Journal of the Korean Data & Information Science Society
2014, **25**(6), 1591–1598

http://dx.doi.org/10.7465/jkdi.2014.25.6.1591
한국데이터정보과학회지

Survival analysis of bank loan repayment rate for customers of Hawassa commercial bank of Ethiopaia

Cheru Atsmegiorgis Kitabo[1] · Jong Tae Kim[2]

[12]Department of Statistics, Daegu University

Received 23 October 2014, revised 12 November 2014, accepted 17 November 2014

Abstract

The reviews of the balance sheet of commercial banks showed that loan item constitutes the largest portion of bank's assets. Although the sector has highest rate of profit, it possesses the greatest risk. Identifying factors that can contribute in lifting-up the loan repayment rate of customers of Hawassa district commercial bank is the major goal of this study. 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 repayment rate. The result from Kaplan-Meier survival estimation revealed that the loan repayment rate is significantly related with loan type, and previous loan experience, educational level and mode of repayment. The log-rank test indicates that the survival probability of loan customers is not statistically different in repaying the loan among groups classified by sex. Moreover, the univariate Cox proportional hazard model result portrayed 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 commercial bank. Hence, banks should design loan strategies giving special emphasis on the significant factors while they are giving loans to their customers.

Keywords: Kaplan-Meier estimation, log-rank, univariate Cox proportional hazard model.

1. Introduction

In a financial system, commercial banking is one of an integral part that plays great role for development of countries. The socio-economic and political development of any nation mainly depends on the quality and quantity of banks which they have. Giving loans to its customers is the principal economic function of commercial banks to fund consumption and investment spending by individuals, business and government units. How well a bank performs its lending function has a great deal to do with the economic health of its region because bank loans support the growth of new businesses and jobs within the bank's trade territory and promote economic vitality.

Almost half the revenue of most banks was collected from their loan account (Rose, 2009). In contrary to this, threat in banking tends to be concentrated in loan portfolio. When a bank

[1] Ph. D student, Department of Statistics, Daegu University, Kyungbook 712-714, Korea.
[2] Corresponding author: Professor, Department of Statistics, Daegu University, Kyungbook 712-714, Korea. E-mail: jtkim@daegu.ac.kr

gets into serious financial trouble, its problem usually springs from loans that have become uncollectible due to mismanagement, illegal manipulation of loans, misguided lending policies or an unexpected economic recession. The current global financial crisis can be the best example to explain the risks of loan system.

Hawassa city is one of the fast growing city in Ethiopia. It is one of the capital cities of the southern region. There are a number of economic activities that have been undertaken in Hawassa city and the nearby areas. All southern regional and zonal head bureaus are located in Hawassa city. Besides, Hawassa city accommodates a number of regional bureaus for international nongovernmental organizations, civic societies, financial institutions and multinational companies. Many private businesses mainly focusing on hotels and tourism, coffee export business, factory product exporters like flour and textiles are actively involving in the economic activity of the country. The huge economic activities undertaking in the city demand loans from banks currently working in the city.

The main objective of this study is to identify the factors that affect the customers' repayments of commercial banks in Ethiopia Hawassa Branch. For this particular study, secondary data collected from the Hawassa district commercial bank of Ethiopia was used to identify factors that affect the loan repayment rate of customers in the branch. A data set actually contains socio-demographic characteristics of the lenders and loan related behavior variables.

Nafiu (2005), Onyiriuba (2004), Berger and Udell (2005), Harhoff and Korting (2008), Nwankwo (2001), and Park (2011) articulated the following causes of loan defaults to include among others: asset value of the unit requesting the loan, lack of supervision, lack of business experience, unrealistic collateral value, business revenue, and outstanding debts of the client. Others include annual turnover; number of years of banking with the customer, lending rate, size loan, maturity period of loan; insider abuse and use of distorted financial statement as a credit analysis base.

It is important that borrowed funds be invested for productive purposes, and the additional incomes generated be used to repay loans to have sustainable and viable production processes and credit institutions. However, failure by businesses to repay their loans on time or to repay them at all has been a serious problem faced.Businesses found in developing countries like agricultural credit administration and especially to businesses who have limited collateral capabilities, poor loan repayment has become a major problem (Okorie *et al.*, 2007).

