COMPETING RISK MODELING OF LOAN DEFAULT IN COMMERCIAL BANK OF ETHIOPIA, HEAD OFFICE, ADDIS ABABA

M.SC. THESIS

BY

REDIAT DEJENE

SUBMITTED TO

COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCE

DEPARTMENT OF STATISTICS

JUNE, 2024

HAWASSA, ETHIOPIA

COMPETING RISK MODELING OF LOAN DEFAULT IN COMMERCIAL BANK OF ETHIOPIA, HEAD OFFICE, ADDIS ABABA

BY: REDIAT DEJENE

ADVISOR: CHERU A. (PhD)

A THESIS SUBMITTED TO DEPARTMENT OF STATISTICS, HAWASSA UNIVERSITY, COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCE, IN PARTIAL FULFILLMENTS OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCES IN STATISTICS (SPECIALIZATION: BIOSTATISTICS)

JUNE, 2024

HAWASSA, ETHIOPIA

DECLARATION

I Rediat Dejene declare that this MSc thesis entitled “Competing Risk Modeling of Loan Default in Commercial Bank of Ethiopia, Head Office, Addis Ababa” is my original work and has never been submitted to any other universities, and all the sources that I have used or quoted have been cited and acknowledged by means of references.

_____	_____	_____
Name of Researcher	Signature	Date

ADVISOR'S APPROVAL SHEET
HAWASSA UNIVERSITY
SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis entitled "**Competing Risk Modeling Of Loan Default In Commercial Bank Of Ethiopia, Head Office, Addis Ababa**" submitted in partial fulfillment of the requirements for the degree of master of science in statistics with specialization in Biostatistics of the graduate program of the College Of Natural and Computational Science, Hawassa University, and is a record of original research carried out by **Rediat Dejene**, under my supervision, and no part of the thesis has been submitted for any other degree or diploma. The guidance and assistance received during the course this investigation have all been duel acknowledged. Therefore, I recommend that it would be accepted fulfilling the thesis requirements.

_____	_____	_____
Name of Advisor	Signature	Date

EXAMINORS APPROVAL SHEET

HAWASSA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

We the undersigned, member of the board of examiners of the final open defense by Rediat Dejene have read and evaluated her thesis entitled “Competing Risk Modeling of Loan Default in Commercial Bank of Ethiopia, Head Office Addis Ababa” and examined the candidate. This is, therefore, to certify that thesis has been accepted in partial fulfillment of the requirements for the degree of Master of Science in Statistics with a specialization in Biostatistics.

_____	_____	_____
Name of Department Head	Signature	Date
_____	_____	_____
Name of Advisor	Signature	Date
_____	_____	_____
Name of External Examiner	Signature	Date
_____	_____	_____
Name of Internal Examiner I	Signature	Date
_____	_____	_____
Name of Internal Examiner II	Signature	Date
_____	_____	_____
SGS approval	Signature	Date

Final approval and acceptance of this thesis is contingent upon the submission of the final copy of the thesis to the school of graduate studies (SGS) through the department/school graduate committee (DG/SGC) of the candidate’s department.

Stamp of SGS date_____

ACKNOWLEDGEMENT

Before anything else, gratitude and acknowledge to the divine God, the galaxy's creator for his divine guidance, blessings, and grace throughout this research journey.

I would like to express my deepest gratitude and appreciation to my advisor Cheru Atsmegiorgis (PhD) for his invaluable guidance, mentorship, and unwavering support throughout this research endeavor. His expertise, encouragement, and constructive feedback have been instrumental in shaping this study and my academic growth. I would like to take this opportunity to acknowledge and thank my instructors as well as the dedicated staff in the statistics department.

I would also like to extend my heartfelt thanks to the Commercial Bank of Ethiopia (CBE), Head office specifically Manager Data Warehouse and Business Intelligence department, for their generous support in providing the necessary data. I also extend my gratitude to Mr. Yonas Gebreegziabher, Manager in Employee Performance Management at CBE, and Kaleb Birhanu, Senior Data Engineer in Manager Data Warehouse and Business Intelligence at CBE, for their special support during data extraction.

Furthermore, I am profoundly grateful to my supportive father Dejene Kinato and my lovely mother Birtukan Shibiru for their unconditional love, unwavering belief in my abilities, and constant encouragement throughout this journey. My special thanks to my beloved sister Mihiret and brothers, Dagim and Yared for their constant support, understanding, and motivation have been my pillars of strength, enabling me to overcome challenges and pursue my academic aspirations.

At last not least, my heartfelt thanks to my beloved friends for their support and involvement have been pivotal, and I am truly grateful for their contributions. Finally I am indebted to all those who have contributed directly or indirectly to the successful completion of this study. Thank you all from the bottom of my heart.

List of Abbreviations and Acronyms

CBE Commercial Bank of Ethiopia

CBs Commercial Banks

CI Confidence Interval

CIF Cumulative Incidence Function

CSH Cause Specific Hazard

ER Early Repayment

NBE National Bank of Ethiopia

NPL Non Performing Loan

PH Proportional Hazard

SHR Sub Hazard Ratio

Contents

List of Tables

List of Figures

Abstract

***Background**: It is equally true that bank loans, as they are profitable, equally risky. The risk of borrowers defaulting on their obligations poses a significant challenge for banks. Traditional credit scoring systems aim to estimate the probability that an applicant will default. However, for the financial institution, it is important to consider not only if but also when the creditor defaults. The major aim of this study was to investigate factors that affect the probability of default in the presence of early repayment, the case of CBE, Head office.*

***Method**: To reach the aim, using a retrospective study design, 1077 customers who took loans from 01 January 2015 to 30 December 2023 were taken from the Data Warehouse and Business Intelligence department at CBE, head office and the data set comprised the bank-specific, customer-specific, and loan-specific variables. The fine-gray model was applied to identify factors affecting two mutually exclusive events, default and early repayment.*

***Result**: From the total of 1077 borrowers, 893 (82.9%) defaulted and 184 (17.1%) were repaid early the corresponding median time was 12.2 and 81.3 months respectively. The result of the Fine-Gray model shows that loan size, monthly and quarterly repayment mode, previous loan experience, purpose of loan for manufacturing, and international trade were significantly associated with default risk, whereas, interest rate, quarterly repayment mode, and previous loan experience were significantly associated with early repayment.*

***Conclusions and Recommendations**: The result of the Fine-Gray model revealed that loan size, quarterly and monthly modes of repayment, significantly increase default risk, and previous loan experience significantly decreases default risk, while, interest rate and previous loan experience significantly increase early repayment risk. It is recommended that the bank should exercise caution when approving loans with large loan sizes, and borrowers without previous loan experience, especially if they opt for monthly or quarterly repayment modes.*

Key words: Credit risk, Default, Early repayment, Survival analysis, Competing risk, Gray test, Fine-Gray model

1 Introduction

1.1 Background of the study

Loans are the most important asset held banks, lending provides the bulk of bank income. It is equally true that bank loans, as they are profitable, equally risky. Bank loans fluctuate and influenced by the changes in economy policy and the economy in general. Therefore, it is very important for the bank to formulate their loan policies in order of minimize risk associated with them (BIRTUKAN TADESSE, 2014).

Credit risk is a prominent and significant risk that has a substantial impact on the viability and overall performance of financial institutions (Bhattarai, 2019). The banking sector's primary goal is to gather public funds to provide credit, support business expansion, and contribute to a thriving economy; however, the risk of borrowers defaulting on their obligations poses a significant challenge for banks, as it has consequences for the overall stability of the financial system (Alrfai et al., 2022).

If the sector or borrower experiences financial difficulties, it could lead to a higher number of defaults, which could have a significant negative impact on the bank's financial health. Therefore, banks need to be careful about the concentration of their loan portfolio and set appropriate risk tolerance levels to mitigate the risk of excessive concentration. Loan defaults continue to be a major challenge that confronts financial institutions in developing countries and this impedes their potential role in sustainable development (Baidoo et al., 2020).

Ethiopia, a developing country in Africa, also experiences similar circumstances. The banking sector plays a crucial role in Ethiopia's economy, and it accounts for around 93 percent of the total capital in the financial sector and contributes 4.2 percent to Ethiopia's national economy (Abate & Kaur, 2023).

The accumulation of non-performing loans (NPLs) in Ethiopian commercial banks can be attributed to multiple factors. These factors include economic downturns, macroeconomic instability, and specific issues related to individual banks that contribute to credit risk (Kotiso, 2018). Ethiopian commercial banks have a significant reliance on loans, which account for a substantial portion of their assets, ranging from 50 to 75 percent, and their primary focus is on providing consumer loans as a means to generate revenue and meet the demand for credit.

However, this heavy dependence on loans exposes them to significant risks associated with non-performing loans (NPLs), which can pose harm to the overall financial system (Assfaw, 2021).

The key determinant of profitability in Ethiopian CBs found to be credit risk, which is measured by loan loss provision to total loan. In Ethiopia, bank management should focus its efforts on the implementation of credit risk management to evaluate credit risk more effectively and avoid problems associated with loan default risk (DIRIBA, 2014).

It is essential for credit scoring to accurately determine whether the lender should grant the borrower a loan based on information about the borrower as well as information about the current economic factors that may induce a borrower to default. A borrower may be fit to pay off the initial loan payments, but there may potentially be many events that occur over the duration of the loan which may inhibit a borrower from making payments. This is why modeling consumer credit risk is of vital interest to financial companies (Hassan et al., 2018).

Credit scoring is a technique mainly used in making consumer credit decisions. Traditional credit scoring systems aim to estimate the probability that an applicant will default. However, for the financial institution it is important to consider not only if but also when the creditor defaults (Wycinka, 2015).

(Thomas et al., 1992) initially proposed the utilization of survival analysis within the credit risk domain as a substitute for logistic regression. The use of survival analysis in this particular scenario provides a significant benefit by allowing the modeling of the time it takes for default to occur, rather than merely predicting whether an applicant will default or not (Thomas, 2002).

Previously there was a study done by (Atsmegiorgis et al., 2014) on survival model of loan repayment rate on cross-sectional data in case of commercial bank of Ethiopia, Hawassa district and revealed different significant factors that associated with survival of repayment rate such as educational level, having previous loan experience, mode of repayment, collateral type and purpose of loan.

Therefore, analyzing different bank specific, loan specific, and borrower specific factors that affect the probability of default in the presence of early repayment in case of Commercial Bank of Ethiopia (CBE), Head office has become rationale of this study.

1.2 Statements of the problem

Credit risk refers to potential loss associated with loans issued to consumers due to borrower's failure to meet contractual obligations. Among other risks (e.g. liquidity buffer, operational and market risks), credit risk is the most significant source of regulatory capital demand. In order to hold adequate capital against credit risk, financial institutions should adopt the use of advanced statistical capital models to calculate capital demand (Marimo, 2015).

Loan default which is attributed to non-repayment of loans approved and disbursed to borrowers, (Enimu et al., 2017); (Yibrie & Ramakrishna, 2017) raises concerns not only because of the effect it has on financial institutions but also on individuals, entrepreneurs and the economy as a whole. There are several negative consequences that arise due to incidence of loan default. According to (Amuakwa-Mensah et al., 2017) reported that loan default which emanates from non-performing loans (NPLs) has the tendency of retarding economic growth, private investment and increasing unemployment, because there will be inadequate funds for entrepreneurs with viable investment ambitions to execute the envisioned projects.

The phenomenon of default on borrowed funds is still one of the critical problems faced by financial institutions in developing countries in their attempt to expand their services and reach a large number of clients (Endris, 2021). NPLs which eventually lead to loan default (due to consistent non-repayment of loan) are endemic in most developing countries and therefore call for urgent attention given its negative consequence (Baidoo et al., 2020).

Consumer credit data is analogous to lifetime data as it concerns the credit status of a cohort of customers with different loan repayment behaviors over a given observation period. In the consumer credit context, survival methods results can be used as input into the computation of credit risk parameters. These play a crucial role in risk management, and include the Probability of Default (PD) and Loss Given Default (LGD) models (Marimo, 2015).

The ability to deal with censored observations is the main advantage of survival analysis. In some credit risk evaluation sources of censored observations include the end of and early repayments during, the follow-up period, and such kind of censoring during follow-up can cause biased estimates of the models' parameters in classical survival analysis. However, competing risk analysis, which is an extension of classical survival analysis for more than one event, allows these problems to be overcome (Wycinka, 2019).

