

**KWAME NKRUMAH UNIVERSITY OF SCIENCE AND
TECHNOLOGY**



**DEFAULT RATE MODELLING OF A BANK IN GHANA USING
COX REGRESSION MODEL.**

By

Yaw Attua

(B.Sc. Actuarial Science)

Index Number: 8731812

A THESIS SUBMITTED TO THE DEPARTMENT OF MATHEMATICS,
KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY IN
PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE
OF MSC ACTUARIAL SCIENCE

May, 2016

Declaration

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

Yaw Attua (PG 8731812)

Student

.....

Signature

.....

Date

Certified by:

Dr. A.Y. Omari-Sasu

Supervisor

.....

Signature

.....

Date

Certified by:

DR. R.K. Avuglah

Head of Department

.....

Signature

.....

Date

Dedication

This thesis is dedicated to my family for all their tremendous support. It is also dedicated to the memory of the founders and pioneers of Actuarial Science.

KNUST

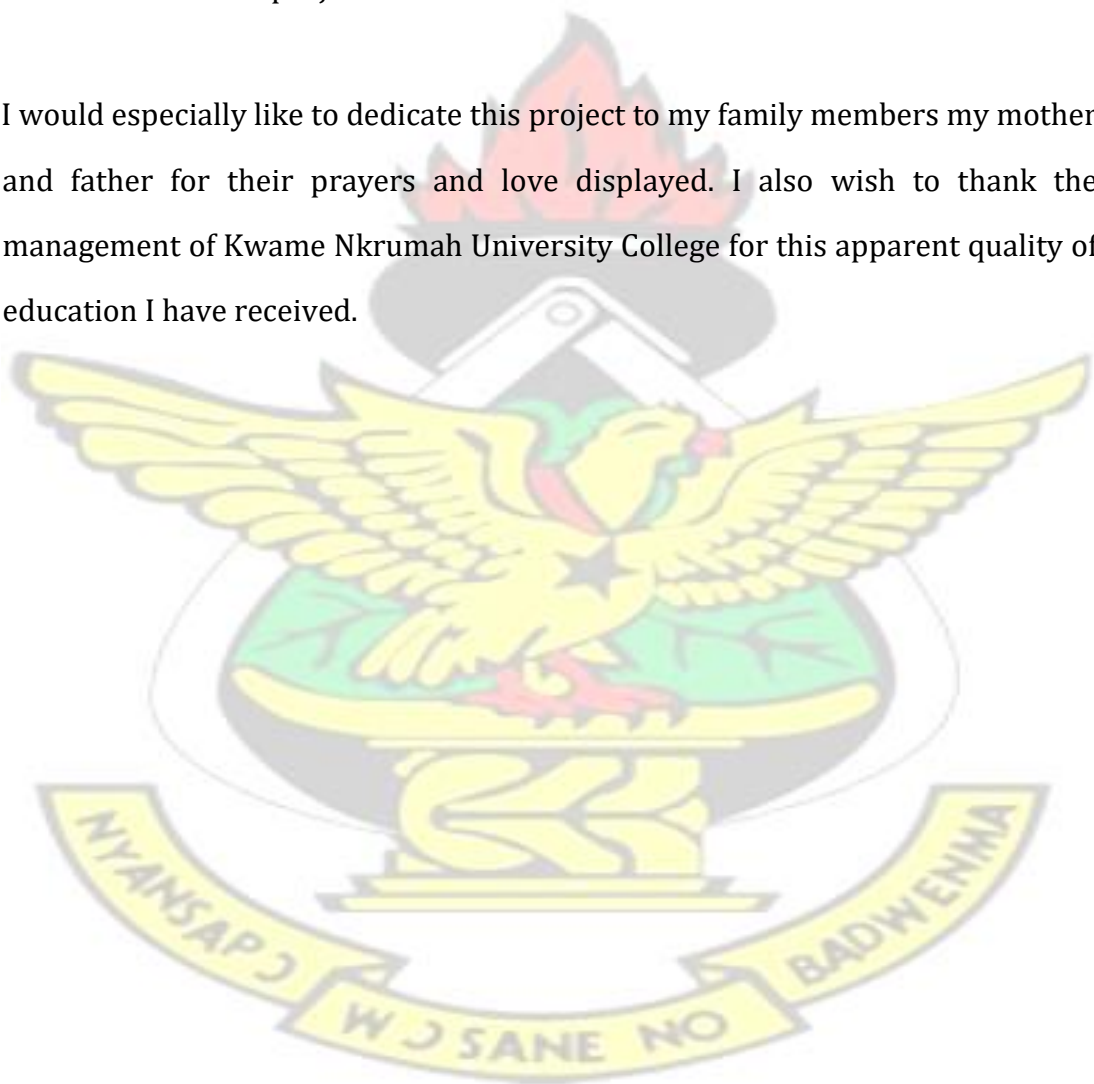


Acknowledgements

I would like to thank God first and foremost for supplication in the course of the thesis. I wish to in addition express my utmost gratitude to Dr. A. Y. OmariSasu for his immeasurable sacrifices during my project.

My gratitude also extends to Ms. Afua Egyima Attua and Mrs. Nana Yaa Kesiwa Asomani both of Zenith Bank for their time and guidance afforded to me within the duration of this project.

I would especially like to dedicate this project to my family members my mother and father for their prayers and love displayed. I also wish to thank the management of Kwame Nkrumah University College for this apparent quality of education I have received.



Abstract

The subject of non-performing loans which in effect affect all banks including small market enterprises' ability to prevent ultimate ruin. "For a credit bureau to function effectively, however, it must be possible to uniquely identify individuals with reasonable certainty. Identification is necessary in order to retrieve a current loan applicant's past credit history from a credit database" (Gine et al., 2011). There arises a fundamental need to devise an easy strategy to reduce the occurrences of non-performing which would be widely accepted. Cox proportionalhazard model with time-varying covariates was used to evaluate certain financial sensors that would definitively predict the bank's loan failures during the years of 2012 to 2015 of Zenith Bank Ahodwo branch. The Breslow approximation to the partial likelihood of the overall cox relations of client is used to estimate the correlation coefficient parameter, betas through differentiation. The models resultant outputs were compared to the actual defaults and it is worthy to note that the model presents a well above average predictor quality. The significance of each covariate is tested statistically and the 4th covariate (the number of dependents) was found to be significant there were insufficient data reject the null hypothesis for all the other covariates. A new way of coarse-classifying of covariates enlisting the help of survival analysis methods is proposed. Also, a good number of diagnostic methods (Cox-Snell, Martingale amongst other residuals) are used to check the aptitude of the model's fitness for applicability to any loan data. These results can be extended to a general case of bank failures under the volatile economic environments.

Contents

Declaration	i
Dedication	ii
Acknowledgements	iii
Abstract	iv
Table of Contents	v
List of Tables	ix
List of Figures	ix
List of Abbreviations	xi
1 INTRODUCTION	1
1.1 Background of Study	1
1.2 Statement of Problem	3
1.3 Objectives of Study	5
1.4 Methodology	5
1.5 Justification	6
1.6 Limitation	6
1.7 Organization of the Study	7
2 LITERATURE REVIEW	8
2.1 Introduction	8
2.2 History and Development of Bank Loan Financing	8
2.2.1 Credit Risk Defined	8
2.2.2 Origin and Evolution of Credit Risk Analysis	9
2.3 Credit Analysis in Developed Countries	10
2.3.1 The Bank for International Settlement (BIS)	11
2.3.2 Credit Market Context	12

2.4	Credit Analysis Techniques Common to the World	15
2.4.1	Empirical Models of Credit Analysis	15
2.4.2	The 5Cs Model	15
2.4.3	CAMELS Model	16
2.4.4	Credit Scoring Model	16
2.4.5	Risk Adjusted Return on Capital (RAROC)	17
2.4.6	Altman's Z-Score	18
2.5	Credit Risk Management in the Ghanaian Economy	18
2.5.1	Credit Services at Zenith Ghana	19
3	METHODOLOGY	20
3.1	Introduction	20
3.2	Description of the Data	20
3.2.1	The Type and Source of the Data	20
3.3	The Method of Data Collection	21
3.4	Survival Models	21
3.4.1	Introduction of Survival Data	21
3.4.2	Covariates and Proportional Hazards	22
3.5	Proportional Hazards Models	23
3.6	Fully Parametric Models	24
3.6.1	Parametric Models for the Hazard Function	24
3.7	The Cox Proportional Hazards Model	25
3.7.1	Introduction	25
3.7.2	Fitting the regression parameters	25
4	RESULTS AND ANALYSIS	30
4.1	Introduction	30
4.2	Preliminary Analysis	30
4.2.1	Sex	31
4.2.2	Disposable Income Brackets	32
4.2.3	Marital Status	33

4.2.4	Number of Dependents	34
4.2.5	Amount of Loan Bracket	35
4.2.6	Age	36
4.2.7	Term of Loans	36
4.3	Regression Analysis	37
4.3.1	Test for Multicollinearity Among Independent Variables.	37
4.3.2	Over all Relationship Between independent and dependent variables	38
4.3.3	Relationship of Individual Independent Covariates to Dependent Variable	38
4.3.4	Strength of Model Relationship	39
4.3.5	Classification of the model	40
4.3.6	Goodness of fit	41
4.3.7	The Generalized Theoretical Model	44
4.3.8	Baseline Hazard Function	45
4.3.9	Final model	45
4.3.10	Cox Regression Model Calculator	45
4.3.11	Respondent Cox Regression Model Estimator.....	46
4.3.12	Type I and Type II Error	46
5	FINDINGS, CONCLUSION AND RECOMMENDATIONS	48
5.1	Introduction	48
5.2	Summary of Findings	48
5.3	Conclusion	51
5.4	Recommendation to Stakeholders	51
5.5	Recommendation for Future Research	53
	References	54
	Appendix A	59
	Appendix B	60

List of Tables

4.1	Cross Table - Payment Status versus Sex	31
4.2	Disposable Income Brackets	32
4.3	Cross table of payment Status versus marital status	33
4.4	Cross Table - Payment Status versus Marital Status	34
4.5	Cross table Payments verses the Number of dependents	35
4.6	Payment Status versus Amount Collected	36
4.7	Cross table - payment Status versus age	36
4.8	Cross table -Payment Status versus loan term	37
4.9	Omnibus Tests of Model Coefficients	38
4.10	Cross Reference Table of Model's Payment Status versus the Actual	46
5.1	Cross table of payment Status	59
5.2	Summary statistics (Events)	60
5.3	Descriptive statistics (Explanatory variables)	61
5.4	Goodness of fit statistics	61
5.5	Test of the null hypothesis H_0 : beta=0	61
5.6	Regression coefficients	62
5.7	Kaplan-Meier table	63
5.8	Mean survival time	64
5.9	Quantiles estimation	64

List of Figures

4.1	Graph of Disposable Income Brackets	32
-----	---	----

4.2	Graphical representation of Cox-Snell Residual	40
4.3	Graphical representation of Martingale's Residual	41

KNUST



List of Abbreviations.

CRM – Credit Risk Management.

SMEs – Small and Medium Enterprises.

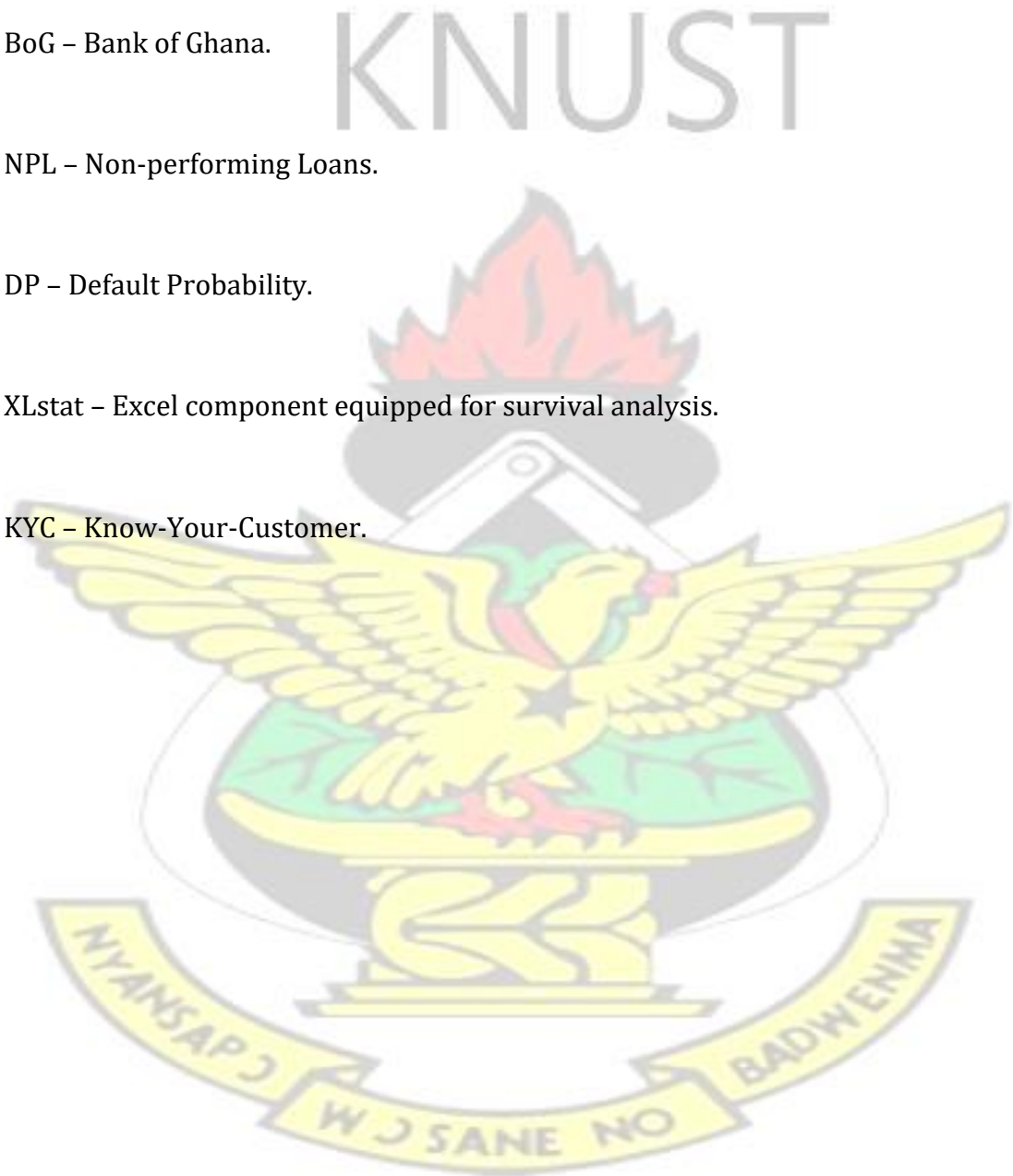
BoG – Bank of Ghana.

