Estimation of a Change Point in the Cox Hazard Model

By

SAAD WAQAS

Department of Statistics

Faculty of Natural Sciences

Quaid-i-Azam University, Islamabad

2023

Estimation of a Change Point in the Cox Hazard Model

By

SAAD WAQAS

A THESIS SUBMITTED IN THE PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF PHILOSOPHY IN STATISTICS

Supervised By

Dr. Ismail Shah

Co-Supervised By

Dr. Sajid Ali

Department of Statistics

Faculty of Natural Sciences

Quaid-i-Azam University, Islamabad

2023

CERTIFICATE

Estimation of a Change Point in the Cox Hazard Model

By

Saad Waqas

(Reg. No. 02222013006)

A THESIS SUBMITTED IN THE PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF M.PHIL. IN

STATISTICS

We accept this thesis as conforming to the required standards

1. _____
Dr. Ismial Shah
(Supervisor)

2. _____
Prof. Dr. Muhammad Aslam
(External Examiner)

3. _____
Dr. Sajid Ali
Co-Supervisor

4 _____
Prof. Dr. Ijaz Hussain
(Chairman)

DEPARTMENT OF STATISTICS
QUAID-I-AZAM UNIVERSITY
ISLAMABAD, PAKISTAN
2023

Declaration

I "SAAD WAQAS" hereby solemnly declare that this thesis titled, "Estimation of a Change Point in the Cox Hazard Model".

- This work was done wholly in candidature for a degree of M.Phil Statistics at this University.
- Where I got help from the published work of others, this is always clearly stated.
- Where I have quoted from the work of others, the source is always mentioned. Except of such quotations, this thesis is entirely my own research work.
- Where the thesis is based on work done by myself jointly with my supervisor, I have made clear exactly what was done by others and what I have suggested

Dated: 11-10-2023 Signature: 

Dedication

I am feeling great honor and pleasure to dedicate this research work to

My grandmother

Whose endless affection, prayers and wishes have been a great source of comfort for me during my whole education period and my life

Acknowledgments

All praises to Almighty Allah (SWT), the light of Heavens and Earths, The One Who put good thoughts in one's mind, turn them into determinations, and then makes the way towards their fulfillment by showering all His Blessings throughout the journey. Best of praises and Peace be upon all the Sacred Messengers and especially for the Last of them is Hazrat Muhammad (SAWW) who is the minaret of knowledge for all the mankind.

It is a matter of great honor for me to express gratitude to my Supervisor, Dr. Ismail Shah, for his supervision and constant support. I pay my heartiest gratitude to my Co-Supervisor Dr. Sajid Ali for his endurance, patience, listening, and intellectual brought up. His generous encouragement and criticism have enabled me to produce this research work.

I offer my deepest sense of gratitude, profound respect, and tribute to all those teacher, especially Dr. Abdul Haq, Prof. Dr. Ijaz Hussain, and Dr. Manzoor Khan, who instilled within me the knowledge and guided me throughout my research.

I would like to express my heartfelt gratitude to my Parents and Family for their unwavering support, boundless encouragement, and endless sacrifices throughout my academic journey. Mr. Tariq Aziz thank you for being a pillar of strength and a source of inspiration. Your support has made a meaningful difference in my life.

I have a good time at QAU with my friends, especially Nouman Iqbal, Raheel Riasat, