

Modelling Credit Risk for Personal Loans Using Product-Limit Estimator

Okumu Argan Wekesa (Corresponding author)

Jomo Kenyatta University of Agriculture and Technology

Department of Actuarial Science and Statistics

P.O BOX 62000 - 00200 Nairobi Kenya

Tel: +254-722-836-593 E-mail: arganokumu@yahoo.com

Mwalili Samuel

Jomo Kenyatta University of Agriculture and Technology

Department of Actuarial Science and Statistics

P.O BOX 62000 - 00200 Nairobi Kenya

Tel: +254-728-327-109 E-mail: mwalili_2000@yahoo.com

Mwita Peter

Jomo Kenyatta University of Agriculture and Technology

Department of Actuarial Science and Statistics

P.O BOX 62000 - 00200 Nairobi Kenya

Tel: +254-721-598-901 E-mail: petermwita@fsc.jkuat.ac.ke

Received: March 18, 2011

Accepted: June 8, 2011

Published: January 5, 2012

doi:10.5430/ijfr.v3n1p22

URL: <http://dx.doi.org/10.5430/ijfr.v3n1p22>

Abstract

A product-limit approach was adopted to estimate time to default for male and female loan applicants. For each group, a sample of 250 applicants was observed for a 30 months. The life of the account is measured from the month it was opened until the account becomes 'bad' or it is closed or until the end of observation. The account is considered bad if payment is not made for two consecutive months in line with the industry practice. If the account does not miss two payments and is closed or survives beyond the observation period, it is considered to be censored. The results showed that there is no significant difference between male and female applicants in terms of their survival times and hazard rates.

Keywords: Survival Analysis, Product – Limit Estimator, Default Risk

1. Introduction

Traditional credit risk models aim at determining a customer's probability of defaulting on loan repayment. This study used survival analysis method which draws its origin from the study of time to death or occurrence of any other event on life data. The technique has also been applied in engineering to model failure time of components and parts. In credit risk modelling, the event of interest is default, thus modelling time to default on loan obligations.

Modelling of credit risk using survival analysis was first introduced by Narain (1992). Thomas et.al (1999) further developed the model. Narain (1992) applied the survival model on 24 months of loan data. The result showed that the survival analysis approach provides more detailed and relevant information for credit management than the conventional approaches. Thomas et.al (1999) applied the technique by using the accelerated life exponential model to 24 months loan data. The results also showed the approach to be superior to conventional credit scoring methods in that a better credit-granting decision could be made if the score was supported by the estimated survival times. The research by Thomas et.al (1999) also did a comparison of exponential, Weibul and Cox non-parametric models with logistic regression and they concluded that survival method was a better modelling tool.

In the literature, a number of techniques have been applied to model credit risk. Orgler (1970) applied regression analysis in a model for commercial loans. Wiginton (1980) was one of the first to publish credit scoring results using logistic regression. It was compared with discriminant analysis. Leonard (1993) also applied logistic regression in evaluating commercial loans. Durand (1941) pioneered the use of discriminant analysis for credit scoring. Decision tree and rule was adopted by Murkowski (1985) and Mehta (1968) for credit scoring.

Other techniques include K-Nearest Neighbour Classifiers which was used by Chatterjee and Barcun (1970) and Henley and Hand (1996). Baensens et. al (2003) studied the use of Bayesian network classifiers to rate borrowers. Linear programming was applied by Hardy and Adrian (1985) and compared it with other statistical approaches.

1.1 Descriptive Methods of Time-to-event

Survival analysis is a statistical method for modelling the time to some events for a population of individuals. For example, events may refer to death in medical application, or recidivism of released prisoners in criminology application, or first bought of a new product by customer in marketing studies. The time to the occurrence is termed as survival time or lifetime. In application to credit risk modelling, the events refer to default of a loan and therefore its lifetime refers to time-to-default T .

