

---

# Comparison of Survival Analysis Approaches to Modelling Credit Risks

Sammy Mungasi, Collins Odhiambo \*

Institute of Mathematical Sciences, Strathmore University, Nairobi, Kenya

## Email address:

sam.mungasi@gmail.com (S. Mungasi), codhiambo@strathmore.edu (C. Odhiambo)

\*Corresponding author

## To cite this article:

Sammy Mungasi, Collins Odhiambo. Comparison of Survival Analysis Approaches to Credit Risks. *American Journal of Theoretical and Applied Statistics*. Vol. 8, No. 2, 2019, pp. 39-46. doi: 10.11648/j.ajtas.20190802.11

**Received:** April 1, 2019; **Accepted:** May 15, 2019; **Published:** June 5, 2019

---

**Abstract:** Credit risk is a critical area in finance and has drawn considerable research attention. As such, survival analysis has widely been used in credit risk, in particular to model debt's time to default mechanisms. In this study, we revisit different survival analysis approaches as applied in credit risk defaulters' data and assess their performance in light of the Kenyan context. In practice, inconsistency in validity of credit risk models used by many company when predicting and analysis of loan default is a common phenomenon that occurs unexpectedly. Loan defaults, often causes major loses to creditors' and can be of great benefit if quantified correctly in advance by using correct models. Here, we address the unbiasedness, analysis and comparison of survival analysis approaches, particularly, the models of credit risk. We carry out data analysis using Cox proportional hazard model and it's extensions as well as the mixture cure and non-cure model. We then compare the results systematically by investigating the most efficient and preferable model that produces best estimates in Kenyan real data setting. Results show, the Cox Proportional Hazard (CPH) model is more efficient in the analysis of Kenyan real data set compared to the frailty, the mixture cure and non-cure model.

**Keywords:** Credit Risk, Default Rates, Cox Proportional Hazard Model, Mixture Cure Model, Hazard Ratio

---

## 1. Introduction

Credit risk is described as the risk of default on a debt that may arise when a borrower fails to make contractual payments. Credit risk arises whenever two counter-parties engage in borrowing and lending [1]. It remains a critical area both in banking and other lending institutions and is of great concern to stakeholders, i.e. borrowers, institutions, and policy regulators. Since the advent of Value at Risk (VaR) models in the 1990's, there has been evolution of risk management practices and credit risk modelling. In this study, we review existing survival models in literature and systematically assess their performance in Kenyan setting.

Currently, Kenya has 42 commercial banks, 8 representative offices of foreign banks, 13 microfinance banks and one mortgage finance company [2]. Technically, financial institutions have been utilizing survival analysis tools to model credit risks [3]. According to the Central bank of Kenya 2016 annual report [4-9], the banking sector's performance remains consistently, resilient. However despite

being resilient, Kenya's public and publicly guaranteed debt increased by 14.3 percent during the financial year 2017/18, with both domestic and external debt increasing at 17.4 percent and 11.6 percent, respectively [6-7]. Public debt portfolio comprised of 49.2 percent and 50.8 percent domestic and external debt respectively by the end of the financial year 2017/18. The ratio of public debt to Gross Domestic Product (GDP) declined marginally to 57 percent at end-June 2018 as the projected rate of economic expansion surpassed the rate of buildup in public debt [5]. There has been an increase in non-performing loans as well. In a Credit Survey Report for the Quarter ended March 2018, the ratio of gross non-performing loans to gross loans increased from 10.66 percent in December 2017 to 11.81 percent in March 2018 [10]. The ratio of core capital to total risk-weighted assets also increased slightly from 16.05 percent in December 2017 to 16.15 percent as at March 2018 [11]. Historically, lending institutions have used different models in managing credit risk [12-17]. Several authors have looked at credit risk scoring and modelling by extending CPH

models [2, 18, 20-24]. One of the rationale for using survival analysis in credit risk setting is that time-to-default can be modeled with other determining features [21]. Given the emerging of many financial services provider firms and the growth in the financial industry, there is need for efficient credit risk measurement techniques. Standards such as the Basel II and International Financial Reporting Standard (IFRS) 9 have furthermore implied the need for effective credit risk measurement models. Kenya recently adopted the IFRS 9, thus credit providers are assessing the impact of it as well as seeking models useful in measuring different credit risk. Whereas other authors have fitted models of loan defaulters, there has not been a systematic comparison on the performance of different models in Kenyan context. In this study we fit different survival models into the Kenyan credit data and systematically compare the performance of models. This work attempts to evaluate and give recommendations on the best survival model in credit risk context in relation to the Kenyan data.

In the next section, we describe the CPH model with its extensions. In Section 3, we discuss results and in Section 4 we discuss our results and draw conclusions.