The study by (Marimo, 2015) applied survival method to predict the two mutually exclusive events, default and Early Settlement (ES) and concluded that the occurrence of these two events over the observation period impacts negatively on profitability for financial institutions. The size of an institution's credit portfolio can decrease as a result of early repayments, which changes the probability of default over time. Prognosis of the probability of default should therefore also take into consideration the prognosis of early repayments (Wycinka, 2019).

To the best of my knowledge, in Ethiopia no study have done on competing risk modeling of loan default in the presence of early repayment. Thus, this study goal is to close that gap. Thereby, the study has tried to respond to the following fundamental research questions.

- What are the major risk factors that affect the probability of loan default in the presence of early repayment?
- What is the estimated probability of default over time, considering the presence of early repayment, and how does it differ among different borrower groups?
- Is there a statistically significant difference in the CIF curves between different groups of covariates?

1.3 Objectives of the study

1.3.1 General objective

- The main objective of the study is to apply a competing risk model in credit data to examine the predictor variables that influence the probability of loan default in the presence of early repayment in case of CBE, Head Office.

1.3.2 Specific Objectives

- To investigate the major risk factors that affects the probability of loan default in the presence of early repayment.
- To estimate the probability of default in the presence of early repayment.
- To test whether there is a significant difference between cumulative incidences of groups of covariates.

1.4 Significance of the study

As output of the analysis, identifying factors that affect probability of loan default in the presence of early repayment will help policy makers to formulate successful credit policies. The study will have positive impact in creating smooth relationship between the borrower and the lender. Other researchers will make use of the research out come because it will help them to identify the factors behind probability of loan default and also will help them to make research on similar issues.

1.5 Scope of the study

The study is delimited to the identification of factors that determine the probability of loan default for borrowers in commercial bank of Ethiopia at Addis Ababa, Head office. It does not include borrowers of commercial bank of Ethiopia in other regions and branches.

1.6 Operational Definition

- **Credit risk:** is the potential financial loss that a lender may face when borrower fails to repay a loan or meeting their repayment obligations.
- **Default:** is failure by a borrower to make the required payments, either partially or in full, within the specified timeframe.
- **Early repayment:** is the act of repaying a loan before the predetermined maturity date
- **Survival analysis:** is a statistical methodology used to analyze the time until an event of interest occurs.
- **Competing risk:** refers to a situation in which multiple events or outcomes can occur, and the occurrence of one event precludes the occurrence of the others.
- **Gray's test:** is a statistical test used in survival analysis to assess the differences in cumulative incidence functions (CIF).
- **Fine-Gray model:** is a statistical model used in survival analysis to examine the association between covariates and the occurrence of specific events in the presence of competing risks.

1.7 Organization of the Study

The remaining part of this thesis is structured as follows. Chapter 2 has presented a review of related literature about credit risk due to NPL or loan default, risk factors for default risk, and literature on competing risk models. Chapter 3 has emphasized the data source, competing risk model for survival data, methods of parameter estimation, and model assumption checking. In chapter 4, briefly discusses the result and major findings of the study. In chapter 5, contains the conclusion, recommendation and limitation of the study.

2 Literature Review

2.1 Overview of loan default

Loan defined as a type of debt instruments, which entails the redistribution of financial assets over time between the lender and the borrower according their agreement. It is also typically, the money which is expected to paid back in regular installments or partial repayments periodically that each installment being of the same amount (Velan, 2017). Additionally, success of loan repayment defined as the ability to repay the loan full as per the loan agreement and loan defaulting is the inability to repay the loan by either failing to complete the loan as per the loan agreement or neglect the loan (Abajihad, 2020).

Credit risk is among the four business risks (capital risk, market risk, credit risk and operational risk) faced by organizations. The last decade or so has witnessed a growing emphasis on management and measurement of this risk. Credit risk in the simplest way may be defined as the "probability of loss from a credit transaction". An integral element of credit risk is the risk of default, that is, failure to meet debt obligations (Gupta, 2017).

Default is usually defined as a violation of debt contract conditions, such as the lack of will or the inability to pay the loan back. In the case of default, the creditor (e.g. a bank or other financial institution) suffers a loss. The repayment ability is examined by checking the stability and sufficiency of income to cover all expenses and by evaluating the riskiness of the client. The riskiness of the client is typically established based on the estimation of the probability of default (PD) conditional to the client's characteristics (Rychnovský, 2018).

Default and ER events impact negatively on the lender because they both cut out a proportion of anticipated interest (Marimo, 2015).

Loan default is considered to have occurred with regard to a particular borrower when either or both of the two following events have taken place.

- The bank considers that the obligor is unlikely to repay his/her credit obligations to the bank in full.
- The obligor is more than 90 days past the due date on any credit obligation to the banking group.

Early Repayment(ER) refers to early closure of loan accounts. The customer "settles" the outstanding amount ahead of the original repayment period. The reasons for early closure of accounts differ from customer to customer. Some customers close accounts by switching to another lender.

Types of Term Loan

The Commercial Bank of Ethiopia has been giving term loans in three times horizons.

- **Short-term Loan**: This is a loan advanced for the purpose of working capital, and is payable within a year.
- **Medium-term Loan**: This is loan given or the purpose of building construction, machinery, equipment, furniture and vehicles. It is paid back within one to five years.
- **Long-term Loan**: The purposes of this type of loans are for construction of factory building and acquisitioning of machinery, for irrigation agriculture and plantation crops, for transportation vehicles and communication equipment and for any other infrastructure related with the project to be financed.

Based on the directive issued by (NBE, 2018) Ethiopian commercial banks have the obligation to categorize their loans into different classifications, namely pass, special mention, substandard, doubtful, and loss.

- **Pass**: Loans falling under this category enjoy complete protection based on the borrower's current financial status and repayment ability, and they are not subject to any form of criticism.
- **Special mention**: Short-term loans that are overdue for a period exceeding 30 days but less than 90 days, as well as medium and long-term loans that are overdue for a period exceeding 6 months but less than 12 months.
- **Substandard**: Short-term loans that are overdue for a period exceeding 90 days but less than 180 days, as well as medium and long-term loans that are overdue for a period exceeding 12 months but less than 18 months.
- **Doubtful**: Short-term loans that are overdue for a period exceeding 180 days but less than 360 days, as well as medium and long-term loans that are overdue for a period exceeding 18 months but less than 3 years..

- **Loss:** Short-term loans that are overdue for a period exceeding 360 days, and medium and long-term loans that are overdue for a period exceeding 3 years.

According to the classification mentioned in the (NBE, 2018) directive, the last three categories of loans, namely substandard, doubtful, and loss, are considered Non-Performing Loans (NPLs) or Default..

2.2 The risk factors that affect probability of default and early repayment

Several studies have explored loan specific, borrowers specific and bank specific factors associated with loan default in the world.

In a study (Mukono, 2015) conducted in Nairobi country, Kenya, an examination was undertaken on the factors that determine the repayment of loans by small and medium enterprises. The researcher utilized a logit regression model, along with descriptive statistical tools and inferential statistics, to analyze the data. The findings from the analyzed data indicate that loan size or amount, loan repayment period, collateral value, number of installments, loan type, purpose of loan, previous loan repayment mode, have a negative influence. Furthermore, lender or firm characteristics, such as interest rate, also negatively affect the loan repayment performance of small micro enterprises.

(Aslam et al., 2020) found that factors contributing to loan default among GB borrowers using binomial logistic regression. The results showed that interest rate and repayment amount were significant factors for loan default. In a study (Kariuki & Ngahu, 2016) conducted study examined the effect of interest rates on loan performance of MFIs in Naivasha Sub-County, Kenya and Interest rates were found to substantially influence loan performance.

The investigation conducted by (Benjamin, 2017) focused on the phenomenon of Microcredit Loan Repayment Default within the realm of Small Scale Enterprises in the Upper West Region. The findings indicated that interest rate, duration of the loan, loan amount all concurrently contribute to the probability and rate of default. The results demonstrate that enterprises that secured loans with higher interest rates display a greater likelihood of defaulting. Additionally, the amount of the loan serves as a positive determinant of the probability of defaulting on loan repayment. Consequently, enterprises that secure larger loan amounts face a 6.1 percent increase in the likelihood of default.

Previous study (Caselli et al., 2021) conducted the determinants of the time to default of approximately 15,000 loans guaranteed by the Italian Central Guarantee Fund from 2007- 2009, used cox proportional model. The result revealed that in regard to loan characteristics, both loan maturity (term of repayment) and size (amount of loan) produce statistically significant coefficients. In particular, the longer the maturity and the smaller the loan size, the longer the loan survival duration. An increase of one month in loan maturity decreases a loan's failure risk by 0.99%. Moreover, larger loans are more likely to experience default than are smaller ones, as an increase in the loan amount by 10 euros increases the failure risk by 1.37%.

The Performance of Loan Repayment Determinants in Ethiopian Micro Finance were studied with the specific reference of Sidama Micro Finance Institution (SMFI), Hawassa, Ethiopia (Pasha & Negese, 2014). Found that loan size and loan and repayment period were influence negatively and significantly borrowers loan repayment performance.

A study (Yibrie & Ramakrishna, 2017) conducted at ACSI by employing multinomial logit model found that, of several variables hypothesized to affect loan repayment performance of borrowers loan size, loan tenure and interest rate were found to be statistically significant factors.

In a study (Atsmegiorgis et al., 2014) conducted in commercial bank of Ethiopia Hawassa district, a sample of 183 customers who took loan from October, 2005 to April, 2012 was taken from the bank record. Kaplan-Meier estimation method and univariate Cox proportional hazard model were applied to identify factors affecting bank loan re- payment rate. The univariate Cox proportional hazard model result revealed that educational level, having previous loan experience, mode of repayment, collateral type and purpose of loan are significantly related with loan repayment rate of customers.

In the study (Marimo, 2015), survival analysis was applied to vehicle finance data obtained from a leading South African financial institution for the period 01 April 2009 to 31 March 2014. Logistic regression and Cox regression methods were compared to determine the superior technique in the presence of competing risks, default and early repayment. The result revealed that the occurrence of these two events over the observation period impacts negatively on profitability, however the default event is worse than the early settlement event. The study found that the customers with lower interest rate and shorter tenure(term) were associated with higher

risk of early settlement, similarly debt interest rate were the major risk factor that influence loan default and implied that the customers with higher interest rate had higher probability of default. And suggested that Survival analysis is superior to Logistic regression by gives more valuable information such as a whole predicted survival function rather than a single predicted survival probability, a better credit granting decision is made if supported by the estimated survival times.

2.3 Overview of credit risk in survival analysis

Credit scoring is often used for application scoring (whether to fail to reject or reject a new applicant) and behavioral scoring (prediction of likelihood of default by already accepted customers). This study is inclined towards behavioral scoring where the prediction of likelihood to default and early repayment is of prime interest. With the emergence of computer usage, statistical methods were developed to help granters identify and model good and bad risks (Hand, D. J., & Henley, 1997).

Traditionally, the estimation of the Probability of Default (PD) for a loan applicant has been carried out using classification techniques like logistic regression. However, alternative approaches have gained significance in the contemporary field of credit scoring. Among these, survival analysis has emerged as a valuable tool. This method allows for the modeling of the time it takes for a customer to default, rather than merely predicting whether default will occur or not. Furthermore, survival analysis enables the estimation of default probabilities over any desired time period, making it a flexible and useful technique (Dirick et al., 2019). In credit scoring we look for differences in application characteristics for customers with different survival times. Also, it is possible that there are two or more types of failure outcome. In consumer credit we are interested, in several possible outcomes when concerned with profitability: early repayment, default, closure, etc.

The concept of using survival analysis to generate credit-scoring models was first proposed by (Thomas et al., 1992), and later expanded upon by (Banasik et al., 1999). (Thomas et al., 1992) used the accelerated life exponential model to analyze 24 months of loan data. The author demonstrated that the suggested model accurately predicted the number of failures at each failure time. Then a scorecard was developed using multiple regressions, and it was demonstrated that a better credit-granting decision could be made if the score was supported by the projected survival times. Thus, it was discovered that survival analysis adds depth to the traditional technique. The

author stated that these methods can be used to any area of credit operations where predictor factors exist and the time to an event is important.