NPL – Non-performing Loans.

DP – Default Probability.

XLstat – Excel component equipped for survival analysis.

KYC – Know-Your-Customer.



Chapter 1

INTRODUCTION

1.1 Background of Study

The bank loans system in Ghana experiences very keen competitive strategies especially as far as the past ten years or so is concerned. This competitive nature is mostly because of the emergence of banks from the sub-Saharan region specifically Nigeria started very revolutionary ways of banking never before envisaged in this country. Since then there has been a heightened rush to pull the acute numbers of the untouched population. The Banks reacts to this by sending direct sales persons to approach prospective customers, growth in bank conglomerate system, the starting up of “cashless” accounts, mobile phone banking and notification systems. As business expansions grew, so did the desire to grant loans to organizations and interested persons. So as to make profit and overtake their counterparts, several banks have meted out loans and advances to clients but not all of such loans get settled. Therefore, brings forth the instance of Non-Performing Loans (NPLs) onto the world of banking and becoming a great concern to banks and regulatory bodies recently.

A non-performing loan is one granted by a bank or financial institutions on which interest and principal to be repaid are not settled within the stipulated time frame or period. As per the International Monetary Fund (IMF), a loan component is considered to be an NPL if it is 90 days elapsed. A loan is in actual fact said to be an NPL also if it happens to have less than 90 days elapsed but repayments are no longer expected. Settlement of principal repaid and interest on the loans create a series of cash flows for any bank it stands to reason that it would be an avenue for most banks to make profits.

Hence an excessively capital intensive NPL portfolio decreases banks potentials for profits and their propensity to in extension lend to other borrowers. This virtually has an adverse influence on the industry as well economy of any nation. According to a country-wide Banking Survey Report released in 2010 and begot by Pricewaterhouse Coopers, between the years 2007 - 2009, the complete income of the banking industry multiplied in two parts from GHS 793 million in 2007 to GHS 1.5 billion in 2009. Over the exact interval however, the increased deterioration of the industry's loan portfolio adversely affected profit intervals. Impairment charges for NPLs over the three-year period rose upwards from GHS 60 million in 2007 to GHS 266 million in 2009. The Central Bank of Ghana also disclosed that the non-performing loans ratio (which measures the ratio of losses on these loans to accumulated loan advances, deteriorated from 16.2% in December 2009 to 17.6% as at December 2010. This increased NPL ratio has contributed to decrease the market share of the Ghana's top 5 banks from 49.5% in 2009 to 45% in 2010 (BOG report 2010). Switching attention from the traditional banking departments, the non-banking institution Micro-finance and Small Loans Center (MASLOC) has summoned several of its defaulting clients in different sectors of the country to the courts.

So let's consider the deterministic factors that are responsible for this heightened default saturation on loans amongst Ghanaian banks. The immensely hard economic environment affecting some small and medium scale enterprises (SMEs) and individuals was a factor. Secondly Government indebtedness is effectively to blame for high interest rates on loans concurrently. Thirdly and most importantly, poor credit profiling of loan applicants by the banks lead eventually to bad loans. A good risk management system will filter out potential borrowers to ensure that unsecured loans are redeemable. This natural competitive nature of banks for borrowers and pressure from governments to lend a great considerable amount, do not undergo enough due-diligence on the

background study of loan clientele before granting loans. Good economic practices has also birthed the concept known as Know-Your-Customer (KYC) requirements which have been relaxed for new customers. All these leads to loan defaults in the long run. The lack of a centralized database system from which credit profiles of potential loan clients can be readily be acquired to ascertain their credit score does not help banks. Every borrower's credit profile should be readily searchable in a centralized system and thus is strictly adhered to and as result cannot be manipulated. This would have acted as a check and balance against loan defaults. In Ghana, the lack of a good address and house numbering system also intensifies the problem as it becomes difficult to trace people to their residences.

Thus banks that will regard risk management dealings with utmost seriousness since it will control their losses resulting to high returns to shareholders. This will also go a long way to benefit the economy in the long run.

1.2 Statement of Problem

Credit risk has been a matter of considerably and constant research interest in Banking and Financial circles. Credit risk is the risk of loss resulting from the failure of a borrower or counterpart to honor its financial or contractual obligations to the Bank (Scotia Bank Annual Report, 2012). Default occurs if the debtor is unable to meet its legal obligation according to the debt contract. A non-performing loan pile up as a result which are a few of such loans granted by a bank or financial institutions on which interest and principal repayments are not settled within the stipulated time frame or period.

To come to the above stated end, it is vital that models used in assessing

the credit risks of granting loan to clients must be critically explored in close conjunction to empirical models of credit analysis. The potency of the credit analysis theory used by the banks in assessing the credit risks of loan applicants and in fact loan grants as a whole has a trickling down effect on the prevention of NPLs and building of strong loan portfolios which also tends to affect the profit structure of the banks. The sustainability of these banking institutes hinges upon their ability to report gains so as to at least be able to cater for the cost of continuing in their business dealings, hence the relevance of this extensive study on an optimal (near perfect) method of the credit assessment models employed.

The main objective of the researcher is to highlight on the numerous applied models used by the banks in credit risk analysis as far as loans are concerned in their credit scoring and dealings, as high rate of non-performing loans may lead to the ruin of the industry. Zenith Bank Ghana Ltd is used as a case study.

"Our vision is clear. Our mission statement is clear. Our reward system is good. Zenith is the best bank in the world. I won't trade it for any place. I look back and I am surprised at our achievements so far. The leap is too great."

Indeed, all of Zenith's performance indicators have been improving remarkably each year. The value of its operating assets grew by 282% between 2007 and 2009. From GHC 140 million in 2007, it increased to GHC 368.3 million a year later and reached GHC 535.1 million at the end of 2009.

The bank's share of industry assets rose from 2.1% in 2007 to 3.1% in 2008 and went up still to 4.1% in 2009. This saw it improve its industry ranking for this performance indicator from sixteenth to eight over the period.

A similarly impressive run was recorded in Zenith's share of industry deposits in the last three years. It accounted for 4.9% of industry deposits last year, a jump from 4.4% and 2.6% in 2008 and 2007 respectively. Overall, Zenith Bank had the 7th largest pool of deposits in the industry last year. The Kaplan Meier/

Nelson Alan as well as the Cox Regression Model and maximum likelihood hazard function would be adopted in this study: in effect this study, seeks to bring out a closer look at credit analysis performed around the Cox model, ascertaining the variance and probability of failure of the loan portfolio revised every single year with the help of Nelson Alen/ Kaplan Mieir estimators and the use of the hazard functions to model risk of every individual client after a loan for significant recommendations to help prevent the instance of failure of the Ghanaian Banking Industry in the long run.

1.3 Objectives of Study

This study is aimed at achieving three main objectives such as:

1. To suggest a simpler more efficient method of determining the credit score of a borrower based on a deterministic value of quantifiable loan applicant characteristics using the Hazard Function and the Cox regression model.
2. To see effects of covariates on default over time and to attempt to transform this data set.

Notwithstanding all these, the researcher allocated time, took up a dogged approach to the many questioning to enable the research work become an absolute success.

1.4 Methodology

The database was gathered from Zenith Bank Ghana Ltd at Ahodwo, Kumasi in Ashanti Region. The data is a secondary data which consists of demographic as well as behavioral tendencies of past (3) years and loan applicants of the bank.

Computations and resulting analysis were performed using excel and xlsat. Cox Hazard Function Model was used to create the predictor credit risk model based on past and current inputs.

1.5 Justification

Efficient credit risk analysis and management is crucial to the long-run success story of any banking institution; since it protect the stakeholders' interests and prevents most financial institutions from essentially going to ruin even inspite of provisions for loan loss. If this is not effectively managed, it may lead to the destruction of the credit market and in effect most financial businesses as occurred in 2008/09 leading to the world economic recession.

It is thus essential because it enables financial companies to ascertain how far the effects of demographic and behavioral characteristics such as age, sex, occupation, marital status and amount collected as well as many others quantify the creditworthiness of a loan applicant. Hence the hazard function and the Cox regression model can be used by managers of banking and loans industries to foresee a much lower defaulting expectancy than obviously being currently experienced each year under review.

1.6 Limitation

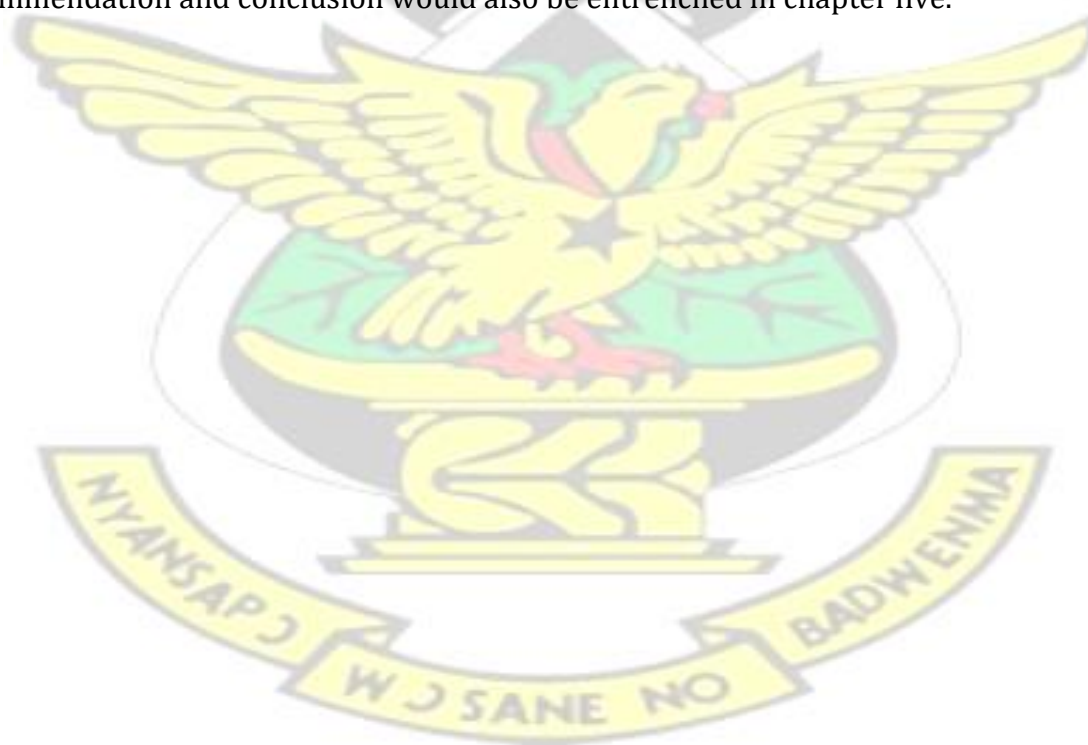
Amongst the many prevailing setbacks in regards to the perpetuation of the study, the following are the most predominant;

1. Unwillingness on the part of relevant the institution(s) in question to giveprivileged information relating to their customers.
2. Restricted access to well detailed data set of variables.

3. Lack of readily available mathematical, actuarial and statistical softwaresfor data analysis.

1.7 Organization of the Study

Chapter one touches on the background of study subject, the objectives, few of the setbacks encounter during the study and the organization of the study. Chapter two unearths the genesis and evolution of credit scoring and also with the help of researches of some of its other previous schools of thought that may have lead to the researcher stumbling into this thesis topic. Chapter three implies the methodology employed to perpetuate the objectives under study would be explicitly theorized. Analysis and interpretation of data gathered would be thoroughly performed out in chapter four. Summary of the outcomes, recommendation and conclusion would also be entrenched in chapter five.



Chapter 2

LITERATURE REVIEW

2.1 Introduction

This chapter shows genesis of credit risk analysis and management while also enlisting on the help previous other works of researchers using their various point of views or schools of thought as references.

2.2 History and Development of Bank Loan Financing

2.2.1 Credit Risk Defined

According to Zhang, University of Michigan (2009), “credit risk evaluations is a delicate area in banking sector and mainly concerns a fair variety of blue chip board members: corporate institutions, clients and standard regulators. A vital component of credit risk is the event of default, which occurs if the debtor is unable to settle h-is/er legal obligations in reference to the debt contract”.

The most of this risk is basically that of the lender and involves loss of principal as well as interest, disruption of lenders cash flows, and unnecessary extra collection costs. To reduce the lender’s credit risk, the lender may perform a credit check on the prospective borrower, may require the borrower to take out appropriate, such as mortgage or seek security or guarantees from third parties. In general, the higher the risk, the higher will be the interest rate that the debtor will be asked to pay on the debt.

There are two compartment of credit risk and can be simply captured by the types of default ratios, i.e. actual and the other being implied. The initial type refers to

the direct occurrences of defaults, sometimes denoted as the physical default ratios in finance. The second constitutes to the risk-neutral default probability formulated from the public credit market data deductions, e.g. public market bond. The study of open stock market credit risk has been inclined to the research of the implied defaults, which is yet a puzzling compartment of risk.

2.2.2 Origin and Evolution of Credit Risk Analysis

The source and for that matter the evolution of Credit can be mapped back to many several hundreds of years. Credit is said to be much older than writing. Savings and credit groups that have existed for several decades include the “susus” to Ghanaians, “chit funds” to Indians, “tandas” to Mexicans, “arisan” in Indonesians, “cheetu” to Sri Lankans, “tontines” in West Africa, “pasanaku” to Bolivians, and loan sharking, as well as many number of savings entities, lending clubs and borrowing societies are found globally.