## 2. Methods

We fits the Cox PH model and it's extensions, that is, penalized splines and frailty model as well as the mixture cure and no-cure model to real Kenyan data set. Assessment is done to ascertain the most effective model in analyzing credit risk.

### 2.1. Data

The data used for the study was obtained from Metropolis Credit Reference Bureau for the period 2014 to 2017. The data comprised of 20,299 individuals and included various covariates namely: the age of the individuals (from 22 years to 75 years), age bracket (18-33, 34-43, 44-53, >54), gender (male and female), marital status (married, single, divorced and widowed), the type of the account (credit card, loan account and current account), status of the loan (active or defaulted), loan amount (ranged from USD 10 to USD 2,290,686) and the loan amount group.

### 2.2. Study Design

We fitted the Cox PH model and its extensions, that is, penalized splines and frailty model as well as the mixture cure and no-cure model to real Kenyan data set. Assessment is done to ascertain the most effective model in analyzing credit risk.

#### 2.2.1. Survival Analysis Framework

The survival function denoted as the probability of not having experienced the event of interest over observed time  $t$ , is  $S(t) = P(T > t)$ . In the context of credit risk, the event of interest is default [25]. The hazard function models the

instantaneous risk. The Cox proportional hazard (CPH) model is more flexible and has widely been used [30].

#### 2.2.2. Penalized Splines

Whenever, the number of knots in a given spline becomes sufficiently large, a fitted function of the spline depicts more variation than justified by the data. Penalized splines are considered as a variant of smoothing spline with more flexible choice of knots, bases and penalties. A smoothness penalty implements the procedure of integrating the square of the second derivative of the fitted spline function. The penalty could also be based on higher-order finite differences of adjacent basis splines (B-splines) [31].

#### 2.2.3. Frailty Model

The Frailty models provide a better way for incorporating random effects in a given model to account for association and heterogeneity that is not observed. Generally, a frailty model can be considered as an unobserved random factor that modifies multiplicatively the hazard function of an individual, group or cluster of individuals. The model is represented by the following hazard given the frailty:

$$\lambda(t|Z, X) = Z\lambda(t|X) \quad (1)$$

Where  $\lambda$  is the hazard function and the frailty  $Z$  is an unobservable random variable varying over the sample which increases the individual risk if  $Z > 1$  or decreases if  $Z < 1$ . The conditional survivor function for the model is presented as:

$$S(t|Z, X) = \exp\left(-Z \int_0^t \lambda(u|X) du\right) = \exp(-Z\Lambda(t|X)) \quad (2)$$

where  $\Lambda(t|X) = \int_0^t \lambda(u|X) du$ .  $S(t|Z, X)$ , represents the fraction of individuals surviving until time,  $t$ , given,  $Z$ , and given the vector of observable covariates  $X$  [31].

#### 2.2.4. Mixture Cure and Non-Mixture Cure Model

Conventionally, mixture cure models have been motivated by the existence of dis aggregated long-term survivors [24]. On the other hand, under non-mixture survival models, event of interest is assumed to occur in the long-run. Both mixture cure and non-cure models are used in the setting where a given fraction of the population will not experience the event of interest. The mixture cure model is therefore, a mixture of distributions where, on one hand, a logit regression model generates a mixing proportion of non-susceptibility; the survival model, on the other hand describes the survival function of the cases susceptible to the event of interest. The models are of particular interest in credit risk modeling as the event of interest here, default, will not occur for a very high proportion of the cases [24]. The survival function of the mixture cure model is given as;

$$S(t|x) = \pi(x)S(t|Y = 1, x) + 1 - \pi(x) \quad (3)$$

where  $Y$  is the susceptibility indicator ( $Y = 1$  if an account is susceptible, and  $Y = 0$  if not). The conditional survival function modeling the cases that are susceptible is given by a

CPH model:

$$s(t|Y = 1, x) = \exp(-\exp(\beta'x) \int_0^t h_0(u|Y = 1)du) \quad (4)$$

In a non-cure mixture context, the Breslow-type estimator is used for estimation of the cumulative baseline hazard similar to the Cox PH model. Excellent summary on non-cure mixture model can be found in [23-26].

### 3. Data Analysis

Data was extracted from Kenya Metropal and comprised of 20,299 individuals. The data was extracted between December 2014 and December 2017 where different individual accounts were tracked over a certain period of time to obtain the accounts where the individuals went into default as well as those that did go into default within the period of the study. The date captured various covariates, that is, time difference, gender, age, type of the product, marital status, year of data retrieval and the loan amount groups. The time difference means for active individuals is 24.17 months, with a lower confidence interval of 24.04 months and an upper confidence interval of 24.30 months. For defaulters, the mean time difference is 24.70 months with a lower confidence interval of 24.45 months and an upper confidence interval of 24.94 months. The mean original amount for active individuals is Kshs 300,077.82 with a lower confidence interval of Kshs 247,589.00 and an upper confidence interval of Kshs 352,566.63. For the defaulters,

the mean original amount is Kshs 195,210.76 with a lower confidence interval of Kshs 137,670.56 and an upper confidence interval of Kshs 252,750.95. The mean age for active individuals is 41 years with a lower confidence interval of 40.83 years and an upper confidence interval of 41.17 years. For the defaulters, the mean age is 38.11 years with a lower confidence interval of 37.81 years and an upper confidence interval of 38.40 years. The time (in years) when data was retrieved is December 2014, 2015, 2016 and 2017. The gender for the individuals under study is male and female, we shall denote male as gender 1 and female as gender 2 for our analysis part. The individuals are grouped into the age brackets of 18-33, 34-43, 44-53 and above 54 years. The amounts of the individuals are under different products namely; current account, loan account, and the credit card. The individuals under study are considered to be married, divorced, single or widowed. The individuals amount is also grouped in ranges starting from 0-50,000, 50,001- 100,000, 100,001-250,000, 250,001-500,000, 500,001-1,000,000 and over 1,000,000. Note that all the covariates have a p-value of less than 0.001 apart from the marital status which has a p-value of 0.006. The data can be summarized in table 1.

We generated age boxplot with respect to active/defaulters clients and results shows evidence that young people are more likely to default loans than older people while there is no much significance in the time difference.

Table 1. Summary of data, variables, covariates and there corresponding counts and p-value.

Time Difference (95% CI)		Mean=24.17 (24.04,24.30)	Mean=24.70 (LCI=24.45, UCI=24.94)	<.001
Original Amount (95% CI)		Mean=300,078 (247,589,352,567)	Mean=195,210.76 (137,671,252,751)	<.001
Age (95% CI)		Mean=41.00 (40.83,41.17)	Mean=38.11 (37.81,38.40)	<.001
Gender	F	= 7857 (82.61%)	= 1654 (17.39%)	<.001
	M	= 7989 (74.05%)	= 2799 (25.95%)	<.001
Age Bracket	>54	= 2150 (86.14%)	= 346 (13.86%)	<.001
	18-33	= 4652 (71.95%)	= 1814 (28.05%)	<.001
	34-43	= 4908 (78.73%)	= 1326 (21.27%)	<.001
	44-53	= 4136 (81.05%)	= 967 (18.95%)	<.001
Loan Groups	A	= 533 (86.67%)	= 82 (13.33%)	<.001
	B	= 848 (84.38%)	= 157 (15.62%)	<.001
	C	= 1527 (84.32%)	= 284 (15.68%)	<.001
	D	= 1473 (78.94%)	= 393 (21.06%)	<.001
	E	= 1657 (75.66%)	= 533 (24.34%)	<.001
	F	= 9808 (76.55%)	= 3004 (23.45%)	<.001
Product Name	Current Ac	= 2419 (78.36%)	= 668 (21.64%)	<.001
	Loan Ac	= 10877 (83.75%)	= 2111 (16.25%)	<.001
	Credit Card	= 2550 (60.37%)	= 1674 (39.63%)	<.001
Status	Divorced	= 43 (75.44%)	= 14 (24.56%)	0.006
	Married	= 10483 (78.60%)	= 2854 (21.40%)	0.006
	Single	= 5311 (77.11%)	= 1577 (22.89%)	0.006
	Widowed	= 9 (52.94%)	= 8 (47.06%)	0.006
	Amount Group	0-50,000	= 9808 (76.55%)	= 3004 (23.45%)
50,001-100,000		= 1657 (75.66%)	= 533 (24.34%)	<.001
100,001-250,000		= 1473 (78.94%)	= 393 (21.06%)	<.001
250,001-500,000		= 1527 (84.32%)	= 284 (15.68%)	<.001
500,001-1,000,000		= 848 (84.38%)	= 157 (15.62%)	<.001
Over 1,000,000		= 533 (86.67%)	= 82 (13.33%)	<.001