The logistic regression method can perform stepwise regression, which involves assessing model fit by adding or removing a potential candidate variable. The approach produces parameter estimates with optimal features such as lack of bias and low variances. Nonetheless, although logistic regression tells us if the client will default/settle early, survival approaches predict not just if but when customers will experience an event (Stepanova & Thomas, 2002). Statisticians developed further probability models to handle lifetime data.

(Banasik et al., 1999) compared performance of exponential, Weibull and Cox's nonparametric models with logistic regression and found that survival-analysis methods are competitive with, and sometimes superior to, the traditional logistic-regression approach. Also, the concept of competing risks was used when considering two conceivable outcomes: default and early settlement. It was noted by (Banasik et al., 1999) that there are various methods for advancing the performance of the most basic survival-analysis models, such as Weibull's, exponential, or Cox's proportional hazards models.

Survival analysis methods have lately got prominence in the credit modeling field. Several authors, for example, (Bellotti & Crook, 2007), have undertaken analyses demonstrating that survival methods tend to get superior to traditional statistical methods for estimating credit risk. The advanced survival methodology makes use of more information than traditional models because it includes censoring and time details that are difficult to include into linear or logistic regression models. Survival approaches make the fewest assumptions to achieve the necessary key analysis. There are no distributional assumptions required for the response variable. Survival strategies are thought to be better than LR. Some researchers, including (Stepanova & Thomas, 2002) and (Bellotti & Crook, 2007) conducted studies to demonstrate the limits of LR in handling survival data and its inferiority to survival analysis.

In recent research (John, 2020) it was proven that in standard survival analysis, as there is only one possible event, borrowers either experience the event or are censored at the time they were last known not to have the event. For each specific event, the analysis is carried out under a competing risk scenario by treating them as an event of interest. Competing events are censored

since they are not an event of interest. Using Kaplan-Meier method introduces bias estimates that increase the probability of the event of interest. The answer is that the uncensored individuals left in the sample do not represent all of those who failed to observe the event. Consider a censored case where the loan did not default during the data collection phase. Cases concerning early repayment and maturity are under censorship. This type of censoring is used in models when the default is the only event of interest. However, in the presence of competing hazards, the likelihood of a specific event occurring at any particular time is critical for normal survival analysis. The cumulative incidence, which considers competing hazards, is a plausible estimation for this matter.

3 Data and Methodology

3.1 Description of the study area

The present investigation has utilized a dataset comprising information on the loan default and early repayment of consumer credit data. This dataset has been obtained from Information Management Bureau in CBE, Head Office. The current study conducted at CBE, Head office, which is found in Addis Ababa City. Commercial Bank of Ethiopia (CBE) is one of the predominant banks operating in Ethiopia. CBE was legally established as a share company in 1963. Since then, it has been playing significant roles in the development of the country. As of June 2021, it had about 1.1 trillion birr in assets and held approximately 67% of deposits and about 53% of all bank loans in the country. Currently CBE has more than 40+ million account holders in its more than 1940+ branches.

3.2 Study design

In this study, a retrospective cohort study design was used to identify factors that contributed to loan default in the presence of early repayment. The data used for the study is a secondary data collected from individual accounts. This information illustrates a longitudinal study where individual customers from the same cohort go through different kinds of behaviors leading to default or early repayment.

3.3 Study population and Period

The study populations have comprised all borrowers who took business loan for different purposes or sectors from CBE, Head Office during the period of January 2015 to December 2023. The loan could be for domestic trade services, manufacturing, agriculture, building and construction and international trade (import or export) purposes.

3.4 Eligibility criteria

- **Inclusion criteria:** The eligible participants for this study were borrowers who took business loan from the bank between Jan 2015 and Dec 2023.
- **Exclusion criteria:** Borrowers who took loan predate January 2015 and borrowers who had maturity date after Dec 2023 were excluded; i.e the data consists of all active accounts between January 2015 and Dec 2023.

3.5 Data Collection Procedure

The current data is a secondary data obtained from CBE, Head Office Addis Ababa. All the information required is extracted from the Data Warehouse and Business Intelligence department for the period o1 January 2015 to 30 December 2023.

3.6 Data Structure for Modeling Competing Risk Events

Based on the inclusion and exclusion criteria a total of 1077 business loan borrowers whose repayment terms varied from 6 to 101 months together with their repayment status during the observation period of up to Dec 2023 were extracted. Application and behavioral variables are provided per account in the data set and some of the variables are used in the regression models as covariates. The repayment status is given per account per month under observation.

The current study investigates the use of competing risks model for a total of 1077 business loan borrowers and individual accounts observed and tracked every month from the point of entry (open date) to the point of exit into either default or ER or to the termination of study at the end of Dec 2023. The repayment status variable observed whether the account defaulted (If the loan's status falls into one of the following categories: substandard, doubtful, or loss) or paid off their loans early (early repayment was repayment before the indicated end date of the loan repayment and is considered a competing risk).

In this study there is no censored observation because of two reason, first if an account neither defaults nor settles early then the account may repay at given time but in the given data set there was no any account that repay in the maturity date, second there is no possibility of withdrawals and loss to follow up censoring in the current data as all the accounts are kept in check to final classification.

The standard way of representing competing risks data for modeling probability of default as event of interest and early repayment as competing risk event is illustrated in the following table.

Table 3. 1 Data Structure for Modeling Competing Risk Events

ID	Time	Status	Mode of repayment	Purpose of loan
1	3	1	Monthly	Agriculture
2	6	1	Quarterly	Domestic trade and service
3	8	2	Monthly	Building and construction
4	15	1	Annually	Manufacturing

ID: Borrowers

Time: Times (in month) at which the event of interest, default or competing event, early repayment occurs.

Status: 1 if the account is defaulted, 2 if the account is repaid early. Mode of repayment and purpose of loan are covariates.

3.7 Study variables

3.7.1 The response variables

The response variable in the study are time to loan default of borrowers starting from the time of the loan granted and time to early repayment. The status of the borrowers is 1 if interested event (loan default) occurred, 2 if early repayment (competing event) occurred.

3.7.2 Predictor variables

Based on the literature review, the covariates considered in the study are

- Loan size (Amount of loan in birr)
- Purpose of loan(International trade(import-export), Manufacturing, Domestic trade and service, Building and construction, Agriculture)
- Interest rate (In %)
- Previous loan experience (Yes, No)
- Mode of repayment (Monthly, Quarterly, Annually)
- Collateral (No, if it is unsecured and Yes, if it is secured loan)

3.7.3 Coding and description of variables

Table 3. 2 Coding for outcome variables and explanatory variables

No	Variables	Categories
1	Loan size	Continuous(in birr)
2	Purpose of Loan	(0) Agriculture, (1) Manufacturing, (2) Domestic trade and service, (3) Building and construction, (4) International trade
3	Interest rate	Continuous (in %)
4	Previous loan experience	(0) No, (1) Yes
5	Mode of repayment	(0) Annually, (1) Quarterly, (2) Monthly
6	Collateral	(0) No, (1) Yes

Outcome Variables

No	Variables	Description
1	Time (in month)	Length of time starting from the time of the loan granted to either of the event occurred
2	Default	Status= 1, if borrowers have default repayment status
3	Early Repayment(ER)	Status= 2, if borrowers repay the loan early(before its maturity date)

3.8 Statistical models

3.8.1 Competing risk analysis of survival data

According to Austin et al. (2016), a competing risk is one in which the occurrence of one event prevents the occurrence of the other or significantly alters the probability of the event of interest. It occurs in research projects where study participants are exposed to multiple events or failure causes that are mutually exclusive. For instance, in follow-up research involving patients with cardiovascular disease, cancer is considered a competing risk because the patient's death was caused by cancer rather than cardiovascular disease. Furthermore, failing to consider the

potential impacts of competing events may result in inaccurate risk assessments, missing reporting requirements, and perhaps incorrect interpretation.

One can choose either the Sub-distribution hazard model (Fine-Gray Model) or the cause-specific hazard model for fitting models in the presence of competing risks.

• For participants who are currently event-free, the effect of the covariates on the outcome's rate of occurrence is estimated using the cause-specific hazard model. The sub-distribution hazard model, also known as the Fine-Gray model, enables us to estimate the impact of covariates on the outcome's absolute risk over time.

Furthermore, when analyzing survival data in a competing risk arrangement, where a member of the risk set is exposed to many causes of failure, standard survival approaches are not appropriate. The Cox proportional hazard model (Cox, 1972) is one of the most often used techniques for evaluating the competing risk data to investigate the impact of covariates on the cause-specific hazard function. When estimating regression parameters under a particular cause, this model treats persons failing for reasons other than the cause of interest as censored observations, which is a severe limitation when applied to competing risk data. (Fine & Gray, 1999) devised an innovative approach to overcome the limitation, utilizing the cumulative incidence function (CIF) to delineate the likelihood of an event occurring prior to a designated time. When a particular cause is of interest, the CIF-based approach, in contrast to the Cox model, does not overlook the other competing risks. Furthermore, the Fine-Gray model is predicated on a proportional hazards model for a competing risk's sub-distribution, in which the study's covariates have a direct impact on the cumulative incidence function.

Given$\Pr(T \leq t, D = k)$, the CIF for the K^{th} failure or cause is given, where D is the type of event that occurred from the potential event at survival time t. $S(t) = P(T \geq t)$ is the survival function, and $F(t) = 1 - S(t) = \Pr(T \leq t)$ is the incidence of the event throughout the follow-up, according to traditional survival analysis. Taking competing risks into consideration, the cumulative incidence function (CIF) enables assessment of the incidence of an event's occurrence, in instead of $1 - S(t)$. With all competing occurrences taken into consideration when making clinical decisions, this makes it possible to estimate incidence in a population.

One important aspect to consider in competing risk analysis is the possibility of multiple event types occurring, where the occurrence of one event prevents the subsequent occurrence of other event types. The probability of experiencing the kth event before time t and before the occurrence of a different type of event is denoted by the function $\mathrm{CIF}_k(t)$. In contrast to the survival function when there are no competing risks, the $\mathrm{CIF}_k(t)$ does not necessarily approach unity as time increases. This is due to the presence of competing events that prevent the occurrence of events of type k. As a result, the CIF or sub-distribution function can be greater than one, which means it does not meet the criteria of being a proper distribution (Pintilie, 2006).

3.8.2 Basic hazard functions in competing risk analysis

In competing risk analysis, there are two commonly used hazard functions: the cause-specific hazard and the sub-distribution hazard (cumulative incidence functions).

The cause-specific hazard function for the kth cause,(Austin et al., 2016) is defined as

$$\lambda_r^{cs}(t) = \lim_{\Delta t \to 0} \frac{Prob(t \leq T < t + \Delta t, D = k | T \geq t)}{\Delta t}, \qquad k = 1, \ldots, D \qquad (1)$$

In our particular case, we consider k to be equal to 1 and 2, representing two specific types of events or failures.

It represents the instantaneous rate at which the kth event occurs for individuals who are currently event-free, meaning they have not yet experienced any of the different types of events. The sub-distribution function for the kth event, (Austin et al., 2016) is defined by

$$\lambda_k^{sd}(t) = \lim_{\Delta t \to 0} \frac{Prob\big(t \leq T < t + \Delta t, D = k \big| T > t \cup (T < t \cap D \neq k)\big)}{\Delta t} \qquad (2)$$

$$F_k(t) = P(T \leq t, D = k) \qquad k = 1, \ldots D$$

and corresponds to the probability of a subject failing from cause k in the presence of all the competing risks. The CIF (Cumulative Incidence Function) is employed to estimate the risk of experiencing a specific event for individuals who have not yet encountered that event. It represents the instantaneous risk of failure from the kth event for individuals who have not yet experienced an event of type k.

The fundamental distinction between the cause-specific hazard and CIF lies in their respective risk sets. The risk set encompasses the individuals or subjects who are under investigation and susceptible to the event. In the CIF, the risk set comprises individuals who are currently alive as well as those who have previously experienced a competing event. In other words, the CIF considers both individuals who are still at risk of the event and those who have already experienced a different event. On the contrary, the risk set of the cause-specific hazard function only includes individuals who are currently event-free and still at risk of the specific event being analyzed.

