey, 2000). The internal ratings can enable banks to understand their level of risk exposure and how they choose to control risk exposures. According to William & Carey (2000), most banks ratings are produced for all commercial or institutional loans (that is, not consumer loans), in the USA and in some cases for large loans to households or individuals for which underwriting procedures are similar to those for commercial loans.

It is well established that the effective use of screening technology greatly reduces the costs of informational asymmetries between borrowers and lenders thereby enhancing the efficiency of the financial intermediation process (Stiglitz & Weiss, 1981). Improvements in screening and monitoring techniques are a valuable alternative to incomplete contracts aimed at reducing moral hazard and adverse selection problems. These informational asymmetries lie at the heart of market failures such as credit rationing (Stiglitz & Weiss, 1981).

1.2 Problem statement

The major cause of serious banking problems continues to be directly related to lax credit standards for borrowers and counterparties, poor portfolio risk management, or a lack of attention to changes in economic or other circumstances that can lead to deterioration in the credit standing of a bank's counterparties[1]. In the literature devoted to credit risk analysis there are not many publications on modeling the credit risk in consumer portfolios or personal credit portfolios. Most of the research deals with measuring credit risk by Probability of Default modeling in portfolios of small, medium and large companies, or even for financial companies (Cao, Vilar, & Devia, 2009).

Commercial banks in Kenya have managed to reduce their exposure to credit risk mainly through screening. Apart from collecting information concerning the borrower, they also conduct a background check on the borrower by reviewing their credit default history, through a platform known as the Credit Reference Bureau (CRB). Commercial banks in Kenya are yet to adopt a more sophisticated, less arbitrary alternative to traditional lending limit controls. This in turn causes failure of banks to incorporate a wider range of factors that affect credit default rates. It also hinders effective credit risk management.

The problem this research is seeking to address is the lack of efficient credit risk management systems in the banking sector. This will be done by coming up with a credit rating system for borrowers using the characteristics of the borrowers. The banks or other credit institutions will be able to determine the likelihood/ probability of default of loan obligors using the model in this research.

---

[1] Overview of the New Basel Capital Accord, 2001

1.3 Research objectives

The main objectives of this research are:

i. To determine the characteristics of loan obligors that are important predictors of credit (default) risk.
ii. To assess, quantitatively, the extent to which characteristics of loan obligors affect their likelihood/probability of default

1.4 Research questions

The study provides answers to the following research questions;

i. Which attributes of loan obligors are important predictors of credit risk?
ii. How does a given set of characteristics of loan obligors affect their probability of default?

It has been found in previous research that characteristics such as gender play a relevant role in determining the likelihood of default of loan borrowers. This research will seek to find out which characteristics of loan obligors are most likely to determine the likelihood of default. The research will also look at how a given group of characteristics can affect the probability of default of the borrower.

1.5 Significance of the research

Determining the probability of default, *PD*, in consumer credits, loans and credit cards is one of the main problems to be addressed by banks, savings banks, savings cooperatives and other credit companies. This is a first step needed to compute the capital in risk of insolvency, when their clients do not pay their credits, which is called default. The risk coming from this type of situation is called credit risk, which has been the object of research for many years.

Financial institutions are increasingly adopting statistical approaches in the management of credit risk. Firms that have taken up the statistical approaches are placed at an advantage since these methods provide accuracy, thus making credit risk management more effective. Firms in industries such as banking, which are widely exposed to credit risk, need to adopt better methods when it comes to credit risk in order to reduce the risk and improve performance of the firm.

This research seeks to determine the cross sectional dependence of default rates on certain characteristics of loan obligors. The findings of this paper will enable banks to build credit scoring models based on the characteristics of loan obligors. The research will also analyze the measures taken by commercial banks in Kenya to mitigate or hedge themselves against credit (default) risk and provide recommendations on how credit risk management can be improved within the banking sector in Kenya.

CHAPTER 2

2.0 Literature Review

2.1 Introduction

Credit portfolios consist of instruments with different maturities. This definition for a credit portfolio also applies to loan portfolios, whereby the maturity/term of loan differs across the different obligors based on specific characteristics or information that the bank collects concerning their obligors. Empirical evidence indicates that long-term credits are riskier than short-term credits. Moreover, maturity effects are stronger for obligors with low probability of default (Basel Committee on Banking Supervision, 2004).

Thomas (2009) found that the Internal Ratings Based (IRB) approach suggested in the New Basel Accord regulations uses a capital allocation formula derived from a Merton style[2] structural model of the credit risk of portfolios of corporate loans. Yet this formula is being applied in the case of consumer loans as well as corporate loans. This has highlighted that although there are a number of well-established credit risk models for portfolios of corporate loans which are widely used by financial organizations, there are no such established consumer credit risk models for portfolios of consumer loans.

The nature of corporate loans and that of consumer loans is very different therefore it would not be an optimal decision to apply the same credit risk model in both cases. For corporate loans, the assets and liabilities of the company play a major role in identifying the probability of default of the company. For consumer loans, probability of default is based on the characteristics of the individual, which banks can identify from the information they collect from their loan obligors before granting them a loan. Other models need to be identified for managing consumer loan portfolios for better management of credit default risk.