2. Methodology

2.1. Description of study area and population

This survey utilized data collected in Hawassa district Commercial bank of Ethiopia which has more than 30 sub-branches in the southern Ethiopia. Actually in this study, all customers of this bank who took loan for different purposes were the target population of the survey. The loan could be for Domestic Trade Services (DTS), Manufacturing (M) and Building and Construction (BC). The target population are all loan customers of the branch.

2.2. Study design

A cross-sectional study design was employed which mainly based on the data collected in the bank about the loans of customers and some related demographic and socio-economic

data characteristics of study participants at a certain cross section. Regarding the sampling techniques, the study participants were selected from the dataset using simple random sampling.

In order to have an optimum sample size, there are a number of points one has to take into account. Some of the issues are objective of the research, design of the research, cost constraint, degree of precision required for generalization and etc. Based on the above information, there are several formulas developed for sample size calculation that conform to different research situations. Accordingly, the following sample size determination formula (Cochran, 1977) is adopted for this study.

$$n = \frac{z^2_{\alpha/2} p(1-p)}{d^2},$$

where n = the sample size needed, $z_{\alpha/2}$ is the upper $\alpha/2$-th quantile of standard normal distribution with $\alpha = 0.05$ significance level. The maximum allowable difference between the maximum likelihood estimate and the unknown population parameter, denoted by d, desired to be 0.021. The parameter p represents the proportion of customers who paid their loan in the stated period. According to the study done in Nigeria by Dermini (2006), among all customers who took loan, 80% repay the loan in the agreed time ($p = 0.8$). Using these values, the optimum sample size required for this study was found to be 183.

2.3. Variables in the study

The response (dependent) variable is continuous and describes the length of time to repay the loan in days. The predictor (covariate) variables which are assumed to influence the survival of loan repayment included in the model are sex (female, male), age of customer, educational status, marital status, purpose of loan, experience of taking loan, type of loan, repayment date, mode of repayment and type of collateral.

2.4. Method of statistical analysis

2.4.1. Kaplan-Meier estimator

An initial step in the analysis of a set of survival data is to present numerical or graphical description of the data for individuals in a particular group. This description includes survival distribution and Kaplan-Meier survival function estimation which are used for the estimation of the distribution of survival time from all of the observations available. The Kaplan-Meier estimator of the survivorship function (or survival probability), $S(t) = P(T \geq t)$, is defined as:

$$S(t) = \Pi^t_{t_j<t} \frac{n_j - d_j}{nj} = \Pi^t_{t_j<t} \left\{ 1 - \frac{d_j}{n_j} \right\},$$

where d_j is the number of individuals who experience the event at time t_j, and n_j is the number of individuals who have not yet experienced the event at that time.

2.4.2. Comparison of survivorship functions

Under the null hypothesis that the two survivorship functions are the same, and assuming that the censoring experience is independent of group, and that the total number of observed events and the sum of the expected number of events is large, Q follows a chi-square

distribution with one degree of freedom. We can use the above test to compare k groups. In this study we use the log-rank test and generalized Wilcoxon test which are special cases of Q. The general form of this test statistic is given by:

$$Q = \frac{(\Sigma_{i=1}^{m} w_i(d_{1i} - e_{1i}))^2}{\Sigma_{i=1}^{m} w_i^2 v_{1i}},$$

where $e_{1i} = n_{1i} \times d_i/n_i, v_{1i} = n_{1i} \times n_{0i}(n_i - d_i)/\{n_i^2(n_i - 1)\}, n_{0i}$ is the number at risk at observed survival time t_i in group 0, n_{1i} is the number at risk at observed survival time in t_i the group 1, d_{0i} is the number of observed repays in group 0, d_{1i} is the number of customers who repay their loan in group 1, n_i is the total number of individuals or risk before time t_i, and d_i is the total number of repays at t_i. The log-rank test, sometimes called the Cox-Mantel test, is the most well known and widely used test statistic for comparing the survival curves of two treatment groups. This test is based on weights equal to one, i.e. $w_i = 1$. Therefore, the log-rank test statistic becomes:

$$Q_{LR} = \frac{\Sigma_{i=1}^{m}(d_{1i} - e_{1i})^2}{\Sigma_{i=1}^{m} v_{1i}}.$$

