

Applying Survival Analysis to Telecom Churn Data

Melik Masarifoglu, Ali Hakan Buyuklu

Department of Statistics, Yildiz Technical University, İstanbul, Turkey

Email address:

memasarif@hotmail.com (M. Masarifoglu), hbuyuklu@yildiz.edu.tr (A. H. Buyuklu)

To cite this article:

Melik Masarifoglu, Ali Hakan Buyuklu. Applying Survival Analysis to Telecom Churn Data. *American Journal of Theoretical and Applied Statistics*. Vol. 8, No. 6, 2019, pp. 261-275. doi: 10.11648/j.ajtas.20190806.18

Received: February 12, 2019; **Accepted:** November 27, 2019; **Published:** November 23, 2019

Abstract: In competitive telecommunication environment, it is imperative to maintain an effective customer retention strategy even while mobile service operators attracting new customers. Not only acquiring new customers costly process, but successful customer retention helps build brand loyalty and good business reputation. Motivated by real mobile service operator data set, we designed and proposed a solution to employ survival analysis technique that estimates customers' survivals and hazards. We aim to examine the impact of: campaign, tariff, tenure, age, auto-payment on survival times and hazards. After hazard ratios and survival experiences determined for each predictor, results enable mobile service operator to target the right customers to incentivize so that they can stay with their current operator. Proactive actions triggered by the results of the survival model is key to customer retention.

Keywords: Customer Retention, Telecom Churn Prediction, Survival Analysis

1. Introduction

Customer churn also known as customer attrition is when a customer ends subscription with their current operator [1]. There are two types of churn, the first is involuntarily churn which involves the customer's account being cancelled by the mobile service operator as a result of death, relocation or non-payment reasons. The second is voluntarily churn which is when customer cancels their subscription due to poor voice quality, limited cellular coverage, network problems, high prices or billing issues [2]. In voluntarily churn, customer defects to a competitor. Only voluntary churn data was examined in this article.

The Prominence of Churn Models:

Finding new customers for mobile service operators in a saturated market is very difficult and costly because of the marketing activities, incentives and campaigns to attract new customers. Competition is intense due to increasing varieties of customer-focused services, campaigns and tariffs offered from mobile service operators. As a result, cost of acquiring a new customer is five times higher than retaining a current customer.

For comparison, below is a list of average costs to acquire one customer for a variety of industries:

1. Travel: Priceline.com: \$7
2. Telecom: Sprint PCS: \$315

3. Retail: Barnesandnoble.com: \$10

4. Financial: TD Waterhouse: \$175

The telecom sector has the highest acquisition cost among other sectors. Due to high acquisition cost, mobile service operators are more focused on retention than customer acquisition. Thus, it is crucial to implement an effective retention strategy (ERS) to retain customers, as retention is far easier; less expensive and more achievable than acquisition. In addition, successfully retained customers acquire more new customers; they provide cash flow and profit and they are less sensitive to new pricing. In comparison with new customers, existing old customers may be more loyal, and churning may take a longer time [3-6].

One of the best practices to implement an ERS is to create a model based on customer behavior that will predict churn with high accuracy. The goal is to detect well in advance, which customers may churn, so that mobile service operators can react early to keep them.

Customer Relationship Management and Churn Models:

Customer Relationship Management (CRM) is the whole set of processes and systems applied to establish long-lived and profitable relationships with specific customers to support business strategies [7]. In telecommunication companies, retention of customers is one of the key activities of the CRM.

The main objective of CRM is to establish a win-win relationship between the customer and the firm. CRM has

four dimensions:

1. Customer Identification
2. Customer Attraction
3. Customer Retention
4. Customer Development

Data mining techniques that are used in these four dimensions are:

1. Association
2. Classification
3. Clustering
4. Forecasting
5. Regression
6. Sequence discovery

In CRM, churn models are classification techniques used for customer retention [8]. The goal of data mining techniques for churn analysis is to:

1. identify who is going to churn and when
2. identify why customers are churning

This paper intends to illustrate how to apply survival analysis techniques to telecom churn data to identify the above items. The rest of the paper is structured as follows. Section 2 defines some basic concepts and terminology of survival analysis as well as describes the application of survival analyses to different fields. Section 3 introduces the most famous and widely used model for survival analyses, the Cox proportional hazard model. In Section 4, a telecom churn data set is described, and a Cox model is applied. Section 5 concludes the paper.

2. Survival Analysis

Survival Analysis (SA) encompasses a variety of statistical procedures used to analyze and quantify time to event data. SA aims to answer the question “How long is it before some event occur?”. In practice, the event of interest may be: finding a job, failure of a component, default of an applicant for credit, corporate failure, divorce, academic promotion, etc. In SA, survival means the subject is free from the event of interest during the observation period [9-12].

The goal of SA is:

1. Estimate and interpret survival function
2. Compare survival functions
3. Asses the relationship between survival and one or more predictors (usually termed as explanatory variables or covariates)

Characteristic of SA include:

1. Outcome variable is time until the event occurs
2. Its ability to manage incomplete data, resulting from the event not occurring during the follow up time (censored data)
3. Framework of SA:
4. Definition of event
5. Specify survival times scale (year, month, week, etc.)
6. Identify the start time

SA models can be based on historical data or can be based on prospective studies where data is collected in real time. SA is also sometimes referred: event history analysis, failure time

analysis, hazard analysis, transition analysis, and duration analysis.

History of SA:

The origins of SA and its history spread far back to the early work on mortality by John Graunt, who published the book “Natural and Political Observations on the Bills of Mortality” in 1662. The concept of “Life Tables” was introduced in this book. A new, modern era in SA started during World War II in USA. Reliability of military equipment was tested using SA. After World War II, SA became popular and spread to other various disciplines [13].

The most influential papers on SA were published by Kaplan and Meier [14] and Cox [15], in which the Kaplan-Meier product limit estimator and Cox proportional hazard model were introduced, respectively.

Application of SA

SA is successfully used in various fields, such as medical research, engineering (reliability analysis), finance, and economics (duration analysis). For example, in finance, a firm’s vulnerability to global financial crisis are analyzed. There are several studies in this field ([16-19]). These research papers aim to identify firms that are at risk of failure in the future. Prior notice can help some corporations to take necessary steps to survive. Additionally, the time to sell or buy stock within the stock market can be analyzed and predicted with SA [20].

SA is also used in banking such as banks’ survival during financial crisis or measure their strength when they enter to a new country’s market [21, 22] and for credit risk analysis [23, 24].

In the telecommunication sector, churn prediction is vital for mobile service operators. It has received a lot of attention and popularity from marketing and academia. There are several articles published that mostly employ machine learning algorithms [1-4, 8]. Few articles have been published using SA techniques [25]. SA is also used to estimate customer life time value (CLTV) which is a measure of customer profitability over time [26, 27].

2.1. Censoring

In theory, researchers would like to know the exact survival time for each subject who enters the study during the observation period. However, this is often impractical. Some survival times are below (or above) some bound. Because of time limitations of the study, researchers are not able to wait indefinitely to see if every subject has the event. For some subjects, the study ends before the event has occurred or some subjects are lost to follow up or drop out of the study. It is important to know the drop out reasons for these subjects. If the reasons are unrelated to the event of interest, then this is called non-informative censoring. When censored data is random and non-informative, there is no issue of bias [28].

Informative censoring means that a subject leaves the study for a reason that is related to the event. Imputation techniques for missing data methods can be applied to deal with informative censoring [29].

There are three general types of censoring: right-censoring,

left-censoring and interval-censoring [30]. We will only focus on right censoring which is the most common type of censoring. Now, let's consider the mathematical definition below and figure 1 to help fully understand right-censoring.

For the i th subject, $i=1, 2, \dots, n$ (number of subjects), let:

1. T_i denotes the survival time
2. C_i denote the censoring time
3. δ_i denote the event indicator where:

$$\delta_i = \begin{cases} 1, & \text{event was observed } (T_i \leq C_i) \\ 0, & \text{event was not observed } (T_i > C_i) \end{cases}$$

4. Y_i denote the observed response where:

$$Y_i = \min(T_i, C_i)$$

In other words, observed survival times will always be less than or equal to true survival times.

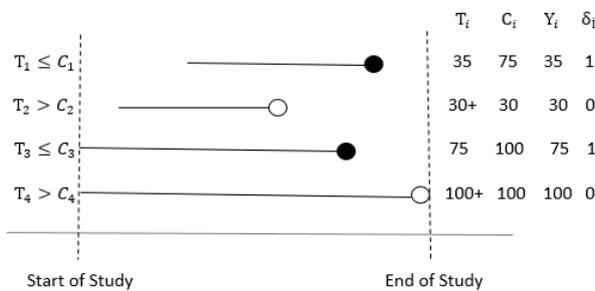


Figure 1. Right censored example.