Default times are subject to random variation and are thus random variables. To describe their randomness, there are five standard ways:

$$\text{Density function (PDF), } f_{20}(t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr(t \leq T \leq t + \Delta t)}{\Delta t}$$

$$\text{Distribution function (CDF), } F_{20}(t) = \int_0^t f_{20}(u) du = \Pr(T \leq t)$$

$$\text{Survivor function, } S_{20}(t) = \int_t^{\infty} f_{20}(u) du = 1 - F_{20}(t) = \Pr(T \geq t)$$

$$\begin{aligned} \text{Hazard function } h_{20}(t) &= \lim_{\Delta t \rightarrow 0} \frac{\Pr(t \leq T \leq t + \Delta t \mid T \geq t)}{\Delta t} \\ &= \frac{f_{20}(t)}{S_{20}(t)} = \frac{-d \ln S_{20}(t)}{dt} \end{aligned}$$

$$\text{Cumulative hazard function } H_{20}(t) = \int_0^t h_{20}(u) du = -\ln S_{20}(t)$$

These five formulations are mathematically equivalent but they highlight different aspects of the default time. The distribution function tells us the probability that default occurs at or before time t . Conversely, survivor function is the probability that default does not occur at or before time t ; in other words, the loan survives (non-default), at least, to time t . The interpretation of hazard function is slightly tricky. It is the “rate” that borrower defaults at time t , conditional on his staying on the books up to that time. Note that hazard is not a probability and thus can be greater than one.

In survival analysis, one must consider a key analytical problem called censoring. In essence, censoring occurs when we have some information about an individual’s survival time, but do not know the exact survival time. There are a number of types of censoring, such as random, interval, left, and right censoring. In credit scoring application, most of the cases are right censoring.

For example, suppose we follow a group of borrowers for 3 years. If we observe borrower A fails to repay at 15th month, he is certainly classified as a default case and his default time is 15. On the other hand, consider borrower B, who repays on time during the whole observed period. We do not know his exact default time but are sure that it must be greater than 36. For such case, borrower B is known as a right censored observation. Another example of right censoring could be when borrower C repays on time from the 1st month to the 12th month. At the 12th month, we do not have future repayment pattern of borrower C. As borrower B, we do not know the exact default time of borrower C, we only know that it must be greater than 12. This is also a right censoring example.

1.2 Statement of the Problem

Since the year 2003, the Kenyan financial market has experienced growing liquidity, which has caused banks to rigorously market various loan products. This has given rise to the need to review the banks’ credit granting criteria to reflect the growing volume of loan portfolio and to respond to the current global credit crunch. However, research on credit risk has

surprisingly received insignificant attention from both practitioners and scholars in Kenya and the larger African continent. Over the years, banks have perpetually used traditional credit scoring techniques to rate loan applicants.

A number of studies have been carried out on the issue of credit risk modelling using different approaches. A limited number of studies have applied survival analysis techniques but none has used product-limit method to analyse credit risk. To this end, the research intended to model probability of servicing loans and hazard rates for both male and female borrowers using this method. Furthermore, existing credit scoring models classify borrowers into different risk categories but cannot provide any information on when the borrower is likely to default. It is more informative for the lender not only to know the probability of defaulting but also when the default is likely to happen. This helps to fairly price risks and improve the focus on ultimate profitability. For instance if the lender knows that a group of loan applicants are bad type, instead of rejecting their applications, it may grant loans to them at higher interest rate, as long as the term of the loan is shorter than the likely time to default. Thus some “bad” applicants can also be viewed as profitable propositions.

1.3 General Objective

The broad objective of this research was to use product- limit survival model to generate default probabilities at various points in time. The study also intended to perform a test of equality of the two risk groups, namely male and female applicants.

1.4 Specific Objectives

The specific objectives included:

1. To estimate time to default using product-limit estimator for each risk group.
2. To determine hazard rate for each risk group on the basis of product-limit estimator.
3. To test the statistical significance of the differences in the survival curves for each risk group based on log-rank tests.

2. Methodology

This study was carried on data provided by a leading commercial bank in Kenya. The loan borrowers included in the study were randomly picked from the bank’s databank comprising 70 branches. The sample was based on a portfolio of personal loans whose maturity was 30 months. Thus the study cohort included loans taken in the month of January, 2007. The sample was constituted by 250 male applicants and 250 female applicants. Window of observation was 30 months (January 1, 2007 to June 30, 2010). In the context of this study number censored included those who made early settlement.