2.4.3. Regression models for survival data

The Cox proportional hazards model

This model was proposed by Cox (1972) and has come to known as the Cox regression model. The model is therefore referred to as a semi-parametric model. Semi-parametric models are models that parametrically specify the functional relationship between the lifetime of an individual and its characteristics but leave the actual distribution of lifetimes arbitrary.

The hazard function:

Cox proportional hazard model is usually written in terms of the hazard model formula. This model gives an expression for the hazard at time t for an individual with a given specification of a set of explanatory variables denoted by X and it is generally given by:

$$h(t, x_i, \beta) = h_0(t) \exp(\beta x_i),$$

where $h_0(t)$ is the baseline hazard function that characterizes how the hazard function changes as a function of survival time X_i is the vector of values of the explanatory variables for the i-th individual at time t and β is the vector of unknown regression parameters that are assumed to be the same for all individuals in the study, which measures the influence of the covariate on the survival experience.

3. Results and discussion

3.1. Descriptive statistics

A total of 183 participants who borrowed loan from Hawassa district commercial bank of Ethiopia were included in this study. Among all the respondents involved in this study, about 80% were in the age range of $35 - 55$ years old. In contrast, there is only one respondent whose age was more than 65 years old. Regarding lenders' educational status, more than

three fourth 76.5% of the customers were 12 grade complete whereas those with diploma and more constitute only 11.4%. With respect to marital status, the dataset collected from the bank has only two categories of married and single. About 93% of the bank customers were married people, whereas 7.2% of the respondents had single marital status.

Table 3.1 The Socio-demographic characteristics of study participants

Variables	Frequency	Percentage
Age		
25-34	18	9.8
35-45	93	50.8
46-55	51	27.8
56-65	20	10.9
≥66	1	0.5
Total	183	100
Educational Status		
8 grade	22	12
9-12 grade	140	76.5
Diploma	8	4.3
Degree and above	13	7.1
Total	183	100
Marital Status		
Married	170	92.8
Single	13	7.2
Total	183	100

3.2. Comparison of survival experience

From the log-rank test (Mantel-Cox test) given in Table 3.2, it can be seen that there is no significant loan repayment rate difference among the levels of the risk factor sex between male and female. In contrary, significant differences were observed among the levels of the risk variables educational level, previous experience, mode of repayment and type of loan. Moreover, the Kaplan- Meier survivor estimates analysis was performed using the survival function curve.

Table 3.2 Comparison of survival experience among levels of demographic, loan related risk behavior variables

		Log-rank test		
Variables	Mean Survival Time	Chi-Square	df	Sig.
Sex		3.37	1	0.066
Female	189.375			
Male	191.14			
Educational status		13.47	2	0.005
8th	143.765			
Grade 9-12	142.981			
Higher education	125.416			
Previous loan experience		25.93	1	0.004
Yes	197.412			
No	168.347			
Mode of repayment		47.9	1	0.038
Monthly	17.758			
Quarterly	57.231			
Semi-annually	174.902			
Annually	297.079			
Types of loan		17.54	1	0.016
Building and Construction	189.137			
Manufacturing	177.915			
Domestic Trade Service	168.167			

3.3. Survival curve analysis of loan repayment

Figure 3.1 depicted that people with huge experience in borrowing have the lower chance of repaying the loan in all granted months considered in this study and this difference is significant (p-value= 0.004) at 5% level of significance.