Although there is less analysis of the credit risk of consumer loan portfolios, some people have developed models that can be used to analyze credit risk for these portfolios. Thomas Lyn developed a Markov chain model based on behavioural scores to establish the credit risk

[2] Merton's (1974) model assumes that a company has a certain amount of zero-coupon debt that will become due at a future time T. The company defaults if the value of its assets is less than the promised debt repayment at time T. The equity of the company is a European call option on the assets of the company with maturity T and a strike price equal to the face value of the debt.

of portfolios of consumer loans. Thomas L. found that in particular, the need for robust models of the credit risk of consumer loan portfolios has been brought into sharp focus by the failure of the ratings agencies to accurately assess the credit risks of Mortgage Backed Securities (MBS) and collateralized debt obligations (CDO) which are based on such portfolios. This is one of the main factors that led to the subprime mortgage crisis of 2007-2009.

Thomas (2009) also found that the need for models of the credit risk of portfolios of consumer loans is clear for capital adequacy provisions. These models are also needed to support the pricing estimates when such consumer loan portfolios are securitized.

2.2 Characteristics of loan obligors

Itoo, Mutharasu, & Jose (2013) found that the borrower's gender, borrower's age, borrower's marital status, the borrower's income, loan rate, loan type, amongst other factors, are significantly positively correlated with the defaulter's outstanding loan amount.

Characteristics of loan obligors should therefore be a great focus for banks and other lending institutions when it comes to management of credit default risk on consumer loan portfolios. Certain characteristics of borrowers or a group of characteristics attributed to those borrowers can play a great role in determining their probability of default. This is the reason why borrowers have to give certain information to the lending institution in order for them to determine whether the borrower is likely to default on payments and for the lending institution to make a decision on how to handle probable risk of default.

2.2.1 Age

A wide range of research has shown that the age of loan borrowers alone is not very significant in determining the likelihood of loan default. However it may be used alongside other factors to determine the probability of default of consumer loans.

The following findings from prior methodology show results of different models that tested the correlation between age and default probability.

Capozza, Kazarian & Thomson (1997) indicated that the borrower's age was negatively correlated with the default probability. Hakim and Haddad (1999) also found that the age of the borrower is significantly negatively correlated with the default probability. Jacobson and Roszbach (2003) indicated the applicant's age was significantly negatively correlated with the unsecured loan default. Cairney and Boyle (2004) also showed that the age of the borrower was significantly negatively correlated with the default risk of credit loans. Kumar

(2010) found that there is no significance between the age of the borrower and mortgage defaults.

Younger customers have high likelihood of defaulting unlike elderly customers (Adem, Gichuhi, & Atieno, 2012). Von Furstenberg and Green (1974) found that inclusion of a situational factor like the age of the borrower improves the performance of the scoring models.

2.2.2 Gender

Gender might not play a significant role in determining the probability of default compared to other factors like occupation. Jacobson and Roszbach (2003) found that an applicant's gender was significantly negatively correlated with the unsecured loan default. Onoja & Emodi (2012) used Tobit model to analyse data on default rates within their sample population in Nigeria and the study found that gender was not a major determinant of loan default.

Goriunov & Katerina on the other hand found that Males default more often than female loan obligors. Adem et al. (2012) also found that male customers have high likelihood of defaulting compared to their female counterparts.

2.2.3 Marital Status

Marital status is considered to be an important factor in creation of credit rating systems according to Von Furstenberg & Green (1974), who found that inclusion of situational factors like marital status of borrower improves the performance of the scoring models.

Scoring models are used to assign scores to rate individuals or businesses according to their likelihood of default. Those with a lower likelihood of default are rated higher than those with a higher likelihood of default. The scoring/rating systems can either be external or internal.

Cairney & Boyle (2004) showed that the marital status was significantly positively correlated with the default risk of credit loans. Adem, Gichuhi, & Atieno (2012) fond that single customers have a higher likelihood of defaulting compared to their married customers.

Marital status (single, married, widowed, divorced) can be used to reflect the level of disposable income for individuals, therefore affecting the probability of default of an individual. Married individuals (especially with children) for example can be assumed to have less disposable income compared to single individuals.

Evans and Winston (2008) found that single, college-educated women managed their credit more prudently than both men in general and married women, in a study conducted in Ghana.

2.2.4 Occupation and income

Adem, Gichuhi, & Atieno (2012) found that financial sector customers have equal likelihood of default as support staff customers and long term loans have less likelihood of defaulting compared to short term loans.

2.3 Traditional credit risk management techniques

Previously, banks only offered secured loans where presence of collateral represented high quality borrowers and thus a lower probability of default for all loans. If banks are protected by a high level of collateral they have less incentive to undertake adequate screening of potential borrowers and loans at the time of the decision (Jimenez & Saurina, 2003).