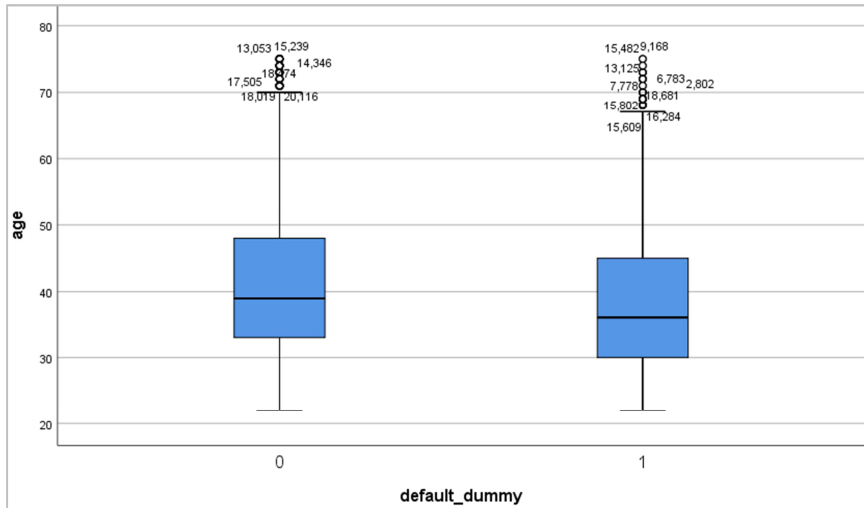


Figure 1. Age distribution dropbox (0 for active clients and 1 for defaulters) with outliers.

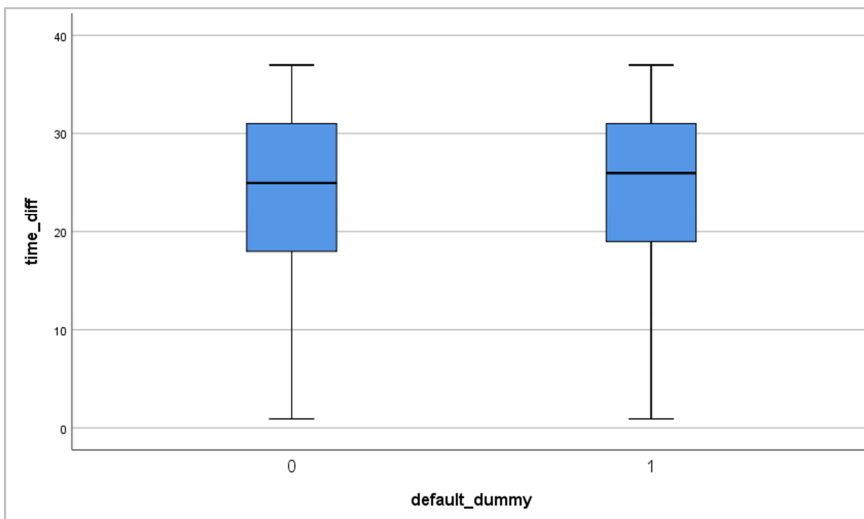


Figure 2. Time distribution drop-whiskers plot: Duration on loan (0 for active clients and 1 for defaulters).

### 4. Results

We fitted the Kaplan-Meier curves and the figure below represents the result.

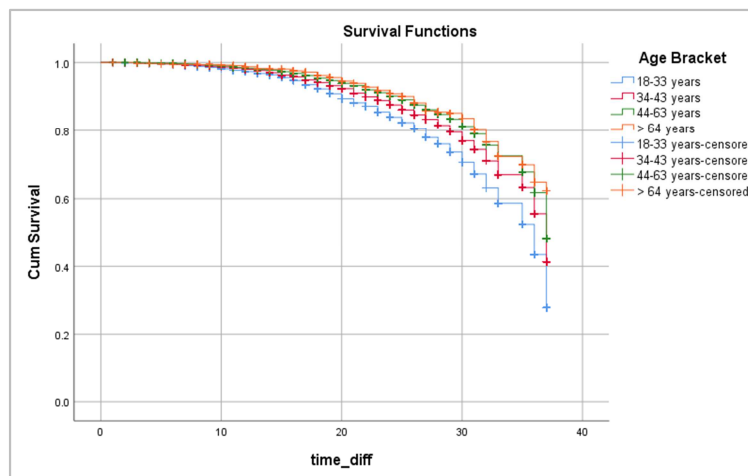


Figure 3. KM curves for age bracket.

By design, there are three assumptions used analysis of Kaplan-Meier; both censored and non-censored patients have the same survival prospects, the survival probabilities are the same for subjects recruited either earlier or late during observation, events happen at specified time.