In the univariate analysis of predictor variables, the cumulative incidence function (CIF) was utilized to assess the association between the outcome variable and each individual predictor. To determine the significance of the association, Gray's test was employed. Gray's test, introduced by (Gray,1988), is a K-sample test that compares the weighted averages of the sub-distribution hazards across different groups for the specific event of interest. The null hypothesis in Gray's test is that there is no difference in the sub-distribution hazards across the groups being compared. In other words, it assumes that the occurrence rates of the event of interest are similar across the groups. The test evaluates whether there is sufficient evidence to reject this null hypothesis and conclude that there is a significant difference in the sub-distribution hazards among the groups.

3.9 Regression models for competing risks

In survival analysis with competing risks, two commonly used regression modeling approaches rely on the hazard functions mentioned earlier: the cause-specific hazards model and the Sub-distribution hazard model, also known as the Fine-Gray model (Porta N. Gómez G. & Malats, 2007). The cause-specific hazards model is employed to estimate the impact of covariates on the rate of occurrence of the outcome among subjects who are currently event-free. It focuses on the hazard function specific to each event type and allows for the analysis of how covariates influence the occurrence rate of a particular event while considering the presence of competing events. On the other hand, the Fine-Gray model (Fine & Gray, 1999) enables the estimation of the effect of covariates on the absolute risk of the outcome over time. It takes into account the cumulative incidence function and allows for the examination of how covariates affect the overall risk of the specific event, considering the presence of competing events.

3.9.1 Cause-specific hazard model

In modeling cause-specific hazards, each hazard is examined individually by treating individuals who fail from other causes as censored observations. This approach is suitable when the objective is to identify factors that influence the rate at which events occur. In the usual regression analysis of competing risks, a Cox proportional hazards (PH) model, as defined by (Prentice et al., 1978), is established for each cause-specific hazard.

$$\lambda_k(t|X) = \lambda_{0k} exp(x\beta_k) \qquad k = 1, \ldots . D$$

Where x is a px1 vector of covariates, βk is a px1 vector of unknown parameters to be estimated for each outcome variable, k is the cause of failure at time t. And λ_{ok} are the baseline cause specific hazard and it is important to thoroughly examine the proportional hazards (PH) assumption for each cause in this scenario, as it is a significant assumption that carries considerable weight.

3.9.2 Fine-Gray model

The Cox proportional hazards (PH) model has commonly been used to analyze competing risks data. However, due to several limitations of this method when dealing with survival data in the presence of competing risks, (Fine & Gray, 1999) introduced a CIF (Cumulative Incidence Function) based proportional hazard model specifically designed for analyzing competing risk data. In this modeling approach, for each cause of interest, the occurrence of the event is considered in the presence of covariates (Mohammad et al., 2017). This CIF-based proportional hazard model is particularly useful when the objective is to determine which covariates influence the probability of an event happening over time. Additionally, the Fine-Gray model provides sub-distribution hazard ratios that indicate the relative effect of covariates on the sub-distribution hazard function or cumulative incidence function (Austin & Fine, 2017). The sub-distribution hazard model is employed to identify factors associated with the incidence of a specific cause within the competing risks framework. Importantly, this method of analysis does not treat individuals failing from other causes as censored observations.

When dealing with competing risks, the selection of an appropriate modeling approach depends on the specific goals of the study. Researchers have recommended using the sub-distribution hazard model when the aim is to estimate incidence rates or predict prognosis in the presence of

competing risks. On the other hand, the cause-specific hazard model is more suitable when the focus is on addressing etiological questions related to the different causes of events (Austin et al., 2016). In this thesis, we have utilized the Fine-Gray modeling approaches to identify the factors that influence the likelihood of loan default in the presence of early repayment. By employing the Fine-Gray model, we can account for the presence of competing events and examine the specific factors associated with the incidence of loan default while considering the occurrence of early repayment.

Using the relationship between the survival, hazard and cumulative incidence function (Leoce, 2016).

$$\lambda(t) = \frac{f(t)}{S(t)} = \frac{f(t)}{1 - F(t)} \qquad (4)$$

The sub-distribution hazard (hazard of the cumulative incidence) for each cause supposed as the hazard for an individual who either fails from cause k or does not, can be written as:

$$\lambda_k(t, X) = \frac{f_k(t)}{1 - F_k(t)} \qquad (5)$$

Under the proportional hazard, Fine-Gray Model can be specified (Leoce, 2016) as:

$$\lambda_k^*(t; X) = \lambda_{0k}^*(t)\exp(X\beta k) \qquad (6)$$

Where λ_{0k^*}(t) is the baseline sub-distribution hazard for the cause of k, x is px1 covariates, such as loan size, purpose of loan, mode of repayment, interest rate,… and exp (βk) is the relative risk probability of k^{th} cause associated with the given covariates. In the current study the Fine-Gray model was fitted separately for each outcome variables (default and early repayment).

3.10 Methods of parameter estimation

The estimation of parameters in the Fine-Gray model follows a partial likelihood approach, which is similar to the standard Cox model. This approach is employed because a proportional hazard assumption is imposed on the sub-distribution hazards (Pintilie, 2006). However, in the Fine-Gray model, the parameters are estimated by incorporating weights into the partial likelihood.

The partial likelihood for the Fine-Gray model is given (Kuk & Varadhan, 2013) as:

$$L(\beta) = \prod_{k=1}^{r} \frac{exp(X\beta_k)}{\sum_{i \in R^*(tk)} Wki\ exp(X\beta_i)} \qquad (8)$$

The product is taken over all r time points, (t1< t2 < … < tr), where r is the total number of the event of interest i.e. ($\sum_{i=1}^{n} indicator\{\in_i = 1\}$). The modified risk set, $R^*(t_k)$ is a set of subjects that are still at risk for the event of interest at time t (i.e., those who did not experience the primary event and are not censored by time t) (Fine & Gray, 1999). Thus, subjects that have experienced other types of events remain in the risk set all the time. Besides, the weight is defined as

$$W_{ki} = \frac{\hat{G}(tk)}{\hat{G}(\min(tk, ti))} \qquad (9)$$

Where t_i = min (T_i,C_i) for i such that $\in i \neq 1$ and t_k is the time of the kth event. $\hat{G}$ is the KM estimate of the survivor function of the censoring distribution ($G(t) = P(C \geq t)$). For individuals who have not experienced any event by a specific time point (t_k), the weight assigned to them is one. However, for those who have encountered a competing event before t_k, the weight is less than one. Consequently, individuals who encounter a competing event at time t_i do not contribute fully to the partial likelihood. The weight assigned to them decreases as the time point (t_k) moves further away from the time of the competing event (ti). In cases where there is a single event of interest, all weights are equal to 1, and the risk set consists only of individuals who are at risk at the specified time point (Kuk & Varadhan, 2013). After formulating the likelihood, the objective is to select parameter values that maximize the likelihood.

3.11 Model Diagnosis

The primary assumption in survival data modeling is the proportionality of hazards. When utilizing the Fine-Gray model, it is crucial for the hazards of the Cumulative Incidence Function (CIF) to be proportional. In contrast, the Cox proportional hazard model requires the cause-specific hazards to exhibit proportionality (Pintilie, 2006). The proportionality assumption is commonly employed in competing risk regression models, where the sub-distribution with covariates x is considered a constant shift on the complementary log-log scale from a baseline sub-distribution function. If the curves representing the hazards do not intersect, it indicates that the model satisfies the proportionality assumption (Zhang, 2017).

3.11.1 Proportionality of the cause-specific hazards

By examining the graph of log (-log(S)) against log (time), we can gain insight into whether the cause-specific hazards can be assumed to be proportional. In this context, S represents the Kaplan-Meier estimate when considering only the event of interest, and both observations without an event and the competing risks are treated as censored. It is possible to create this plot for each level of a covariate (Pintilie, 2006).

3.11.2 Proportionality of the hazard of the CIF

To assess the proportionality assumption in competing risks regression, one can plot log (-log (1-F)) against log (time), where F represents the Cumulative Incidence Function (CIF) for the event of interest. In the present study, we have employed the Fine-Gray model and examined the assumption of proportionality for the CIF.

Statistical software

The competing risk data was analyzed using the survival package and Cmprsk package of the R statistical software (version 4.3.3).

4 Results and Discussion

4.1. Descriptive result

4.1.1. Summary Statistics for categorical variables

The data set consisted of 1077 customers, among those borrowers 14 had annually repayment mode, 205 had quarterly repayment mode and 858 had monthly repayment mode. Of total of borrowers, 723 were took for the purpose of domestic trade and service, 113 were took for manufacturing purpose, 113 were took for agriculture, 97 were took for international trade purpose, and 31 were took for building and construction purpose. Out of 1077 borrowers, 769 had no previous loan experience and 308 had previous loan experience and also 147 had no collateral (not secured loan) and 930 had collateral (secured loan). All the results have been summarized in **Table 4.1** below.

Table 4. 1 Descriptive statistics summary of categorical variables

Categorical variables	Levels	No	Percent (%)
Purpose of loan	International trade	97	9.0%
	Manufacturing	113	10.5%
	Building and construction	31	2.9%
	Agriculture	113	10.5%
	Domestic trade and service	723	67.1%
Previous loan experience	NO	769	71.4%
	YES	308	28.6%
Mode of repayment	Annually	14	1.3%
	Quarterly	205	19%
	Monthly	858	79.7%
Collateral	No	147	13.6%
	Yes	930	86.4%

4.1.2. Summary statistics for categorical variables based on each type of event with corresponding median time

The data set consisted of 1077 customers whose term of repayment varied from 6 to 101 months together with their repayment status during the observation period of up to Dec 2023, of the total of borrowers, 893(82.9%) defaulted and 184(17.1%) were repaid early, the median time was 12.2 and 81.3 months respectively. Of the 723 borrowers with loans for domestic trade and service, 79.4% defaulted and 20.6% were repaid early with their corresponding median time of 79.4 and 48.8 months respectively. Among 113 borrowers with loans for agricultural purposes, 96.5% were defaulted and 3.5% were repaid early with their corresponding median time of 11.7 and 48.1 months respectively. Among 113 borrowers with loans for manufacturing purposes, 95.6% were defaulted and 4.4% were repaid early with their corresponding median time of 17.1 and 74 months respectively.

Out of 769 borrowers with no previous loan experience 97.8% defaulted and 2.2% were prepaid with their corresponding median time of 11.6 and 85.0 months respectively. Out of 308 borrowers with previous loan experience, 45.8% were defaulted and 54.2% were repaid early with their corresponding median time of 34.0 and 18.3 months respectively. Among 858 borrowers who had monthly repayment mode, 81.5% defaulted and 18.5% were repaid early with their corresponding median time of 11.3 and 24.5 months respectively. Among 205 borrowers who had a quarterly mode of repayment, 90.2% defaulted and 9.8% were repaid early with their corresponding median times 36.6 and 81.3 months respectively. Among 14 borrowers with annually mode of repayment, 64.3% defaulted and 35.7% were repaid early with their corresponding median time of 95.4 and 96.1 months respectively. Of 930 borrowers with secured loans, 82.5% defaulted and 17.5% were repaid early with their corresponding median time of 12.2 and 77.7 months respectively. Among 147 borrowers with no collateral, 85.7% defaulted and 14.3% were repaid early with their corresponding median time of 12.2 and 91.3 months respectively. All the results have been summarized in Table 4.2 below.

Table 4. 2 Distribution of loan characteristics of borrowers

			Repayment Status			
			Default	Med time	ER	Med time
Covariates	Categories	No	%	Months	%	Months
Purpose of loan	International trade	97	76.3	36.5	23.7	83.1
	Manufacturing	113	95.6	17.1	4.4	74.0
	Domestic trade and service	723	79.4	11.5	20.6	48.8
	Building and constructions	31	90.3	36.5	9.7	91.3
	Agriculture	113	96.5	11.7	3.5	48.1
Previous loan experience	No	769	97.8	11.6	2.2	85.0
	Yes	308	45.8	34.0	54.2	18.3
Mode of repayment	Monthly	858	81.5	11.3	18.5	24.5
	Quarterly	205	90.2	36.6	9.8	81.3
	Annually	14	64.3	95.4	35.7	96.1
Collateral	No	147	85.7	12.2	14.3	91.3
	Yes	930	82.5	12.2	17.5	77.7

4.1.3. Summary statistics for continuous variables

The descriptive statistics for continuous variables indicated that the mean and median follow up period of the study were about 20 and 12 months respectively. The mean amount of loan was 6,274,404 birr and the median was 2,900,000 birr. Similarly, the mean and median of interest rate was 12.9 and 14 respectively. All the results had been summarized in **Table 4.3** below.