Ngare (2008) found that:

> In most banks in Kenya, credit risk management was organized in units within the credit management department with persons responsible for credit risk management reporting to the credit manager. Most banks did not have an autonomous credit risk management department. Qualitative loan assessment methods were found to be the most prevalent methods in making credit granting decisions while liquidity run on the borrower, credit concentration and adverse trading by the borrower were the main sources of credit risk among the banks in Kenya. In addition, most banks were found to use loan diversification, bank guarantees and bank covenants to mitigate against credit risk.

Saunders & Allen (2002), state that it is hard to draw the line between traditional and new approaches, especially because many of the better ideas of traditional models are used in the new models.

2.4 Developments in credit risk modelling

Gordy (2000) found that important advances have been made in modeling credit risk. Consumer credit modeling has been shaped by the decisions it sought to support, particularly the initial decision of whether to accept a new applicant for credit. Application scoring or credit scoring initially sought to improve these decisions and to make them more consistent (Thomas, Oliver, & Hand, 2005).

According to Gakuri, Ngugi, & Ndwiga (2012), companies recognize how credit concentrations can adversely impact financial performance. As a result, a number of institutions are actively pursuing quantitative approaches to credit risk measurement. The banking industry is also making significant progress toward developing tools that measure credit risk in a portfolio context. They are also using credit derivatives to transfer risk efficiently while preserving customer relationships. Portfolio quality ratios and productivity indicators have been adopted. The combination of these developments has vastly accelerated progress in managing credit risk in a portfolio context.

Recent years have witnessed significant advances in the design, calibration and implementation of portfolio models of credit risk (Gordy, 2002). Large commercial banks and other financial institutions with significant credit exposure rely increasingly on models to guide credit risk management at the portfolio level. Models allow management to identify concentrations of risk and opportunities for diversification within a disciplined and objective framework. This offers a more sophisticated, less arbitrary alternative to traditional lending limit controls.

More widespread and intensive use of models is encouraging a more active approach to portfolio management at commercial banks, which has contributed to the improved liquidity of markets for debt instruments and credit derivatives (Gordy, 2002).

Commercial banks in Kenya however are yet to make advances in the use of statistical models for credit risk in their risk management systems. They still make use of credit risk management practices such as: thorough loan appraisal, asking for collateral and checking the credit history of the borrowers, credit rationing, loan securitization, and loan syndication.

Focusing on the use of qualitative methods over quantitative methods of risk management puts Kenyan commercial banks at a disadvantage in managing credit risk because quantitative models make it possible to numerically establish which factors are important in explaining default risk, evaluating the relative degree of importance of the factors, improving the pricing of default risk, screening out bad loan applicants and calculating any reserve needed to meet expected future loan losses, unlike qualitative techniques.

Variability of interest rates is one factor that may affect probability of default and increase credit risk in the bank's loan portfolios but this may not be the primary cause of default rates. Other factors may have a great impact on credit risk for commercial banks and they need to be incorporated in determining the probability of default of individuals or businesses. These factors include characteristics of obligors such as career, marital status, gender, age and history of default (which can be obtained from the credit reference bureaus) among other factors.

The Kenya Banks' Reference Rate (KBRR) is a uniform base lending rate across the banking sector in Kenya to enable consumers to compare the pricing of loan products. Publication of information on interest rates for the banking sector is expected to increase transparency, competition, enhance credit access and lower the overall cost of credit to borrowers. One of the main reasons this platform was created is because lack of comparable information on pricing of loan products in the banking sector can expose borrowers to high interest rates leading to default on their loan product.

Thomas (2009) highlighted that although there are a number of well-established credit risk models for portfolios of corporate loans which are widely used by financial organizations, there are no such established consumer credit risk models for portfolios of consumer loans. There appears to be a disconnect between the models of credit risk of portfolio of consumer loans needed for these more recent portfolio level decisions and the widely used credit scoring model of the credit risk of individual loans which have long been used for acceptance and operating decisions.

Traditionally, consumer credit modeling has modeled each loan/customer in isolation, but the lenders are really interested in the characteristics of their portfolios of retail loans. This interest has been reinforced by the emphasis on internal-ratings-based modeling in the new Basel Capital Accord, which regulates banks' lending. This allows banks to insert their own parameter estimate for the probability of default and loss given default into a corporate credit-risk-type model to estimate the distribution of default loss for segments of their consumer loan portfolio.

Stein (2005) however found that more powerful models can set prices more accurately than weaker ones. Lenders using more powerful models will have an advantage over those that do not as they are able to align their fees and loan terms more exactly and thus avoid over or under charging clients. The use of credit scoring models by financial institutions has increased dramatically. With this increase has come a need among users to understand the economic value of the models and to use this information to integrate them into traditional lending practices in a profitable manner.

CHAPTER 3

3.0 Methodology

3.1 Research design

The study is a quantitative research about individual loans default rates as a result of a set/group of characteristics that the borrower possesses.