**4.1. Cox Proportional Hazard (CPH) Model with Time Independent**

We fitted the CPH model (time, Status) regressed against

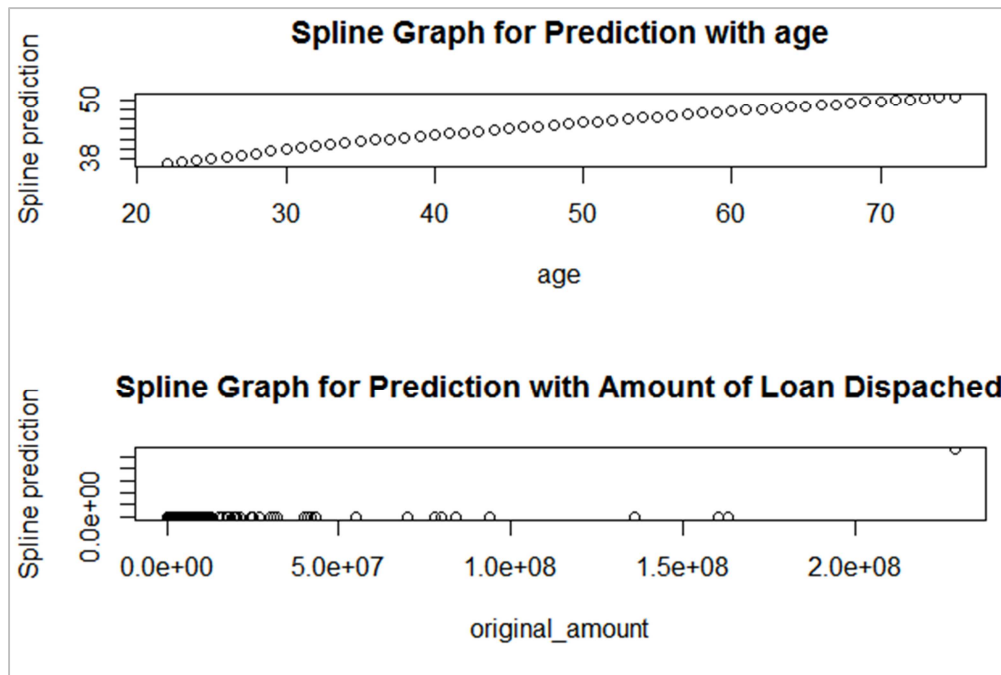
*Table 2. Cox PH Model with time independent.*

Covariates	coef	exp (coef)	se (coef)	z	Pr (> z )	exp (coef)	Exp (-coef)	LCI (95%)	UCI (95%)
gender2	-3.190e-01	7.269e-01	3.122e-02	-10.218	< 2e-16	0.7269	1.3757	0.6837	0.7728
age_bracket (34-43)	-2.667e-01	7.659e-	3.622e-02	-7.363	1.8e-13	0.7684	1.3014	0.7157	0.8250
age_bracket (44-53)	-4.569e-01	6.333e-01	3.985e-02	-11.463	< 2e-16	0.6321	1.5821	0.5845	0.6835
age_bracket (>54)	-6.933e-01	4.999e-01	5.879e-02	-11.792	< 2e-16	0.5039	1.9844	0.4491	0.5655
Divorced	-2.756e-01	7.591e-01	2.680e-01	-1.028	0.3038	1.0000	1.0000	1.0000	1.0000
Single	4.053e-02	1.041e+00	3.140e-02	1.291	0.1967	1.3194	0.7579	0.7803	2.2310
Widowed	-1.233e-01	8.840e-01	3.546e-01	-0.348	0.7280	1.3618	0.7343	0.8045	2.3052
Loan Ac	-8.604e-01	4.230e-01	3.286e-02	-26.180	< 2e-16	NA	NA	NA	NA
Current Ac	-4.777e-01	6.202e-01	4.583e-02	-10.423	< 2e-16	0.4375	2.2858	0.4101	0.4667
Original Amount	-3.098e-08	1.000e+00	1.468e-08	-2.110	0.0349	0.6339	1.5775	0.5793	0.6936

gender, age bracket, loan principal amount, marital status and product name). The number of customers, n= 20,282, number of events= 4, 445 (Loan Defaulters). 17 observations were missing such information and were therefore deleted. Concordance= 0.635 (se = 0.005 ), Rsquare= 0.049 (max possible= 0.98), Likelihood ratio test= 1019 ( p=<2e-16), Wald test = 1041 (p=<2e-16) and Score (logrank) test = 1094 (p=<2e-16). See table 2.

**4.2. Penalized Splines for Cox PH Model**

We tested model nonlinearities in the covariate by using penalized splines. B-splines are piecewise-defined polynomials joined at knots, with cubic B-splines being the most frequently used basis functions. Results shows linearity in age as a prediction though amount of loan dispatched was non-linear.



**Figure 4.** Penalized Spline graphs for defaulter’s prediction with age and amount of loan advanced as covariates. The y-axis (spline prediction in each case illustrates the realized estimated defaulter frequencies, obtained from B-splines. For the two variables (age and loan advanced), the data have been sorted and grouped into 50 equal-sized groups.