Number who made early loan settlements:

Males: 19 customers

Females: 8 customers

Number who defaulted:

Males: 8

Females: 12

Times in (months) at which borrowers made early settlement of their loan accounts or defaulted were follows:

Early settlements: Males: 3,3, 6,6,6,6,8,9, 12,12,12,15,15,18,18,18,18,24,25

Females: 4, 4,12,18,18,20,20,22

Defaults: Males: 4, 5, 5,7,10,10,13,16,

Females: 2,3,3,7,11,11,16,21,21,25,25,27

The product limit estimator was used to assess the gender based survival for personal loan borrowers. The data was analysed using SPSS version 10.

2.1 The Product - Limit Estimator

This function estimates survival rates and hazard from data that may be incomplete.

The survival rate is expressed as the survivor function (S):

$$S(t) = \frac{\text{number of individuals surviving longer than } t}{\text{total number of individuals studied}}$$

- where t is a time period known as the survival time, time to failure or time to event (such as death); e.g. 5 years in the context of 5 year survival rates. Some texts present S as the estimated probability of surviving to time t for those alive just before t multiplied by the proportion of subjects surviving to t .

This is univariate method which generates the characteristic “stair step” survival curves. It also called Kaplan-Meier estimator. The survival curves for the risk groups were compared using log-rank test.

Hypothesis test:

H_0 : the curves are statistically different

H_1 : the curves are statistically the same.

2.3 Model Derivation

Suppose $t_{(1)} < t_{(2)} < \dots < t_{(r)}$ are the ordered failure times. Let n_j denote the number of individuals alive (at risk of failure) just before time $t_{(j)}$, including those who will fail at time $t_{(j)}$. If an observation is censored at the same time $t_{(j)}$, that one or more failures occurs, the censoring is assumed to occur after any failures and n_j includes the censored observations. Let d_j denote the number of failures at time $t_{(j)}$. The conditional probability that an individual fails in the time interval from $t_{(j)} - \Delta$ to $t_{(j)}$, given survival upto time $t_{(j)} - \Delta$, is estimated as

$$\frac{d_j}{n_j}$$

The conditional probability that an individual survives beyond $t_{(j)} - \Delta$, given survival upto time $t_{(j)} - \Delta$, is estimated as

$$\frac{n_j - d_j}{n_j}$$

In the limit as $\Delta \rightarrow 0$,

$$\frac{n_j - d_j}{n_j}$$

becomes an estimate of the conditional probability of surviving beyond $t_{(j)}$ given survival upto $t_{(j)}$.

For $t_{(k)} \leq t < t_{(k+1)}$, the probability of surviving beyond time t is

$$\begin{aligned} S(t) &= P\{T > t\} = P\{T > t \text{ and } T > t_{(k)}\} \\ &= P\{T > t \mid T > t_{(k)}\} \cdot P\{T > t_{(k)}\} \\ &= P\{T > t \mid T > t_{(k)}\} \cdot P\{T > t_{(k)} \mid T > t_{(k-1)}\} \cdot P\{T > t_{(k-1)}\} \\ &= P\{T > t \mid T > t_{(k)}\} \cdot P\{T > t_{(k)} \mid T > t_{(k-1)}\} \cdot P\{T > t_{(k-1)} \mid T > t_{(k-2)}\} \cdots \\ &\quad \cdots P\{T > t_{(1)} \mid T > t_{(0)}\} \cdot P\{T > t_{(0)}\} \\ &\approx \prod_{j=1}^k P\{T > t_j \mid T > t_{(j-1)}\} \end{aligned}$$

Where $t_{(0)} = 0$ and $t_{(r+1)} = \infty$.

The Product-limit estimator of the survivor function at time t for $t_{(k)} \leq t < t_{(k+1)}$ is

$$\hat{S}(t) = \prod_{j=1}^k \frac{n_j - d_j}{n_j}$$

Also

$$\hat{S}(t) = \prod_{t_j \leq t} \left(\frac{n_j - d_j}{n_j} \right) = \prod_{t_j \leq t} \left(1 - \frac{d_j}{n_j} \right)$$

Variance of the Product-limit estimator (Greenwood's formula)