The main variables used in this study are categorized into two: dependant variable, which is credit default risk and explanatory variables which include: checking status, loan duration (months), credit history, credit amount, savings status, employment duration, installment commitment, marital status, gender, other parties, property magnitude, age, other payment plans, housing, existing credits, skills, number of dependants and nationality.

The data was collected from a sample of one commercial bank in Germany then a statistical approach was used to link the credit default risk to the characteristics of the loan borrowers. This enabled a numerical quantification of the extent to which certain characteristics of an individual will determine their likelihood of default. A generalized linear model is used in this study to determine whether certain characteristics of loan borrowers are a significant determinant of credit default.

3.2 Population and sampling

According to a research done by Schildbach (2015), the German banking market has always been highly segmented and featured a three-pillar structure consisting of private, cooperative and public-sector banks. Public development banks and other special purpose banks play a role as well. Schildbach (2015) continues to explain that, within the pillar of the private commercial banks a distinction can be made between domestic institutions and banks that are majority-owned by foreign entities. The cooperative banks in turn comprise locally focused credit cooperatives and the two regional institutions of the credit cooperatives. Finally, the public-sector pillar is composed of savings banks and Landesbanks (development banks are not included).

A set of data which includes a description of one German bank credit data set was used for this research. The German data set includes 1000 individual loan borrowers, eighteen characteristics they possess and their loan status (good or bad).

3.3 Data collection

Cross sectional secondary data was used to conduct this research. The data was collected from an online source that consisted of information from German banks.

There was limited information from Kenyan Banks, which was the original source of data; therefore the available data from a German bank was used. This was one of the challenges faced while carrying out the research.

3.4 Model specification

To achieve the main objective of this study, a generalized linear model for the binomial family was used to relate the response variable, which is credit risk, to the independent variables, also referred to as predictors or covariates. The variable credit risk is bivariate, meaning that only two outcomes can be obtained i.e. credit risk or no credit risk.

The basic reasons for using GLMs are: to determine which of the selected variables or factors are significant predictors of credit risk and to quantify the relationship between the predictors and the credit risk. The response variable in this research is credit (default) risk and the independent variables are the characteristics of the defaulters. The data set in this research contains the following characteristics of loan obligors: checking status, loan duration (months), credit history, credit amount, savings status, employment duration, installment commitment, marital status, gender, other parties, property magnitude, age, other payment plans, housing, existing credits, skills, number of dependants and nationality

3.4.1 The Model

The distribution of the data belongs to the exponential family – Binomial distribution.

A linear predictor, which is a function of the covariates, was chosen as follows:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_{18} x_{18} + \varepsilon_i \qquad (1)$$

The parameters β_0 and β_1 will be estimated. x_i represents the explanatory variables where $i = 0,1,2\ldots n$. A link function, which connects the mean response to the linear predictor, was also selected based on the distribution of the data.

The variable y of the linear predictor is binary and represents the probability of default. The probability takes either of the values zero or one to represent 'no default' and 'default' respectively.

In equation (1) the independent variables x_i from the Bank data set are:

$x_1 = checking\ status,$

$x_2 = durationmonths,$

$x_3 = credithistory,$

$x_4 = credit\ amount,$

$x_5 = savings\ status,$

$x_6 = employment\ duration,$

$x_7 = installment\ commitment,$

$x_8 = marital\ status,$

$x_9 = gender,$

$x_{10} = other\ parties,$

$x_{11} = property\ magnitude,$

$x_{12} = age,$

$x_{13} = other\ payment\ plans,$

$x_{14} = housing,$

$x_{15} = existing\ credits,$

$x_{16} = skills,$

$x_{17} = numdependants,$

$x_{18} = nationality$

The model fitted above is a saturated model containing all the variables. Possible interactions between the explanatory variables will be analysed.

Combining the mean of the distribution and the linear predictor gives the general relationship

$$g(\mu) = y \cdots\cdots\cdots (3)$$

where g is the link function and y is the linear predictor.

The mean is given by:

$$E[y] = \mu = g^{-1}(y) \cdots\cdots\cdots (4)$$

In the explanatory variables, some of the characteristics of an obligor constitute factors with two levels (categories). Therefore they will have a separate linear predictor given by:

$$y = \propto_i + \beta x \cdots\cdots\cdots (5)$$

Where $i = 1(male) or\ 0(female)$ for the gender effect

and $i = 1(married)\ or\ 0(single)$ for the marital status, and so on.

The three parameters that would need to be estimated in the above equation (5) are$\propto_i\ and \beta$.

The parameters in the GLM will be estimated using maximum likelihood estimators (MLE). The log-likelihood function depends on the parameters in the linear predictor through the link function. Therefore, the M.L.E. estimates of the parameters can easily be found by maximizing the log likelihood, l, with respect to the parameters in the linear predictor.

3.4.2 Link function

Since the data follows a binomial distribution, the logit link function will be used for the dependant variable. It is given by: $g(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$

In our model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots\cdots + \beta_{18} x_{18} + \varepsilon_i$$

where **Y** is the dependent variable and X_i is the set of independent variables (characteristics of loan obligors) and $\boldsymbol{\beta}$ is the set of regression coefficients and $\boldsymbol{\varepsilon_i}$ is the error term.