2.2. Model

SA has two main functions known as the survival and hazard function. Let's introduce both functions.

Survival Function:

Let's assume for now that T is non-negative and a continuous random variable denoting the elapsed time until an event occurs with probability density function $f(t)$ and cumulative distribution function $F(t)$. $F(t)$ is defined as

$$F(t) = \Pr\{T < t\} \tag{1}$$

$S(t)$ is defined to be the complement of the c.d.f. [31], can be obtained from

$$S(t) = 1 - F(t) \tag{2}$$

$$= \Pr\{T \geq t\} \tag{3}$$

$$= \int_t^\infty f(x)dx \tag{4}$$

The survival function simply indicates the probability that the event of interest has not occurred by time t . Let's take the derivative of $S(t)$ and express $f(t)$ in terms of $S'(t)$:

$$S'(t) = -F'(t) \tag{5}$$

$$= -f(t) \tag{6}$$

As t ranges from 0 to ∞ , the three characteristics of survival function are [32]:

1. $S(t)$ is non-increasing in t

2. At $t=0$, all the subjects are alive which means $S(0)=1$
3. At $t=\infty$, all the subjects have the event which means $S(\infty)=0$

Hazard Function:

The hazard function is defined as [33]:

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr\{t \leq T < t + \Delta t | T \geq t\}}{\Delta t}, t \geq 0 \tag{7}$$

The numerator of expression (7) is the probability of failure between time t and $t+\Delta t$, given that it has not occurred up to time t , while the denominator Δt is the length of interval. Taking the limit of this expression as length of the interval goes to zero is called instantaneous rate of failure. Let's take a detailed look at the expression (7) [30, 34].

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr\{t \leq T < t + \Delta t | T > t\}}{\Delta t}, t \geq 0 \tag{8}$$

$$= \lim_{\Delta t \rightarrow 0} \frac{P(t < T \leq t + \Delta t, T > t)}{\Delta t} \cdot \frac{1}{P(T > t)} \tag{9}$$

$$= \lim_{\Delta t \rightarrow 0} \frac{P(t < T \leq t + \Delta t) \cap P(T > t)}{\Delta t} \cdot \frac{1}{P(T > t)} \tag{10}$$

$$= \lim_{\Delta t \rightarrow 0} \frac{P(t < T \leq t + \Delta t)}{\Delta t} \cdot \frac{1}{P(T > t)} \tag{11}$$

$$= \frac{1}{P(T > t)} \cdot \lim_{\Delta t \rightarrow 0} \frac{P(t < T \leq t + \Delta t)}{\Delta t} \tag{12}$$

$$= \frac{1}{P(T > t)} \cdot \lim_{\Delta t \rightarrow 0} \frac{F(t + \Delta t) - F(t)}{\Delta t} \tag{13}$$

$$= \frac{f(t)}{S(t)} \tag{14}$$

The formula is derived for the hazard function. Let's use expression (6) and plug in (14)

$$\lambda(t) = -\frac{S'(t)}{S(t)} \tag{15}$$

The result is

$$\lambda(t) = -\frac{d}{dt} \log S(t) \tag{16}$$

Integrate from 0 to t and use $S(0)=1$

$$S(t) = \exp \left\{ - \int_0^t \lambda(x) dx \right\} \tag{17}$$

Expression in the parenthesis called the cumulative hazard function and defined as

$$A(t) = \int_0^t \lambda(x) dx \tag{18}$$

Expected value of survival times can be found by taking the integral of the survival function (Darth Vader Rule) [35]

$$E(x) = \mu = \int_0^\infty S(x) dx \tag{19}$$

2.3. Regression Versus Survival Analysis

In SA, researchers may be interested in modeling the risk of event any time, conditional on having survived up to that moment. Further, they may also be interested in what factors benefit the time to event. SA can also handle censored data.

Logistic regression, researchers may be interested in modeling the risk of event whether it occurs in some fixed period of time. Additionally, they may also be interested in what factors benefit the risk of event occurring. Although, it is not mathematically

correct to use logistic regression with censored data, it still can be applied but it gives us limited information since it avoids censoring ([33, 36, 37]). Properties and advantages of models are summarized in table 1 and table 2.

Table 1. Properties of models.

Technique	Predictor Variable	Outcome	Censored Data
Regression	Linear	Categorical or Continuous	Normal Distribution
	Logistic	Categorical or Continuous	Binary
Survival Analysis	Categorical or Continuous	Survival times often far from Normal Distribution	Yes

2.4. Kaplan-Meier

It is a non-parametric method and known as Product Limit Formula. In this method S(t), which is the probability of surviving beyond t, is computed for each t.

Let $t_{(1)} < t_{(2)} < t_{(3)} < \dots < t_{(k)}$ denotes the observed times of event of the n subjects in the study, $m_{(i)}$ is number of deaths at each of $t_{(i)}$, $r_{(i)}$ is number of subjects who are event free at time $t_{(i)}$. S(t) can be estimated by

$$\hat{S}(t) = \prod_{t_{(i)} < t} \left\{ 1 - \frac{m_{(i)}}{r_{(i)}} \right\} \quad (20)$$

We not only estimate $\hat{S}(t)$, but it is also used to make

comparison between two or more groups.

Log-Rank Test

Survival curves are compared statistically using log-rank test. The purpose of this test is whether there is no statistical difference between the survival experience of two or more groups. Hypothesis of log-rank test defines as

$$H_0: S^1(t) = S^0(t) \text{ all } t \leq \tau$$

$$H_1: S^1(t) \neq S^0(t) \text{ some } t \leq \tau$$

Where τ is the largest time interval that at least one subject is at risk [31].

Table 3 is illustrated how to simply estimate hazard function and survival function.

Table 2. Equations & outcome of models.

Technique	Equation with one covariate	Outcome
Regression	Linear	$y = \beta_0 + \beta_1 X$
	Logistic	$\log \frac{p}{1-p} = \beta_0 + \beta_1 X$
Survival Analysis	$\lambda(t) = \lambda_0 \exp(\beta_0 + \beta_1 X)$	Hazard Ratio

Table 3. Survival analysis main functions example.

Time	Start	Fail	Censored	Hazard	Cumulative	
					Hazard Function	Survival Function
1	100	15	5	$\frac{15}{100} = 0.15$	0.15	$1 - 0.15 = 0.85$
2	$100 - 15 - 5 = 80$	10	3	$\frac{10}{80} = 0.125$	$0.125 + 0.15 = 0.275$	$0.85 * (1 - 0.125) = 0.61$
3	$80 - 10 - 3 = 67$	12	8	$\frac{12}{67} = 0.18$	$0.18 + 0.275 = 0.455$	$0.61 * (1 - 0.18) = 0.50$

3. The Cox Proportional Hazard Model

The Cox Proportional Hazard Model (CPHM) is a widely used statistical technique for quantifying the risk of observing the event of interest during the observation time. This technique is also used to assess simultaneously the effect of several covariates on survival time.

The hazard function, $\lambda(t|X)$, is defined as;

$$\lambda(t|X) = \lambda_0(t)e^{\beta^T X} = \lambda_0(t)e^{\sum_{i=1}^p \beta_i X_i} \quad (21)$$

Where X_i is a vector of covariates (explanatory or predictor variables), and β is a vector of regression coefficients. The baseline hazard, $\lambda_0(t)$, is the hazard when all covariates are zero. It is not estimated and specified for the CPHM, it is consequently called semi-parametric [12].

3.1. Hazard Ratio (HR)

The CPHM estimates the ratio of hazard values between

two levels, say X and X^* , with different covariate values. It is estimated by

$$HR = \frac{\lambda_0(t)e^{\sum_{i=1}^p \beta_i X_i}}{\lambda_0(t)e^{\sum_{i=1}^p \beta_i X_i^*}} \quad (22)$$

$$= e^{\sum_{i=1}^p \beta_i (X - X^*)} \quad (23)$$

Let us assume the covariate values of two different subjects differ on only one level. That is:

$$X = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ \vdots \\ X_k + 1 \\ \vdots \\ X_p \end{pmatrix}, X^* = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ \vdots \\ X_k \\ \vdots \\ X_p \end{pmatrix}$$

The result for the ratio is

$$\frac{\lambda(t,X)}{\lambda(t,X^*)} = e^{\beta k}, \text{ for all } t \geq 0 \tag{24}$$

Assumptions of Cox's Method ([38, 39]):

1. Non-Informative censoring
2. Proportional hazard
3. No tied events

CPHM does not require any specific distribution for the survival times. However, it assumes a proportional hazard assumption (PHA) which means the hazard ratio, in expression (22) does not depend on t. The survival curves for two groups must have hazard functions that are proportional for all values of t and additionally the hazard ratio does not vary with time. Graphically if the hazard curves cross, PHA may be violated.