As per Paddy Jonathan, optimistically early dating of the practice of risk analysis is experienced in the time of ancient Mesopotamia, later known as Greece and Rome, the former pertaining to scares total practice and the latter to the history of philosophy. Covello and Mumpower (1985), said, “defaulters faced penalties of being sold by creditors into slavery, but his spouse and wards may only be sold for a three-year duration”. Mapping out the practices of most places, credit default was a punishable crime with punishment varying from geo-area to area. Some of the common punishments including death, mutilation, torture, imprisonment or enslavement. The punishments had a tendency of being visited upon debtors and their dependents. Debt could be transferable to family members or political entities.

Gestel et al (2009) mapped a more linear and elaborated statements of development via linguistics of the term “risk”. The code that embodies the legal school of thought from nearly 4,000 years ago in Mesopotamia, Hammurabi’s Code, fails to lay down basic guidelines of lending, it did however state that failure to settle a debt as per agreed terms was and is a crime. The code unequivocally states that a defaulter should be treated identically to theft and fraud. Considering the penalties associated with credit and many dangers lenders faced in recouping their funds it fills one in today new age world with a mere wonderment reasons anyone would even consider borrowing or lending money in ancient times with these risks.

2.3 Credit Analysis in Developed Countries

According to the Casu et al (2006), “authors of Introduction to Banking, credit risk management is explained as a management tool which allows us to maximize a bank’s risk-adjusted rate of return by maintaining credit risk exposure within acceptable parameters”. The author proceeds to explain that “the need for financial institutions like banks to manage credit risk arising from individual creditors, individual transactions and the risk inherent in their entire portfolio”. The essential functions of credit risk management according to Raghavan (2003) are to “identify, measure, and more importantly monitor the profile of the bank”. By this definition, credit risk management helps financial institutions be able to check, track and evaluate their various activities in order to prevent and correct credit risk. The author also describes risk management as a system which is “a proactive action in the present for the future”.

2.3.1 The Bank for International Settlement (BIS)

The Bank for International Settlements is an international organization of central banks which came into existence to “foster cooperation among central banks and other agencies in pursuit of monetary and financial stability” (Wikipedia online, 2015). BIS as is expected isn’t accountable to any government and/or governmental agency”. It works via subcommittees, the secretariats it hosts, and through its annual General Meeting of all members. The BIS also provides banking services, but only to central banks, or to international organizations like itself.

Based in Basel, Switzerland, the BIS was established by The Hague agreements of 1930. As an organization of central banks, the BIS seek to make monetary policy more predictable and transparent among its 58 member central banks. While monetary policy is determined by each sovereign nation, it is subject to central and private banking scrutiny and potentially to speculation that affects foreign exchange rates and especially the fate of export economies. Two aspects of monetary policy have proven to be particularly sensitive, and the BIS therefore has two specific goals: to regulate capital adequacy and make reserve requirements transparent. Capital adequacy policy applies to equity and capital assets. These can be overvalued in many circumstances.

Accordingly, the BIS requires bank capital/asset ratio to be above a prescribed minimum international standard, for the protection of all central banks involved. The BIS’ main role is in setting capital adequacy requirements. From an international point of view, ensuring capital adequacy is the most important problem between central banks, as speculative lending based on inadequate underlying capital and widely varying liability rules causes economic crises as “bad money drives out good”(Gresham’s Law). The BIS sets requirements on two categories of capital, Tier 1 capital and Total capital. Tier 1 capital is the book value of its stock plus retained earnings. Tier 2 capital is loan loss reserves plus

subordinated debt. Total capital is the sum of Tier 1 and Tier 2 capital. Tier 1 capital must be at least 4% of total risk-weighted assets. Total capital must be at least 8% of total risk-weighted assets. When a bank creates a deposit to fund a loan, its assets and liabilities increase equally, with no increase in equity. That causes its capital ratio to drop. Thus the capital requirement limits the total amount of credit that a bank may issue. It is important to note that the capital requirement applies to assets while the bank reserve requirement applies to liabilities.

2.3.2 Credit Market Context

This section of my thesis utilizes publicly retrievable information on loans and how their made out to first-time clients on the peer-to-peer lending website Prosper.com between February of '07 and November of '09. This website links borrowers and lenders in the United States to each other as far as small loans are concerned. Prospective borrowers' issues out a request for a loan and also specify the interest rate limit they are willing to reach. Prosper pulls the borrower's credit history from a credit bureau and posts it with the loan request. Interested lenders may then choose to invest in the prospective loan applicant. If fair amount of lenders invests to the point that the entire amount is fully funded, then lenders compete with one another by bidding down the interest rate below the borrower's preset limits. When the bidding ceases, the most miniscule of interest rate is the winning bid becoming the interest rate on the loan and Prosper act as a clear house so as to speak and facilitates the loan transaction. If too little or no lenders are attracted and the loan is not appropriately funded, the loan request is denied and without necessitating any money exchange.

The website's peer-to-peer format is a magnet to many high-risk borrowers who are unable to secure credit from traditional lending establishments. "As a result,

prosper loans have a relatively high default rate of over 35% of the approximately 17,600 loans in a sample as at the end of 2013 were in default. Although the default probability for a Prosper loan is elevated, it is similar to default rates in other high-risk credit markets” (e.g., S. Miller, 2012 stated from Adams et al. (2009)). Obviously, mitigating loans that could eventually default is crucial for the profitability of lending institutions. Loans that default in the year one have a standard recovery rate of only 10.5%. Therefore, timing to default is very essential to the loss determination. Meanwhile, loans that go bad after the year two have a standard recovery rate of 70.2%. The median time to default for bad loans is 413 days from loan origination.

There is a large and growing body of literature that uses Prosper data to explore various aspects of credit markets. Many of these studies focus on the role of soft information unique to peer-to-peer lending. “Ravina (2012) and Pope & Sydnor (2011) find evidence that borrowers receive preferential loan terms if they post photographs of attractive people; (Duarte et al., 2009) find that lenders are influenced by the trustworthiness conveyed in the posted photograph. Freedman & Jin (2008) and Lin, Prabhala, & Viswanathan (2012) focus on the impact of Prosper online social groups on loan outcomes such as interest rate, funding probability and loan performance. Iyer et al. (2013) and Miller (2012) analyze how lenders use a wide range of information to infer the quality of potential borrowers. Ramcharan and Crowe (2013) use Prosper data to analyze how changes in house prices across states – perfect interest rates, delinquency, and credit rating for Prosper borrowers; Rigbi (2011) analyzes the effect of usury laws on access to credit, interest rates, and default probability, using an exogenous change in usury laws that apply to Prosper borrowers” (Miller, 2012).

In this research probe, I adopt a similar literature line and methodology characteristics as other studies using data for survival analysis, but focus on how

results are affected by modeling default more functionally using Cox proportional hazard. To achieve this, I use all of the verified financial information available to the credit officers at Zenith Bank, and self-reported employment status, not including other intangible or subjective inside data, such as the outward appearance or trustworthiness portrayed by the photo included in the client's application and compare their decision criterion to my test results to see there's any correlation.

The credit ratings and scores provide the individual costumers' or corporations' creditworthiness. The concluding statements centered which method is more forthright considering statistical models of the expected default probabilities, as well as judgment by rating/scoring specialists as far as their advantages and limitation are concerned.

Some rating and scoring basic the credit market related to the thesis include;

"The letters Aaa and Baa (in Figure 1.1) are terms often used in the Moody's risk rating system, which use Aaa, Aa, A, Baa, Ba, B, Caa, Ca, C to stand in for the dispensation for default running through the lowest to the highest" as detailed by Hamilton (2005).

"FICO, adapted by Fair Isaac Corporation, is the best-known consumer credit score and it's the most frequently used by United States of America banks and other credit card or mortgage lending instruments. It ranges from 300 (very poor) to 850 (best), and meant to represent the creditworthiness of a borrower i.e. the propensity of he or she will repay the debt. However, for the same client, the three main U.S. credit bureaus often report inconsistent FICO scores based on their personal proprietary models" (Zhang, 2009).

"For the reader interested in a more detailed discussion on the evolution of these techniques over the past 30 years we mention the works by Altman and Saunders

(1998), Saunders (1999), Crouhy et al. (2000), Hand (2001), Hamerle et al. (2003), Hanson & Schuermann (2004), Wang et al. (2005), and Chen et al. (2006)" (Miller, 2012).

2.4 Credit Analysis Techniques Common to the World

2.4.1 Empirical Models of Credit Analysis

Major Banks in Ghana employ actuarial models in credit analysis to improve the control of the uncertainties in credit decisions. The application of such models to credit analysis certainly reduces the reliance on discretion and improves objectivity in credit decisions. "Credit Analysis refers to the process of deciding whether or not to offer credit to a particular customer. It usually consists of two (2) steps: gathering relevant information and determining creditworthiness" (Ross et al., 2008). Literature and practice are superfluous in methods and approaches for credit assessment, both for individual and business loans.

2.4.2 The 5Cs Model

A common approach, which is applied for any number or array of groups of credit, is the "5C's Model". "For this approach different components, which determine the credit rating of a borrower, are classified, weighted off against each other and used in order to arrive at an informed final credit decision. The five "C's" are Character, Capacity, Capital, Conditions and Collateral" (Ross et al, 2008).

Character assessment implies that a lender ultimately has to make sure that the borrower is a trustworthy, prudent and reliable person who honors his obligations. That is, his or her willingness to meet the stipulated credit

obligations. The following factors may be used to analyze and predict this trustworthiness:

1. Degree, to which the borrower has come to honored past obligations, creditobligations or other financial obligations: that is payment of other regular bills.
2. Marital status.
3. Employment history.
4. Steadiness as far the residency.
5. Reputation: by references from employees, business partners, other banks,and or lending facilities etc.

2.4.3 CAMELS Model

A very popular rating model developed in the United States is the CAMELSModel. Each letter of the CAMELS stands for a different factor of the risk. This is displayed as follows:

1. C : Capital and Capital Adequacy.
2. A : Asset composition and quality.
3. M : strategic and operating management.
4. E : quality, structure and size of earning.
5. L : liquidity
6. S : sensitivity to market risk in particular interest rate and foreign exchange risks.

Each of these risk components are analyzed via ratios and qualitative analysis. Various components then receive component rating ranging from “A” to “E”, which are later combined to a similar composite rating.

2.4.4 Credit Scoring Model

Credit Scoring is also a viable option for credit analysis. It involves the process of computing the probability of default when issuing consumer credit. A credit scoring system basically looks at quantitative and qualitative characteristics of a prospective borrower and assigns, depending on the levels, to which the borrower’s characteristics meet certain stratification, and scores for the fulfillment of a single characteristic. A scoring model often computes 10-15 different characteristics.

Single scores are as result summed up to achieve a combined score. If the combined score is in excess of a pre-defined combined score limit or “cut-off score” if you will, the model implies, that the borrower is a “good credit risk”, and that the credit should be invariably extended, and vice versa for “bad credit risks”. For example, a particular successful application is for credit cards (Liu, 2005).

2.4.5 Risk Adjusted Return on Capital (RAROC)

According to Ieraci (2007), “RAROC is a risk-adjusted profitability measurement and management system for measuring risk-adjusted financial performance and for providing a constant view of profitability across business”. RAROC aids in accepting and/or rejecting decisions, loan pricing, structuring (for instance in the existence of collateral coverage), compare profitability across business sectors and settlement of business units. The RAROC equation is expressed below:

$$RAROC = \frac{(Total\ Revenue - Overhead - Expected\ Loss - Taxes)}{(Marginal\ Credit\ Capital)}$$

where,

Time Horizon = One year forward estimate of profitability.

Total Revenue = Expected 1st yr Spread Revenue + Upfront Fees.

Overhead = Non-Interest Expense (fixed charge applied per segment).

Expected Loss = Obligor PD * Facility LGD * Loan.

Exposure Taxes = Jurisdiction specific tax payable on loan income. Marginal

Credit Capital = 1st yr credit capital based on one factor VaR model

(similar to Basel II).

Quantification of credit risk of financial institutions with RAROC model consists of the following process. The initial stage is the loan officer along with internal credit analysts to estimate the propensity for default. The next stage in the process is to ascertain the bank's exposure given default. This involves estimating the outstanding associated with a credit line given a default. Exposure estimates are spread with the kind of credit and initial credit rating and are based on historical data. The next stage in the process is to compute the severity of loss given default.

2.4.6 Altman's Z-Score

"The Altman's Z-Score is a statistical measure that computes the distance a data point is from the mean of the data set (the standard deviations from the mean point). In a more financial sense, Z-score is the output from a credit-strength test that gauges the likelihood of bankruptcy" (Altman, 1993). The application of empirical credit assessment models is thought of having three major advantages.

2.5 Credit Risk Management in the Ghanaian Economy

As with most developing countries that have pursued economic and structural reforms, Ghana has been undergoing a process of financial sector restructuring

and transformation as an integral part of a comprehensive program for some time (Bank of Ghana Consultation Paper, 2007). “Unlike most central banks, which contended the global crisis through the lessening of monetary policy, the BoG was compelled to tighten policy in 2009. The finance, insurance, real estate and business services contributed about 5% of the GDP in Ghana” (Mhango, 2010).

Domestic credit growth was influenced by the change in fiscal policy in 2009 as well as the global economic recession. However, as public sector credit growth slowed, the global crisis emerged and further subdued the private sector’s appetite for credit therefore the need for credit risk management. The slowdown in economic activity in 2009 depressed incomes and reduced borrowers’ abilities to service their debt. “The author further spells that as a result of the economic crisis, the ratio of nonperforming loans increased to 9.6% at the end of March 2009, with above average ratios in the forestry, mining and construction sectors. The NPL ratio continued to increase during 2010 and reached an alarming 20% in February 2010. In response to the rising NPL ratio, commercial banks tightened their lending standards as a means of limiting their risk exposure” (Mhango, 2010).