The LOGIT model uses a variant of the cumulative logistic function:

$$Y = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_{18} x_{18}+\varepsilon_i)}}$$

The logistic function ensures that the $(0 \leq Y \leq 1)$ requirement is met, and this is demonstrated as follows:

As $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$ approaches positive infinity, then $Y = \frac{1}{1+e^{-\infty}} = \frac{1}{1} = 1$

As $\beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$ approaches negative infinity, then $Y = \frac{1}{1+e^{\infty}} = \frac{1}{\infty} = 0$

In the model, Y_I is nonlinearly related to X_i.

Logit is estimated by the Maximum Likelihood Estimator (MLE), which aims at finding the β that maximizes the likelihood of the sample data set being observed. MLE is an iterative process since it performs a number of iterations until the best estimate is obtained from the definition given above.

When estimating a logit model, we first determine the ODDS RATIO given as:

$$\textbf{Odds ratio} = \frac{P_i}{1-P_i}$$

The odds ratio tells us the likelihood that something will occur as opposed to it not occurring. Y_i acts as a dummy variable with the options 0 for no default and 1 for default. Pi can be used to represent Y_i as the dummy variable.

Having got the odds ratio we obtain the natural logarithm of the odds ratio and regress it against the variables. Thus, the logit model is as follows:

$ln\left(\frac{P_i}{1-P_i}\right) = pr(P_i = 1) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_{18} x_{18} + \varepsilon_i$ where P_i is the dummy variable.

3.4.3 Assumptions of a logistic regression

- The true conditional probabilities are a logistic function of the independent variables.
- No important variables are omitted.
- No extraneous variables are included.
- The independent variables are measured without error.
- The observations are independent.
- The independent variables are not linear combinations of each other

CHAPTER 4

4.0 Results and Analysis

The data used to generate the results had 1000 observations i.e. loan obligors in a German Bank. 18 variables/characteristics in total were considered but the model narrowed down to six significant variables. The data follows a binomial distribution. Therefore a generalised Linear Model (GLM) with a logit function was used to generate the results on STATA. Below are the findings.

4.0.1 Significant variables

The statistically significant variables are: checking status, duration in months, employment duration, instalment commitment and other payment plans. Their p-value is less than 5% (critical value).

Variable	**p> |z|**	**Coefficients**
Checking status	0.000	0.812049
Duration in months	0.001	-0.267933
Employment duration	0.004	0.194156
Instalment commitment	0.001	-0.2582487
Other payment plans	0.001	-0.6542078
Gender	0.001	0.3483756
constatnt		2.05397

Given the coefficients for the significant variables, our saturated model is going to be as follows:

$$y = 2.05397 + 0.812049x_1 - 0.267933x_2 + 0.1941565x_3 - 0.2582487x_4 - 0.6542078x_5 + \varepsilon$$

The coefficient (or parameter estimate) for the variable, checking status, is 0.812049. This means that for a one-unit increase in the amount in the checking account, we expect a 0.812049 increase in the log-odds of the dependent variable, credit risk, holding all other independent variables constant.

Duration in months: if duration increases by one month, we expect a 0.267933 decrease in the log-odds of credit risk, holding all other independent variables constant.

Employment duration: If employment duration increases by one year, we expect a 0.194156 increase in the log-odds of credit risk, holding all other independent variables constant.

Installment commitment: if the instalment commitment increases by one euro, we expect a 0.2582487 decrease in the log-odds of credit risk, holding all other independent variables constant.

Other payment plans: for one additional payment plan, we expect a 0.6542078 decrease in the log-odds of credit risk, holding all other independent variables constant.

4.0.2 Odds Ratio (OR)

	Variable	**Odds Ratio(OR)**
x_1	Checking status	2.252519
x_2	Duration in months	0.9735625
x_3	Employment duration	1.214286
x_4	Instalment commitment	0.7724031
x_5	Other payment plans	0.5198537
x_6	gender	1.416764

Given the odds ratios, we can conclude the following:

Gender: Since the Odds Ratio is greater than 1, we can conclude that the control (F=0) is better than the intervention (M=1). This means that male loan obligors are more likely to default on loans with an odds of 1.417 compared to their female counterparts.