For $t_{(k)} \leq t < t_{(k+1)}$,

$$\bar{Var}(\hat{S}(t)) = (S(t))^2 \sum_{j=1}^k \frac{d_j}{n_j(n_j - d_j)}$$

Derivation

Recall

$$\begin{aligned} \hat{S}(t) &= \prod_{j=1}^k \frac{n_j - d_j}{n_j} \\ \log(\hat{S}(t)) &= \log\left(\prod_{j=1}^k \frac{n_j - d_j}{n_j}\right) \\ &= \sum_{j=1}^k \log((n_j - d_j)/n_j) \\ &= \sum_{j=1}^k \log(p_j) \\ \text{Var}(\log(\hat{S}(t))) &= \text{Var}\left(\sum_{j=1}^k \log(p_j)\right) \\ &= \sum_{j=1}^k \text{Var}(\log(p_j)) \end{aligned}$$

Applying delta estimation technique,

$$\begin{aligned} \text{Var}(\log(p_j)) &\approx \left(\frac{1}{\pi_j}\right)^2 \frac{\pi_j(1-\pi_j)}{n_j} \\ &= \left(\frac{1}{\pi_j}\right) \frac{1-\pi_j}{n_j} \end{aligned}$$

Thus,

$$\begin{aligned} \text{Var}(\log(\hat{S}(t))) &\approx \sum_{j=1}^k \left(\frac{1}{\pi_j}\right) \frac{1-\pi_j}{n_j} \\ \Rightarrow \text{Var}(\hat{S}(t)) &\approx [S(t)]^2 \text{Var}(\log(S(t))) \\ &= [\hat{S}(t)]^2 \sum_{j=1}^k \left(\frac{1}{\pi_j}\right) \frac{1-\pi_j}{n_j} \end{aligned}$$

Substituting $p_j = (n_j - d_j) / n_j$ for π_j , is

$$\bar{Var}(\hat{S}(t)) = (S(t))^2 \sum_{j=1}^k \frac{d_j}{n_j(n_j - d_j)}$$

Standard error of estimation (for large sample) is given by

$$se(\hat{S}(t)) = S(t) \sqrt{\sum_{j=1}^k \frac{d_j}{n_j(n_j - d_j)}}$$

Hazard function estimation for $t_{(j)} < t < t_{(j+1)}$ is given by

$$\hat{h}(t) = \frac{d_j}{n_j(t_{(j+1)} - t_{(j)})}$$

Hazard function estimation for Product-limit estimator

For

$$\begin{aligned} t_{(j)} &< t < t_{(j+1)} \\ \hat{h}(t) &= \frac{d_j}{n_j(t_{(j+1)} - t_{(j)})} \end{aligned}$$

Notation

d_j = number of defaults at time t_j which can take values 0,1,2,...,n

So d_j is a binomial distribution with parameters;

p_j = the probability of defaulting at time t_j

n_j = number of borrowers at risk just before time t_j

In terms of this study, the data set can be represented as follows:

<Table 1 about here>

<Table 2 about here>

4. Results

4.1 The Research Findings

The model outputs were as follows:

For gender factor 0 (female):

Number of Cases: 13 Censored: 6 (46.15%) Events: 7

<Table 3 about here>

<Table 4 about here>

For gender factor 1(males):

Number of Cases: 15 Censored: 4 (26.67%) Events: 11

<Table 5 about here>

<Table 6 about here>

Gender 1 refers to male borrowers and 0 refers to female borrowers. Event 1 denotes loan default and 0 denotes censored state.

The survival data output on gender 1 implies that out of the 250 male loan applicants for loans maturing in 30 months, 11 defaulted and 4 settled their loan accounts before maturity. Mean survival time of 15 means that on average, a male applicant will take 15 months to default. Same interpretation can be attributed to data on female applicants. The following summary was also generated:

<Table 7 about here>

<Table 8 about here>

By observation of survival curves, it can be seen that the two curves are similar. This is confirmed by the test statistic (log rank 0.17), which shows that the two survival distributions are statistically the same for 95% confidence interval. Thus it is not meaningful to classify borrowers on the basis of gender since it does not affect credit risk.

<Figure 1 about here>

<Figure 2 about here>

<Figure 3 about here>

The survival curves generated also give the same indication that there is no significant difference in the survival curves for male and female borrowers.

5. Conclusions and Recommendations

5.1 Conclusions

The research findings show that there is no significant difference between male and female borrowers in terms of their time to default on loan obligations. This implies that gender does not affect credit risk. Mean survival times would guide credit granting process on the average maturity for loans that may minimize on default losses and optimize profitability.