**4.3. Frailty Model**

The frailty model incorporates an excellent way to add random effects, association and unobserved heterogeneity into survival models. The unobserved random

proportionality factor that modifies the hazard function of an individual, or of related individuals is displayed. A frailty model is a random effects model for time variables, where the random effect (the frailty) has a multiplicative effect on the hazard. We fitted time and status regressed

against gender, age bracket, marital status, and product name and frailty random factor for age assuming a Gaussian distribution

*Table 3. Frailty model estimates.*

Covariates	Value	Std. Error	z	p
r (Intercept)	3.56981	0.07783	45.87	<0.001
gender2	0.08900	0.00908	9.80	< 0.001
age_bracket (34-43)	0.06853	0.01431	4.79	<0.001
age_bracket (44-53)	0.12367	0.01555	7.95	<0.001
age_bracket (>54)	0.18903	0.01928	9.80	< 0.001
Married	-0.07917	0.07723	-1.03	0.31
Single	-0.08895	0.07738	-1.15	0.25
widowed	NA	0.00000	NA	NA
Loan_Ac	0.22697	0.01000	22.70	< 0.001
Current_Ac	0.12829	0.01339	9.58	< 0.001
Log (scale)	-1.24457	0.01242	-100.18	< 0.001

Results show, scale= 0.288 Weibull distribution, loglik

*Table 4. Mixture Cure Model estimates for age.*

covariates	data mean	estimate	LCI (95%)	UCI (95%)	Standard error	exp (est)	LCI 95%	UCI 95%
Theta	NA	1.02e-02	NA	NA	NA	NA	NA	NA
shape	NA	3.44e+00	NA	NA	NA	NA	NA	NA
scale	NA	2.51e-06	NA	NA	NA	NA	NA	NA
age	4.04e+01	-2.28e+00	NA	NA	NA	1.02e-01	NA	NA

For gender covariate we fitted (time, status) against gender. The distribution was assumed to be Weibull proportional hazard with link function as logistic. N = 20,299, all events were 4,453, censored events: 15,846, Total time at risk: 492,930, Log-likelihood = -22,298.5, degree of freedom = 4 and AIC = 44,605.

*Table 5. Mixture Cure Model estimates for gender.*

Covariates	Data mean	estimate	LCI (95%)	UCI (95%)	Standard error	exp (est)	LCI 95%	UCI 95%
theta	NA	3.01e-05	3.45e-14	1.00e+00	NA	NA	NA	NA
shape	NA	3.47e+00	3.38e+00	3.55e+00	4.31e-02	NA	NA	NA
scale	NA	2.62e-06	1.96e-06	3.50e-06	3.87e-07	NA	NA	NA
gender2	4.69e-01	9.16e+00	-1.14e+01	2.97e+01	1.05e+01	9.51e+03	1.09e-05	8.30e+12

For age covariate we fitted (time, status) against age bracket. The distribution was assumed to be Weibull proportional hazard with link function as logistic. N = 20,299, events: 4,453, censored events were 15,846, Total time at risk: 492930, Log-likelihood = -22, 276.01, degree of freedom = 6 and AIC = 44,564.03.

*Table 6. Mixture Cure Model estimates for age bracket.*

covariates	data mean	estimate	LCI (95%)	UCI (95%)	Standard error	exp (est)	LCI 95%	UCI 95%
theta	NA	1.20e-04	6.09e-09	7.03e-01	NA	NA	NA	NA
shape	NA	3.49e+00	3.40e+00	3.57e+00	4.35e-02	NA	NA	NA
scale	NA	2.58e-06	1.93e-06	3.46e-06	3.84e-07	NA	NA	NA
age_bracket (34-43)	3.07e-01	7.06e+00	-2.83e+00	1.69e+01	5.05e+00	1.16e+03	5.88e-02	2.29e+07
age_bracket (44-53)	2.51e-01	7.67e+00	-2.22e+00	1.76e+01	5.05e+00	2.15e+03	1.09e-01	4.24e+07
age_bracket (>54)	1.23e-01	8.64e+00	-1.25e+00	1.85e+01	5.05e+00	5.66e+03	2.87e-01	1.12e+08

**4.5 Non-mixture Cure Model**

The rationale behind non-mixture cure models is that observation stage it is assumed that an individual is left with some levels of possible occurrence of event and has the potential of experiencing the event of interest. Number of events is assumed to have a Poisson distribution. For age

(model) = -21, 804.5, Loglik (intercept only) = -22, 316.2, Chi-square value= 1023.41 on 20 degrees of freedom, p= 3.7e-204, with number of Newton-Raphson Iterations: 719 and n=20,282. 17 observations that were missing were deleted.

**4.4. Mixture Cure Model**

During survival, it often happens that a certain subjects never experience the event of interest. These event times are usually considered infinite and the subjects are said to be cured. Survival models that take this feature into account are commonly referred to as cure models. Here we fitted age covariate i.e. (time, Status) against age. The distribution was assumed to be Weibull proportional hazard with link function as logistic. N = 20,299, Events: 4,453, Censored: 15,846, total time at risk: 492,930, log-likelihood = - 22,349.14, 4 degrees of freedom and AIC = 44,706.28.

covariate, we fitted (time, status) against age. The distribution was assumed to be Weibull proportional hazard with link function as logistic. N = 20,299, number of events were 4,453, censored events were 15,846, total time at risk: 492,930, Log-likelihood value yielded -22,247.71, degree of freedom = 4 and AIC = 44503.42.