Table 4.3 Descriptive Statistics for Continuous Variables

Variables	No	Mean	Median	Min	Max	SD
Time(in months)	1077	19.73	12.17	3.17	98	16.2
Loan size (in birr)	1077	6,274,404	2,900,000	10000	55,000,000	8691536.69
Interest rate (in %)	1077	12.96	14	8.50	16.5	1.90

4.1.4 Graphical descriptions of categorical variables by two competent events (CIF plot)

The nonparametric estimates of the cumulative incidence functions for default and ER events for the whole study population are given in **Figure 4.1** below. The plot indicated that the estimated probability of default is 50% during 15 months and 60% during 20 up to 38 months and become 80% and held constant during 40 up to 98 months. The estimated 98 month probability of early repayment without any prior default is 10% during 15 months and 15% during 16 up to 20 months and become around 20% and held constant during 22 up to 98 months.

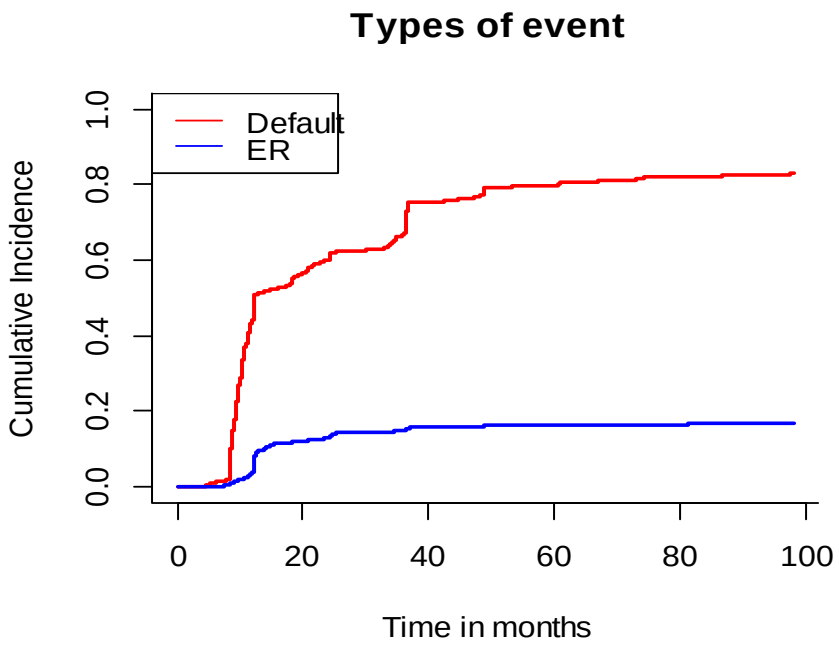


Figure 4.1 Plot of Cumulative incidences of default and early repayment

The estimated probability of default is higher for borrowers having a monthly loan repayment mode when compared to both with annually and quarterly mode of repayment as shown in **Figure 4.2** below. Similarly **Figure 4.3** indicated that the estimated probability of ER for borrowers having a monthly loan repayment mode is higher than both having quarterly and annually repayment mode.

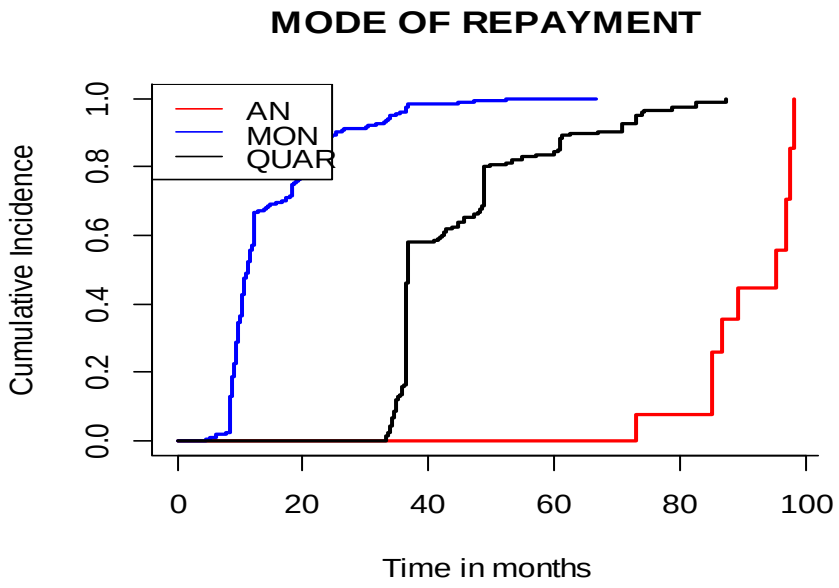


Figure 4.2 plot of cumulative incidence of default for mode of repayment

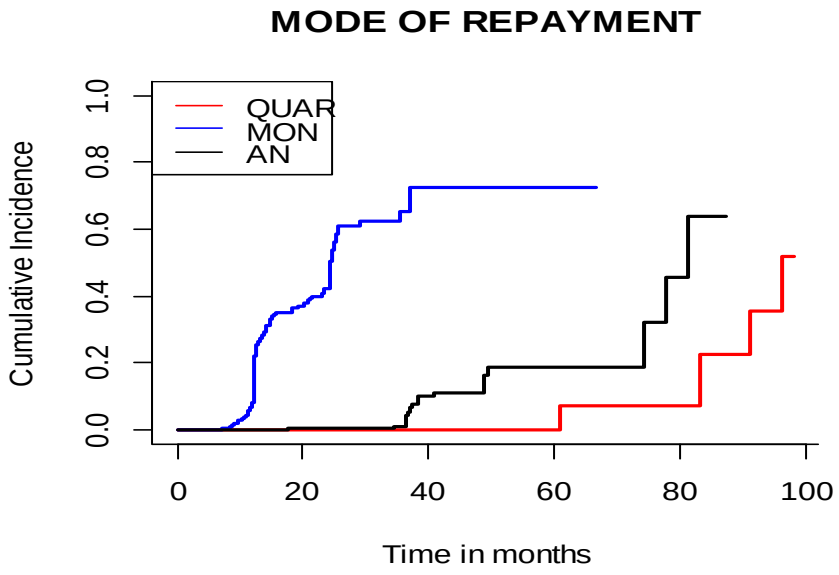


Figure 4.3 ***plot of cumulative incidence of early repayment for mode of repayment***

The estimated probability of default is higher for borrowers with no previous loan experience when compared with borrowers with previous loan experience as shown in **Figure 4.4** below. Similarly from the plot in **Figure 4.5**, we can observe that the estimated probability of ER for

borrowers with previous loan experience is higher than those borrowers with no previous loan experience. The rest of CIF plots are in the Appendix-A.

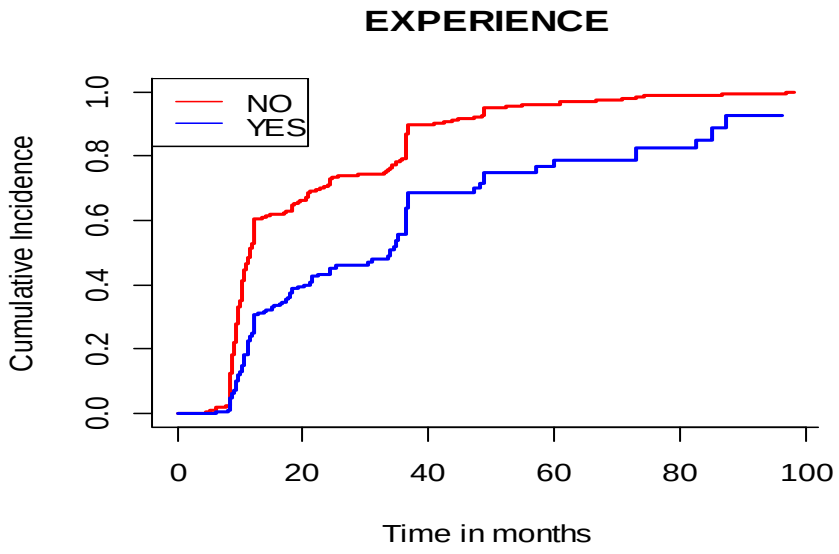


Figure4.4 ***plot of cumulative incidence of default for experienced in loan***

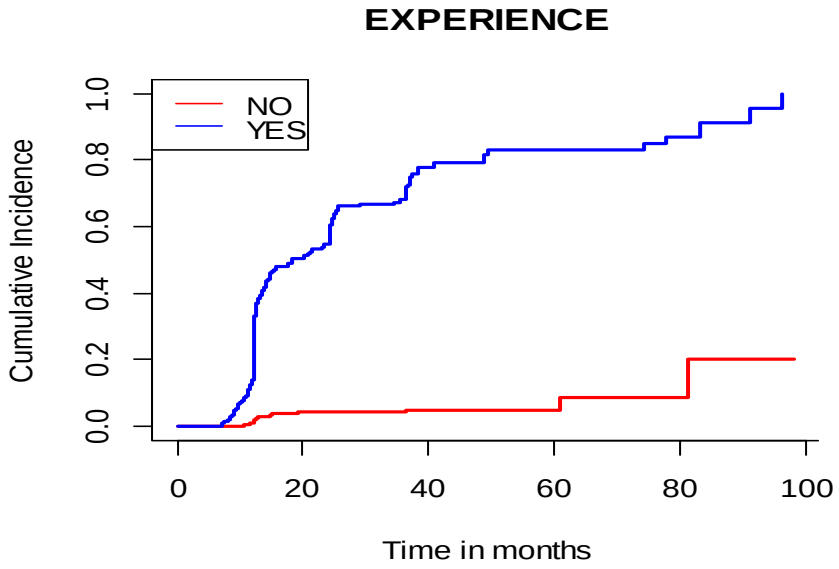


Figure 4.5 plot of cumulative incidence of early repayment for experienced in loan

4.2 Analytical Results

4.2.1. Univariate analysis of competing risk Model

A cumulative incidence function was used for univariate analysis of each prognostic factor and Gray's test was used to assess the association between each significant predictors and the outcome variables considered in the study. Gray's test compares the weighted averages of the sub-distribution hazards across groups for the event of interest. Then, all the potential covariates that were significant at 5% level of significance in the univariate analysis obtained from Gray's test were added to the competing-risks model for the multivariate analyses.

The results of the Gray's test revealed that statistically significant differences in the cumulative incidence function among various groups of covariates such as, the purpose of the loan, previous loan experience, and mode of repayment. These findings indicate that the covariates examined have a significant impact on the likelihood of defaulting on the loan or repaying it early at the follow-up period. However, According to the results of the Gray's test, there is no significant difference in the cumulative incidence of default (p=0.214) or early repayment (p=0.43) between borrowers with secured and unsecured loans within the collateral group. This suggests that the presence or absence of collateral does not have a substantial impact on the cumulative incidence rates of default or early repayment as illustrated in **Table 4.4** below.

Table 4.4 Gray's Test Results for Default and Early repayment status

Covariates	Categories	Default		Early Repayment		
		Gray's test	P-value	Gray's test	P-value	Df
Purpose of loan	International trade	12.2	0.016	67.76	<0.0001	4
	Manufacturing					
	Domestic trade and service					
	Building and construction					
	Agriculture					
Previous loan experience	No	85.95	0.000	301.1	0.000	1
	Yes					
Mode of repayment	Annually	192.8	0.000	156.6	0.000	2
	Quarterly					
	Monthly					
Collateral	No	1.54	0.214	0.61	0.43	1
	Yes					

4.2.2. Sub-distribution Hazard Regression Fine and Gray model

The aim of this study was to determine the risk factors that affect the probability of loan default in the presence of early repayment and for this purpose Fine-Gray model was used to identify the significant risk factors that associated with the probability of loan default and early repayment at 5% level of significance. The SDH regression for multivariable analysis was fitted using Fine-Gray model and the covariates that were significant at 5% level of significance in the univariate analysis obtained from Gray's test were added to the multivariate analysis. The covariate, collateral is not included in multivariate analysis because it was not significant in Gray's test. Sub-distribution hazard ratio (SHR) and 95% CI for SHR obtained from the Fine-Gray model fit results are presented in **Table 4.5** and **Table 4.6** below.