2.5.1 Credit Services at Zenith Ghana

“Zenith Bank (Ghana) Limited, a privately owned African financial services provider, was incorporated in April 2005 under the Ghana Banking Law 2004 (Act 673) as a Subsidiary of Zenith Bank PLC and commenced universal banking operation in September 2005. It has been consistently proactive with heavy investment in staff, technology and infrastructure. In September, 2005 it was unheard of for a bank with a Nigerian parentage to flourish in a market which is oligopolistic in nature, and worse still Ghanaians at that time were unwilling to do business with a bank whose roots are firmly Nigerian because of their

perception of the '419' factor. During that period, three of the banks operating in the country controlled more than 60 percent of the total market share, leaving the remaining 40 plus to be shared among the remaining 19 banks at that time" (Quoted from the Zenith Home page). It has, true to form, always employed sound industry practices to expand its market share and to gain profitability and cost-structure competitive advantage.

Chapter 3

METHODOLOGY

3.1 Introduction

In this chapter, we would also discuss basic concepts on survival analysis entrenched in the depths of Cox PH Maximum Likelihood as an avenue for credit analysis of bank loans.

3.2 Description of the Data

3.2.1 The Type and Source of the Data

There is a right censored well detailed information of clients of Zenith Ghana Ltd, Ahodwo whom are currently patronizing their loan product facilities. This information is somewhat random given discrepancy in the times of acquisition of these loans but highly effective in the perpetuation of this research. The data used in our paper spans the period from 2012 to 2015, and falls into three main categories: data on bank loans, data used to construct client-specific balance sheet variables, and data used to construct takeover vulnerability.

In order to prevent a possible selection bias, we check that there are no systematic differences between Zenith Bank observations that could be matched and those that could not. We find that the distribution across types of loans, the probability of default each type of loan, the average time to default and how the potential loss estimate of individual loans impacts the Bank.

3.3 The Method of Data Collection

The researcher used a qualitative approach to collect data needed. In-depth semistructured interviews were used to gather extensive data from management and credit staff of Zenith Bank Ghana Ltd, Ahodwo.

3.4 Survival Models

3.4.1 Introduction of Survival Data

The objective of survival analysis is data in the form of times from a well-deemed time beginning to an end point. The end point could be the occurrence of some particular events or a particular duration point. In scientific research an easy example of this is an individual specimen in an experimental study, such as a clinical trial in comparison to several different kinds of treatments. The outcome of interest is the duration until an event occurs; that is, an analysis of the time until a specific event occurs. Such events include the time to default, after which there could be a default relapse-free survival time, time to virtual bankruptcy, time to device failure, and time to regain of full health. More generally, “survival times can also be observed in other application areas, such as the time taken by an individual to complete a task in a psychological experiment, the storage times of seeds held in a seed bank, or the lifetimes of industrial or electronic components” (Portnoy, 2003). One reason that survival data are not amenable to standard statistical procedures used in data analysis is that it generally is not

symmetrically distributed. Also, time survived is always positive while a normal distribution is up time values as members of all real numbers. Thus, the normal distribution presumption is invalid. This setback could be solved firstly by changing the data to display a positive and more symmetric distribution, however, a better approach is to adopt a different distributional model for the raw data.

The basic characteristics of survival data that makes standard methods inappropriate is that survival times are often censored. An individual survival time is said to be censored when the end-point of interest has not been observed. “Credit risk modeling in finance is closely related to survival analysis in statistics, including both the first-passage-time structural models and the duration type of intensity-based models. In a rough sense, default risk models based on the actual credit performance data exclusively belong to survival analysis, since the latter by definition is the analysis of time-to-failure data. Here, the failure refers to the default event in either corporate or retail risk exposures; we find that the survival models have at least four-fold advantages in approach to credit risk modeling:

1. Credibility in parameterizing the default intensity,
2. Credibility in incorporating various types of covariates,
3. Effectiveness in modeling the credit portfolios, and
4. being straightforward to enrich and extend the family of credit risk models”(Portnov, 2003).

The main aim of this paper is to introduce an alternative approach for modeling credit risk. More specifically, we will focus on estimating PD for consumer credits and personal credits using survival analysis techniques.

3.4.2 Covariates and Proportional Hazards

“Estimates of the lifetime distribution, whether parametric or non-parametric, are limited in their ability to deal with some important questions in survival analysis, such as the effect of covariates on survival.

A covariate is any quantity recorded in respect of each life, such as age, sex, type of treatment, level of medication, severity of symptoms and so on. If the covariates partition the population into a small number of homogeneous groups, it is possible to compare Kaplan-Meier or other estimates of the survivor function in respect of each population, but a more direct and transparent method is to construct a model in which the effects of the covariates on survival are modeled directly: a regression model. In this section, we will assume that the values of the covariates in respect of the i^{th} life are represented by a $1 \times p$ vector, z_i .

The covariates can be:

1. Direct measurements (eg age or weight)
2. Indicator or dummy variables (eg 0 for a male and 1 for a female or 0 for new treatment and 1 for placebo)
3. A quantitative interpretation of a qualitative measurement (e.g severity of symptoms from 0 to 5 with 0 representing no symptoms and 5 representing extreme severity)” (From the acted ct4 accessed in 2013).

The covariates used in this study are; Sex, Disposable Income Brackets , Marital Status, Number Of Dependents, Amount Of Loan Bracket ,Age, Term Of Loans And Period Of Repayment.

3.5 Proportional Hazards Models

“The most widely used regression model in recent years has been the proportional hazards(PH) model. Proportional hazards models can be constructed using both parametric and non-parametric approaches to estimating the effect of duration on the hazard function. In PH models the hazard function for the i^{th} life under observation, $\lambda_i(t, z_i)$, may be written:

$$\lambda_i(t, z_i) = \lambda_\theta(t)g(z_i), \quad (3.1)$$

where; $\lambda_\theta(t)$ is a function only of duration t , and $g(z_i)$ is a function only of the covariate vector” (From the acted ct4 accessed in 2013).

3.6 Fully Parametric Models

3.6.1 Parametric Models for the Hazard Function

“In a fully parametric PH model, the strong assumption is made that the lifetime distribution, and hence the hazard, belongs to a particular family of parametric distributions, and the regression problem is reduced to estimating the parameters from the data” (From the acted ct4 accessed in 2013).

The pdf of the future lifetime random variable T -axis

$$f_t(t) = t\mu_{x+t} = \mu_{x+t} \exp\left(-\int_0^t \mu_{x+s} ds\right) \quad (3.2)$$

The “hazard” referred as just the force of mortality. You may see the hazard function (from age x) written as $h(t)$ or as $h_x(t)(t)$.

Note also that since $F_t(t) = P(T_x \leq t) =_t q_x = 1 -_t p_x$, we can write:

$$\mu_{(x+t)} = h_x(t) = \frac{fT_x(t)}{1 - FT_x(t)} \quad (3.3)$$

Distributions commonly used are the exponential (constant hazard), Weibull (monotonic hazard), but the Gompertz-Makeham (exponential hazard) and loglogistic (“humped” hazard) will be exempted from this study.

“A common feature of all these papers is that they use parametric or semi parametric regression techniques for modeling the time to default (duration models), including exponential models, Weibull models and Cox’s proportional hazards models, which are very common in this literature. The model established for the time to default is then used for modeling PD or constructing the scoring discriminate function” (From the acted ct4 accessed in 2013).

3.7 The Cox Proportional Hazards Model

3.7.1 Introduction

“The Cox hazard regression model deduced by Cox (1972) has been widely used in the analysis of survival data in medicine and has in recent times become popular in finance. A vital aspect of the Cox model is that at any time t , the ratio of the hazard rates of a type j credit event for two different issuers does not involve the baseline hazard function” (From the acted ct4 accessed in 2013).

The Cox regression model is often referred to as the “proportional hazard” model because, if the covariates are all calendar time-independent, the ratio of hazard rates is constant over time. However, our covariates are time-dependent, so the ratio of hazard rates does change with t . The Cox model is a semi-parametric proportional hazards model under which the force of mortality (or hazard function) for an individual for the i th life is given by:

Definition:

$$\lambda(t; z_i) = \lambda_o(t) \exp(\beta^T z_i^T) \quad (3.4)$$

$\lambda_o(t)$ is the force of mortality that is proportional to the baseline hazard (From the acted ct4 accessed in 2013).

3.7.2 Fitting the regression parameters.

The regression parameters are estimated by maximizing the partial likelihood:

$$L(\beta) = \prod_{j=1}^k \frac{\exp(\beta^T z_j^T)}{\sum_{i \in R(t_j)} \exp(\beta^T z_i^T)} \quad (3.5)$$

Breslow's Approximation To The Partial Likelihood.

If there are ties in the data, i.e. the death times are not distinct, then Breslow's approximation to the partial likelihood can be used:

$$L(\beta) = \prod_{j=1}^k \frac{\exp(\beta^T S_j^T)}{[\sum_{i \in R(t_j)} \exp(\beta^T z_i^T)]^{d_j}} \quad (3.6)$$

One needs only differentiate this partial likelihood relation with respect to find a specific β_z i.e. The maximum partial likelihood estimator which the researcher used xl-stat to compute.

Distribution of the maximum partial likelihood estimators of the regression parameters

"The maximum partial likelihood estimator of the vector of parameters, which we denote by $t(\hat{\beta})$ has the following asymptotic properties:

1. It has an asymptotic multivariate normal distribution.
2. It is asymptotically unbiased.

3. Its variance matrix is equal to the negative of the inverse of the observed information matrix, i.e. the negative of the inverse of the matrix of second derivatives of the log-likelihood, evaluated at the point $\hat{\beta}$ (From the acted ct4 accessed in 2013).

If $\hat{\beta}$ is the maximum partial likelihood estimator of β , then the asymptotic variance of $\hat{\beta}$ is given by:

$$Var(\hat{\beta}) = [I^{-1}(\beta)]_{(\beta=\hat{\beta})} \quad (3.7)$$

As $\hat{\beta}$ is asymptotically normally distributed, a 95% confidence interval for β_j (the j th parameter) is:

$$\hat{\beta}_j \pm 1.96 \sqrt{var \hat{\beta}_j}$$

Model testing

We can compare two models using a likelihood ratio test. Suppose we want to compare a model with p covariates against an extended model with an extra q covariates.

The null hypothesis for this test is:

$$H_0: \beta_{(p+1)} = \beta_{(p+2)} = \dots = \beta_{(p+q)} = 0 \quad (3.8)$$

Survival function for typical a life;

$$S(t) = \exp\left\{-\int_{s=0}^t h(s)ds\right\} \quad (3.9)$$

The test statistic is:

$$-2(\ln L_p - \ln L_{(p+q)}) \quad (3.10)$$

where $\ln L_p / \ln L_{(p+q)}$ denotes the maximized log-likelihood of the model with p and $p + q$ covariates respectively.

If the null hypothesis is true, then the test statistic should be a realization of a X_q^2 random variable. So we reject the null hypothesis at the 5% significance level if the value of the test statistic is greater than the upper 5% point of X_q^2 .

Model's Goodness of Fit

As in other standard procedures, when the proportional hazards model is fitted, one should examine model fit. Residuals are the most popular diagnostic, and they are usually some form of measure of discrepancy between fitted and predicted values. Therefore, if the model is adequate, the plot of residuals should not show any unexpected patterns. Among the simpler methods are:

Cox-Snell residuals (Cox and Snell 1968) are “defined as follows:

$$r_{ci} = \exp \hat{H}_i(t_i) = -\log \hat{S}_i(t_i), \quad (3.11)$$

Where $\hat{H}_i(t_i)$ represents the estimated cumulative baseline hazard, $\hat{H}_i(t_i)$ represents the estimated cumulative hazard for the i^{th} individual at time (t_i) , and $\hat{S}_i(t_i)$, is the estimated survivor function of the i^{th} individual at time (t_i) ”.

It can be illustrated (Collett, 1994) that $-\log \hat{S}_i(t_i)$, has an exponential distribution with unit mean, regardless of the form of $S(t)$ is. If the model fitted is sufficient, then the estimated survival function $\hat{S}_i(t_i)$ will be close and will have similar properties to $S_i(t_i)$. Hence $r_{ci} = -\log \hat{S}_i(t_i)$ will be a set of observations from an exponential distribution with unit mean. To test that the residuals have unit exponential distribution, the Kaplan-Meier estimate of these values is computed. Thus, $\log(-\log \hat{S}_i(r_{ci}))$, is plotted against $\log(r_{ci})$. A straight line exhibiting unit

slope and zero intercept indicates that the fitted model is correct” (Stepanova & Thomas, 2002).

“**Martingale residual** (Therneau et al. 1990) is an extension of the Cox-Snell residual:

$$r_{Mi} = \delta_i - r_{ci}. \quad (3.12)$$

It can be explained as the lapse between the observed number of defaults for an individual in the interval $(0, t_i)$ denoted δ_i and the expected number of failures according to the model. r_{Mi} is plotted against the rank order of time”. Ideally, it should not exhibit any pattern if the model is adequate (Stepanova & Thomas, 2002).

“**The Schoenfeld residual** is the lapses between the observed value of the covariate x_i and its expected value, conditional on the risk set R_i . This implies that clients who are unlikely to default (i.e., their covariate value is similar to those in the risk set) will have a small absolute value of the residual. Individuals who are likely to default relative to those at risk will have a large absolute value of the residual” (Schoenfeld, 1982).

“The main distinguisher of this residual from all the many others is that it has a vector of values for each individual, a value for every covariate. Schoenfeld residuals are plotted against rank order of time for their corresponding covariate. These plots are then used in investigating whether the specific covariate needs a transformation or whether there is an indication of time dependency” (Stepanova & Thomas, 2002).

In closing argument, we intend to estimate in the next chapter the cause-specific hazard function to study how the risks of different types of credit events depend on covariates.

KNUST



Chapter 4

RESULTS AND ANALYSIS

This chapter consists of data analysis (preliminary and Cox regression analysis) and discussion of results.

4.1 Introduction

Quantifying the probability of default, PD, in consumer credits, loans and credit cards is one of the most fascinating conundrums to be solved by banks, savings banks, savings cooperatives and other credit outlets.

4.2 Preliminary Analysis

This section gives preliminary analysis of the data and it involves exploration of the data by means of summary statistic and charts.

The purpose is to determine if there exist any correlation between the many different features or simply, covariates and the likelihood of default on loans. We want to know the credit worthiness of a client base on number of quantifiable borrower's characteristics. The data is a secondary data which consist of both demographic and behavioral characteristics of past (3 years) and prospective customers of the bank.

