Analysis of Credit Risk on Microfinance Loans Using Survival Analysis Techniques: A Case Study

Kennedy Wekulo Manyuru
Department of Mathematics
Catholic University of Eastern Africa
Nairobi, Kenya
kennedymanyuru@yahoo.com

Dr. Karanjah Anthony
School of Science
Maasai Mara University
Narok, Kenya
karanjah@mmarau.ac.ke

Dr. Francis Njui
School of Mathematics
University of Nairobi
Nairobi, Kenya
francis.njui@uonbi.ac.ke

Abstract: A study was conducted on 577 people who took loans from microfinance institutions in Bungoma, Kakamega and Vihiga County. The study was conducted in 72 months starting January 2012 and ending December 2017.
Survival analysis technique was used to model and analyse the probability of default on loan borrowers. In this study, we were able to find out that age had no significant effect on default rate. Male clients were at a higher risk of defaulting than female. For occupation, farmers had the highest probability of default and nurses had the least probability of default. Loan amount had a small significant effect on default rate. However, it was observed that loans of between 100000 and 300000 had the least default rate and therefore it was safer for microfinance institutions to give out small loans. For counties, Bungoma had the highest default rate and Kakamega had the least.

***Keywords**- Survival analysis; Probability of default; Kaplan Meier survival estimate; Cox proportion hazard model.*

*****

I. Introduction

During recent years, financial institutions have devoted important resources to build statistical models to measure the potential losses and their loans portfolios, Robinson(200) and Otero(1999). A stable financial system is an essential part of a growing and prosperous society.
A financial crisis can severely damage the economic output of a society for an extended period of time. For a financial institute to grow, it should be making a profit out of the collection of interest from loans meted to clients .A bank can easily make a loss if a sizeable number of its clients default, https://file.scirp.org/pdf/OJBM_2018012315471945.pdf.
In Kenya the small scale financial institutions are so many and closer to people as compared to banks. In most cases banks have a well organized system of dealing with defaults, they have also huge capital and they can survive longer if some of its clients default.
A microfinance institution will collapse after a short period of time if a sizeable number of its clients will default, given that it has a small capital.
Microfinance institutions, just like banks, need to manage the default rate or credit risk inherent in the entire portfolio as well as the risk in individual credit or transactions. The effective management of credit risk is a critical component of a comprehensive approach to risk management and essential to the long term success of any banking organization. In order to access the credit risk, certain risk parameters must be estimated. They include probability of default (PD) and Loss Given Default (LGD).
Default is defined based on Basel II, which states that; a loan is considered to have been defaulted if either one or both of the following events have occurred.

I. The bank considers that the client is unlikely to meet its credit obligations to the banking group in full, without recourse by the bank to actions such as realizing the security (if held) and.
II. The client is past due more than ninety (90) days on any material credit obligation to the banking group.

This research therefore, focused on developing Cox proportional hazard model that could be used by microfinance institutions to predict the probability of default on a loan given to a client, Sarah Miller (2010).

II. Survival Analysis

Survival analysis is a class of statistical methods used for analysing the expected duration of time until one or several events of interest occur. In this study, the event was the time in months a loan borrower takes to default. The response outcome variable was the time since a client took a loan until the time he/she defaults.
A study was conducted on 577 people who took loans from microfinance institutions in Bungoma, Kakamega and Vihiga County. The study was conducted in 72 months starting January 2012 and ending December 2017.
Survival time on every loan borrower was measured in months, from the time a bank client was given a loan up to the time he or she defaulted. The survival times of those clients who repaid the loan in full before the end of observation period were censored. Those clients who repaid the loan in full on the required time were also censored.
Survival analysis techniques used to analyse the data included; Kaplan Meier survival function and log rank test. Cox proportional hazard regression model was used to fit the data.

ISSN: 2454-4248
14 – 18

Kaplan Meier survival function was used to estimate the probability of loan borrowers in the population whose life times exceeded the specified time.
The log rank test was used to test the significance of the differences in the survival rates of loan borrowers by age, gender, occupation, county of birth and loan amount at 5% significance level.
Cox proportion hazard model was fitted to the data set obtained from the loan borrowers from the three counties to assess the relationship between the various covariates to survival.