The study found that the covariates such as, purpose of loan for manufacturing and international trade, monthly and quarterly modes of repayment, previous loan experience and loan amount were statistically significant risk factors associated with default risk at 5% level of significance.

In contrary the covariate purpose of loan for domestic trade and service, and for building and construction, and interest rate were not statistically significant.

The sub-hazard ratio (SHR) of default for loan size was 1.097 with a 95% confidence interval (1.035, 1.163). This implies that for every one-unit increase in the loan amount, the risk of default also increases by 9.7%, assuming all other variables are held constant. The borrowers with quarterly repayment mode had 1.57 times higher risk of default than those with annually repayment mode with a 95% confidence interval of (1.106, 2.23), keeping other variables constant. The borrowers with monthly repayment mode had 5.17 times higher risk of default than those with annually repayment mode with a 95% confidence interval of (3.56, 7.52), keeping other variables constant. The borrowers who had previous loan experience are 85.7% less likely to default than those who did not have previous loan experience with a 95% CI of (0.108, 0.189), keeping other variables constant. The borrowers with the purpose of the loan, categorized as manufacturing and international trade, were significantly associated with the default risk; the sub-hazard ratios (SHR) were 0.68 and 0.76, respectively. This indicates that the risk of default for these categories is 32% and 24% lower among borrowers who took loans for agricultural purposes with a 95% CI of (0.55, 0.85) and (0.61, 0.95), respectively by keeping other variables constant.

According to the Fine-Gray model result for early repayment, covariates such as previous loan experience, quarterly repayment mode, and interest rate were statistically significant risk factors associated with risk of early repayment; on the contrary, all the purposes of loan categories, monthly repayment mode, and loan size were not statistically significant risk factors associated with risk of early repayment (ER).

The SHR of ER for interest rate was 1.45 with a 95% CI (1.205, 1.758) and this implied that as the interest rate increase by one percent the risk of early repayment increases by 45%, by keeping other variables as constant. The borrowers with previous loan experience were 29.7 times more likely to repay the loan early than those with no previous loan experience, with a 95% CI (17.73, 49.91), by keeping other covariates constant. The borrowers with quarterly repayment mode were 65% less likely to have early repayment than those with annually repayment mode, with a 95% CI (0.160, 0.785), keeping other variables constant.

Table 4.5 Results of the Fine-Gray Model for Default status

Covariates	Categories	Coef	SHR	95% CI	P-value
Purpose of loan	Agriculture(Ref)	-	-	-	-
	Manufacturing	-0.374	0.688	0.556, 0.850	0.00053
	Domestic trade and service	0.012	1.012	0.849, 1.206	0.890
	Building and constructions	-0.217	0.805	0.614, 1.056	0.120
	International trade	-0.268	0.765	0.614, 0.953	0.017
Previous loan experience	No(Ref)	-	-	-	-
	Yes	-1.946	0.143	0.108, 0.189	0.000
Mode of repayment	Annually (Ref)	-	-	-	-
	Quarterly	0.450	1.570	1.106, 2.228	0.012
	Monthly	1.640	5.170	3.565, 7.523	<0.0001
Loan size	Conti (in birr)	0.0925	1.097	1.035, 1.163	0.002
Interest rate	Conti (in %)	-0.0013	0.990	0.970, 1.028	0.930

***Table 4.6* Results of the Fine-Gray Model for Early Repayment status**

Covariates	Categories	Coef	SHR	95% CI	P-value
Purpose of loan	Agriculture(Ref)				
	Manufacturing	-0.041	0.959	0.254 3.616	0.950
	Domestic trade and service	0.383	1.467	0.535 4.026	0.460
	Building and constructions	0.258	1.295	0.372 4.513	0.680
	International trade	0.638	1.893	0.649 5.521	0.240
Previous loan experience	No(Ref)	-	-	-	-
	Yes	3.392	29.75	17.73 49.91	0.000
Mode of repayment	Annually (Ref)	-	-	-	-
	Quarterly	-1.035	0.355	0.160 0.785	0.0021
	Monthly	-0.351	0.704	0.344 1.438	0.340
Loan size	Conti (in birr)	-0.132	0.876	0.706 1.086	0.230
Interest rate	Conti (in %)	0.375	1.455	1.205 1.758	0.0001

4.2.3 Model diagnosis

The log (-log (1-CIF)) vs log (time) plot is a commonly used graphical method to assess the assumption of proportional sub-distribution hazards for the event of interest. In the provided figures in Appendix B, this plot was used to evaluate the assumption of proportionality for the covariates of previous loan experience, mode of repayment, and purpose of the loan.

The plot of log (-log (1-CIF)) vs. log (time) revealed evidence supporting the assumption of proportionality for the default event with the covariates of previous loan experience and mode of repayment. There is no evidence of crossing or overlapping hazards in the plot, indicating that both previous loan experience and mode of repayment satisfy the proportional hazards (PH) assumption. The lines in the plot are nearly parallel, further supporting the satisfaction of the PH assumption. On the other hand, the proportional hazard assumption was violated for the covariate group of the purpose of the loan, as indicated by the plot presented in Appendix B. The violation of the proportional hazards assumption suggests that the proportional hazards assumption does not hold for purpose loan covariates, indicating a potential time-varying effect or interaction with other factors.

4.3 Discussion

This study was primarily conducted to investigate and pinpoint the risk factors that impact loan default in the presence of early repayment on business loan borrowers at CBE, Head Office. The result showed that about 82.9% of borrowers had default loan status and 17.1% borrowers were repaid early and this indicated that there was higher default risk. This implied that majority of the borrowers were unable to repay the loan in a given installments and this leads to decrease the credit score for borrowers and decrease the profit of the institution. The lender is likely to suffer more in the case of default than in early settlement as there is a possibility of losing a fraction of capital in addition to the interest amount whereas in early repayment, the lender cannot lose more than the original capital amount issued (Marimo, 2015).

In this study the Gray's test (Gray, 1988) were used to test whether there was a significant difference in cumulative incidence function curve between groups of covariates for a given events, default or early repayment. The result showed that there was a significant difference in cumulative incidence between groups of covariates such as, purpose of loan, mode of repayment and also previous loan experience for both default and early repayment events at 5% level of

significance. However, there was no significant difference in cumulative incidence for collateral groups, secured and unsecured for both default and early repayment events at a 5% level of significance, and as a result, the variable collateral was not included in the multivariate analysis of sub-hazard distribution, moreover, the CIF plot was presented for all covariates to visualize the group difference for both events, default and early repayment.

The study showed that the SHR of loan default for loan size was 1.097, which implied that the increase in loan size was significantly associated with an increase in default risk. This finding is consistent with previous studies done by (Pasha & Negese, 2014) and (Caselli et al., 2021), while inconsistent with the study done by (Yibrie & Ramakrishna, 2017). This could be because high loan amounts are more likely to default due to the increased financial burden they impose on borrowers. Borrowers with large loan amounts may face more significant payment commitments or debt burdens. If their income or cash flow is insufficient to cover the increasing payment requirements, the likelihood of loan default increases.

The SHR of early repayment for interest rate was 1.45, which indicated that the increase in interest rate is associated with a decrease in early repayment. This finding is inconsistent with the study done by (Marimo, 2015). High interest rates on business loans may put borrowers at risk for early repayment because they can lower their overall interest costs and possibly strengthen their financial position by making extra payments or repaying the loan early. Additionally, high interest rates may encourage borrowers to repay the loan early in order to minimize the total interest paid over the loan term.

The result of the present study confirmed that previous loan experience is a significant risk factor for the probability of default. The borrowers who had previous loan experience have 85.7% times lower risk of default than those who had no previous loan experience. This finding is consistent with (Garomsa, 2017), (Atsmegiorgis et al., 2014). This might be due to borrowers with previous loan experience in the business loan sector who tend to exhibit a better awareness of loan terms, conditions, and potential hazards associated with commercial lending. They are aware of the consequences of defaulting on a commercial loan, including the impact on their business's creditworthiness and the chance for future financing. This awareness enables customers to make informed decisions, manage risks, and fulfill their loan obligations, resulting in a lower likelihood of loan default.

Borrowers with previous loan experience are 29.75 times more likely to return their loans early than borrowers without previous loan experience. This finding is consistent with (Atsmegiorgis et al., 2014). This might occur due to borrowers with previous loan experience are better informed and aware of the impact of interest charges on their financial well-being. Because they understand the value of lowering interest expenses, they are more likely to prioritize early repayment, which increases their chances of repaying loans ahead of schedule as compared to borrowers with no previous loan experience.

The SHRs of loan default for quarterly and monthly repayment modes were 1.57 and 5.17, respectively. This implied that the default risk for the borrowers who had quarterly and monthly repayment modes was 1.57 and 5.17 times higher than those borrowers with annually repayment modes, respectively. This finding is consistent with (Garomsa, 2017), (Caselli et al., 2021) and (Atsmegiorgis et al., 2014). This might be due to monthly and quarterly repayment schedules, which require borrowers to make more frequent payments, putting a burden on cash flow. If the business sees income changes or experiences temporary financial difficulties, it may struggle to meet these more frequent payment requirements. The higher payment frequency may raise the probability of default.

The SHR of early repayment for borrowers with a quarterly repayment mode was 0.355, and this implied that the borrowers who had a quarterly repayment mode were 64.5% less likely to repay the loan early than those with annually repayment mode. This may be since borrowers with quarterly repayment schedules must allocate cash more often, which may limit their capacity to build surplus funds for early payback. Borrowers with annual repayments, on the other hand, have more time to save and may be able to make a lump-sum early repayment more easily.

The study revealed that borrowers in the manufacturing and international trade categories had significantly lower default risk compared to agricultural borrowers. This result could be explained maybe by the fact that the manufacturing and international trade sectors are typically associated with more stable economic conditions. Stable economic environments can provide businesses with a better ability to generate consistent revenue and manage their financial obligations, resulting in lower default risk. In contrast, the agricultural sector may be more influenced by factors such as commodity price fluctuations and weather-related risks, which can increase default risk.

5 Conclusions and Recommendation

5.1 Conclusions

This study was primarily conducted to investigate and pinpoint the risk factors that impact loan default in the presence of early repayment by using a competing risk modeling approach among business loan borrowers at CBE, Head Office, spanning from January 01, 2015, to December 30, 2023. The result showed that about 82.9% of borrowers were defaulted and 17.1% borrowers were repaid early; with the corresponding median time were 12.2 and 81.3 months respectively.

From the current study, we can conclude that loan size, previous loan experience, quarterly repayment mode, monthly repayment mode, and purpose of loan for manufacturing and international trade are significant risk factors associated with default risk at 5% level of significance. The predictor variables such as quarterly repayment mode, interest rate, and previous loan experience, are significant risk factors associated with early repayment risk at 5% level of significance.

The study revealed that the increase in loan size significantly increased default risk. Borrowers with previous loan experience were associated with a significantly lower risk of default than those without previous loan experience. Borrowers with quarterly and monthly repayment modes had a significantly higher risk of default than those with annually repayment modes. Borrowers in the manufacturing and international trade categories had significantly lower default risk compared to agricultural borrowers. The study revealed that the increase in interest rate was significantly associated with increased risk of early repayment. Borrowers with previous loan experience were significantly associated with a higher risk of early repayment than those with no previous loan experience. Borrowers with a quarterly repayment mode had a significantly lower risk of early repayment than those with annually mode of repayment.

5.2 Recommendations

In light of the study's findings, the following suggestions are put forth for policymakers, borrowers, and the banking sector.

Implement robust monitoring mechanisms to track the financial health of borrowers with larger loan sizes throughout the repayment period. Regularly assess their financial statements, cash

flow, and repayment behavior. Promptly identify any signs of financial distress and take proactive measures to address potential defaults.