5.2 Recommendations and Suggestions for Further Research

This method of credit risk modelling is quite reliable as it does not make assumptions about loan default distribution unlike parametric methods. However, given that product-limit is a univariate method, it may be more informative to adopt multivariate techniques like Cox model to model credit risk. Thus further research can be conducted on the same data set using other survival techniques.

References

- Agresti, A. (2002). *Categorical Data Analysis*. Wiley, New York. <http://dx.doi.org/10.1002/0471249688>
- Allison, P. D. (1999). *Logistic Regression Using SAS System : Theory and Application*. SAS Institute Inc., Cary, NC.
- Baesens, B. T., Van Gestel, M., Stepanova, D. & Van den Poel, J. V. (2005). Neural network survival analysis for personal loan data. *Journal of the Operational Research Society* 56 1089-1098. <http://dx.doi.org/10.1057/palgrave.jors.2601990>
- Baesens, B., T. Van Gestel, S., Viaene, M., Stepanova, J. & Suykens, J. V. (2003). Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the Operational Research Society* 54 627-635. <http://dx.doi.org/10.1057/palgrave.jors.2601545>
- Bamber, D. (1975). The area above the ordinal dominance graph and the area below the receiver operating graph. *Journal of Mathematical Psychology* 12 387-415. [http://dx.doi.org/10.1016/0022-2496\(75\)90001-2](http://dx.doi.org/10.1016/0022-2496(75)90001-2)
- Banasik, J. J. N. & Crook, L. C. T. (1999). Not if but when will borrowers default. *Journal of the Operational Research Society* 50(12) 1185-1190.
- Boyle, M. J. N., Crook, R. & Hamilton, L. C. Thomas. (1992). Methods for credit scoring applied to slow payers. L. C. Thomas, J. N. Crook, D. B. Edelman, ed. *Credit Scoring and Credit Control*. Oxford University Press: Oxford, 75-90.
- Chatterjee, S., Barcun, S. (1970). A nonparametric approach to credit screening. *Journal of the American Statistical Association* 65 150-154. <http://dx.doi.org/10.2307/2283581>
- Copas, J. B. & Heydari, F. (1997). Estimating the risk of reoffending by using Exponential mixture models. *Journal of the Royal Statistical Society Series. A* 160 237-252. <http://dx.doi.org/10.1111/1467-985X.00059>
- Cox, D. R. (1975). Partial Likelihood. *Biometrika* 62(2) 269-276. <http://dx.doi.org/10.1093/biomet/62.2.269>
- D'Agostino, R. B., Nam, B. H. (2004). Evaluation of the performance of survival analysis models: discrimination and calibration measures. N. Balakrishnan, C. R. Rao, ed. *Advances in Survival Analysis*. Elsevier, London, 1-25.
- Davis, R. H., Edelman, D. B. & Gamberman, A. J. (1992). Machine-learning algorithms for credit-card applications *IMA Journal of Management Mathematics* 4(1) 43-51. <http://dx.doi.org/10.1093/imaman/4.1.43>
- Desai, V. S., Conway, D. G., Crook, J. N. & Overstreet, G. A. (1997). Credit-scoring models in the credit-union environment using neural networks and genetic algorithms. *IMA Journal of Management Mathematics* 8(4) 323-346. <http://dx.doi.org/10.1093/imaman/8.4.323>
- Durand, D. (1941). *Risk Elements in Consumer Instalment Financing*. National Bureau of Economic Research, New York.
- Farewell, V. T. (1986). Mixture models in survival analysis: are they worth the risk? *The Canadian Journal of Statistics* 14(3) 257-262. <http://dx.doi.org/10.2307/3314804>
- Gamel, J. W., McLean, I. W. & Rosenberg, S. H. (1990). Proportion cured and mean log survival time as functions of tumour size. *Statistics in Medicine* 9 999-1006. <http://dx.doi.org/10.1002/sim.4780090814>
- Hand, D. J. & Henley, W. E. (1997). Statistical classification methods in consumer credit scoring: a review. *Journal of the Royal Statistical Society Series. A* 160 523-541. <http://dx.doi.org/10.1111/j.1467-985X.1997.00078.x>
- Hanley, J. A. & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143 29-36.
- Hardy, W. E. & Adrian, J. L. (1985). A linear programming alternative to discriminant analysis in credit scoring. *Agribusiness* 1(4) 285-292. [http://dx.doi.org/10.1002/1520-6297\(198524\)1:4<285::AID-AGR2720010406>3.0.CO;2-M](http://dx.doi.org/10.1002/1520-6297(198524)1:4<285::AID-AGR2720010406>3.0.CO;2-M)
- Henley, W. E. (1995). *Statistical Aspects of Credit Scoring*. Ph.D. thesis. Open 83 University, Milton Keynes, U.K.
- Henley, W. E. & Hand, D. J. (1996). A k-nearest-neighbour classifier for assessing consumer credit risk. *Statistician* 65 77-95. <http://dx.doi.org/10.2307/2348414>
- Henley, W. E. & Hand, D. J. (1997). Construction of a k-nearest-neighbour credit-scoring system. *IMA Journal of Management Mathematics* 8 305-321. <http://dx.doi.org/10.1093/imaman/8.4.305>
- Kolesar, P. & Showers, J. L. (1985). A robust credit screening model using categorical data. *Management Science* 31(2) 123-133. <http://dx.doi.org/10.1287/mnsc.31.2.123>