III. Data

Secondary data on age, gender, occupation, loan amount and county of birth was obtained from various microfinance institutions in Bungoma County, Kakamega County and Vihiga County. Primary data on default status was obtained by interviewing the credit controller personnel of the various microfinance institutions. The study was carried out on 577 loan borrowers from the three counties beginning January 2012 and ending December 2017 a period of 72 months. The data set contained the client description variables, personal information about the client (age, sex, occupation and county of birth) and a description of loan amount. Time to default was the dependent variable. The data used in this study was randomly generated using sampling techniques for variables of interest. It was organized in microsoft excel spreadsheet. Age was categorized into four groups of 20-29years, 30-39years, 40-49years and 50-59years. Categories of gender were male and female. Occupation was categorized into nurses, teachers, farmers and business person. County of birth were categorized into Kakamega, Bungoma and Vihiga. Loan amount was categorized into four groups of 100000-300000, 300001-500000, 500001-700000 and 700001-1000000.

IV. Results

SPSS software package was used to analyse the data set of 577 loan borrowers. Kaplan Meier survival function was used to estimate the probability of default, log rank was used to test the significance of the differences in the survival rates of loan borrowers by age, gender, occupation, county of birth and loan amount at 5% significance level and Cox proportion hazard model was fitted to the data set obtained from the loan borrowers from the three counties to assess the relationship between the various covariates to survive.

Kaplan Meier Survival Estimates for Age Groups

It was observed that for age group 20-29, 154 were given loans, 55 defaulted and 99 repaid the loan in full within time. The survival rate for the age group was 64.3%. For age group 30-39, 260 were given loans, 83 defaulted and 177 repaid the loan in full within required time. The survival rate for the age group was 68.1%. For age group 40-49, 141 were given loans, 47 defaulted and 94 repaid the loan in full within time. The survival rate for the age group was 66.7%. For age group 50-59, 22 were given loans, 7 defaulted and 15 repaid the loan in full within time. The survival rate for the age group was 68.2%. The log rank test was used to compare the survival rates of the different age groups, and it observed that there was no significant difference in their survival rates, with a p-value of 0.874 which was higher than the 0.05 level of significance.

The Kaplan Meier survival curves for the age groups is as shown in the figure below

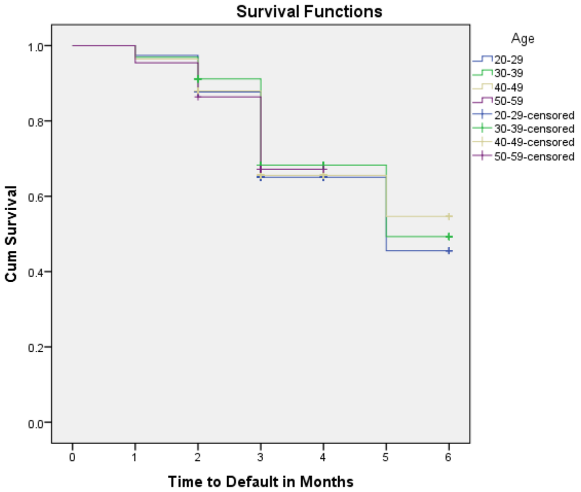

Figure 1

From the survival curves for age groups, it was observed that the survival rate was higher during the first 24 months after taking a loan and thereafter it started diminishing. Age group 50-59 survived up to 48 months and the rest of the age groups survived up to 72 months. Between 12 months and 60 months, the survival curves are very close suggesting that there was no significance difference in their survival rates during this period. During the last 12 months, the curves are observed to be a bit spread suggesting that there was a significant difference in their survival rates. During this period, age group 40-49 had the highest survival rate followed by age group 30-39 and age group 20-29 had the least survival rate.

Kaplan Meier Survival Estimates for Gender

For gender, out of 249 female clients who took the loan, 65 defaulted and 184 repaid the full loan within time and for male clients, out of 328, 127 defaulted and 201 repaid the loan in full within time. Female clients had a survival rate of 73.9% and male had a survival rate of 61.3%. From the log rank test to compare the survival rates of the male and female clients, it was observed that there was a significant difference in their survival rates, with a p-value of 0.001 which was less than the 0.05 level of significance. The female clients had a significantly higher survival rate than the male clients.