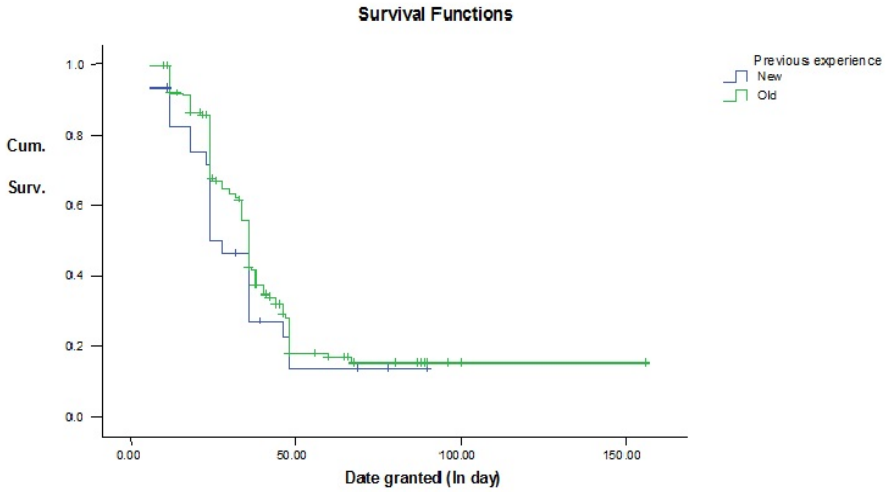

Figure 3.1 Survival function showing loan repayment of customers with experience and no experience

3.4. Survival function showing loan repayment of customers among different modes of repayment

The survival function Figure 3.2 below portrayed that people who took loan to repay yearly basis have higher likelihood to repay the loan in the stated dates by the bank. In contrary to this, those people who borrowed money to repay in quarter (Q) time had very small likelihood to repay the loan to the bank. The analysis showed that this difference in the likelihood of repaying the loan among the different mode of repayment was significant with 5% level of significance.

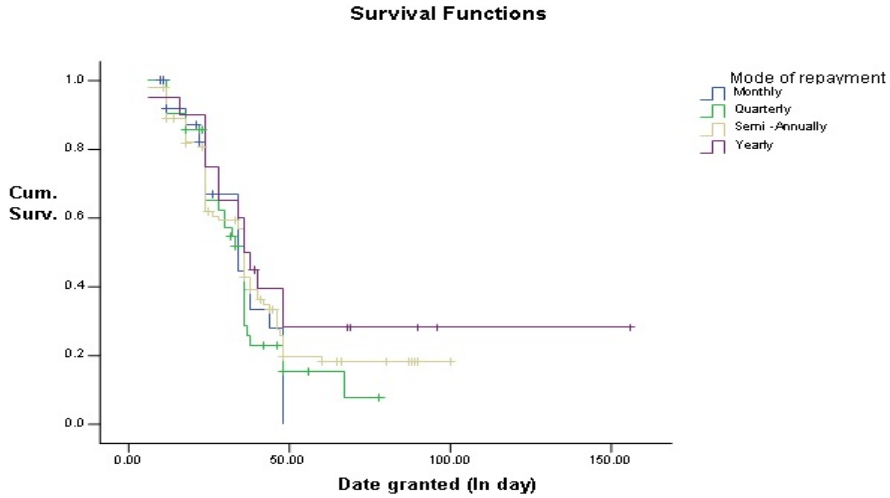

Figure 3.2 Survival function showing loan repayment of customers among different modes of repayment

Table 3.3 Univariate Cox proportional hazards analysis on the time to repay the loan