- Kuk, A. Y. C. & Chen, C. H. (1992). A mixture model combining logistic regression with proportional hazards regression. *Biometrika* 79(3) 531-541. <http://dx.doi.org/10.1093/biomet/79.3.531>
- Larson, M. G. & Dinse, G. E. (1985). A mixture model for the regression analysis of competing risks data. *Applied Statistics* 34 201-211. <http://dx.doi.org/10.2307/2347464>
- Leonard, K. J. (1993). Empirical Bayes analysis of the commercial loan evaluation process. *Statistics & Probability Letters* 18(4) 289-296. [http://dx.doi.org/10.1016/0167-7152\(93\)90017-D](http://dx.doi.org/10.1016/0167-7152(93)90017-D)
- Makowski, P. (1985). Credit scoring branches out. *Credit World* 75 30-37.
- Maller, R. A. & Zhou, S. (1994). Testing for sufficient follow-up and outliers in survival data. *Journal of the American Statistical Association* 89 1499-1506. <http://dx.doi.org/10.2307/2291012>
- Maller, R. A. & Zhou, X. (1996). *Survival Analysis with Long-Term Survivors*. Wiley, New York.
- Mehta, D. (1968). The formulation of credit policy models. *Management Science* 15(2) 30-50. <http://dx.doi.org/10.1287/mnsc.15.2.B30>
- Myers, J. H. & Forgy, E. W. (1963). The development of numerical credit evaluation systems. *Journal of the American Statistical Association* 58 799-806. <http://dx.doi.org/10.2307/2282727>
- Narain, B. (1992). Survival analysis and the credit granting decision. L. C. Thomas, J. N. Crook, D. B. Edelman, ed. *Credit Scoring and Credit Control*. Oxford

Table 1. Female applicants' data

t_i	d_i	n_i	c_i
0	0	250	0
2	1	250	0
3	2	248	0
4	0	245	2
7	1	241	0
11	2	234	0
12	0	223	1
16	1	211	0
18	0	195	2
20	0	177	2
21	2	157	0
22	0	136	1
25	2	114	0
27	1	89	0

Table 2. Male applicants' data

t_i	d_i	n_i	c_i
0	0	250	0
3	0	250	2
4	1	247	0
5	2	243	0
6	0	238	4
7	1	236	0
8	0	229	1
9	0	221	1
10	2	212	0
12	0	202	3
13	1	190	0
15	0	177	2
16	1	162	0
18	0	146	4
24	0	128	1
25	0	104	1

Table 3. Estimated survival time for female applicants

	Survival Time	Standard Error	95% Confidence Interval
Mean.	16	3	(10, 23)
Median.	13	4	(6, 20)

Table 4. Estimated survival percentiles for female applicants

Percentiles	25.00	50.00	75.00
Value	27.00	13.00	10.00
Standard Error	0.00	3.73	2.50

Table 5. Estimated survival time for male applicants

	Survival Time	Standard Error	95% Confidence Interval
Mean.	15	2	(11, 20)
Median.	16	5	(6, 26)

Table 6. Estimated survival percentiles for female applicants

Percentiles	25.00	50.00	75.00
Value	25.00	16.00	7.00
Standard Error	0.00	4.92	3.04