Table 7. Non-Mixture Cure Model estimates for age.

Covariates	data mean	estimate	LCI (95%)	UCI (95%)	Standard error	exp (est)	LCI 95%	UCI 95%
theta	NA	3.60e-37	2.59e-49	5.02e-25	NA	NA	NA	NA
shape	NA	3.45e+00	3.36e+00	3.54e+00	4.41e-02	NA	NA	NA
scale	NA	5.11e-08	3.29e-08	7.94e-08	1.15e-08	NA	NA	NA
age	4.04e+01	8.91e-01	5.79e-01	1.20e+00	1.59e-01	2.44e+00	1.79e+00	3.33e+00

For age categories we fitted (time, status) against age bracket. The distribution was assumed to be Weibull proportional hazard with link function as logistic. N = 20299, Events: 4453, Censored: 15846, Total time at risk: 492930, Log-likelihood = -22232.87, degree of freedom = 6 and AIC = 44477.75.

Table 8. Mixture Non-Cure Model estimates for age bracket.

Covariates	Data mean	estimate	LCI (95%)	UCI (95%)	Standard error	exp (est)	LCI 95%	UCI 95%
theta	NA	1.54e-238	3.19e-11	0.00e+00	NA	NA	NA	NA
shape	NA	3.45e+00	3.37e+00	3.54e+00	4.29e-02	NA	NA	NA
scale	NA	5.80e-09	2.54e-10	1.33e-07	9.26e-09	NA	NA	NA
age_bracket (34-43)	3.07e-01	-2.95e-01	-3.66e-01	-2.25e-01	3.61e-02	7.44e-01	6.93e-01	7.99e-01
age_bracket (44-53)	2.51e-01	-4.21e-01	-4.99e-01	-3.43e-01	3.98e-02	6.56e-01	6.07e-01	7.10e-01
age_bracket (>54)	1.23e-01	-7.36e-01	-8.51e-01	-6.21e-01	5.87e-02	4.79e-01	4.27e-01	5.38e-01

## 5. Discussion and Conclusions

In this study, we reviewed the performance of several survival analysis techniques in credit risk analysis. A real-life data sets from Kenya Metropol from December 2014 to December 2017 was used, and we used Akaike Information Criteria (AIC) as the main evaluation measures to assess model performance. The survival models used were the CPH model, penalized splines, frailty and the mixture cure and non-cure model. From the study it's clearly evidenced that all the models were significant in the analysis of a Kenyan real data set. However, the penalized splines model only fitted age as a loan default predictor. The mixture non-cure model also fitted age and age bracket as the loan default predictors. Furthermore, the CPH model seemed to have outperformed the other models in comparison though the other models did not perform significantly different in most cases. The models revisited collectively have the advantage of not requiring the survival function to go to zero when time goes to infinity; a situation that is seldom and appropriate for credit risk data. Stepanova et al. [25], in their evaluation of credit scoring models found that survival analysis models outperformed other scoring models. Survival analysis is advantageous in that the time to default can be modeled, and not just whether an applicant will default or not [18, 24-28]. This study conquers with other studies [18, 30-33] in that the CPH model outperformed the frailty, penalized spline, and mixture cure and non-cure models. The individual in the age bracket of 18 years to 33 years are more likely to default while the males are more likely to default a loan compare to female. Individuals with credit card account are more likely to default followed by the ones with a current account then the loan account individuals. Those individuals that are single are also more like to default a loan. The study concludes that the CPH model is more efficient in the analysis of Kenyan real data set compared to the frailty, penalized spline and the mixture cure and non-cure model. However, it would be appropriate to further extend the mixture cure and non-cure

model and study the performance of these models in comparison with a CPH model and some of its extensions. Comparison between CPH model, its extensions, and mixture and non-mixture models assuming different distributions was assessed using the AIC, where a lower AIC value indicates better model fit.

Table 9. Results of AIC for four survival analysis models.

Model	AIC
Cox PH Model with time independent	39,747
Frailty Model	42,1001
Mixture Cure model	44,503
Non- Mixture Cure model	44,478

Additionally, this study also points out that there was a challenge in finding an appropriate evaluation measure that is evidenced across all the methods for survival analysis comparison.