There are methods that can be applied to test proportionality:

1. Schoenfeld residuals are used to assess CPHM fit for survival data. They should be independent of time. One of the diagnostic test can be employed is to plot Schoenfeld residuals against time. If there is a random pattern in the plot, the PHA has not been violated [40].
2. Plot $-\log(-\log S(t))$ against time, if the curves look parallel, PHA might be valid [31]
3. Include interaction term between the covariate and time in the model and test whether it is significant. If it is significant, the interaction term should be kept in the model [40]

Even while PHA fails to hold, CPHM still can be used because of its great strength and ability to enclose variables that change over time. When there are time dependent variables, the model is time dependent Cox and the form of hazard function changes accordingly

$$\lambda(t|X(t)) = \lambda_0(t)e^{\beta^t X + \beta^t X(t)} \tag{25}$$

Where X is the set of fixed covariates and X(t) is the time dependent covariates. However, fixed covariate form of hazard function is much simpler than the hazard function with time dependent covariates ([41, 42]). Further, time dependent Cox model does not have some characteristics of fixed covariate model. For example, survival curves cannot be estimated in time dependent Cox model. Time dependent covariates should be used in caution in order to avoid any bias [43]. Time dependency for some variables might be arose from long term study times [46]. An example of time dependent covariates such as body weight, income, marital status, marketing promotions, hypertension status and location. Examples of fixed covariates are: sex, race, income [44]. Time dependent variables can be handled both in SAS [41] and R [45].

3.2. Inference

The CPHM Partial likelihood function is given by

$$L(\beta) = \prod_{X_{i \text{ uncensored}}} \left[\frac{e^{X_i \beta}}{\sum_{j \in R(t_i)} e^{X_j \beta}} \right]^{\delta_i} \tag{26}$$

Where δ_i denotes the event indicator, $t_{(1)} < t_{(2)} < t_{(3)} <$

$\dots < t_{(k)}$ are the ordered, observed event times and $R_{(t)}$ is the risk set for time $t_{(i)}$, which is the set of all subjects surviving just before time $t_{(i)}$ [15]. Additionally, the numerator has uncensored data while the denominator has both censored and uncensored data. Expression (26) is called partial likelihood because the baseline hazard function is unspecified.

Below is a simple example. The observed event times are shown for four subjects, in which the third subject is censored: 2, 3, 4+, 5 (subject 1-4)

The partial likelihood function is expressed as:

$$L(\beta) = \frac{e^{X_1 \beta}}{e^{X_1 \beta} + e^{X_2 \beta} + e^{X_3 \beta} + e^{X_4 \beta}} \times \frac{e^{X_2 \beta}}{e^{X_2 \beta} + e^{X_3 \beta} + e^{X_4 \beta}} \times \frac{e^{X_4 \beta}}{e^{X_4 \beta}} \tag{27}$$

Expression (26) has the assumption that data has no ties. Ties are said to be present in the data when two or more subjects share the same survival time. No ties can be observed if the time is measured on a perfect continuous scale. However, in real life applications, this is unrealistic, and time is measured as a discrete variable. For instance, if the time scale is month, more than one subject can fail in the same month. We usually record: day, week, month or year of the survival times, and as a result, multiple failures can occur at the same time. There are several useful methods introduced to estimate the likelihood with tied event times. Such methods are:

1. Cox's Exact Discrete Method
2. Breslow Method
3. Efron Method

Let's describe and write the partial likelihood function for each method.

3.2.1. Cox's Exact Discrete Method

The Exact Discrete Method assumes that tie events occur from inaccurate measurement of a continuous time scale [15]. Consequently, Cox's partial likelihood function needs to be revised to handle ties. Let's introduce the revised partial likelihood function to handle ties.

Suppose time $t_1 < t_2 < \dots < t_n$ are ordered observed event times and $i = 1, 2, \dots, m$ following d_i as events occur at t_i , set of $D(t_i) = \{i_1, \dots, i_{d_i}\}$ are observations which have the event at t_i . At t_i , all the arrangements of d_i are $d_i!$ permutations assigned to Q_i and P is one element of Q_i so defined as $P = \{p_1, \dots, p_{d_i}\}$ and let's define $R(t_i; P; k) = R(t_i) - \{p_1, \dots, p_{k-1}\}$. The partial likelihood function is as follows:

$$L(\beta) = \frac{1}{d_i!} e^{x_{i+} \beta} \sum_{P \in Q_i} \prod_{k=1}^{d_i} [\sum_{j \in R(t_i; P; k)} e^{x_j \beta}]^{-1} \tag{28}$$

Where $x_{i+} = \sum_{j=1}^{d_i} x_j$. All possible arrangements of ties should be considered ([47, 32]). Now, let's illustrate how to deal with ties with example data found in table 4:

Table 4. Example of ties event.

Id	T	Event	X
1	3	1	12
2	3	1	6
3	4	0	8
4	6	1	9
5	10	1	7
6	10	1	11

Where id is subject's number, T is time in weeks, event is whether failure occurs (1 as yes, 0 as no) and X is the covariate value. In week three and ten, there are ties with two failures. Which id fails first is unknown, so all possible arrangements should be considered.

$$\Pr\{\text{ID 1 failing in week 3}\} = 12\Pr\{\text{ID 1 failing in week 3}\} = \frac{1}{2} \quad (29)$$

$$L_{12} = \frac{1}{2} \left(\frac{e^{12\beta}}{e^{12\beta} + e^{6\beta} + e^{9\beta} + e^{7\beta} + e^{11\beta}} \right) \left(\frac{e^{6\beta}}{e^{6\beta} + e^{9\beta} + e^{7\beta} + e^{11\beta}} \right) + \frac{1}{2} \left(\frac{e^{6\beta}}{e^{12\beta} + e^{6\beta} + e^{9\beta} + e^{7\beta} + e^{11\beta}} \right) \left(\frac{e^{12\beta}}{e^{6\beta} + e^{9\beta} + e^{7\beta} + e^{11\beta}} \right) \quad (32)$$

$$L_4 = \left(\frac{e^{9\beta}}{e^{9\beta} + e^{7\beta} + e^{11\beta}} \right) \quad (33)$$

$$L_{56} = \frac{1}{2} \left(\frac{e^{7\beta}}{e^{7\beta} + e^{11\beta}} \right) \left(\frac{e^{11\beta}}{e^{11\beta}} \right) + \frac{1}{2} \left(\frac{e^{11\beta}}{e^{7\beta} + e^{11\beta}} \right) \left(\frac{e^{7\beta}}{e^{7\beta}} \right) \quad (34)$$

If the number of ties is large, the Exact method can involve extensive calculations due to a large number of permutations for each tie.

3.2.2. Breslow Method

The Breslow approximation is based on the assumption that time is continuous, and the hazard rate is constant between the time interval (t_i, t_{i+1}) [48]. Another assumption is that any subject, who has a censored time in the interval (t_i, t_{i+1}) , is said to be censored at the beginning of the interval t_i . The

$$L_{12} = \frac{e^{\beta X_1} e^{\beta X_2}}{[e^{\beta X_1} + e^{\beta X_2} + e^{\beta X_4} + e^{\beta X_5} + e^{\beta X_6}]^2} = \frac{e^{12\beta} e^{6\beta}}{[e^{12\beta} + e^{6\beta} + e^{9\beta} + e^{7\beta} + e^{11\beta}]^2} \quad (37)$$

$$L_4 = \frac{e^{\beta X_4}}{e^{\beta X_4} + e^{\beta X_5} + e^{\beta X_6}} = \frac{e^{9\beta}}{e^{9\beta} + e^{7\beta} + e^{11\beta}} \quad (38)$$

$$L_{56} = \frac{e^{\beta X_5} e^{\beta X_6}}{[e^{\beta X_5} + e^{\beta X_6}]^2} \quad (39)$$

$$= \frac{e^{7\beta} e^{11\beta}}{[e^{7\beta} + e^{11\beta}]^2} \quad (40)$$

In comparison to the Exact method, the Breslow method can be much simpler to run.

3.2.3. Efron Method

The Efron method is used as the default method in R [50]. SAS, on the other hand, uses the Breslow approximation [51]. Even though Efron method is the better approach to estimating β 's than the Breslow method, nearly all statistical software uses the Breslow approximation. Implications of ties to the estimations are [52]:

1. When no ties data exist, three methods have the same results
2. When a few ties data exist, any method can be used
3. As the number of tied events increases in percent, the worse performance is obtained from Breslow
4. Efron estimations' are biased toward zero.
5. Efron method always estimates better than the Breslow method. The Efron method should be chosen if the Exact method is too time consuming.