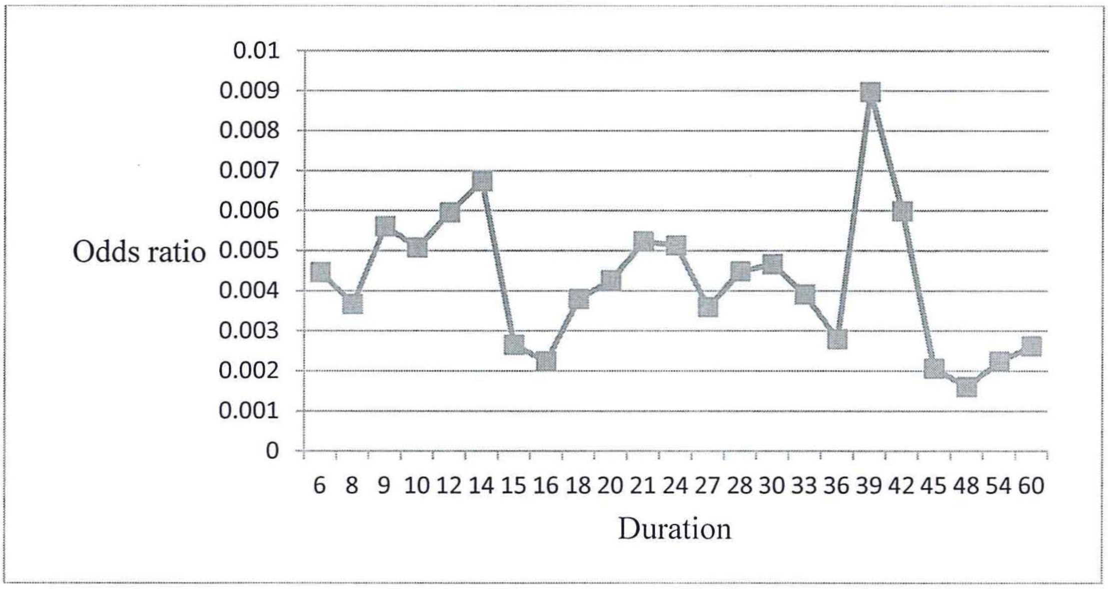
Other payment plans: Since the Odds Ratio is less than 1, we can conclude that the intervention (yes =1) is better than the control (no = 0). This means that loan obligors with other payment plans are less likely to default on loans with an odds of 0.5198537 compared to those that do not have any other payment plan.

Checking status: Customers with no checking status are the most likely to default on loans with an odds of 7.362 followed by those with an amount greater than 200, with an odds of

3.609. Those with a checking status of 0<=X<200 are the least likely to default on loans. This is summarised in the table below:

Code	Amount	P>|z|	Odds Ratio
0	<0	0.718	1
1	0<=X<200	0.027	1.467488
2	No checking	0.000	7.361702
3	>=200	0.000	3.608841

Duration (months): The results show that loan obligors with a longer duration of the loan repayment term are less likely to default with lower odds compared to those with a shorter duration as summarized in the table below. Given the odds ratios, we can conclude that there is a higher likelihood (or more cases) of default for a loan duration of up to one year (12 months) but for loan durations that exceed one and a half years (18 months), less cases or likelihood of default are observed. This is indicated in the graph below:



Employment duration:

Employment	Odds Ratio	Std. Err
Unemployed	0.277141	0.4457962
<1	0.8509157	0.259921
1<=X<4	1.332594	0.3838934
4<=X<7	2.04142	0.6525131
>=7	1.741587	0.5225784

The table shows that, given the odds ratio, individuals who have been employed for a longer period are more likely to default compared to those who have less experience in the job market.

Instalment commitment - Installment rate in percentage of disposable income:

Installment commitment	Odds ratio	Std. Err
1	1.1698341	0.594
2	0.9086022	0.225
3	0.8296296	0.220
4	0.6645702	0.147

Individuals who have committed a lower percentage of their disposable income towards repayment of the loans are more likely to default with a higher odds compared to the others.

4.1 Model diagnostic tests

Pearson and deviance residuals

Residuals represent the difference between the data and the model. The Pearson residuals are comparable to standardized residuals used for linear regression models. Deviance residuals represent the contribution of each observation to the overall deviance. In our model, the deviance is 1016.024 and the pearson residual obtained is 1041.133

The goodness of fit of the model

The Hosmer – Lemeshow test was used to analyse the goodness of fit of the model. The 'estat gof' command on STATA produces a chi-square of 180.89 with 6 df, yielding a *p*-value of 0.000 (significant). The HL statistic was used based on 10 groups.

The model is therefore a satisfactory fit to the data.

Model selection

The estimated model describes the data well since the Likelihood under the current model (L_m) is close to the likelihood under the saturated model (L_s) and the average is almost equal to one i.e. $\frac{L_s}{L_m} = \frac{508.00197}{520.06057} = 0.9768$

CHAPTER 5

5.0 Discussion, conclusions and Recommendations

5.1 Introduction

This chapter gives an overview of the findings obtained in this research and provides further details on these findings. A discussion is carried out on how credit risk management models can be applied in the banking sector in Kenya and the benefits that will be realised. Finally, a general conclusion and recommendation is given with regard to credit risk management in the banking sector, focusing on Kenya.

5.2 Summary

The research questions the study sought to answer were:

i. Which characteristics of loan obligors are important predictors of credit risk?
ii. How does a given set of characteristics of loan obligors affect their probability of default?

The study found five significant characteristics/attributes of loan obligors that would affect their likelihood of loan repayment: gender, other payment plans, checking status, duration of loan and instalment commitment.