Assumptions and notations include that, there are no deaths, the event would be a survival till the full payment term of the loan. Corporate loans were disregarded for the intents and purposes of this study's optimal efficiency. As an observational plan an individual is censored when they default. Non-informative type 1 censoring are more technical ways of describing our data's censoring mechanism.

The source of data was collected from Zenith Ghana Ltd Bank at Ahodwo, Kumasi in Ashanti Region. XLstat an extension of excel was used to undertake the sample analysis which included descriptive statistic and analysis of association.

4.2.1 Sex

In Table 4.1, it was observed that, out of 137 candidates who applied for loans from that branch of the bank, 69 appeared to be male computing to 51% of the total loanees, while 68 were female marking 49% of the total number of loan bearers. 54 and 45 respectively of male and female non-defaulters paid back all money owed normally. This accounted for 54.54% as far as the males and 45.45% as far as females are concerned of the total number of non-defaulter. It is observed that 38 clients defaulted representing 27.74% of the total borrowers with the female folk ahead at 16.98% of the overall number of borrowers while 10.95% of those were male. It is evident in fact that a dominant amount of females defaulted, 23 represented as 60.53% as compare to the overall number of defaulters that of the males which was 15 accounting for 39.47%.

Table 4.1: Cross Table - Payment Status versus Sex

Count of Default Paid in Bill	Sex		Total
	female	Male	
0	23	15	38
1	45	54	99
Total	68	69	137

4.2.2 Disposable Income Brackets

These disposable income bracket were broken up into strata from 0-5, in order of severity. Where order 0 would denote disposable salaries in the regions of 100-999, order 1 would denote disposable salaries in the regions of 1000-1999, order 2 would denote disposable salaries in the regions of 2,000-2,999, order 3 would denote disposable salaries in the regions of 3,000-3,999, order 4 would denote

disposable salaries in the regions of 4,000-4,999 and order 5 would denote disposable salaries on the upwards of 5000 baring no limit.

Table 4.2: Disposable Income Brackets

Count of Defaulted/Paid In Full (Censoring)	Column Labels						
Row Labels	0	1	2	3	4	5	Grand Total
0	24	10	3			1	38
1	21	51	17	6	4		99
Grand Total	45	61	20	6	4	1	137

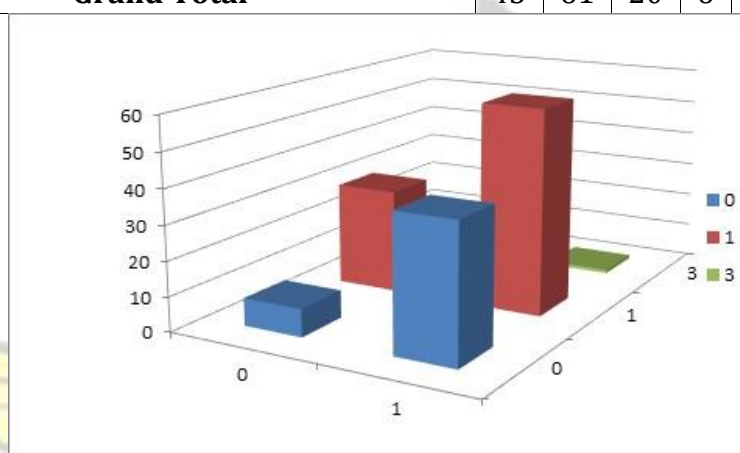


Figure 4.1: Graph of Disposable Income Brackets

In Figure 4.1, we figure out that most of the people who collected loans were in the salary brackets of 1 and 0 orders respectively. These brackets account for the most amounts of default 24 for 0 and 10 for 1 in (Appendix A).

Others may argue that as far income and employment is concerned it may be important to discuss whether a burrower is self-employed or not but the researcher believed that what is relevant in actuarial circled as far as money earned goes simpler to concentrated on more quantifiable values.

4.2.3 Marital Status

Table 4.3: Cross table of payment Status versus marital status

Count of Defaulted/Paid In Full (Censoring)	Column Labels		
Row Labels	0	1	Grand Total
0	8	30	38
1	39	60	99
Grand Total	47	90	137

The marital status was fairly divided into Single (coded as 0) and married (coded as 1). Clients whom were married borrowers are observed to be more 90 (65.69%) as compared to the single borrowers 47 (34.31%). The probability of defaulters represented 78.94% of the borrowers who are married. It is observed to be more than probability of defaulter representing 28.47% of the overall number of burrowers. The proportion of defaulting borrowers amongst whom were single is 24.49 % which is below the proportion of people who defaulted. It is evident from observation that more of the loan clients who defaulted were married, 27 marriage borrowers representing 69.23% out of 39 clients defaulted. Whereas married clients represented 64.23% of total number of borrowers, just 61.32% of the married customers settled their loans normal (did not default) as shown in

Table 4.1.

Table 4.4: Cross Table - Payment Status versus Marital Status

			Marital Status		Total
			Single	Married	
Payment State	Non default	Count	39	60	99
		% of non-defaulting	39.39	60.61	100
		% of overall number of marital status	82.98	66.67	
		% of Total	28.47	43.8	
	Default	Count	8	30	38
		% of defaulting	21.05	78.95	100
		% of overall number of marital status	17.02	33.33	
		% of Total	100	100	
Total		49	88	137	

4.2.4 Number of Dependents

Clients who had one and no dependents were dominant (38 each), which represented 27.74% of borrowers, followed by those who two dependents and then borrowers with exactly three dependents (33 and 12 clients respectively). Customers with two dependent defaulted most, since 34.21% of customers within the total number of defaulter 13 clients were reported to have defaulted. A more thoroughly detailed representation of distribution of the data according to clients' dependents variety is found below:

Table 4.5: Cross table Payments verses the Number of dependents

Count of Defaulted/ Paid In Full (Censoring)	Column Labels											
Row Labels	0	1	2	3	4	5	7	8	9	10	12	Grand Total
0	9	6	13	3	3	1		1		1	1	38
1	29	32	20	9	2	5	1		1			99
Grand Total	38	38	33	12	5	6	1	1	1	1	1	137

The percentage of default within each number of dependent for zero, one, two, three and more than three are 23.68% ,16.79% ,39.39% ,23.07% and 23.57% respectively.

4.2.5 Amount of Loan Bracket

Clients who obtained amounts within ranges GHS 100.00-GHS 999.00 were ranked 0, GHS 1,000.00-GHS 4,999.00 were ranked 1, GHS 5,000.00-GHS 9,999.00 were ranked 2, the upsides of GHS 10,000.00(rank 3).A bulk amount of the clients (41.61%) took out loans within ranges of GHS 1,000 - GHS 4,999 further denoted as order 1. (refer to Figure 4.1.5 Below). This was followed by range of loan amounts taken by clients GHS 100 -GHS 999 representing 35.04%. Only 7 people collected on the upwards of GHS 10,000 representing 5.11%. Almost every clients who took an amount of loan above GHS10,000 and range of amounts from

GHS 5,000-GHS 10,000 paid their debt regularly. Customers who borrowed GHS 100.00-GHS 999 had highest propensities to default with 21 representing 55.26% of its members who took out the loan defaulted. The next highest category too default was the loans range of amounts GHS 1,000.00-GHS 4,999.00 defaulting more than those who collected above GHS 10,000.00. There were no loan amount category that has no defaulting clients.

Table 4.6: Payment Status versus Amount Collected

Defaulted/Paid In Full (Censoring)	Loan Amount				
Row Labels	0	1	2	3	Grand Total
0	21	13	3	1	38
1	27	44	22	6	99
Grand Total	48	57	25	7	137

4.2.6 Age

The study observed 137 loan clients, 41 of these were between ages 30-39 making a proportion of 29.93% of the total number of borrowers. This was followed by 28 clients who were in the 20-29 years. All in all, 11 defaulters were recorded within the ages of 40-49 years representing 28.95% followed by customers 30-39 years 10 (26.32%). The age ranges of 70-79 years had the no defaulters. 18.42% of the customers within the age 20-29 years defaulted. After that the age ranges of 50-59 years had least number of defaulters 2(5.26%). Of the 26 clients between 60-69 years, 8 of them defaulted accounting for 21.05% (refer to Appendix A).

Table 4.7: Cross table - payment Status versus age

Defaulted/Paid In Full (Censoring)	AGE ORDER						
	0	1	2	3	4	5	Grand Total
0	7	10	11	2	8	0	38
1	21	31	18	9	18	2	99
Grand Total	28	41	29	11	26	2	137

4.2.7 Term of Loans

A dominant number of the clients, 75 (54.74%) took out loans with within range of six months due to a six month loan promotion (refer to Figure 4.1 Below). This was followed by range of loan term taken by clients of a year representing 22.63%. Only 2 people took loans with terms on the upwards of 4 years representing 1.46%. Clients who patronized the six month loan promotion observed the highest number of defaults representing 42.11%. This was follow by the yearly loan, 10 representing 26.32%. A detailed representation of this distribution is structured below in Table 4.8.

Table 4.8: Cross table -Payment Status versus loan term

Defaulted/ Paid In Full (Censoring)	TERM OF LOANS								
Row Labels	0	1	2	3	4	5	6	7	Grand Total
0	16	1	10	3	2	3	1	2	38
1	59	9	21	4	2	3	1		99
Grand Total	75	10	31	7	4	6	2	4	137

4.3 Regression Analysis

This portion highlights behavioral and demographic factors that go a long way to significantly influencing characteristics which may act as a predictor of individual who are most likely to default on loans they have acquired. All that is observed in the preliminary analysis is further reanalyzed under the rigorous use of statistical tools such as chi-square goodness-of-fit test and test for independence covariates. The cox regression model was enlisted to ascertain the credit worthiness of a loan client.

4.3.1 Test for Multicollinearity Among Independent Variables

Multicollinearity in the model is unearthed by taking note of the standard error for coefficients (ith beta). Any covariate with a corresponding standard error

larger than 2.0 implies numerical problems, such as multicollinearity amongst the independent variables. Independent covariates that indicate numerical problems were dropped before analysis. Hence, none of the independent covariates in this experiment had a standard error larger than 2.0.

4.3.2 Over all Relationship Between independent and dependent variables

Let the Hypothesis be stated as follows.

H_0 : there exist no difference between the model with only a constant and the model with independent variables.

H_1 : a relationship exists between the dependent and independent variables.

Table 4.9: Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	-2	32.29671	8	0.000
	Log(Likelihood)			
	Score	36.64014	8	0.000
	Wald	29.65959	8	0.000

The occurrence of a pattern between the dependent variable and combination of independent variable is based on the statistical significance of the model chisquare at step1 after the independent variable have been added to the analysis (Table 4.9).

4.3.3 Relationship of Individual Independent Covariates to Dependent Variable

$$H_0: \beta_1 = \beta_2 = \dots = \beta_l$$

$$H_1: (\beta_1 \neq \beta_2 \neq \dots \neq \beta_l)$$

In this instance we consider that the regression parameters β_5 and β_7 are positive (which are 0.43 and 0.422 respectively), this would imply that the hazard rate (e.g. the force of mortality) increases with these j th covariate, i.e. there is a positive correlation between hazard rate and covariate. For instance, a loan client who has a significantly increased no of dependents is more likely to default on loan's contractual obligation, we would expect to discover the regression parameter corresponding to the covariate representing weight to be positive.

In this instance we consider that the regression parameters $\beta_1, \beta_2, \beta_3, \beta_4, \beta_6$ and β_8 are negative (which are -0.287, -0.169, -0.877, -0.047, -0.162 and -0.056 respectively), their dependent variable (e.g. the force of mortality) would decline with their correspondent independent variable, i.e. there is a negative correlation between hazard rate and covariate. For instance, a loan client who has an elevated disposable income are less likely to default.

If-so-facto it follows as the absolute value of the j -th regression parameter if extensive eg. ($\beta_3 = 0.877$), the hazard rate is significantly affected by the j -th independent variable, i.e. the stronger the correlation between dependent and independent variable. Adversely if the magnitude of the j -th regression parameter is small (e.g. $\beta_4 = 0.047$), the dependent is not significantly affected by the j -th independent variable, ie there is a weak correlation between dependent and independent variables.

The significance of the independent variables is tested statistically and the 4 th (the number of dependents) covariate was found to be significant there were insufficient data reject the null hypothesis for all the other covariates.

4.3.4 Strength of Model Relationship

Cox-Snell residuals were evaluated as proposed in the third chapter to ascertain whether Cox-Snell residuals exhibits unit exponential distribution, the KaplanMeier estimate of the survivor function is deduced and log-log transformation of the resulting values is drawn up against log of the matching residual (Figure 4.2).

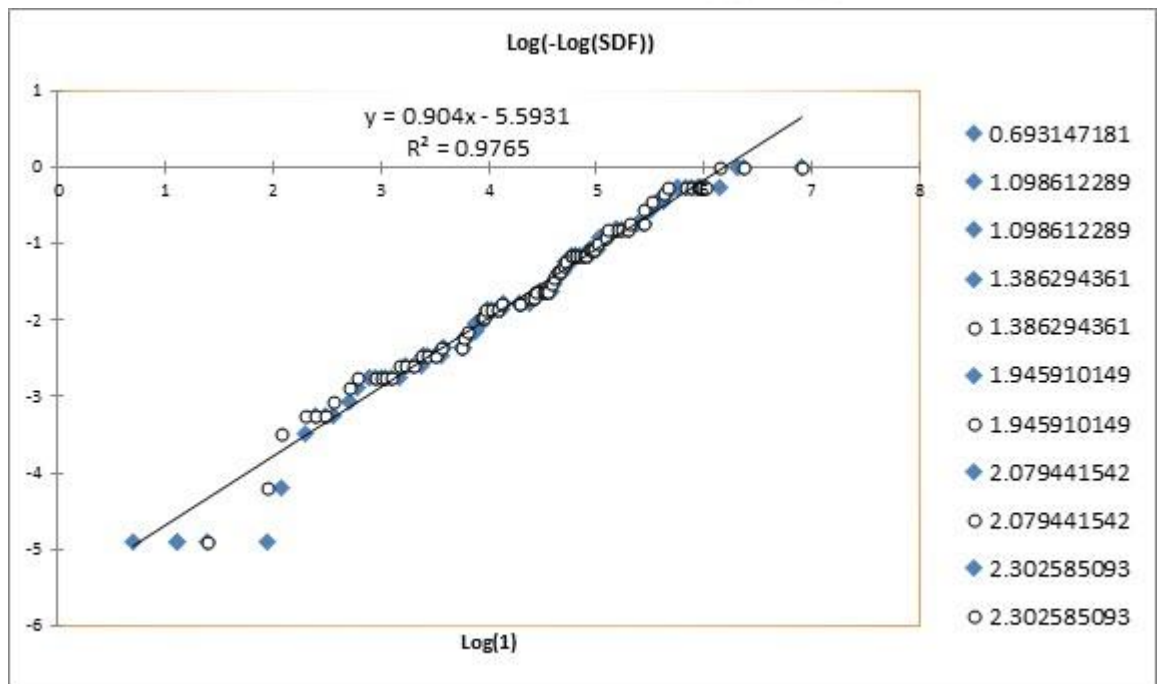


Figure 4.2: Graphical representation of Cox-Snell Residual

The curve of plotted points significantly resemble a straight line as seen in Fig 4.2, with a unit slope and zero intercept if the observations with the lowest residuals are ignored. There are only a few of those, and they correspond to the loans with the shortest lifetime'one month. The value is 0.9765 this gives an impression that about 98% of the variation of payment status is explained by the independent variables. The R^2 model is therefore useful in predicting default. A straight curve with near unit slope (0.904 to be exact) and zero intercept of 6.24 implies that the model is fitted adequately.