The bank should conduct more rigorous loan assessments for borrowers opting for monthly and quarterly repayment modes. Consider additional factors such as cash flow stability, income consistency, and debt-to-income ratios to evaluate the borrower's ability to meet frequent repayment obligations. Offer flexible repayment options that cater to the specific needs of borrowers with higher default risks. Provide alternatives repayment schedules to align with their income streams and cash flow patterns. Flexibility in repayment can reduce the strain on borrowers and improve their ability to make timely payments, thus decreasing the probability of default.

Customize repayment structures that align with the cash flow patterns and revenue cycles in the agricultural sector. Recognize that farmers may have irregular income streams and seasonal cash flows. Offer flexible repayment options, such as aligning repayments with harvest or sale seasons, to better accommodate the borrower's cash flow capabilities.

If early repayment reduces the bank's profitability, it may be effective to impose penalties on borrowers who opt to repay their loans ahead of schedule. Implementing penalties for borrowers who opt to repay their loans early may assist the bank in recovering the interest that would have been generated if the loans were paid according to the original timetable.

In general, the bank should exercise caution when approving loans with higher loan sizes, and borrowers without previous loan experience, and if they opt for monthly or quarterly repayment modes. Thorough risk assessment, tailored repayment structures, and proactive monitoring are essential to mitigate default risks associated with these factors.

5.3 Limitation of the study

This research has certain drawbacks. The first limitation is that we only acquired the data from the Commercial Bank of Ethiopia, head office. However, to enhance the generalizability of the findings and ensure their external validity, it is recommended that future researchers expand the sample to encompass borrowers from various branches of CBE or other financial institutions operating in Ethiopia. Secondly, the dataset only comprised of business borrowers, as such, including other types of loan categories may help to understand the problem better. By doing so,

a more comprehensive outlook can be obtained regarding the application of competing risk modeling in predicting loan default within the commercial banking sector of the country.

While this study has meticulously incorporated significant factors such as loan size, interest rate, loan purpose, previous loan experience, and collateral, it is crucial to acknowledge the potential existence of other variables such as the borrower's credit score, industry-specific economic indicators, business performance metrics, or the borrower's financial statements that could influence loan default and early repayment behavior, yet were not considered in this analysis. Therefore, it is recommended to explore the inclusion of additional covariates to capture a more comprehensive range of factors that might impact loan default and early repayment behavior. Furthermore, the behavioral variables utilized in this study were fixed at estimation; future research could explore implementing dynamic behavioral factors to improve prediction accuracy.

References

Abajihad, M. (2020). *Factors Affecting Loan Repayment Performance of Borrowers of Microfinance institutions in Case of Southwest Oromia, Jimma Zone*.

Abate, M. T., & Kaur, R. (2023). Banking Sector in Ethiopia: Origin and Present State. *EPH - International Journal of Business & Management Science*, *9*(2), 1–13. https://doi.org/10.53555/eijbms.v9i2.134

Alrfai, M. M., Salleh, D. B., & Waemustafa, W. (2022). Empirical Examination of Credit Risk Determinant of Commercial Banks in Jordan. *Risks*, *10*(4). https://doi.org/10.3390/risks10040085

Amuakwa-Mensah, F., Marbuah, G., & Ani-Asamoah Marbuah, D. (2017). Re-examining the Determinants of Non-Performing Loans in Ghana's Banking Industry: Role of the 2007–2009 Financial Crisis. *Journal of African Business*, *18*(3), 357–379. https://doi.org/10.1080/15228916.2017.1308199

Aslam, M., Kumar, S., & Sorooshian, S. (2020). Predicting likelihood for loan default among bank borrowers. *International Journal of Financial Research*, *11*(1), 318–328. https://doi.org/10.5430/ijfr.v11n1p318

Assfaw, A. M. (2021). Determinants Of Financial Performance Of Private Commercial Banks In Ethiopia. *Academy of Accounting and Financial Studies Journal*, *25*(Special Issue 4), 1–17.

Atsmegiorgis, C., Jong, K., & Kim, T. (2014). *Survival analysis of bank loan repayment rate for customers of Hawassa commercial bank of Ethiopaia*. *25*, 1591–1598.

Austin, P. C., & Fine, J. P. (2017). Practical recommendations for reporting Fine-Gray model analyses for competing risk data. *Statistics in Medicine*, *36*(27), 4391–4400. https://doi.org/10.1002/sim.7501

Austin, P. C., Lee, D. S., & Fine, J. P. (2016). Introduction to the Analysis of Survival Data in the Presence of Competing Risks. *Circulation*, *133*(6), 601–609. https://doi.org/10.1161/CIRCULATIONAHA.115.017719

Baidoo, S. T., Yusif, H., & Ayesu, E. K. (2020a). Improving loan repayment in Ghana : Does

financial literacy matter ? *Cogent Economics & Finance*, *8*(1), 1–20. https://doi.org/10.1080/23322039.2020.1787693

Baidoo, S. T., Yusif, H., & Ayesu, E. K. (2020b). Improving loan repayment in Ghana: Does financial literacy matter? *Cogent Economics and Finance*, *8*(1). https://doi.org/10.1080/23322039.2020.1787693

Banasik, J., Crook, J. N., & Thomas, L. C. (1999). Not if but when will borrowers default. *Journal of the Operational Research Society*, *50*(12), 1185–1190. https://doi.org/10.1057/palgrave.jors.2600851

Bellotti, T., & Crook, J. (2007). Credit scoring with macroeconomic variables using survival analysis. *49th Annual Conference of the Operational Research Society 2007, OR49, April*, 79–92.

Benjamin. (2017). Microcredit Loan Repayment Default among Small Scale Enterprises: A Double Hurdle Approach. *Ghana Journal of Development Studies*, *14*(1), 146. https://doi.org/10.4314/gjds.v14i1.8

Bhattarai, B. P. (2019). Effect of Credit Risk Management on Financial Performance of Commercial Banks in Kenya. *Strategic Journal of Business & Change Management*, *9*(4), 87–103. https://doi.org/10.61426/sjbcm.v9i4.2420

BIRTUKAN TADESSE, F. E. and H. M. (2014). *AN ASSESSMENT OF LENDING PRACTICE: THE CASE OF COMMERCIAL BANK OF ETHIOPIA, Head Office*. ST. MARY'S UNIVERSITY IN ADDIS ABEBA.

Caselli, S., Corbetta, G., Cucinelli, D., & Rossolini, M. (2021). A survival analysis of public guaranteed loans: Does financial intermediary matter? *Journal of Financial Stability*, *54*, 100880. https://doi.org/10.1016/j.jfs.2021.100880

Cox, D. R. (1972). Regression Models and Life-Tables Authors (s): D . R . Cox Source : Journal of the Royal Statistical Society . Series B (Methodological), Vol . 34 , No . 2 Published by : Wiley for the Royal Statistical Society Stable URL : http://www.jstor.org/stable. *Journal of the Royal Statistical Society.1*, *34*(2), 187–220.

DIRIBA, T. (2014). *DETERMINANTS OF COMMERCIAL BANKS PERFORMANCE IN ETHIOPIA; PERFORMANCE IN TERMS OF PROFITABILITY; PROSPECTS, CHALLENGES AND POLICY IMPLICATIONS*.

Dirick, L., Bellotti, T., Claeskens, G., & Baesens, B. (2019). Macro-Economic Factors in Credit Risk Calculations: Including Time-Varying Covariates in Mixture Cure Models. *Journal of Business and Economic Statistics*, *37*(1), 40–53. https://doi.org/10.1080/07350015.2016.1260471

Endris, A. K. and E. (2021). *Factors Affecting Loan Repayment Rate from ACSI among Smallholder Farmers: In the case of Habru District, Eastern Amara Regional State, Ethiopia*. https://doi.org/https://doi.org/10.21203/rs.3.rs-684555/v1

Enimu, S., Eyo, E. O., & Ajah, E. A. (2017). Determinants of loan repayment among agricultural microcredit finance group members in Delta state, Nigeria. *Financial Innovation*, *3*(1). https://doi.org/10.1186/s40854-017-0072-y

Fine, J. P., & Gray, R. J. (1999). A Proportional Hazards Model for the Subdistribution of a Competing Risk. *Journal of the American Statistical Association*, *94*(446), 496–509. https://doi.org/10.1080/01621459.1999.10474144

Garomsa, A. (2017). *Assessment of factors affecting loan repayment performance of borrowers an empirical study on selected microfinance institutions in oromia region*.

Gray, R. J. . (1988). A Class of K-Sample Tests for Comparing the Cumulative Incidence of a Competing Risk Author (s): Robert J . Gray Source : The Annals of Statistics , Vol . 16 , No . 3 , (Sep ., 1988), pp . 1141-1154 Published by : Institute of Mathematical Statistics. *Statistics*, *16*(3), 1141–1154.

Gupta, V. (2017). A Survival Approach to Prediction of Default Drivers for Indian Listed Companies. *Theoretical Economics Letters*, *07*(02), 116–138. https://doi.org/10.4236/tel.2017.72011

Haller, B. (2014). (2014). The analysis of competing risks data : A guided tour. *Statistics in Medicine*, *132*(4), 1–41.

http://www.biostat.umn.edu/~tracyb/ph7402/notes/CompetingRisksTutorial.pdf

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

Hassan, M. K., Brodmann, J., Rayfield, B., & Huda, M. (2018). Modeling credit risk in credit unions using survival analysis. *International Journal of Bank Marketing*, *36*(3), 482–495. https://doi.org/10.1108/IJBM-05-2017-0091

John, O. (2020). *Competing Risks of Default and Prepayment of Mortgage Market*.

Kariuki, M. W., & Ngahu, S. (2016). Effect of Interest Rates on Loan performance of Microfinance Institutions in Naivasha Sub-country, Kenya. *International Journal of Economics, Commerce and Management*, *IV*(4), 549–566.

Kotiso, M. S. (2018). Factors Affecting Default Risk of Commercial Banks : Evidence from Ethiopian Banking Industry. *Research Journal of Finance and Accounting*, *9*(1), 1–10.

Kuk, D., & Varadhan, R. (2013). Model selection in competing risks regression. *Statistics in Medicine*, *32*(18), 3077–3088. https://doi.org/10.1002/sim.5762

Leoce, N. M. (2016). *Prognostic Modeling in the Presence of Competing Risks: An Application to Cardiovascular and Cancer Mortality in Breast Cancer Survivors* (Issue May).

Marimo, M. (2015). *Survival Analysis of Bank Loans and Credit Risk Prognosis Master of Science Mathematical Statistics*. 1–26.

Mohammad, K. A., Fatima-Tuz-Zahura, M., & Bari, W. (2017). Fine and Gray competing risk regression model to study the cause-specific under-five child mortality in Bangladesh. *BMC International Health and Human Rights*, *17*(1), 1–8. https://doi.org/10.1186/s12914-017-0112-8

Mukono. (2015). *Determinants of Loan Repayment By Small and Medium Enterprises in Nairobi County, Kenya Ann Mukono a Research Project Submitted in Partial Fulfillment of the Requirements for the Award of the Degree of Master of Science in Finance, School of Business, Uni*. *November*.

NBE. (2018). *Licensing and Supervision of Banking Business. Asset Classification and Provisioning (5th Replacement).*

Pasha, S. A., & Negese, T. (2014). Performance of Loan Repayment Determinants in Ethiopian Micro Finance-An Analysis. *Eurasian Journal of Business and Economics*, *7*(13), 29–49. http://www.ejbe.org/EJBE2014Vol07No13p029PASHA-NEGESE.pdf

Pintilie, M. (2006). *competing risk: a practical perspective*.

Porta N. Gómez G., C. M. L., & Malats, N. (2007). Competing risks Methods. *Department of Statistics and Operations Research, Universitat Politècnica de Catalunya*, *Technical*(1). http://upcommons.upc.edu/e-prints/handle/2117/2201

Prentice, R. L., Kalbfleisch, J. D., Peterson, A. V., Flournoy, N., Farewell, V. T., & Breslow, N. E. (1978). The Analysis of Failure Times in the Presence of Competing Risks. *Biometrics*, *34*(4), 541. https://doi.org/10.2307/2530374

Rychnovský, M. (2018). SURVIVAL ANALYSIS AS A TOOL FOR BETTER PROBABILITY OF DEFAULT PREDICTION. *Acta Oeconomica Pragensia*, *2018*(1), 34–46. https://doi.org/10.18267/j.aop.594

Stepanova, M., & Thomas, L. (2002). Survival analysis methods for personal loan data. *Operations Research*, *50*(2), 277–289. https://doi.org/10.1287/opre.50.2.277.426

Thomas, L. C., Crook, J. N., Edelman, D. B., Mathematics, I. of, & Applications, I. (1992). *Credit scoring and credit control : based on the proceedings of a conference on credit scoring and credit control, organized by the Institute of Mathematics and Its Applications and held at the University of Edinburgh in August 1989*. Clarendon Press. https://cir.nii.ac.jp/crid/1130000793955658624.bib?lang=en

Wycinka, E. (2015). *TIME TO DEFAULT ANALYSIS IN PERSONAL CREDIT SCORING*. *January 2015*.