The following were the major findings for each attribute:

Male loan obligors are more likely to default on loans with an odds of 1.417 compared to their female counterparts. Customers with no amount in their checking accounts are the most likely to default on loans with an odds of 7.362 followed by those with an amount greater than 200 Euros in their checking account, with an odds of 3.609. Those with an amount between 0 and 200 Euros in their checking account are the least likely to default on loans.

Loan obligors with other payment plans are less likely to default on loans with an odds of 0.5198537 compared to those that do not have any other payment plan. Loan obligors with a longer loan repayment term are less likely to default compared to those with a shorter duration; there is a higher likelihood (or more cases) of default for a loan duration of up to one year (12 months) but for loan durations that exceed one and a half years (18 months), a lower likelihood of default are observed. Individuals who have been employed for a longer period of time (years) are more likely to default compared to those who have less experience in the job market.

Finally, individuals who have committed a lower percentage of their disposable income towards repayment of the loans are more likely to default with a higher odds compared to the others.

5.3 Discussion

Based on the findings stated above, we can determine how a certain group of characteristics affects the borrower's probability of default; P(Y=1). For example what is the probability of default of a borrower who possesses the following group of characteristics: male, has other payment plans, more work experience, low loan repayment duration and low checking status?

$$P(Y=1) = \frac{e^{(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_5 x_5+\varepsilon_i)}}{1+e^{(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_5 x_5+\varepsilon_i)}}$$

$$= \frac{\exp\left(2.054+3.484*1+0.812*0-0.268*\frac{6}{12}+0.194*7-0.258*1-0.654*0\right)}{1+\exp\left(2.054+3.484*1+0.812*0-0.268*\frac{6}{12}+0.194*7-0.258*1-0.654*0\right)}$$

$$P(Y=1) = 0.9985$$

This shows that the individual is highly likely to default on their loan. In the same way, different sets of characteristics can be used to determine the probability of default of different borrowers. This will enable banks to rate individuals' credit risk based on their characteristics as a method of credit risk management.

Some of the research that obtained similar findings include: Goriunov & Katerina and Adem et al. (2012) who found that male customers have high likelihood of defaulting compared to their female counterparts. Various reasons can be given to explain why this is the case according to different research carried out.

Adem, Gichuhi, & Atieno (2012) found that long term loans have a lower likelihood of default compared to short term loans. This could be because the concentration of the burden of loan repayment is not spread or smoothed over time, leaving no room for additional obligations or financial setbacks. These are however bound to occur and as a result, the borrower delays or fails to repay the loan amount.

Customers with low or no money in their checking accounts are the most likely to default on loans. The checking account status in the data used for this research also accounts for the individual's salary commitment for at least one year. If the bank keeps data on their customer's checking account status, it will clearly indicate the individual's probability of loan repayment given their instalment commitment over the year.

Other findings were that Loan obligors with other payment plans are less likely to default on loans compared to those that do not have any other payment plan. Individuals who have been employed for a longer period of time (years) are more likely to default. Individuals who have committed a lower percentage of their disposable income towards repayment of the loans are more likely to default compared to the others.

5.4 Conclusions and Recommendations

This research found that gender, loan duration, employment duration, instalment commitment and other payment plans are important predictors of credit risk.
The likelihood of loan repayment can be determined quantitatively, given the characteristics of loan obligors.

Financial institutions, especially in the banking sector, which are adversely affected by credit risk, should actively pursue quantitative approaches to credit risk measurement and management. Various tools such as generalized linear models can aid in better measurement of credit risk and as a result better management of the stated risk. Characteristics of loan obligors is one among many other risk factors that can be analysed by banks to come up with credit rating/scoring models. The use of such quantitative techniques will not only lead to less default rates but will also pave way for innovation in the banking sector that relates to programmes and models that can be used as tools for minimizing credit risk.

Bibliography

Adem, O., Gichuhi, W., & Atieno, O. (2012). *Parametric modelling of bank loan defaults in Kenya.* Mombasa.

Altman, E. E., & Saunders, A. (1998). Credit risk measurements: Developments over the last 20 years. *Journal of Banking and Finance* .

Beck, T., Behr, P., & Guttler, A. (2009). *Are women better loan officers?*

Cao, R., Vilar, J., & Devia, A. (2009). Modelling consumer credit risk via survival Analysis. 3-30.

Coyle, B. (2000). *Framework for credit Risk Management.* Chartered Institute of Bankers.

Dervitz, A. *Business cycle, credit risk and economic capital determination by commercial banks.*

Dietsch, M., & Petey, J. (2002). The credit risk in SME loans portfolios: Modeling issues, pricing, and capital requirements. *Journal of Banking and Finance* , 303-322.

Gakuri, R. W., Ngugi, J. K., & Ndwiga, P. M. (2012). Effects of Credit risk management techniques on the performance of unsecured bank loans employed in commercial banks in Kenya. *Internation Journal of Business and Social research , 2* (4), 221-236.