4.3.5 Classification of the model

This implies about how the model identifies Defaulting persons from Non-Defaulting individuals, can be analyzed using the classification table. As show in Table 4.10, the model is able to predict Default and Non-Default adequately (57.89% and 82.83% respectively). Ultimately, the model is 75.91% accurate of identifying Defaulters from Non-defaulters.

4.3.6 Goodness of fit

“**Martingale residuals** is plotted against rank order of time. The scatter of the points within a band increases with rank order of time. It is expected because calculation of r_{Mi} involves an estimated survivor function that close to one for earlier times, hence its logarithm is close to zero and r_{Mi} will be clustered around one for uncensored and around zero for censored observations” (Stepanova & Thomas, 2002).

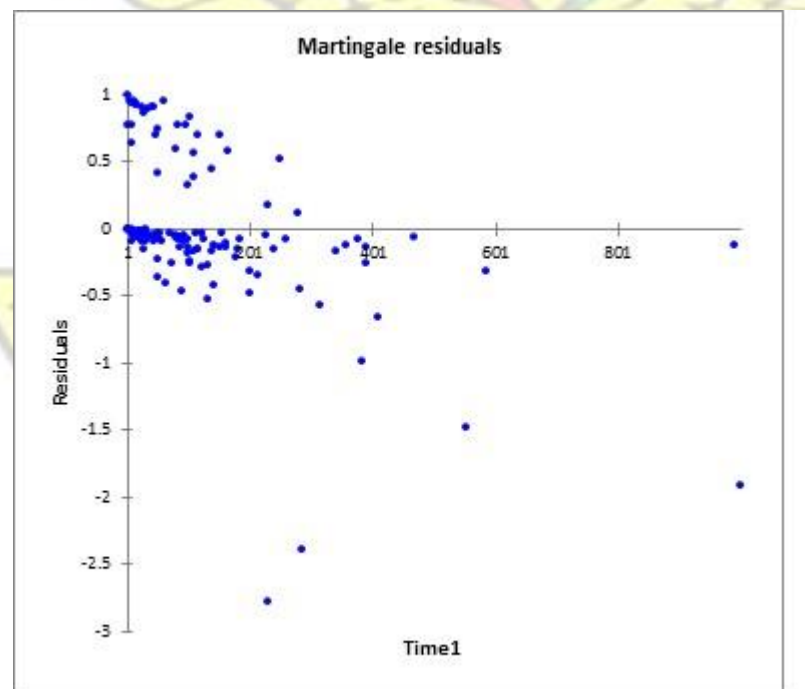


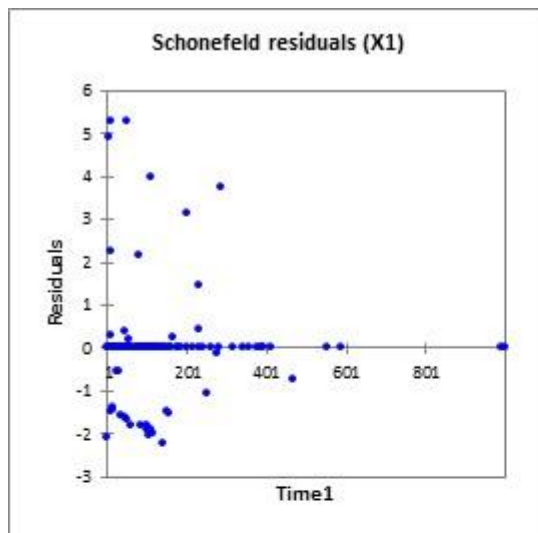
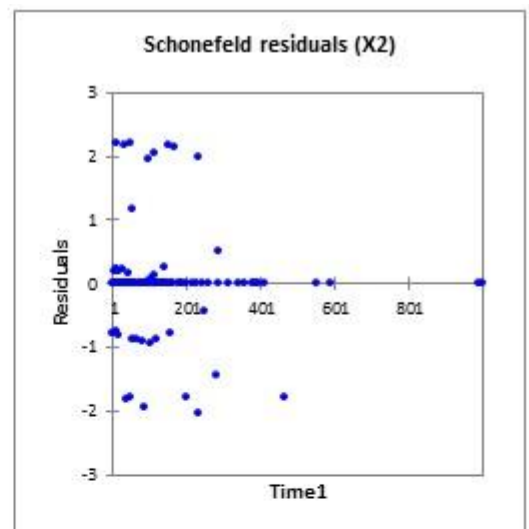
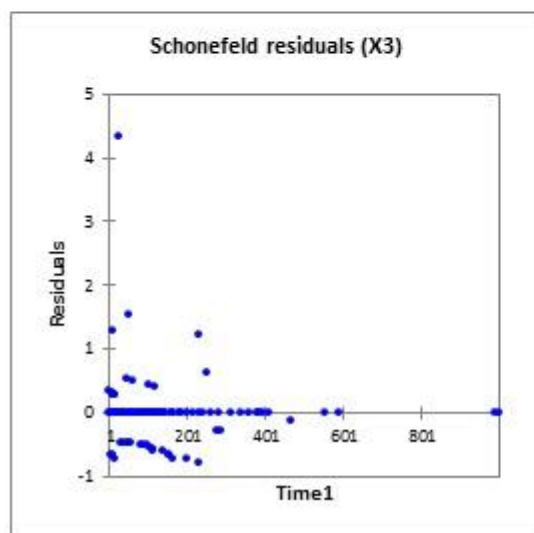
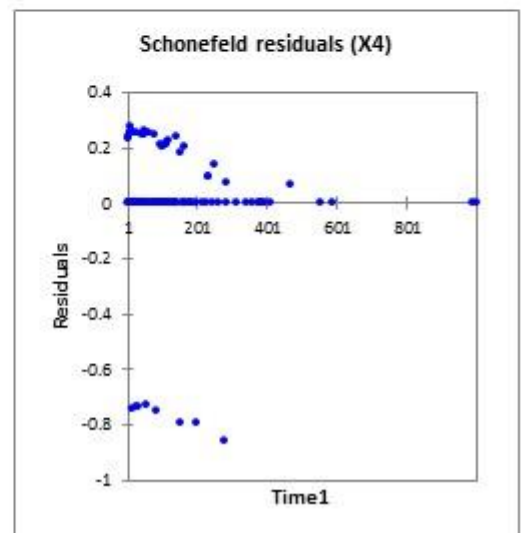
Figure 4.3: Graphical representation of Martingale’s Residual

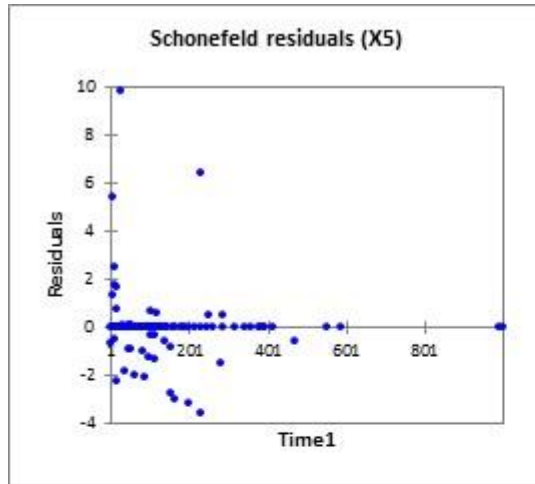
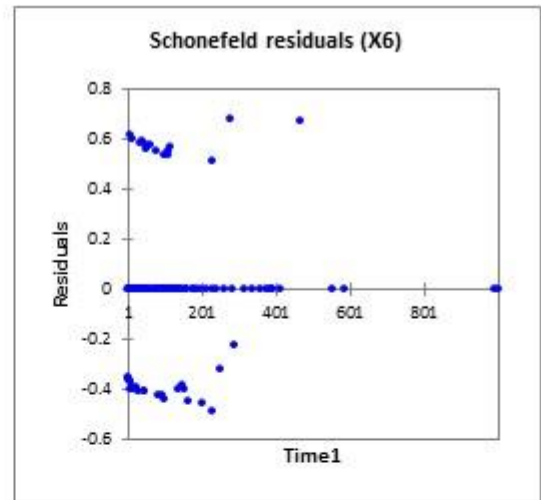
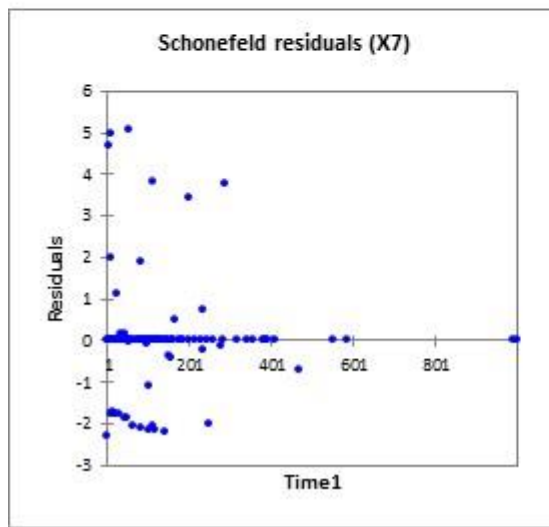
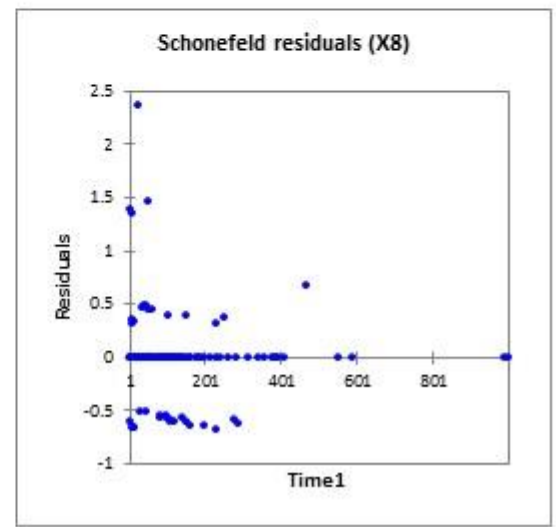
Ideally, it should not exhibit any pattern if the model is adequate.

Schoenfeld residuals (sample plot is in Appendix C) were “plotted to investigate the following:

1. Whether any covariates need to be transformed;
2. Whether the effect of a covariate on the survival time changes over time”(Stepanova & Thomas, 2002).







This diagnostic is very tedious when the size of covariates is as large as in our data. The plots do not show any signs of time dependency, necessitate any or transformation.

4.3.7 The Generalized Theoretical Model

The cox regression model for individual covariates is represented below,

$$\lambda_t = \lambda_0 \exp^X(\beta_i x_{ij}) \quad (4.1)$$

$$\lambda_t = \lambda_0 \exp -0.287(0 - 7) - 0.169(0 - 5) - 0.877(0 - 5) + 0.430(0 - 12) - 0.162(0 - 1) + 0.422(0 - 7) - 0.056(0 - 3).$$

4.3.8 Baseline Hazard Function

The baseline hazard function would be a six month loan taken by a single female without dependents between the ages of 20-29 whose disposable income and loan amount is within range of 100-999.

4.3.9 Final model

The best-fit cox regression model for individual characteristics based their parametric coefficient significance is formulated below;

$$\lambda_t = \lambda_0 \exp 0.430(0 - 12) \quad (4.2)$$

4.3.10 Cox Regression Model Calculator

It is modeled to compute the probability of default. Substituting the values of the independent variables which are in the model into the cox regression model, the probability of default can be calculated.

Example: A client whose age is 44 years, monthly disposable income of about GHS 800, married, female, with four (4) dependents and sought out a loan of about GHS 990.00 will variably have a default rate 0.999. Therefore, definitely has to be refused a loan (Refer to Appendix A).

4.3.11 Respondent Cox Regression Model Estimator.

It is structured to compute the bank's default rate and for that matter gives the rates of default of clients who collected a loan. Table 4.9 gives the result of 137

clients who borrowed money from the bank with their individual survival rates as well as their average probability of defaulters.

Table 4.10: Cross Reference Table of Model's Payment Status versus the Actual

Defaulted/Paid In Full (Censoring)			
	0	1	Grand Total
0	22	16	38
1	17	82	99
Grand Total	39	98	137

4.3.12 Type I and Type II Error

Let the Hypothesis be stated as follows:

H_0 : The borrower will default on a loan.

H_1 : The borrower will not default on a loan.

From Appendix 4.2.5, the number of defaulters and non-defaulters correctly predicted by the model appeared to be 22 and 82 respectively. The number of predicted defaulters that eventually turn out did not default was 17 and the number of predicted non-defaulters that in contrast defaulted was 16. The total number of defaulters and non-defaulters predicted by the model were 39 and 88 respectively.