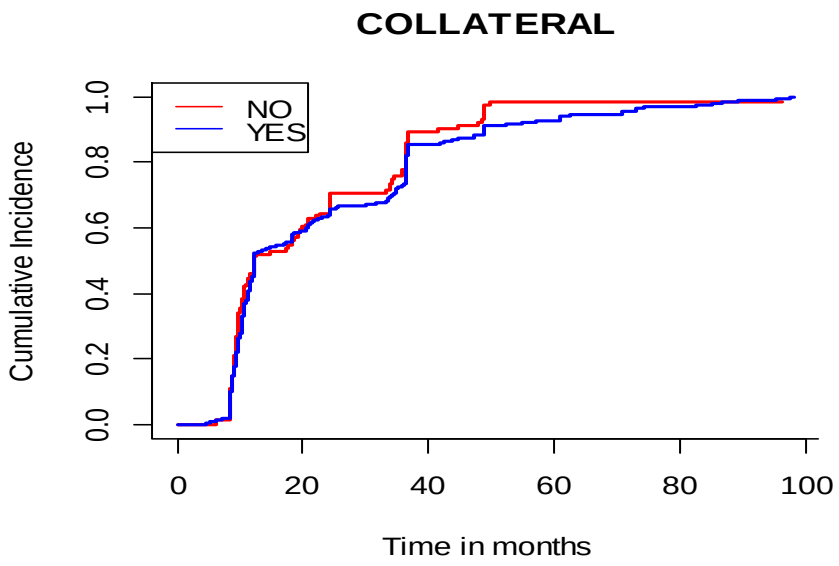
Wycinka, E. (2019). Competing risk models of default in the presence of early repayments. *Econometrics*, *23*(2), 99–120. https://doi.org/10.15611/eada.2019.2.07

Yibrie, O., & Ramakrishna, R. (2017). Determinants of Loan Repayment Performance in ACSI. *International Journal of Advanced Research in Management and Social Sciences*, *6*(4), 151–170.

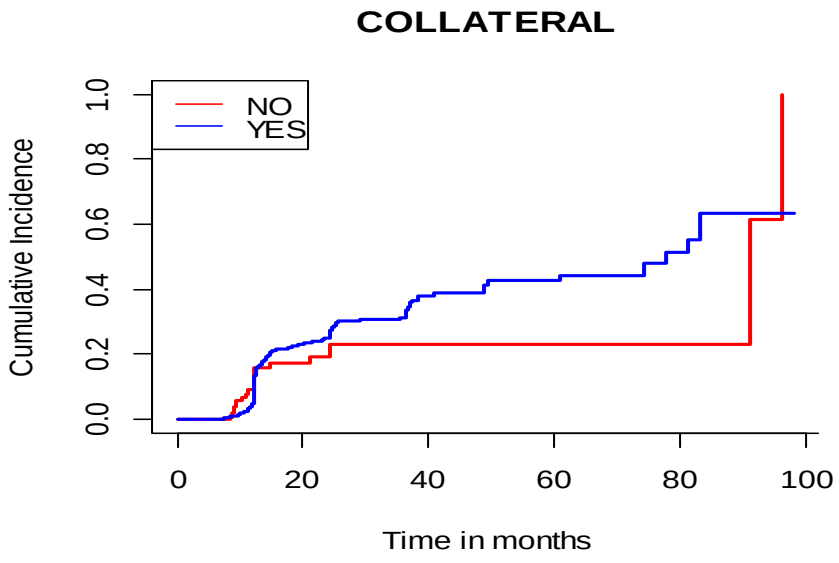
Zhang, Z. (2017). Survival analysis in the presence of competing risks. *Annals of Translational Medicine*, *5*(3). https://doi.org/10.21037/atm.2016.08.62

Appendix-A

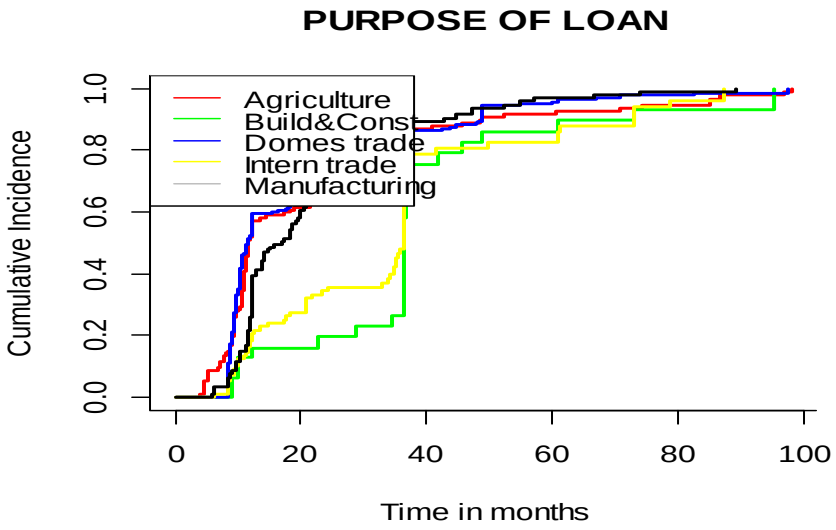
CIF plot of default for Collateral group



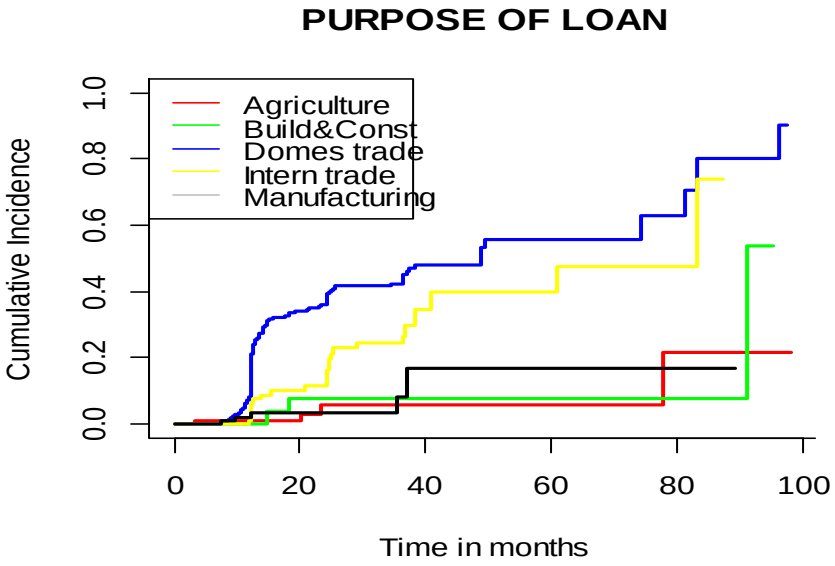
CIF plot of early repayment for Collateral group



CIF plot of default for by purpose of loan



CIF plot of early repayment for by purpose of loan



Appendix-B

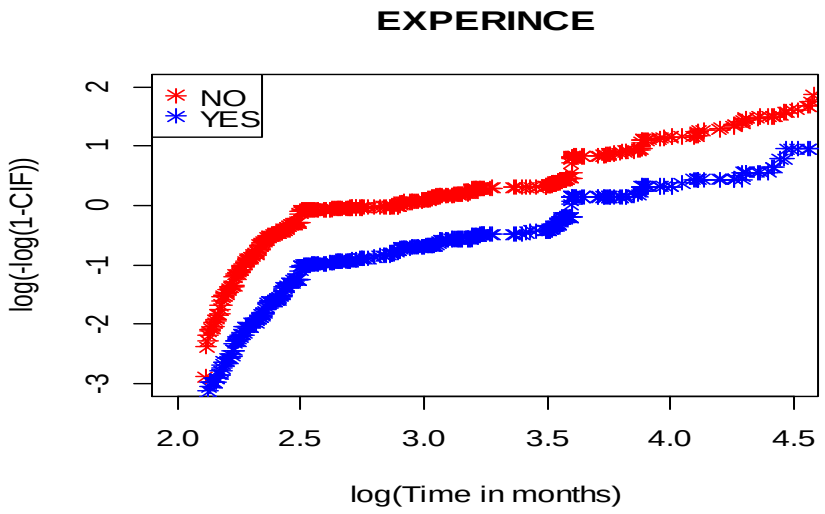


Figure 4.6 Plot of the Proportionality of the hazard of the CIF for experience in loan

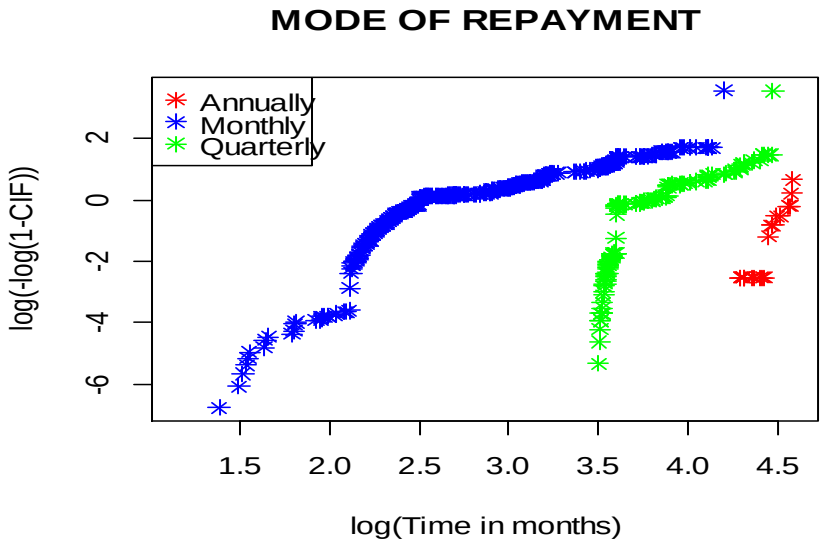


Figure4.7 Plot of the Proportionality of the hazard of the CIF for mode of repayment

PH assumption plot for purpose of loan group

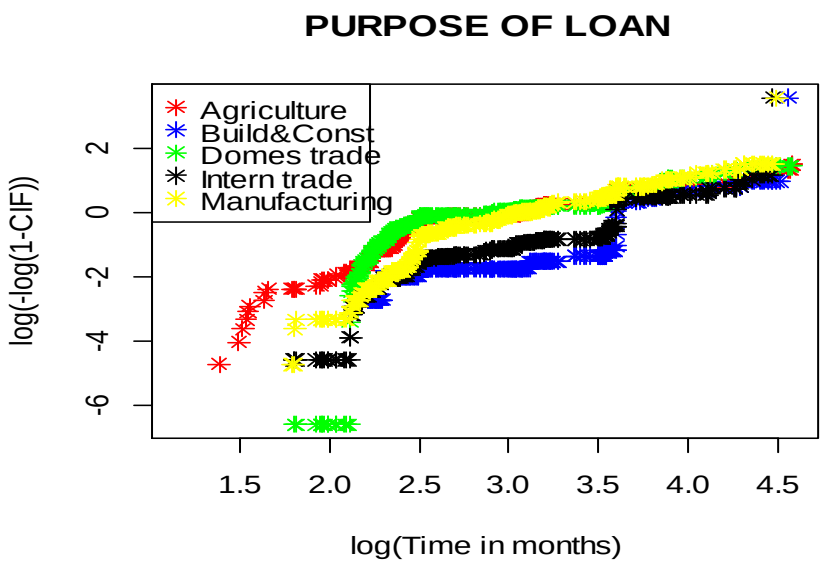


Figure 4.8 Plot of the Proportionality of the hazard of the CIF for purpose of loan