Table 7. Survival analysis for time to default

	Total	Number of events	Number censored	Percent censored
Gender 0	13	7	6	46.15
Gender 1	15	11	4	26.67
Overall	28	18	10	35.71

Table 8. Test Statistics for Equality of Survival Distributions for GENDER

	Statistic	df	Significance
Log Rank	.17	1	.6780

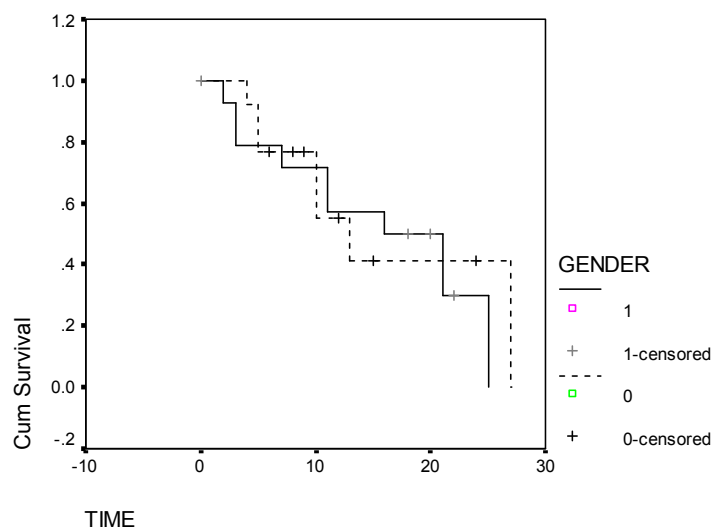


Figure 1. Survival Functions

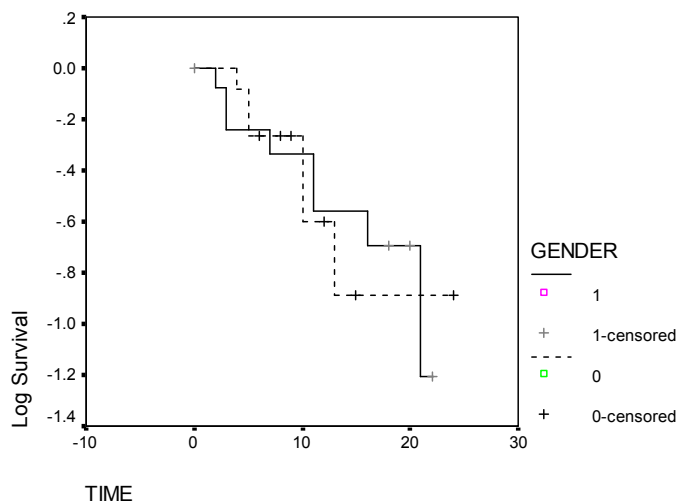


Figure 2. Log Survival Function

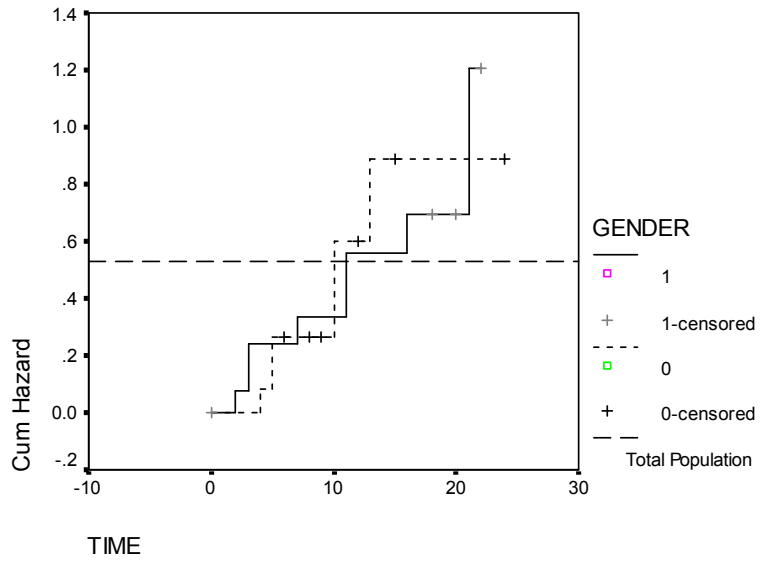


Figure 3. Hazard Function