The adjusted Cox's partial likelihood function for the Efron approximation is:

$$\Pr\{\text{ID 2 failing in week 3}\} = \frac{1}{2} \quad (30)$$

Adjusted Cox's partial likelihood function can be obtained by

$$L(\beta) = L_{12} \times L_4 \times L_{56} \quad (31)$$

Where each of L_{12}, L_4 and L_{56} can be estimated by

adjusted Cox's partial likelihood function can be obtained by:

$$LL(\beta) = \prod_{i=1}^k \frac{e^{[\sum_{j \in D_j} X_j] \beta}}{[\sum_{j \in R_i} e^{(X_j \beta)}]^{d_i}} \quad (35)$$

Where D_i is a set with the collection of failure times called d_i , at time t_i , and R_i is a risk set containing the subjects who were at risk at time t_i [49]. Using the same data from Table 4, the likelihood is:

$$LL(\beta) = L_{12} \times L_4 \times L_{56} \quad (36)$$

Where each of L_{12}, L_4 and L_{56} can be estimated by

$$L(\beta) = \prod_{i=1}^k \frac{e^{[\sum_{j \in D_j} X_j] \beta}}{\prod_{i=1}^{d_i} [\sum_{j \in R_i} e^{X_j \beta} \frac{t-1}{d_i} \sum_{j \in D_i} e^{X_j \beta}]} \quad (41)$$

Where D_i is a set with collection of failure times called d_i , at time t_i , R_i is a risk set contains the subjects who were at risk at time t_i [53]. Again, using the data found in Table 4, the likelihood is:

$$L(\beta) = L_{12} \times L_4 \times L_{56} \quad (42)$$

Where each of L_{12}, L_4 and L_{56} are estimated by: (Before defining L_{12} , let us call)

$$M_X = e^{\beta X_1} + e^{\beta X_2} + e^{\beta X_4} + e^{\beta X_5} + e^{\beta X_6} \quad (43)$$

$$L_{12} = \frac{e^{\beta X_1} e^{\beta X_2}}{M_X [M_X - \frac{1}{2}(e^{\beta X_1} + e^{\beta X_2})]} \quad (44)$$

it follows

$$M_\beta = e^{12\beta} + e^{6\beta} + e^{9\beta} + e^{7\beta} + e^{11\beta} \quad (45)$$

then

$$L_{12} = \frac{e^{\beta X_1} e^{\beta X_2}}{M_\beta [M_\beta - \frac{1}{2}(e^{\beta X_1} + e^{\beta X_2})]} \quad (46)$$

Let's find

$$L_4 = \frac{e^{\beta X_4}}{e^{\beta X_4} + e^{\beta X_5} + e^{\beta X_6}} \quad (47)$$

$$= \frac{e^{9\beta}}{e^{9\beta} + e^{7\beta} + e^{11\beta}} \tag{48}$$

$$L_{56} = \frac{e^{\beta X_5} e^{\beta X_6}}{[e^{\beta X_5} + e^{\beta X_6}][e^{\beta X_5} + e^{\beta X_6} - \frac{1}{2}(e^{\beta X_5} + e^{\beta X_6})]} \tag{49}$$

$$= \frac{e^{7\beta} e^{11\beta}}{[e^{7\beta} + e^{11\beta}][e^{7\beta} + e^{11\beta} - \frac{1}{2}(e^{7\beta} + e^{11\beta})]} \tag{50}$$

Newton-Rabson method can be used to optimize parameters of expressions 26, 28, 35 and 41 but it does not converge when there are collinearity problems in the design matrix [54].

4. Application to Telecom Data

Churn models use historical data to predict customers' behavior. The historical data set used in this study was acquired from a mobile service operator.

In this study, advertisement and marketing departments in

mobile service operator are interested in exploring the impact of selected predictors' (see table 5) on survival times (time till customer churn). With this information, they can target specific customers at risk of churn with incentives, campaigns and tariffs.

For this study, a sample of 10365 customers were randomly selected from the customers data. These customers were followed for one year from January 1st, 2015 to December 31st, 2015. Out of 10365, 2654 customers churned during this time period. It is therefore assumed that the remaining 7711 customers are right censored at the end of the time period. This churn number (2654 customers are 25.6% of total customers) is expected to have as an average monthly churn rate for top mobile service operators is 2% which indicates a loss of a quarter of their customers every year [3]. (Asian mobile service operators can lose up to 40% over a year).

In R, package called "Survival" was used to perform analysis in this data.

Table 5. Variables and descriptions.

Variable	Description
Dummy for Churn	-1: Indicates customer who churned -0: otherwise
Dummy for Campaign	-1: Indicates customer who signed up for a campaign. in January 2015 -0: Otherwise
Tariff	Mobile phone tariff which customer signed up for in January 2015. Tariffs vary by amount of mobile data, number of minutes and text. Tariff has four levels: 1, 2, 3, 4
Tenure	Customer's subscription age in months in January 2015. Minimum tenure is 6 months
Age	Customer's age in years in January 2015. Age has for categories: 1: $20 \leq age < 30$ 2: $30 \leq age < 40$ 3: $40 \leq age < 50$ 4: $50 \leq age < 60$
Auto payment	-1: Direct bank account or credit card authorization to pay monthly bills in January 2015 -0: Otherwise

4.1. Kaplan-Meier

Figure 2 illustrates survival plots of Kaplan-Meier product limit estimates of customers. Figure 3 illustrates the pdf and cdf of the survival times. The distribution is left-skewed (non-normal) since most of the churns occurred later in the

follow up and most of the non-churners have censored data at 12 months. Because of skewness, survival median time can be used as summary statistics which represents the time at which survival probability is 0.5. Since less than 50% of the customers churned, the survival median time cannot be estimated.

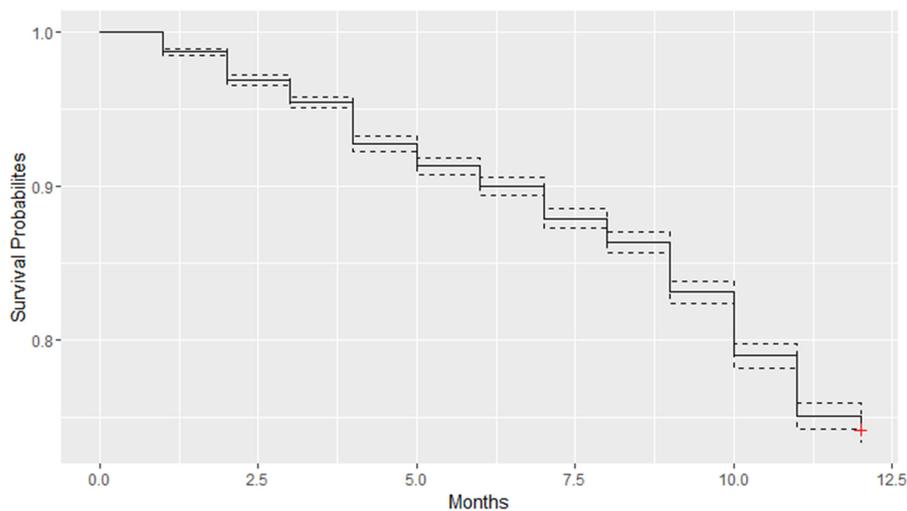


Figure 2. Kaplan Meier estimate of survival times.

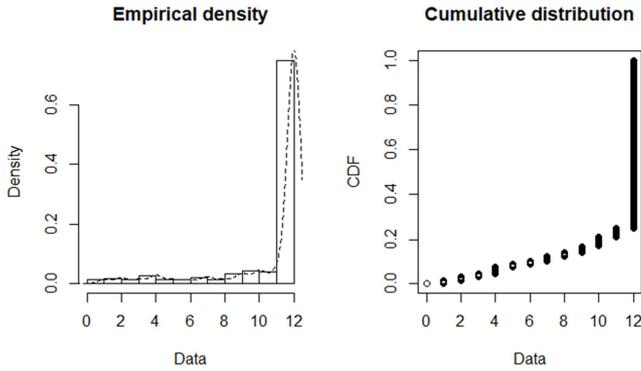


Figure 3. PDF and CDF of survival times.

Figure 4 illustrates hazard function plot of churn data set for each month.

Figure 5 illustrates the survival probabilities of Kaplan-Meier product limit estimates of two groups of customers: one with campaign and one without the campaign. This graph and log rank test result enable mobile operators to assess whether the campaign encourages customers stay.

Figure 5 shows that customers with the campaign have higher survival probabilities until 10 months. After 10 months, those without the campaign have a higher survival probability. Since two survival curves cross each other, the PHA is violated. Further analysis needs to be done with respect to the campaign variable to evaluate its effectiveness.