False positive Rate=False Default/Non-defaulters= $17/99=0.17$

True Positive=True Defaulters/Defaulters= $22/38=0.58$

Type I error= $1-0.58=0.42$

Type II error= $0.1717=0.17$

Sensitivity= True Positive/ (True Positive +False Non-default) = $39/ (39+16)$
= 0.71

Specification= True Negative/ (False Negative +True Negative) = $98/ (98+17)$ =
 0.85

Positive Predictive value=True Positive/ (True Positive + False P) = $39/ (39+17)$
= 0.70

Negative Predictive value= $TN/ (TN+FN) =98/ (98+16) =0.86$

Chapter 5

FINDINGS, CONCLUSION AND RECOMMENDATIONS

5.1 Introduction

This portion ends the research and provides recommendation that would be essential to quantify default rate or for that matter creditworthiness in the financial industry as it is an immensely significant applications of quantitative analysis in business.

This work highlights two extensions that expands on the modern-day's application of Cox's proportional hazards model relevant to creating credit-scoring models. Firstly, it elaborates on a fresh coarse-classifying approach for the covariates in credit scoring. And also it elaborates on how the residual tools can enable researchers analyze the fitness of the model, and discusses pros and cons of each of these tools.

5.2 Summary of Findings

Sex is critical matter of discussion in ascertaining the probability of default. There is a common notion that females default less often on loans possibly because of their more heightened affinity risk aversion. Other researches may hold that the default rate of females is lower than that of males even though most financial institutions do not usually consider this notion in their scoring sheets. It is observed that 21.74% of male clients defaulted whilst 33.82% of female clients had payment issues.

As far as this work goes the research shows that 53.33% of the borrowers in the disposable income levels of order 0 (ranging from 100-999) defaulted while 16.39% of the disposable income levels of order 1 (1000-1999), with the rest being minute and randomly interspersed among the rest of the orders. This indicates that those with a lot of competing opportunity cost for their income found it difficult meet to the requirement of their loan contract. Whereas those with higher disposable income appeared to have little figures and proportions of default if not none as far the higher ranking orders are concerned.

This study goes on to analyze whether a client's marital status has any form of influence on default rate since it is a point of consideration as far as determining the trustworthiness of a borrower is concerned. Common sense suggests that married borrowers are more credit-worthy than the single since the responsibility of family will somehow lower the probability of going back on the loan agreements. Contrary to most credit default studies testifying that the rate of default amongst singles applicants being higher than that of the married, the events of this study is different from these preexisting literature. Married customers defaulted than the single as 33.33% of married defaulted while 17.02% of the single clients defaulted. The deviation from the norm may not be as a result of irresponsibility or the absence of maturity when considering the married clients, but may be due to the geographic area's economic standard of living and the difficulties in sustaining families in the long run.

Dependents refers to the number of persons that the borrower has to provide for. With every addition of the count of dependents, so does added strain on the borrower's income as a result of additional expenditure possible involving basic necessities as housing, cloths and etc., in the grand scheme of things. In this instance considering from zero to one dependent reduces the default percentages from 23.68% to 15.79% which goes against any logical stand point. The events of

this research indicates, the greater the count(s) of dependent, the greater the disposable income adjustment to counteract the consequent increase in the rate of default. Also most people with no dependents had low disposable income. The percentage of default within each number of dependent for zero, one, two, three and more than three are 23.68%, 16.79%, 39.39%, 23.07% and 23.57% respectively.

One would have presumed that older one is the less likely to default due to older fellow's affinity to risk aversion. The risk of default does therefore reduce in the later stages of life from this study. It is observed that 40-49 year had the dominant figures of defaults. Approximately 37.93% of clients in this strata defaulted on their loan settlements. The oldest class appeared to have no defaults. The risk of default fluctuates randomly all throughout the age groups from 20-29 years, 30-39 years, 50-59 years, to 60-69 years with their respective rate of default within their group being 25%, 24.39% 18.18% and 30.77%.

The amount of loan bracket obtained can be seen to influence the levels of credit risk of a borrower. It is observed from the sample survey that the range of amounts denoted by orders 0(GHS 100-GHS 999.00) to 5(GHS 50000-) had declining numbers of default as the orders and amount increased. Client who obtained amounts within ranges GHS1 00.00-GHS 999.00 (ranked 0), GHS 1,000.00-GHS 4,999.00 (ranked 1), GHS 5,000.00-GHS 9,999.00 (ranked 2), upwards of GHS 10,000.00(rank 3) had the following percentages of default 43.75%,22.81%,12% and 14.29% respectively within their individual ranks.

This cox mockup of a credit assessment model processes a 75.91% accuracy rate of identifying defaulting customers in the midst of non-defaulting clients. For

instance, if a loan applicant computed to be a very likely defaulter, this client holds

57.89% potential of eventually defaulting and if a client is computed as a nondefaulter, the individual possesses a 88.89% probability of actually not defaulting.

Type I and Type II errors were found to be 0.42 and 0.17 respectively.

5.3 Conclusion

In spite of the fact that the sample is just a fraction of percentage of the Bank's total clientele, the results are promising: of the 137 borrowers that branch of the bank had, 69 of them turned out to be men and 68 of them turned out to be women. Uncharacteristically, the default rate of the male folk is greater than that of the females. The married loan clients defaulted more than singles. Default rate fluctuates without any pattern with every increase in dependent(s) count. Persons with increasing disposable income decreases their chances of potential default significantly. In this thesis is a proposal for a less than complicated credit risk assessment. Clearly this Cox Regression model makes it more than possible to deduce credit Risk models (based on the partial likelihood proportional hazard functions) to quantify the credit worthiness of a borrower.

5.4 Recommendation to Stakeholders

As per the conclusion of this study, the following recommendations were drawn up fine tune the process of Credit Risk Management and effective mean to continue to improve the quality of constant yet reliable service to clients and shareholders of Zenith Bank Ghana Ltd, and the country's banking sector at large.

1. Currently the loan process is being mostly scrutinized by a more centralized system at some significant costs to the local that is Zenith Bank. It is hence prudent for the banks to hire qualified personnel to quantify credit risk. They can impute relevant data of clients in the entire bank, of changes in financial environment, of consumption trends and of consumer loans to be compared with the centralized formula. The personnel would occasionally have to compute and adjust the weights of each marginal risk factor and monitor the frequency of overdue settlements on loans in a time fashion.
2. The bank must on occasion consider organize training for clients with small businesses and employees of self-employers. This course will go a long way to direct clients on how to utilize their bank funded capital so that the loans granted will be used for the purpose of which it was obtained.
3. The bank should also consider combined Cox Regression Model, the Nelson-Alen / the Kaplan Meier used in this thesis and the observations from the analysis of cross tabulation to secure a failsafe method of operation for quantifying the various degrees' credit risks. This will in the grand scheme of things quantify the creditworthiness of any individual.
4. A state of the art database of all personal loan cases must be established by the bank. By enlisting the help of this database system to suit the future policy of personal consumer loans, the plans of credit extension are improved and the import could be a reference point for creating revolutionary new financial products. The bank could continue to expand on additional characteristic factors determining the credit risk of an individual applying for a loans.
5. Adding to in the model, the errors will be evened out and accuracy rates of the model will effectively be bettered.

6. With the creation of this standardized model of assessing credit risks, the bank can further specify the application forms to focus on the covariates influencing credit risk of applicants awaiting the grant of loans.
7. Application of loans facility can even be set up on-line, this process can shorten application-approval time of personal loans and the efficient serving of customers and free up more business the way of the bank.

5.5 Recommendation for Future Research

Users of credit information use scoring models that tend to group clients as “good” or “bad”. The problem here is that the scoring system does not take into account the society and political environment of area under survey and also the state of the economy. In following papers, it could be interesting to study how default behavior changes when there are hits in the economy and societal, political and geographical environment that the research is been conducted.

This study used Cox regression model as its credit scoring Model. Different methods such as linear discriminant analysis, k nearest neighbor, kernel density estimation, and decision tree score also viable topics for future studies. Apart from the independent covariates (sex, Age, Disposable Income, Marital Status, Number of Dependents, Amount of loan Collected) used in the thesis, other independent covariates such as number of years with purpose of the loan, frequency paid, current employer, years at current address, home ownership, home phone number given and term of the loan could also be used by future researcher to see if they prove to equally contribute to quantify the credit score of a borrower.

This work only studies applicable data of the case study bank and the rationale behind the non-selection of some fairly viable covariates. Further studies can proceed in this general direction to expand on the use of the time-independent

models to overcome the restrictions of the proportional hazards and thus birth a sounder predictive strength in a model of assessing credit risks.

References

Acted CT 4 the combined pack Chapter 8 and Pages 375 - 483

Adams, W., Einav, L., & Levin, J. (2007). Liquidity constraints and imperfect information in subprime lending (No. w13067). *National Bureau of Economic Research*.

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The journal of finance*, 23(4), 589-609.

Altman, E. I., & Hotchkiss, E. (2010). Corporate financial distress and bankruptcy: Predict and avoid bankruptcy, analyze and invest in distressed debt (Vol. 289). *John Wiley & Sons*.

Altman, E. I., Haldeman, R. G., & Narayanan, P. (1977). Zeta tm analysis a new model to identify bankruptcy risk of corporations. *Journal of banking & finance*, 1(1), 29-54.

Amidu, M., & Hinson, R. (2006). Credit risk, Capital Structure and lending decisions of banks in Ghana. *Banks and Bank systems*, 1(1), 93-101.

Annual Report and Annual Meeting Scotia bank annual report (2012).

Basel Committee on Banking Supervision, "Principles for the Management of Banking", (1999), Pg. 5-27, <http://www.bis.org/publ/bcbs54.pdf> [Accessed 17/10/2015].

BOG Annual Report of 2010 (BoG, 2013) september-financial%20stability%20report%20-%20september%202010.

Breslow, N. E. (1975). Analysis of survival data under the proportional hazards model. *International Statistical Review/Revue Internationale de Statistique*, 4557.

Brown A (2004). "The Origins and Evolution of Credit Risk Management", *Risk History*, <http://www.eraider.com/images/articles/RiskHistory1.pdf> [Accessed 15/11/2015].

Casu (2006). *Introduction to Banking, credit risk management*. Pg :282.

Covello, V. T., & Mumpower, J. (1985). Risk analysis and risk management: an historical perspective. *Risk analysis*, 5(2), 103-120.

Cox, D. R. (1972). iRegression models and life tablesi (with discussion). *Journal of the Royal Statistical Society, B*, v34.

Cox, D. R., & Snell, E. J. (1968). A general definition of residuals. *Journal of the Royal Statistical Society. Series B (Methodological)*, 248-275.

Duarte, J., Siegel, S., & Young, L. A. (2010). Trust and credit. In *AFA 2010 Atlanta Meetings Paper*.

Freedman, S., & Jin, G. Z. (2008). Do social networks solve information problems for peer-to-peer lending? evidence from prosper. com.

Gine, X., Goldberg, J., & Yang, D. (2011). Credit Market Consequences of Improved Personal Identification: Field Experimental Evidence from Malawi (No. w17449). *National Bureau of Economic Research*.

Hamilton, D. (2005). Moodys senior ratings algorithm and estimated senior ratings. Moody's Investors Service.

Hsieh, N. C. (2005). Hybrid mining approach in the design of credit scoring models. *Expert Systems with Applications*, 28(4), 655-665.

https://en.m.wikipedia.org/wiki/Bank_for_International_Settlements

Iyer, R., Khwaja, A. I., Luttmer, E. F., & Shue, K. (2015). Screening peers softly: Inferring the quality of small borrowers. *Management Science*.

Jonathan Paddy (2012). Credit Risk Management in Banking Industry. Pages 1-50.

Lin, M., Prabhala, N. R., & Viswanathan, S. (2013). Judging borrowers by the company they keep: friendship networks and information asymmetry in online peer-to-peer lending. *Management Science*, 59(1), 17-35.

Mhango, Y. (2010). Ghana: Annual economic outlook and the eve of oil production. Standard Bank Group Economics.

Miller, S. (2010/2012). Information and default in consumer credit markets: Evidence from a natural experiment. *Journal of Financial Intermediation*, 24(1), 45-70.

Miller, S. (2015). Risk Factors for Consumer Loan Default: A Censored Quantile Regression Analysis.

Pope, D. G., & Sydnor, J. R. (2011). What's in a Picture? Evidence of Discrimination from Prosper. com. *Journal of Human Resources*, 46(1), 53-92.

Portnoy, S. (2003). Censored regression quantiles. *Journal of the American Statistical Association*, 98(464), 1001-1012.

PriceWaterhouse (2010). Banking in Ghana Survey Report.

Raghavan, R. S.(2003). "Risk Management in Banking", pg. 841 - 851, <http://icai.org/resourcefile/11490p841-851.pdf> [Accessed 26/11/2015].

Ramcharan, R., & Crowe, C. (2013). The impact of house prices on consumer credit: evidence from an internet bank. *Journal of Money, Credit and Banking*, 45(6), 1085-1115.

Ravina, E. (2012). Love & loans: the effect of beauty and personal characteristics in credit markets. Available at SSRN 1107307.

Rigbi, O. (2013). The effects of usury laws: Evidence from the online loan market. *Review of Economics and Statistics*, 95(4), 1238-1248.

Rocky Ieraci (2008). Model Validation A RAROC Case Study, Standard & Poor's Risk Solutions a division of The McGraw-Hill Companies, Pg 7-13.

Ross, S. A., Westerfield, R., & Jordan, B. D. (2008). Fundamentals of corporate finance. Tata McGraw-Hill Education.

Schoenfeld, D. (1982). Partial residuals for the proportional hazards regression model. *Biometrika*, 69(1), 239-241.

Stepanova, M., & Thomas, L. (2002). Survival analysis methods for personal loan data. *Operations Research*, 50(2), 277-289.

Takang & Ntui (2008). Bank Performance and Credit Risk Management, Msc. Projecdt at University of Skovde: School of Technology and Society.

Therneau, T. M., Grambsch, P. M., & Fleming, T. R. (1990). Martingale-based residuals for survival models. *Biometrika*, 77(1), 147-160.

Tonywelcomes.blogspot.com/2013/04/non-performing-loans-in-ghanas-banking.html accessed in September 2015.