The Kaplan Meier survival curves for gender is as shown in the figure below

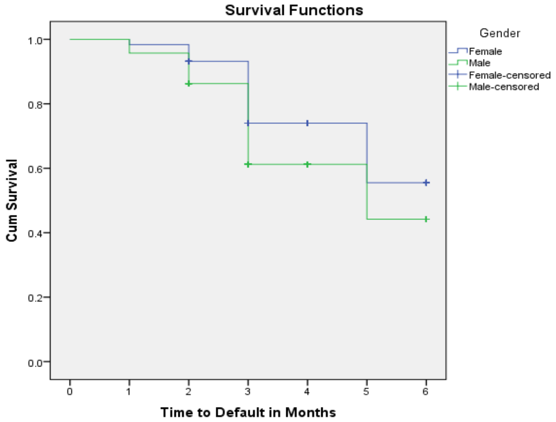

Figure 2

From the survival curves for gender, it was observed that in the first 24 months, curves were very close together suggesting that there was no significant difference in their survival rate. Between 24months and 72 months, the curves spread out suggesting that there was a significant difference in their survival rate. Throughout the entire period, female had a higher survival rate than male as indicated by the plot.

Kaplan Meier Survival Estimates for Occupation

For occupation, out of 173 teachers who took the loan, 38 defaulted and 135 repaid the loan in full. For nurses, those who took the loan were 169 out of which 32 defaulted and 137 repaid the loan in full within time. 135 farmers took the loan out of which 73 defaulted and 62 repaid the loan in full within time. For businessperson, 100 were loaned, 49 defaulted and 51 repaid the loan in full.

It was observed that teachers had a survival rate of 78%, nurses had 81.1%, farmers had 45.9% and businesspeople had a survival rate of 51%. Nurses had the highest survival rate and farmers had the least. From the log rank test to compare the survival rates of the different occupation, it was observed that there was a high significant difference in their survival rates, with a p-value of 0.000 which was less than the 0.05 level of significance.

The Kaplan Meier survival curves for the occupation is as shown in the figure below

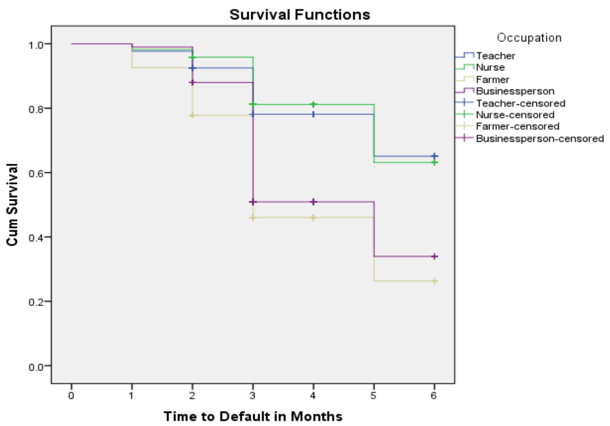

Figure 3

From the plot of occupation survival curves, nurses and teachers had a higher survival rate than farmers and businessperson as indicated by the survival curves. During the first 24 months the curves are fairly close suggesting that during this period there was no significant difference in their survival rate. Between 24 months and 72 months the curves were observed to spread out suggesting that there was a significant difference in their survival rate during this period.

Kaplan Meier Survival Estimates for Loan Amount

For loan amount 100000-300000, out of the 59 clients, 17 defaulted and 42 repaid the loan in full and it was observed that their survival rate was 71.1%. For loan amount 300001-500000, out of the total 156 clients, 55 defaulted and 101 repaid the loan in full. The survival rate of this group was 64.7%. For loan amount 500001-700000, 180 clients took loans and out of which 61 defaulted and 119 repaid the loan in full. The survival rate of this group was 66.1%. For loan amount 700001-1000000, 182 were given loans, 59 defaulted and 123 repaid the loan in full and the survival rate was 67.6%. From the log rank test to compare the survival rates of the different loan amounts, it was observed that there was a small significant difference in their survival rates, with a p-value of 0.043 which is slightly less than the 0.05 level of significance.