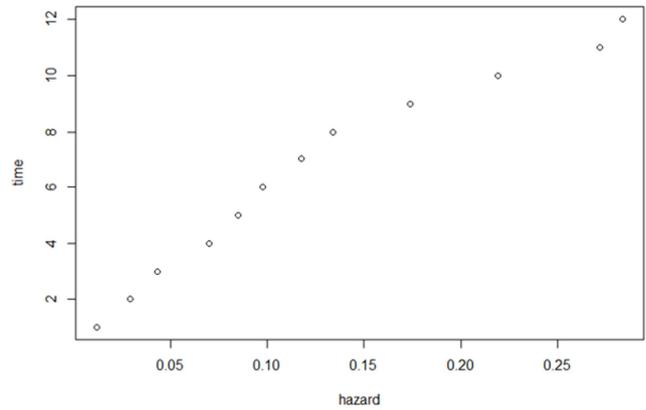


Figure 4. Hazard function of survival times.

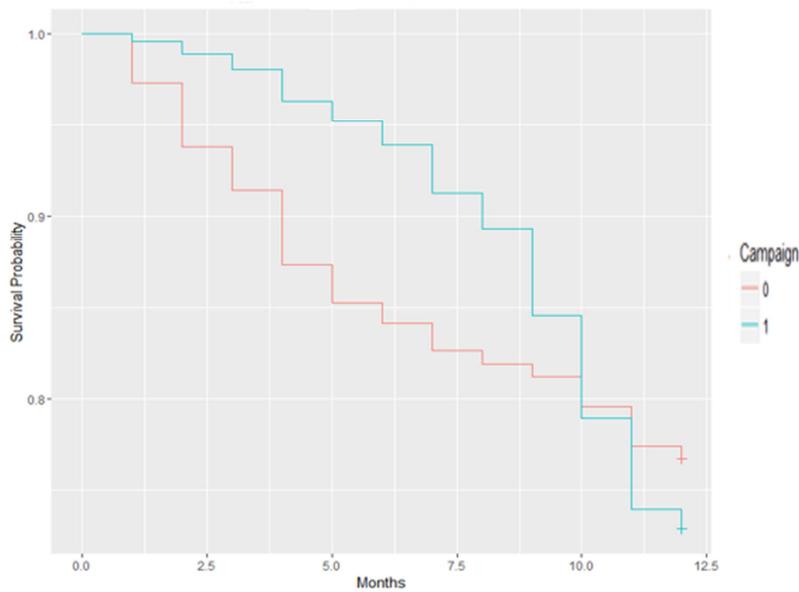


Figure 5. Kaplan Meier Estimate by campaign.

In order to quantify survival curves between two groups statistically, the log-rank test is still used 70% of the time even

though PHA does not hold [55]. Because of the test in the table 6, we conclude that there is a difference between the groups.

Table 6. Log-rank test for campaign.

	N	Observed	Expected	$(O - E)^2 / E$	$(O - E)^2 / V$
Campaign=0	4018	934	996	3.9	6.44
Campaign=1	6347	1720	1658	2.34	6.44

Chisq=6, on 1 degrees of freedom, p=0.011

Figure 6 illustrates the survival probabilities of the Kaplan-Meier product limit estimates of four groups of customers: with Tariff 1, Tariff 2, Tariff 3, Tariff 4. This graph and the log rank test results enable mobile service operator to assess how tariffs affect survival probabilities. Figure 6 shows tariff 1 has the best survival of all tariffs; Figure 6 also

approximately informs us that PHA may be hold, since no curves cross each other. But a straightforward approach is to employ numerous tests to validate PHA.

As a result of the test in the table 7, we conclude that there is a difference between the groups.

Table 7. Log-rank test for tariff.

	N	Observed	Expected	$(O - E)^2 / E$	$(O - E)^2 / V$
Tariff=1	3685	781	962	34.06	55.3
Tariff=2	2608	658	671	0.259	0.359
Tariff=3	1616	544	396	55.41	67.418
Tariff=4	2456	671	625	3.4	4.603

Chisq=96, on 3 degrees of freedom, p=0

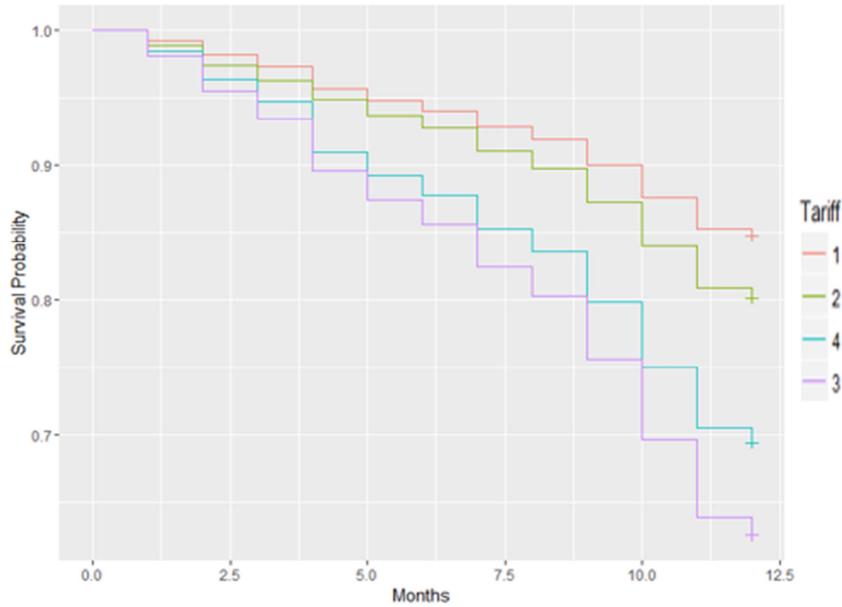


Figure 6. Kaplan Meier Estimate by tariff.

Figure 7 illustrates the survival probabilities of Kaplan-Meier product limit estimates for two stratified groups of customers: one with tenure less than or equal to one year and one with tenure greater than one year. Figure 7 shows how tenure affects survival; customers whose tenure is less than or

equal to one year have lower survival probabilities than those with more than one year. Additionally, PHA appears to be valid.

As a result of the test in the table 8, we conclude that there is a difference between the groups.

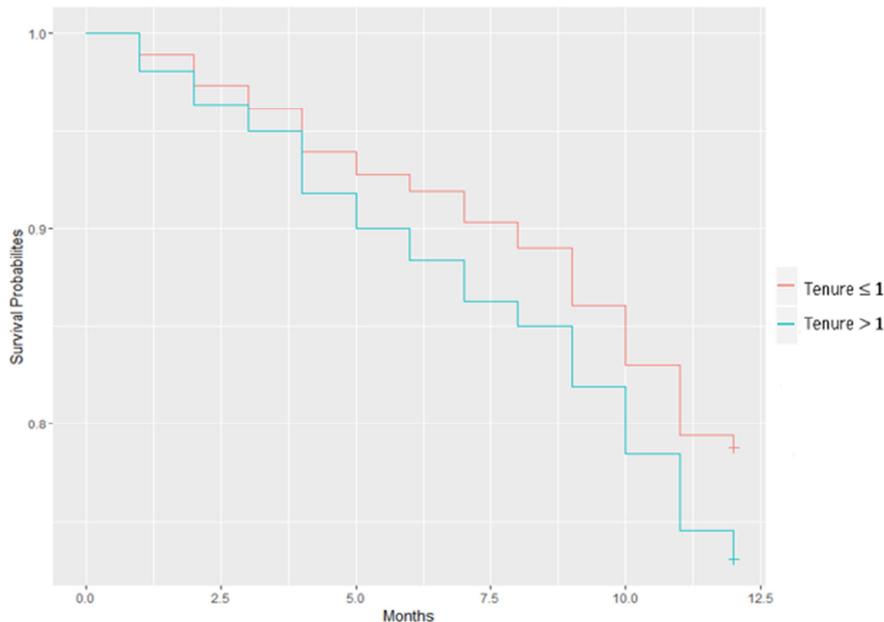


Figure 7. Kaplan Meier Estimate by tenure.

Table 8. Log-rank test for tenure.

	N	Observed	Expected	$(O - E)^2/E$	$(O - E)^2/V$
Tenure≤1	4640	2233	1046	1347	2305
Tenure>1	5725	421	1608	876	2305

Chisq=2305, on 1 degrees of freedom, p=0

4.2. The Cox Proportional Hazard Model

4.2.1. Proportional Hazard Assumption

The package “Survival” in R contains Cox.zph function was used to test proportionality of all the covariates in the model. For each covariate, Cox.zph estimates the p-values by examining the correlation between the scaled Schoenfeld residuals and time [50, 51, 56]. P-values less than 0.05 indicates that PHA is not satisfied.