## References

- [1] Allen, L., Delong, G., & Saunders, A. (2004). Issues in the credit risk modeling of retail markets. *Journal of Banking & Finance*, 28 (4), 727-752.
- [2] Gudmundsson, R., Ngoka-Kisinguh, K., & Odongo, M. T. (2013). The role of capital requirements on bank competition and stability: The case of the Kenyan banking industry. Kenya Bankers Association-KBA Centre for Research on Financial Markets and Policy Working Paper Series.
- [3] Wienke, A. (2010). *Frailty models in survival analysis*. Chapman and Hall/CRC.
- [4] Bellotti T. and Crook J. (2013). Forecasting and stress testing credit card default using dynamic models. *International Journal of Forecasting* 29 (4): 563 – 574.
- [5] Bellotti T and Crook J (2014). Retail credit stress testing using a discrete hazard model with macroeconomic factors. *Journal of the Operational Research Society* 65 (3): 340–350.

- [6] CBK. (2013a). Credit survey Report. Nairobi: Central Bank of Kenya.
- [7] CBK. (2013b). Risk Based Supervisory Framework. Nairobi: Central Bank of Kenya.
- [8] CBK. (2015). The Kenya financial sector stability report. Nairobi: Central Bank of Kenya.
- [9] CBK. (2016). Bank Supervision Annual Report. Nairobi: Central Bank of Kenya.
- [10] CBK. (2017). Bank Supervision Annual Report 2017. Nairobi: Central Bank of Kenya.
- [11] CBK. (2018). Credit Survey Report for the Quarter ended March 2018. Nairobi: Central Bank of Kenya.
- [12] Gupta, V. (2017). A survival approach to prediction of default drivers for India listed companies. *Theoretical Economics Letters*, 07 (02), 116-138.
- [13] Gaynor M. and Town R. (2011). Competition in Health Care Markets. Chapter for the Handbook of Health Economics, Volume 2. T. McGuire, M. V. Pauly, and P. Pita Barros, Editors 2011.
- [14] Jacobson, T., & Roszback, K. (2003). Bank lending policy, credit scoring and value at risk. *Journal of Banking and Finance*, 27 (4), 615-633.
- [15] J-K Im, DW Apley, C Qi and X Shan (2012). A time-dependent proportional hazards survival model for credit risk analysis. *Journal of the Operational Research Society*, 63 (3): 306-321.
- [16] Kagri, H. S. (2011). Credit risk and performance of Nigerian banks. Ahmadu Bello University, Zaria.
- [17] Kithinji, A. M. (2010). Credit risk management and profitability of commercial banks in Kenya.
- [18] Lore, D., Gerda, C., and Bart, B. (2017). Time to default in credit scoring using survival analysis. *Journal of the Operational Research Society*, 68, 652–665
- [19] Marimo, M. (2015). Survival analysis of bank loans and credit risk prognosis (Doctoral dissertation).
- [20] Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *The Journal of finance*, 29 (2), 449-470.
- [21] Narain B. (1992). Survival analysis and the credit granting decision. In: Thomas LC, Crook JN and Edelman DB, editors, *Credit Scoring and Credit Control*, pp. 109–121. Clarendon Press: Oxford.
- [22] Nevo A. (2001). Measuring Market Power in the Ready-to-Eat Cereal Industry. *Econometrica* 2001; 69 (2); 307-342.
- [23] Obuda, F. (2016). Analysis of credit risk on bank loans using cox proportional hazard model. (Unpublished master's thesis). University of Nairobi.
- [24] Tong, E., Mues, C., & Thomas, L. (2012). Mixture cure models in credit scoring. *European journal of operational research*, 218 (1), 132-139.
- [25] Stepanova M. and Thomas L. (2002). Survival analysis methods for personal loan data. *Operations Research Quarterly* 50 (2): 277–289.
- [26] Cao, R., Vilar, J. M., & Devia, A. (2009). Modelling consumer credit risk via survival analysis. *Sort: statistics and operations research transactions*, 33 (1), 0003-30.
- [27] Anthony, W., & Othieno, F. (2016). Semi-Markovian credit risk modeling for consumer loans: Evidence from Kenya. *Journal of Economics and International Finance*, 8 (7), 93-105.
- [28] Zhang, A. (2009). *Statistical Methods in Credit Risk Modeling*.
- [29] Bellotti, T. & Crook J. (2009). Credit scoring with macroeconomic variables using survival analysis. *The Journal of the Operational Research Society* 60 (12): 1699–1707.
- [30] Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34 (2), 187-202.
- [31] Li, Y., & Ruppert, D. (2008). On the asymptotics of penalized splines. *Biometrika*, 95 (2), 415-436.
- [32] Duchateau, L., & Janssen, P. (2007). *The frailty model*. Springer Science & Business Media.
- [33] Goel, M. K., Khanna, P., & Kishore, J. (2010). Understanding survival analysis: Kaplan-Meier estimate. *International journal of Ayurveda research*, 1 (4), 274.