Variables	β	SE	Wald	df	Sig.	HR	95% CI for HR
Sex							
Female (**Ref**)			8.35	1	0.073		
Male	1.054	0.139	7.56	1	0.061	0.89	(0.34, 1.67)
Educational Status							
Less than grade 8 (**Ref**)			19.87	3	0		
Grade 8-12	2.294	0.444	5.16	1	0	0.657	(0.55, 3.23)
Diploma	2.01	0.486	4.13	1	0.045	0.342	(0.13, 2.97)
Degree and above	1.067	0.5	2.13	1	0.067	0.967	(-0.12, 1.87)
Marital Status							
Non-Married (**Ref**)			7.44	1	0.054		
Married	0.57	0.009	63.33	1	0	0.634	(0.19, 0.79)
Loan purpose							
To purchase car (**Ref**)			34.11	2	0.009		
To purchase house	-2.78	1.34	-2.07	1	0.357	1.23	(1.11, 3.78)
For working	5.79	3.69	1.56	1	0.851	2.99	(2.74, 7.92)
Previous experience							
No (**Ref**)			19.91	1	0.006		
Yes	59.11	9.44	6.26	1	0.049	18.15	(4.66, 27.44)
Collateral type							
Vehicle (**Ref**)			109.11	1	0		
Building	-2.15	0.42	34.38	1	0	1.081	(0.04, 0.19)
Site	2.29	0.44	26.73	1	0	9.993	(4.15, 23.7)
Commerce	0.06	0.01	6	1	0	1.006	(1.04, 1.21)
Mode of Repayment							
Monthly (**Ref**)			256.11	1	0.023		
Quarterly	-2.18	1.28	1.7	1	0.563	1.567	(1.49,1.88)
Semi-annually	44.12	21.44	2.05	1	0.231	13.441	(8.19, 16.12)
Annually	13.21	2.1	6.29	1	0.031	19.121	(13.22, 20.13)

3.5. Single covariate analysis

Single covariate Cox proportional hazards model analysis is an appropriate procedure that is used to screen out potentially important variables that can significantly affect the Hazard of loan repayment. The relationship between each covariates and survival time of loan repayment are presented in Table 3.3 below. As can be seen from the Table 3.3, the variables such as, educational level, loan purpose, previous experience of loan, collateral type and mode of repayment are significantly affecting the survival of loan repayment. But the covariates sex and marital status are not significantly related with the survival of loan repayment at 5% significant level.

4. Concluding remarks

This research aimed to identify the major factors that can affect the loan repayment rate of Hawassa district commercial bank customers. Based on the analysis output, majority (78.6%) of the customers were in the age range of 35 to 55. Regarding education status of customers, more than three fourth were high school graduates. The bank provides different type of loans for its customers. Greater proportion of customers took their loan for domestic trade services whereas few customers took the loan for manufacturing products. Based on purpose of loan classification, greater proportion of the loan was taken for working purpose. In contrast to this, very few customers took loan for purchasing vehicle and machines. From univariate Cox proportional hazard analysis, we can conclude that rate of loan repayment is significantly related with educational level, having previous loan experience, mode of repayment, collateral type and purpose of loan. In contrary to the above findings, the predictor variables such as sex and marital status are not significantly related to survival experience of loan repayment rate of customers.

References

Berger, A. N. and Udell, G. F. (2005). *Relationship of lending and lines of credit in small firms finance*, Ph.D Thesis, Federal University of Technology, Owerri.

Cochran, W. G. (1977). *Sampling techniques*, John Wiley & Sons, New York.

Harhofff, D. and Korting, T. (2008). Lending relationship in Germany-empirical evidence from survey data. *Journal of Banking and Finance*, **22**, 1317-1353.

Nafiu, B. A. (2005). Loan defaults and consequences for bank performance. *Journal of Nigeria Deposit Insurance Corporation*, **5**, 2-4.

Nwankwo, B. A. (2001). *Determinants of repayment behaviour in the centenary rural development bank in Uganda,* Ph.D Thesis, Ohio State University, OH.

Okorie, J. T. O., Adeyemo, R. and Agbonlahor, M.U. (2007). An Empirical analysis of microcredit repayment in southwestern Nigeria. *Journal of Humanity & Social Sciences*, **2**, 63-74.

Onyiriuba, A. (2004). Major determinant of agricultural smallholder loan lepayment in developing economy: Empirical evidence from Ondo state Nigeria. *Journal of Socio-Economics*, **21**, 223-234.

Park, B. J. (2011). Study on time -varying herd behavior in individual stock. *Journal of the Korean Data & Informaion Science Society*, **22**, 423-426.

Rose, P. S. (2009). *Commercial bank managemnet*, 4th Ed., Irwin & McGraw-Hill, Boston.