According to table 9, global p-value<0.05 implies evidence of non-proportionality of the churn data set since campaign gives strong evidence of non-proportionality. Subsection 4.2.3 detailed how to handle covariates that do not satisfy the PHA. However, in

the following models, campaign was excluded from analyses [32].

Table 9. Proportional hazard assumption p-values.

Variable	rho	chisq	p
Campaign	-0.55	5.92	0.015
factor(Tariff)2	0.025	1.27	0.259
factor(Tariff)3	0.023	1.08	0.299
factor(Tariff)4	0.015	0.44	0.504
Tenure	0.006	0.08	0.765
factor(Age)2	0.026	1.36	0.259
factor(Age)3	0.009	1.17	0.299
factor(Age)4	0.003	0.02	0.88
Autopayment	0.033	2.18	0.130
Global	NA	6.59	0.01

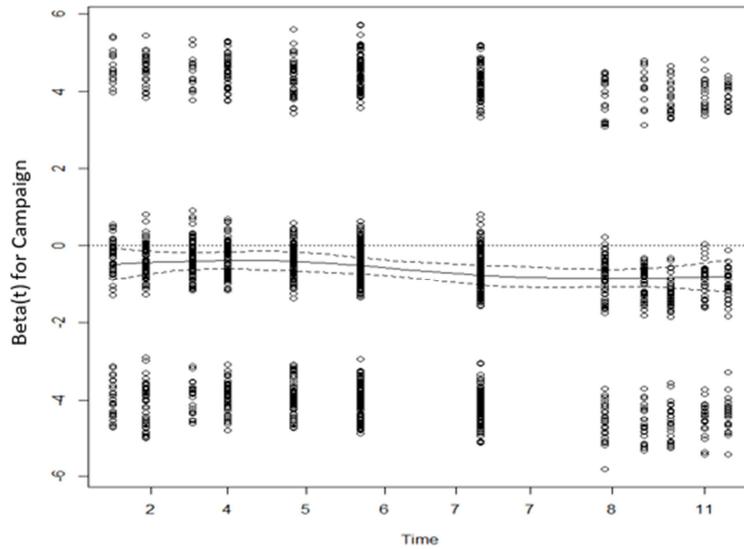


Figure 8. Campaign vs residual.

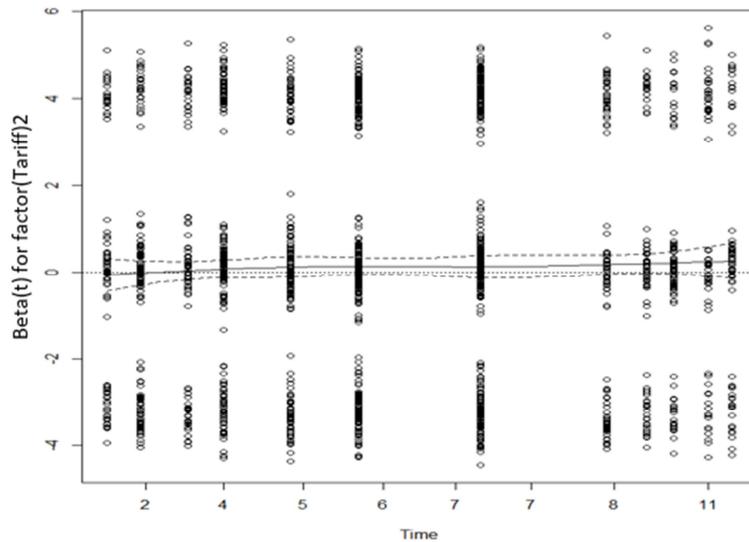


Figure 9. Tariff 2 vs residual.

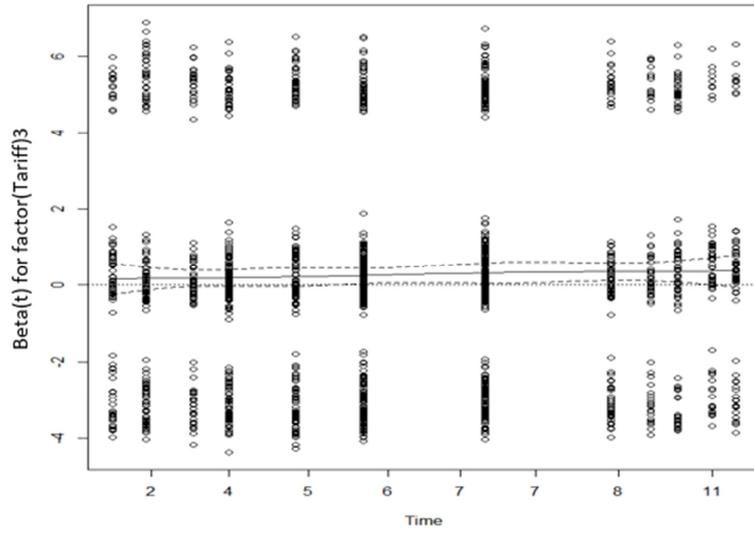


Figure 10. Tariff 3 vs residual.

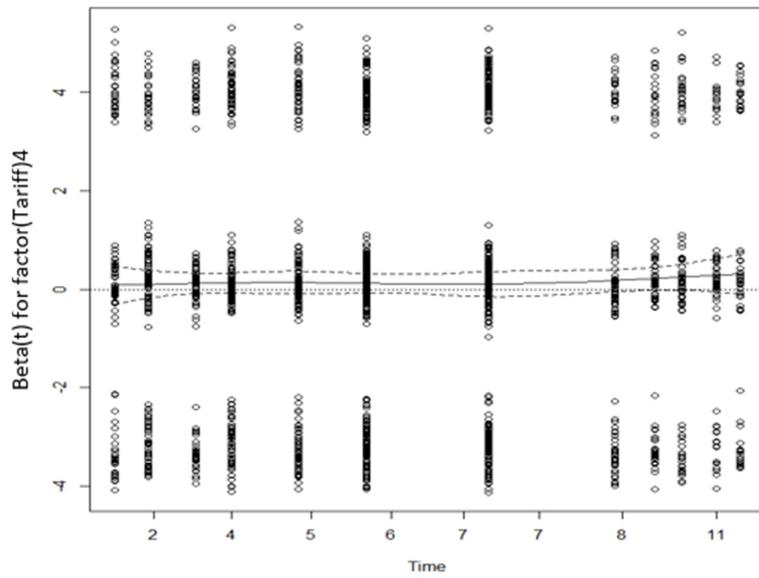


Figure 11. Tariff 4 vs residual.

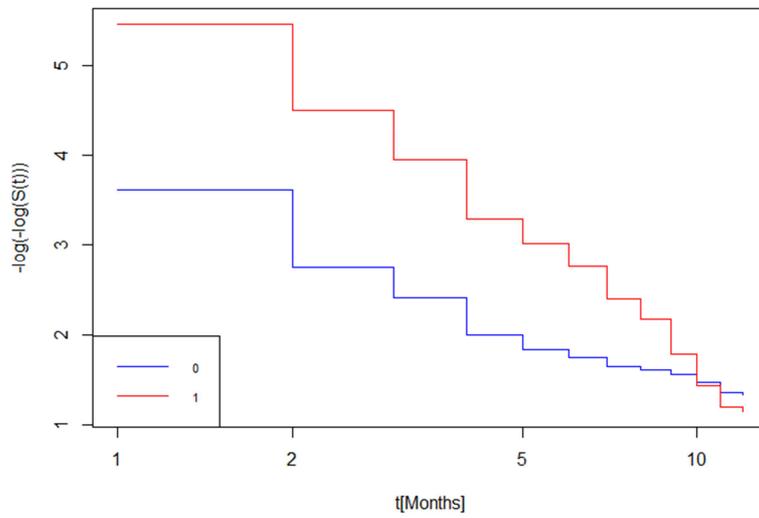


Figure 12. Campaign proportional hazard assumption.