The Kaplan Meier survival curves for the loan amount is as shown in the figure below

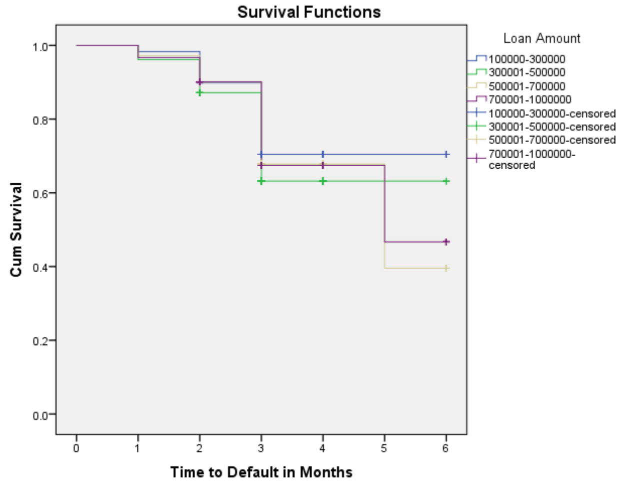

Figure 4

The plot of loan amount survival curves indicated that for the first 60 months, there was a small significant difference in their survival rate since the curves are close. Thereafter, the curves were observed to spread out during the final 12 months. This suggests that there was a big significant difference in their survival rate during the last 12 months of loan repayment.

Kaplan Meier Survival Estimates for County of Birth

For the county of birth, it was observed that out of 180 clients from Vihiga County, 59 defaulted and 121 repaid the loan in full in time. For Bungoma County, out of 201 clients, 80 defaulted and 121 repaid the loan in full in required time. For Kakamega County, out of 196 clients, 53 defaulted and 143 repaid the loan in full. Clients in Kakamega County were observed to have the highest rate of survival of 73%, Vihiga

County clients had a survival rate of 67.2% and Bungoma County clients had a survival rate of 60.2%. The log rank test for comparing the survival rates of the different counties of birth suggested that, there was a significant difference in their survival rates, with a p-value of 0.029 which was less than the 0.05 level of significance.

The Kaplan Meier survival curves for the county of birth is as shown below

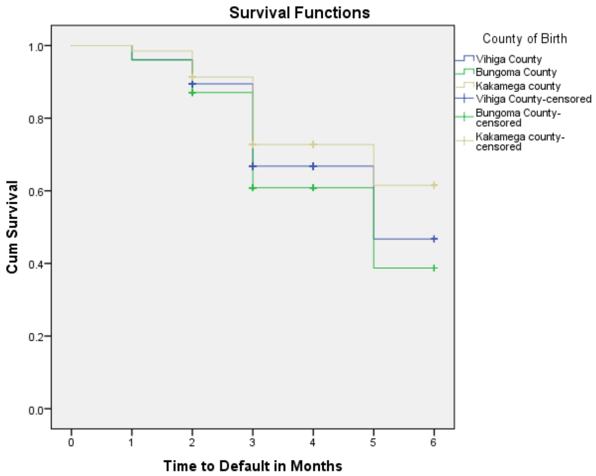

Figure 5

From the survival curves of the county of birth, it was observed that during the first 36 months after taking loans, there was no significant difference in their survival rate as suggested by the closeness of the curves. Between 36 months and 72 months, the curves were observed to spread suggesting that there was significant difference in their survival rate.

Analysis of variables in the Cox Proportion Hazard model

SPSS software was used to analyse the data and the results are as shown in table 2 below (last page).

Interpretation of Cox proportional hazard ratio

Age

Since Exp(β) =0.962 which is less than 1, it means that an individual was less likely to default with a unit increase in age.

Gender

With Exp(β) =1.684 which is greater than 1, it means that there is a difference in the risk between gender whereby men were 39.7% likely to default and women were 26.1% likely to default. The risk of defaulting was higher in men than in women.

Occupation

Exp(β) =1.452 which is greater than 1. This means that the risk in occupation increases. Farmers were at the highest risk of defaulting at 54.1%, followed by businesspeople who were 49% likely to default, teachers were 22% likely to default and nurses were 18.9% likely to default.

Loan amount

Exp(β) =1.014 which is slightly greater than 1. This means that the risk in loan amount was small. Individuals with loan amount 100000-300000 were 28.8% likely to default, 300001-500000 were 35.3% likely to default, 500001-700000 were 33.9% likely to default and 700001-1000000 were 32.4% likely to default.

County of birth

For the county of birth, the hazard ratio, Exp(β)=1.014 is slightly greater than 1, which means that the risk in the county of birth increases. Individuals in Bungoma County were 39.8% likely to default, Vihiga County were 32.8% likely to default and Kakamega were 27% likely to default.