Van Gestel, T., & Baesens, B. (2008). Credit Risk Management: Basic Concepts: Financial Risk Components, Rating Analysis, Models, Economic and Regulatory Capital: Basic Concepts: Financial Risk Components, Rating Analysis, Models, Economic and Regulatory Capital. OUP Oxford.

Van Gestel, T., & Baesens, B. (2009). Credit risk management: basic concepts: financial risk components, rating analysis, models, economic and regulatory capital. *Oxford University Press*.

Zhang, A. (2009). Statistical Methods in Credit Risk Modeling (Doctoral dissertation, The University of Michigan).

Appendix A

Table 5.1: Cross table of payment Status

Defaulted/Paid In Full	Days After Final of Pmt	Mortality rate	Months of Pmts	Age	Disposable Income Bracket	Marital Status	Dependents	Sex	Loan Term	Loan Amt Brkts	Loan Term in Years	Probability of the Event of Default
Individual Betas			-0.287	-0.169	-0.877	-0.047	0.43	0.162	0.422	-0.056		$\exp\{-\beta_0 - \sum \beta_i x_{ij}\}$
1	1	0.000674	0	1	1	1	2	0	0	2	0.5	0.064512643
0	1	0.000568	4	0	0	0	0	1	4	1	1.5	0.914874643
1	2	0.001265	7	2	0	1	4	0	7	0	3	0.999999259
0	3	0.001154	0	2	1	1	2	0	0	0	0.5	0.000507947
0	4	0.000674	0	1	1	1	1	1	0	1	0.5	1.71557E-90
0	7	0.000577	0	0	0	0	0	1	0	0	0.5	0.187564917
1	7	0.015176	2	4	2	1	8	1	2	1	1	0.988790816
0	7	0.000788	0	1	1	1	3	0	0	0	0.5	0.884843184
0	8	0.00057	0	0	0	0	0	1	0	1	0.5	0.18694461
1	8	0.001541	4	2	0	1	5	0	4	0	1.5	0.999999506
0	8	0.000577	0	0	2	0	2	1	0	2	0.5	0
1	8	0.000706	7	1	1	1	2	0	7	2	3	0.983027015
0	10	0.00058	6	0	0	0	0	1	6	1	2	0.984141405
1	10	0.002122	0	2	1	1	4	0	0	1	0.5	0.970612781
0	11	0.001902	2	2	2	0	5	0	2	2	1	0.879757115
0	12	0.000626	0	1	1	0	1	1	0	1	0.5	3.50083E-68
0	12	0.000788	2	1	1	1	2	1	2	1	1	0.086894503
1	13	0.001265	0	2	1	1	3	0	0	0	0.5	0.660478802
0	13	0.00084	0	1	0	0	0	0	0	0	0.5	0.191114771
0	15	0.000902	2	2	1	1	2	0	2	1	1	0.080220693
1	15	0.002122	0	2	1	1	4	1	0	1	0.5	0.922330895
1	16	0.000626	0	1	0	0	0	0	0	0	0.5	0.175661284
0	18	0.001541	0	2	1	1	3	1	0	0	0.5	0.304006636
0	18	0.000647	0	1	1	1	2	1	0	1	0.5	0.000112555
0	18	0.001057	2	2	1	1	2	0	2	0	1	0.084524667
0	19	0.001057	0	2	2	0	3	1	0	1	0.5	2.9769E-161
0	19	0.00084	0	1	1	1	2	0	0	0	0.5	0.073989519
0	20	0.013544	0	4	3	0	7	1	1	3	0.75	0.415031717
0	20	0.003756	1	3	2	1	3	1	2	1	1	8.23876E-15
0	21	0.00974	0	2	2	0	1	0	0	2	0.5	0
0	21	0.004221	0	3	3	1	5	1	0	2	0.5	8.8314E-152
0	22	0.019012	0	4	4	1	9	0	0	3	0.5	0.123630786
0	24	0.001541	0	2	3	1	4	0	2	2	1	7.80374E-09
1	24	0.002122	1	2	5	1	12	0	3	3	1.25	0.999726369
0	25	0.021255	0	4	2	0	3	1	0	1	0.5	2.04704E-33

Appendix B

XLSTAT 2015.5.01.23164 - Cox proportional hazards model - on 11/6/2015 at

4:06:04 PM Time data: Workbook = DemoCox Mock up.xlsx / Sheet = Sheet1 /

Range = Sheet1!B3:B139 / 137 rows and 1 column Status indicator: Workbook =

DemoCox Mock up.xlsx / Sheet = Sheet1 / Range = Sheet1!A3:A139 / 137 rows

and 1 column Quantitative: Workbook = DemoCox Mock up.xlsx / Sheet

= Sheet1 / Range = Sheet1!D3:K139 / 137 rows and 8 columns

Event code: 1

Censored code: 0

Significance level (%): 5

Ties handling: Breslow's method

Stop conditions: Iterations: 100 Convergence: 0.000001

Table 5.2: Summary statistics (Events)

Total observed	Total failed	Total censored	Time steps
----------------	--------------	----------------	------------

137	38	99	101
-----	----	----	-----

Table 5.3: Descriptive statistics (Explanatory variables)

Variable
Observations
Obs.without missing data
Obs.without missing data
Minimum
Maximum
Mean
Std. deviation

X11370137071.0951.866
X21370137051.7961.436
X31370137051.0221.018
X41370137010.6570.476
X513701370121.7231.992
X61370137010.5040.502
X71370137071.2631.828
X8137013730.9340.859

Table 5.4: Goodness of fit statistics

Statistic	Independent	Full
Observations	38	38
DF	0	8
-2 Log(Likelihood)	298.201	265.905
AIC	298.201	281.905
SBC	298.201	295.005
Iterations	1	5

Table 5.5: Test of the null hypothesis H_0 : $\beta_0 = 0$

Statistic	DF	Chi-square	Pr > χ^2
-2 Log(Likelihood)	8	32.29671	<0.0001
Score	8	36.64014	<0.0001
Wald	8	29.65959	<0.0001

Table 5.6: Regression coefficients

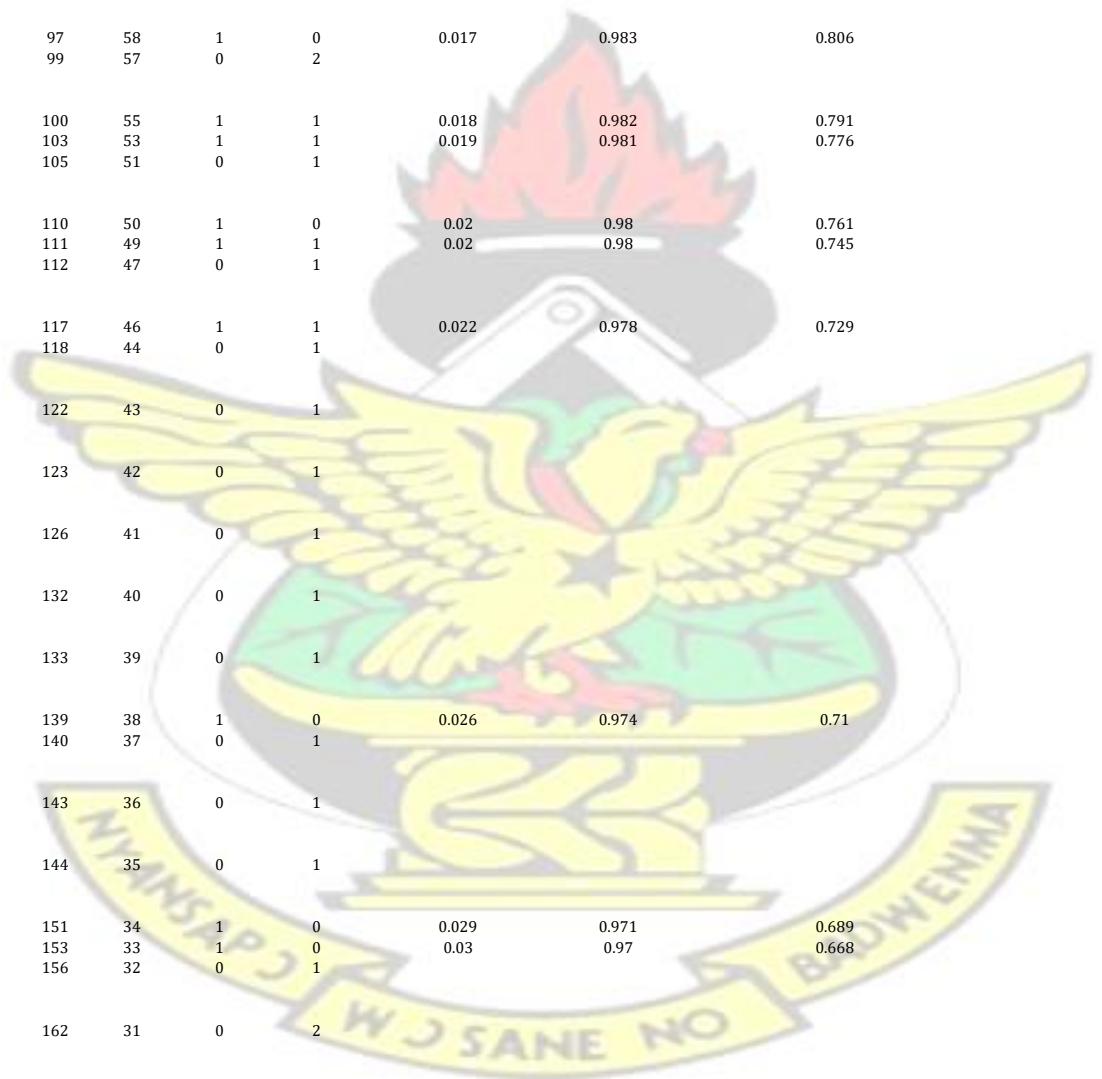
Variable	Value	Standard error	Wald Chi-Square	$Pr > \chi^2$	Hazard ratio	Lower bound(95%)	Upper bound(95%)
X1	-0.287	0.24	1.428	0.232	0.75	0.468	1.202
X2	-0.169	0.147	1.323	0.25	0.845	0.633	1.126
X3	-0.877	0.346	6.423	0.011	0.416	0.211	0.82
X4	-0.047	0.482	0.009	0.922	0.954	0.371	2.456
X5	0.43	0.105	16.769	<0.0001	1.537	1.251	1.888
X6	-0.162	0.367	0.194	0.66	0.851	0.415	1.746
X7	0.422	0.255	2.738	0.098	1.524	0.925	2.512
X8	-0.056	0.309	0.033	0.857	0.946	0.516	1.732

Table 5.7: Kaplan-Meier table

1	At risk	Failed	Censored	Proportion failed	Survival rate	Survival distribution function
1	136	0	1			
2	135	1	0	0.007	0.993	0.993
3	134	0	1			
4	133	0	1			
7	132	1	2	0.008	0.992	0.985
8	129	2	2	0.016	0.984	0.97
10	125	1	1	0.008	0.992	0.962
11	123	0	1			
12	122	0	2			
13	120	1	1	0.008	0.992	0.954
15	118	1	1	0.008	0.992	0.946
16	116	1	0	0.009	0.991	0.938
18	115	0	3			
19	112	0	2			
20	110	0	2			
21	108	0	2			
22	106	0	1			
24	105	1	1	0.01	0.99	0.929
25	103	0	4			
27	99	0	1			
29	98	1	0	0.01	0.99	0.919
30	97	0	2			
31	95	0	2			
33	93	0	1			
35	92	1	0	0.011	0.989	0.909
36	91	0	1			
42	90	0	1			
43	89	1	0	0.011	0.989	0.899
44	88	0	1			
45	87	1	0	0.011	0.989	0.889
48	86	1	0	0.012	0.988	0.878
49	85	0	1			
51	84	1	2	0.012	0.988	0.868
52	81	0	3			
53	78	1	0	0.013	0.987	0.857
54	77	0	2			
56	75	0	1			
59	74	0	1			
61	73	1	0	0.014	0.986	0.845
63	72	0	1			

72	71	0	1			
73	70	0	1			
80	69	1	1	0.014	0.986	0.833
82	67	0	1			
83	66	0	1			
84	65	1	0	0.015	0.985	0.82
87	64	0	2			
90	62	0	1			
92	61	0	1			
95	60	0	2			
97	58	1	0	0.017	0.983	0.806
99	57	0	2			
100	55	1	1	0.018	0.982	0.791
103	53	1	1	0.019	0.981	0.776
105	51	0	1			
110	50	1	0	0.02	0.98	0.761
111	49	1	1	0.02	0.98	0.745
112	47	0	1			
117	46	1	1	0.022	0.978	0.729
118	44	0	1			
122	43	0	1			
123	42	0	1			
126	41	0	1			
132	40	0	1			
133	39	0	1			
139	38	1	0	0.026	0.974	0.71
140	37	0	1			
143	36	0	1			
144	35	0	1			
151	34	1	0	0.029	0.971	0.689
153	33	1	0	0.03	0.97	0.668
156	32	0	1			
162	31	0	2			
164	29	1	0	0.034	0.966	0.645
177	28	0	1			
182	27	0	1			
186	26	0	1			
200	25	0	1			
201	24	1	0	0.042	0.958	0.618

KNUST



216	23	0	1			
228	22	0	1			
231	21	2	0	0.095	0.905	0.559
242	19	0	1			
250	18	1	0	0.056	0.944	0.528
260	17	0	1			
278	16	1	0	0.063	0.938	0.495
283	15	0	1			
287	14	1	0	0.071	0.929	0.46
314	13	0	1			
340	12	0	1			
357	11	0	1			
378	10	0	1			
384	9	0	1			
389	8	0	1			
392	7	0	1			
411	6	0	1			
467	5	1	0	0.2	0.8	0.368
553	4	0	1			
587	3	0	1			
991	2	0	1			
999	1	1	0			
				1	0	0

Table 5.8: Mean survival time

Mean survival time	Standard deviation	Lower bound (95%)	Upper bound (95%)
485.868	71.024	346.663	625.073

Table 5.9: Quantiles estimation

Quantile	Estimate	Lower bound (95%)	Upper bound (95%)
75%	999	467	999
50%	278	201	999
25%	111	84	201
		64	

KNUST