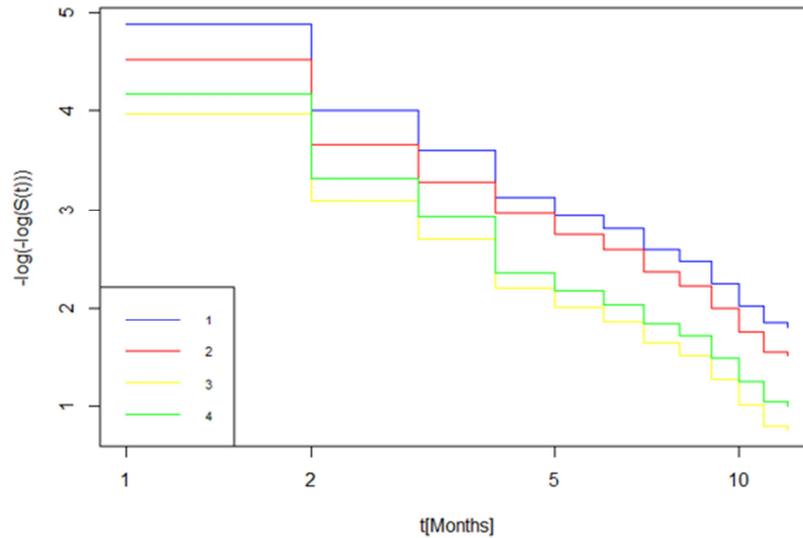


Figure 13. Tariff proportional hazard assumption.

Figure 8, which plots the calculated Schoenfeld residuals for campaign against time, further supports a violation of the PHA since the slope of the line is not zero. In addition, figure 12 plots the $-\log(-\log S(t))$ against time for the campaign predictor variable. The curves do not look parallel and cross each other which again indicates a violation of the PHA.

In the figure 9, figure 10 and figure 11, the calculated Schoenfeld residuals for tariff are plotted against time. The slope of the line is zero which is the evidence of proportionality. Furthermore, figure 13 plots the $-\log(-\log S(t))$ against time for the tariff predictor variable. The curves look parallel and do not cross, which indicates the PHA might be valid.

4.2.2. Analysis of Maximum Likelihood Estimates

Table 10 illustrates the results of two Cox models. Due to a limited amount of computer resources and a large number of ties, R could not provide any results for exact method. Table 10 shows that some coefficients are different. Not only are coefficients different in both methods, but also the AIC is lower in Efron method [56, 57].

Table 10. Maximum Likelihood Estimates for Breslow & Efron Methods.

Variable	Breslow	Efron
factor(Tariff)2	0.17	0.18
factor(Tariff)3	0.35	0.37
factor(Tariff)4	0.27	0.25
Tenure	-0.038	-0.043
factor(Age)2	-0.13	-0.112
factor(Age)3	-0.19	-0.18
factor(Age)4	-0.18	-0.165
Autopayment	-0.24	-0.223
AIC	48100	48005

Where Tariff 1 is the reference category for Tariff variable and

1: $20 \leq age < 30$ is the reference category for the age variable.

4.2.3. Result Analysis

For dealing with ties, the default Efron approximation in R was used and the results are shown in table 12. The corresponding hazard function is

$$\lambda(t|X) = \lambda_0(t)e^{\beta X}$$

$$\text{Where } \beta X = 0.18X_1 + 0.37X_2 + 0.25X_3$$

$$-0.043X_4 - 0.112X_5 - 0.18X_6$$

$$-0.165X_7 - 0.223X_8 \tag{51}$$

All the predictor variables have profound effect on the risk of churn. For tenure, the hazard ratio is 0.95, which means a one month increase in tenure, holding all other variables constant, is associated with a 5% decrease in the hazard at any time over a year. For age, holding all other variables constant, a customer in category 2 is 11% less likely to churn than a customer in category 1 at any time over a year.

Clearly, the Cox model describes the relationship between variables and the risk of churn. The hazard ratio by different levels of each predictor provides valuable information to be used by a CRM department. According to table 12, tariff 1 has the lowest hazard ratio with respect to other tariffs. This demonstrates that customers with tariff 2, 3, 4 deserve special attention.

The tenure predictor also has a significant impact on churn. As customers' tenure increases, the more inclined the customer is to stay with the current operator.

For the age predictor, the older the customer's age, the less likely they are to churn. Researchers should analyze 'why' young customers decide to churn. If the reason is high monthly bills, then this enables operators to develop new tariffs with low monthly bills for young customers. Young customers may use more data than voice. Thus, a tariff or a campaign related to data usage could be more effective than voice usage. Operators should recognize needs and preferences for different age groups to increase survival [4].

For the auto payment predictor, the customers who pay their bill automatically from their bank account have higher survival probabilities. A CRM department may or may not take corresponding actions.

In this study variables were selected to ensure the proportional hazard assumption holds. Some telecom churn predictors can change over time and were not included in this study such as number of minutes, number of calls and data usage. However, we still analyzed Schoenfeld residuals to check whether the PHA holds for each variable, and only the campaign variable failed to satisfy this assumption. Intervals of time should be used to code campaign in [58]. Table 11 is an example.

Table 11. Time dependent covariate coding in R.

Subject	Time 1	Time 2	Churn	campaign
1	1	2	0	0
1	2	3	0	0
1	3	4	1	1

Time 1 is a start time and Time 2 is the end time for the first interval in month. Between month 1 and month 2, the customer did not churn and did not have the campaign. Since campaign depends on time, campaign should be coded for every time range during the study time.

Table 12. Maximum Likelihood Estimates-Efron Method.

Variable	coef	exp(coef)	se(coef)	z	Pr(> z)
factor(Tariff)2	0.18	1.194	0.053	3.187	0.001**
factor(Tariff)3	0.37	1.45	0.055	6.66	2.74e-11***
factor(Tariff)4	0.25	1.28	0.052	3.73	0.000189***
Tenure	-0.043	0.95	0.00079	-54.488	2e-16***
factor(Age)2	-0.112	0.89	0.0609	3.551	0.000384***
factor(Age)3	-0.18	0.83	0.0566	6.255	3.97e-10***
factor(Age)4	-0.165	0.84	0.0545	8.108	5.55e-16***
Autopayment	-0.223	0.8	0.0559	-3.988	6.65e-05***

Signif.codes 0'***' 0.001 '***' 0.01 '**' 0.05 '.' 0.1 '.' 1

5. Conclusion

In liberalized mobile telecom market, churn management is very crucial. In order to stay competitive in this market, possible churners should be identified from existing customers to implement selective or personalized marketing practices.

Usually conventional data mining methods such as Logistic Regression, Neural Networks and Decision Trees are applied to identify churners for telecom churn management. These models are applied to evaluate the risk of churn for each customer for fixed time period. Further, these models identify the factors effecting customer decisions.

In this research, we presented a survival analysis technique to build predictive models to identify churners. Unlike the propensity churn scores or odds ratio, Kaplan-Meier method and the Cox model are employed to quantify and describe the risk until the customer churns. Survival curves and hazard ratios are determined for each level of a predictor which enables mobile service providers to take effective retention actions.

Acknowledgements

We are thankful to the Editor and referees, who provided expertise that greatly assisted and improved the research significantly.

References

[1] Owczarczuk, M., "Churn models for prepaid customers in the cellular telecommunication industry using large datamarts," Expert Systems with Applications 37, 4710-4712, 2010.

[2] Jahanzep, S., and Jabeen S., "Churn management in the Telecom industry of Pakistan: A comparative study of Ufone and Telenor," The Journal of Database Marketing Customer Strategy Management 14 (2), 120-129, 2007.

[3] Hung, S. Y., Yen, D. C., and Wang H. Y., "Applying data mining to telecom churn management," Expert Systems with Applications 31, 515-524, 2006.

[4] Keramati, A., and Ardabili S. M. S., "Churn analysis for an Iranian mobile operator," Telecommunications Policy 35, 344-356, 2011.

[5] Safko, L., "The Fusion Marketing Bible: Fuse Traditional Media, Social Media, & Digital Media to Maximize Marketing," McGraw-Hill Education - Europe 35, 344-356, 2011.

[6] Keramati, A., Marandi-Jafari, R., Aliannejadi, M., Ahmadian, M., Moza_ari, M. and Abbasi U., "Improving churn prediction in telecommunication industry using data mining techniques," Applied Soft Computing 24, 994-1012, 2011.

[7] Ling, R. and Yen D. C., "Customer relationship management: An analysis framework and implementation strategies," Journal of Computer Information Systems 41, 82-97, 2001.

[8] Ngai, E. W. T, Xiu, L. and Chau, D. C. K., "Application of data mining techniques in customer relationship management: A literature review and classification," Expert Systems with Applications 36 (1), 2592-2602, 2009.

[9] DeMaris, A, "Regression with Social data: Modeling Continious and Limited Response Variables," Wiley Series in Probabilty and Statistics, 381-385, 2005.

[10] Langova K., "Survival Analysis for Clinical Studies," Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub 152 (2), 303-307, 2008.

[11] Lawless, J. F., Statistical Models and Methods for Lifetime Data, John Wiley 36, 122-132, 2003.