A. Analysis of regression residue

Regression residuals were used to test the proportional hazards assumption, the results of which are given in table 1 below.

Table 1: R square of Cox Proportion Hazard Model

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.155[a]	.024	.016	.940

B. Interpretation of regression residue and fitting the Cox Proportion Hazard Model model

From the analysis of the regression residue, it was observed that R squared was equal to 0.024 and adjusted R squared was equal to 0.016. This meant that there was 2.4% variation in the independent variables. Since R^2=0.024 which is between 0 and 1, it implies that this model is fairly of best fit. The model can be adjusted by 1.6%. to be of best fit.

Regression Standardized Residue Plot

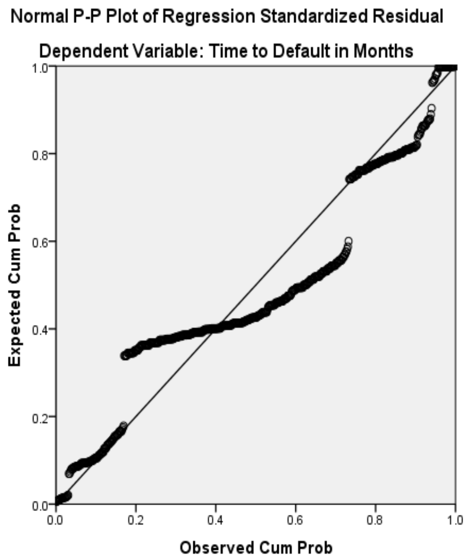

Figure 6

The generated plot of regression standardized residue indicated that few points lie on the line. This suggests that this model is fairly of best fit, Nelson W (2000).

C. Assumption of the Cox PH model

It was observed in the fitted Cox PH model of the credit data set that none of the predictor variables under this study had any significant effect on the survival time. However, the model used was significant and the proportional hazard assumption was not violated by any of the covariates included in this study.

V. Conclusion

From the data analysis, age of the client did not have a significant effect on the default rate. However, clients in age group 20-29 were the greatest defaulters. Clients in age group 50-59 were the least defaulters because of may be being financially stable.
For gender, there was a significant effect on the default rate. Male clients had a higher probability of default than female.
Occupation had a significant effect on default rate. Farmers were the greatest defaulters may be because poor farming methods or unpredictable weather that let to total losses. Businesspeople were also great defaulters because of poor investment. Teachers and nurses were the least defaulters because of their permanent jobs and hence were financially stable.
Loan amount had small significant effect on default rate. However clients who took loans of between 100000 and 300000 were observed to have the least probability to default. Clients who took loans of between 300001 and 500000 were observed to have the highest probability to default.
For the county of birth, clients from Bungoma County were observed to have the highest probability of default followed by clients from Vihiga County and clients from Kakamega County had the least probability of default. This difference in default rate could be due to the difference in Gross Domestic Product (GDP) with Kakamega possibly having a higher GDP than Vihiga and Bungoma County.
From the above findings, we concluded that Cox regression proportional hazard model can make it possible to deduce credit risk models based on the hazard ratio functions to quantify credit worthiness of a loan borrower from microfinance institution.

Table2: Hazard ratio, exp(β), by which each group is at risk of defaulting

	β	SE	Wald	df	Sig.	Exp(β)	95.0% CI for Exp(β)	
							Lower	Upper
Age	-.039	.090	.184	1	.668	.962	.806	1.149
Gender	.521	.154	11.527	1	.001	1.684	1.247	2.276
Occupation	.373	.066	32.405	1	.000	1.452	1.277	1.651
Loan Amount	.014	.074	.034	1	.854	1.014	.877	1.172
County of Birth	.014	.093	.024	1	.878	1.014	.846	1.216

Exp (β) is the ratio by which each group is at risk of defaulting loan repayment (hazard ratio)

REFERENCES

[1] Basel 11.b.c.o.b.s.2006 – International convergence of capital measurement and capital standards: A revised framework. Switzerland:s.n
[2] Nelson, W. (2000). Theory and application of Hazard plotting for censored failure Data. Technometrics, 42(1), 12 – 25.
[3] https://file.scirp.org/pdf/OJBM_2018012315471945.pdf
[4] Robinson (2001) and Otero (1999): Research on perspectives underpinning approaches to the practice of the microfinance in the development context.
[5] Sarah miller (2010); Effect of a loan borrower characteristics on probability of default.