Gordy, M. B. (2000). A comparative anatomy of credit risk models. *Journal of Banking and Finance* , 119-149.

Gordy, M. B. (2002). A Risk-Factor Model Foundation for Ratings-Based Bank Capital Rules.

Goriunov, D., & Katerina, V. *Loan default prediction in Ukrainian Retail Banking.*

Gweyi, M. O. (2013). Credit Risk Mitigation Strategies Adopted By Commercial Banks in Kenya. *International Journal of Business and Social Science* , 77.

Itoo, R. A., Mutharasu, A. S., & Jose, A. F. (2013). Effect of loan value and collateral on value of Mortgage default. *International Journal Latest trends in Fin. Eco. Sc. , 3* (15), 635.

Jimenez, G., & Saurina, J. (2003). *Collateral, type of lender and relationship banking as determinants of credit risk.* Spain: Bank of Spain.

Jorion, P. (2009). *Financial Risk Manager Handbook.* Hoboken, New Jersey: John Wiley & Sons, Inc.

Kolapo T. Funso, A. R. (2012). Credit risk and commercial banks' performance in Nigeria: A Panel data approach. *Australian journal of business and management research* , 31.

Lesson 3: Logistic Regression Diagnostics. (n.d.). Retrieved September 19, 2015, from www.ats.ucla.edu: http://www.ats.ucla.edu/stat/stata/webbooks/logistic/chapter3/statalog3.htm

Maria Psillaki, I. E. (2010). Evaluation of credit risk based on firm performance. *Europen journal of operational research* , 2.

Mishkin, F. S. (2004). *The Economics of Money, Banking and Financial Markets.* Pearson.

Ngare, E. M. (2008). *A survey of credit risk management practices by commercial banks in Kenya.* Nairobi: University of Nairobi.

Onoja, A., & Emodi, A. (2012). *Analysis of gender and poverty effects on loan defaults rate among arable crop farmers in Rivers state, Nigeria.*

Saunders, A., & Allen, L. (2002). *Credit Risk Measurement: New approaches to Value at Risk and other Paradigms* (2 ed.). New York: John Wiley and sons Inc.

Schildbach, J. (2015, September 21). German Banks Lending: Market share developments in individual sectors. *Deutsche Bank Research* , p. 2.

Segoviano, M. A., & Padilla, P. Applications to stress testing under data restricted environments. In *Portfolio Credit Risk and Macroeconomic Shocks* (pp. 485-511).

Stein, R. M. (2005). The relationship between default prediction and lending profits. *Journal of Banking and Finance* , 1213-1236.

Stiglitz, J. E., & Weiss, A. (1981). Credit rationing in markets with imperfect information. *American Economic Review* , 912-927.

Thomas, L. C., Oliver, R. W., & Hand, D. J. (2005). A Survey of the Issues in Consumer Credit Modelling Research. *Journalf the operational research society* , 1006-1015.

Thomas, L. (2009). *Modelling the credit risk for portfolios of consumer loans: Analogies with corporate loan models.* Southampton: Quantitative Financial Risk Management Centre, School of Management, University of Southampton.

Thomas, L., & Scherer, W. (2005). *Behavioral scoring and the dynamics of consumer credit assessment.* Virginia.

Volswein, F., & Szelest, B. (1995). Individual and campus characteristics associated with loan defaulters. *Journal of the association for institutional research* .

Vong, A., & Patricio, A. P. (2005). Credit Risk Assessment in the Macau Banking. *Euro Asia Journal of Management* , *15* (2).

William, T. F., & C. S. (2000). Credit Risk Rating system at Large U.S. Banks. *Journal of Banking and Finance* , 167-201.

APPENDICES

APPENDIX 1

Binomial Distribution

Let Z~Binomial (n,μ) such that $Y = \frac{z}{n}$ (or equivalent Z= nY). The distribution of Z is: $f_Z(z;\theta,\varphi) = \binom{n}{z}\mu^z(1-\mu)^{n-z}$

And the distribution for Y is consequently:

$$f_Y(y;\theta,\varphi) = \binom{n}{ny}\mu^{ny}(1-\mu)^{n-ny}$$

In the format of equation (1) we have:

$$f_Y(y;\theta,\varphi) = exp\left[n(ylog\mu + (1-y)\log(1-\mu)) + log\binom{n}{ny}\right]$$

$$= exp\left[n\left(ylog\left(\frac{\mu}{1-\mu}\right) + \log(1-\mu)\right) + log\binom{n}{ny}\right]$$

In the format of equation (1) this has:

$$\theta = log\frac{\mu}{1-\mu} \rightarrow \mu = \frac{e^\theta}{1+e^\theta}$$

$$\varphi = n$$

$$b(\theta) = \log(1+e^\theta)$$

$$c(y,\varphi) = log\binom{n}{ny}$$

$$a(\varphi) = \frac{1}{\varphi}$$

$$E[Y] = \mu \; and \; Var[Y] = \frac{\mu}{n}(1-\mu)$$