- [12] London, D., "Survival Models and Their Estimation," Actex Publications 2, 50-75, 1998.
- [13] Jerenz, A., "Revenue Management and Survival Analysis in the Automobile Industry," Gabler Edition Wissenschaft, 70-80, 2008.
- [14] Kaplan, L. E., "Nonparametric estimation from incomplete observations," *Journal of American Statistical Association* 53, 457-81, 1958.
- [15] Cox, D. R., "Regression Model and Life-Tables," *Journal of the Royal Statistical Society* 34, 187-220, 1972.
- [16] Lee, M., "Business Bankruptcy Prediction Based on Survival Analysis Approach," *International Journal of Computer Science Information Technology* 6 (2), 2014.
- [17] Pereira, J., "Survival Analysis Employed in Predicting Corporate Failure: A Forecasting Model Proposal," *International Business Research* 7 (6), 2014.
- [18] Kumar, K., and Gepp A., "Predicting Financial Distress: A Comparison of Survival Analysis and Decision Tree Techniques," *Procedia Computer Science* 54 (2), 396-404, 2015.
- [19] Iwasaki, I., "Global financial crisis, corporate governance, and firm survival: The Russian experience," *Journal of Comparative Economics* 42 (1), 178-211, 2014.
- [20] Bu, Z., Liu, L., Cao, J. and Wu Z., "Survival analysis method for stock market prediction, Behavioral, Economic and Socio-cultural Computing," 10, 1109, 2015.
- [21] Leung, M. K., Rigby, D. and Young, T., "Entry of foreign banks in the People's Republic of China: a survival analysis," *Journal of Applied Economics* 10 21-31, 2010.
- [22] Evrensel, A. Y., "Banking crisis and financial structure: A survival-time analysis," *International Review of Economics & Finance* 17 (4), 589-602, 2008.
- [23] Tsujitani, M. and Baesens, B., "Survival Analysis for Personal Loan Data Using Generalized Additive Models," *Behaviormetrika* 39 (1), 1109, 9-23, 2012.
- [24] Baesens, B., Gestel, TV., Stepanova, M., Poel, D. V and Vanthienen J., "Neural network survival analysis for personal loan data," *Journal of the Operational Research Society* 56 (9), 1089-1098, 2005.
- [25] Singh, G., "Predicting Customer Churn in the Telecommunications Industry-An Application of Survival Analysis Modeling Using SA," *SAS Conference Papers-Sugi* 27 114 (27), 2002.
- [26] Rud, O. P., "Data Mining Cookbook," John Wiley & Sons 56 (9), 12-25, 2001.
- [27] Barrios, E. and Lansanganb J. R. G., "Forecasting Customer Lifetime Value: A Statistical Approach," *Philippine Management Review* 19, 23-34, 2012.
- [28] Ranganathan, P. and Pramesh C. S., "Censoring in survival analysis: Potential for bias," *Perspectives in Clinical Research* 3 (1), 40, 2012.
- [29] Shihcorresponding, W. J., "Problems in dealing with missing data and informative censoring in clinical trials," *Curr Control Trials Cardiovasc Med* 35 (1), 21-31, 2011.
- [30] Hosmer, JR. DW and Lemeshow S., "Applied Survival Analysis: Regression Modeling of Time to Event Data," New York: John Wiley & Sons, 1999.
- [31] Kleinbaum, D. G. and Lemeshow Klein M., "Survival Analysis: A self-learning text", Springer, 2005.
- [32] Kalbeisch, J. D. and Prentice R. L., "The Statistical Analysis Failure Time Data," A John Wiley & Sons Publication, 2002.
- [33] Lee, E. T. and Wang J. W., "Statistical Methods for Survival Data Analysis," A John Wiley & Sons Publication, 2003.
- [34] Hogg, R. V., McKean, W. J. and Craig A. T., "Introduction to Mathematical Statistics," Pearson Education, 2005.
- [35] Muldowney, P., Ostaszewski, K. and Wojdowski W., "The Darth Vader rule," *Tatra Mountains Mathematical Publications* 52 (1), 2012.
- [36] Kutner, H. M., Nachtsheim, J. C., Neter, J. and Li W., "Applied Linear Statistical Models," Macgraw-Hill 52 (1), 2005.
- [37] Kuhn, M. and Johnson K., "Applied Predictive Modeling," Springer, 2013.
- [38] Abadi, A., Yavari, P, Dehghani-Arani, M., Alavi-Majd, H., Ghasemi, E., Amanpour, F. and Bajdik C., "Cox Models Survival Analysis Based on Breast Cancer Treatments," *Iranian Journal of Cancer Prevention* 7 (3), 124-129, 2014.
- [39] Lane, W. R., Looney S. W. and Wansley J., "An application of the cox proportional hazards model to bank failure," *Journal of Banking & Finance* 10 (4), 511-531, 1986.
- [40] Therneau, T. M. and Grambsch P. M., *Modeling Survival Data: Extending the Cox Model* Presented at the 2006 Midwest SAS Users Group (MWSUG), Dearborn, Michigan, October, 22-24, 2000.
- [41] SAS Institute Inc (2008). SAS/STAT Software, Version 9.2. Cary. URL <http://support.sas.com/documentation/92/>.
- [42] Bellera, A. C., MacGrogan, G., Debled, M., Lara, C. T., Brouste, V. and Mathoulin-Pelissier S., "Variables with time-varying effects and the Cox model: Some statistical concepts illustrated with a prognostic factor study in breast cancer, *BMC Medical Research Methodology* 10 (20), 2010.
- [43] Fisher, L. D. and Lin D. Y., "Time-dependent covariates in the Cox proportional-hazards regression model," *Annual Review of Public Health*, 20 (145), 57, 1999.
- [44] Powell, T. M. MS and Bagnell M. E., "Your Survival Guide to Using Time Dependent Covariates," *SAS Global Forum*, 168, 2012.
- [45] Therneau TM (2014). survival: A Package for Survival Analysis in S. R package version 2.37-7, URL <http://CRAN.R-project.org/package=survival>.
- [46] Thomas, L., and Reyes, E. M. Tutorial: Survival Estimation for Cox Regression Model with Time-Varying Coefficients Using SAS and R, *Journal of the American Statistical Association* (<http://www.jstatsoft.org/>) 61 (1), 2014.
- [47] Borucka, J., "Methods for Handling Tied Events: In the Cox Proportional Hazard Model," *studia oeconomica posnaniensia* 2 (2), 263, 2014.
- [48] Breslov, N., "Covariance Analysis of Censored Data," *Biometrics* 30, 89: 99, 1974.

- [49] Hertz-Picciotto, I. & Rockhill, B., "Validity and Efficiency of Approximation Methods for Tied Survival Times in Cox Regression," *Biometrics* 53, 1151-1156, 1997.
- [50] R Development Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>, URL <http://www.R-project.org/>, 2011.
- [51] SAS Institute Inc., *SAS/GRAPH 9.2 Reference*, Second Edition. Cary, NC: SAS Institute Inc., SAS Institute Inc., 2010.
- [52] Allison, P. D., "Fixed Effects Regression Methods for Longitudinal Data Using SAS," SAS Institute Inc., 127-137, 2005.
- [53] Efron, B., "The efficiency of Cox's Likelihood Function of Censored Data," *Journal of the American Statistical Association* 72, 557-565, 1977.
- [54] Gonzalez, D., Pina, M. and Torres, P., "Estimation of Parameters in Cox's Proportional Hazard Model: Comparisons between Evolutionary Algorithms and the Newton-Raphson Approach Mexican International Conference on Artificial Intelligence-Springer Link, 513-523, 2008.
- [55] Li, H., Han, D., Hou, Y., Chen, H. and Chen Z., "Statistical Inference Methods for Two Crossing Survival Curves: A Comparison of Methods, *PLOS One* 10 (1), PMID: PMC4304842, 2015.
- [56] Fox, J and Weisberg S. Cox, "Proportional-Hazards Regression for Survival Data in R", An Appendix to An R Companion to Applied Regression, Second Edition, 2011.
- [57] SAS Documentantation SAS/STAT (R) 9.22 User's Guide/Procedures/The PHREG Procedure, SAS, 2010.
- [58] Therneau, T., Crowson, C. and Atkinson E., *Using Time Dependent Covariates and Time Dependent Coe_cients in the Cox Model* https://cran.r-project.org/web/packages/survival/vignettes/time_dep.pdf, 2017.
- [59] Sharma, A. and Panigrahi P. K. A, "Neural Network based Approach for Predicting Customer Churn in Cellular Network Services," *International Journal of Computer Applications* (0975-8887) 27 (11), 2011.
- [60] Lu, J, "Modeling Customer Lifetime Value Using Survival Analysis: An Application in the Telecommunications Industry SAS" User Groups International Proceedings Seattle Washington, 120 (28), 2003.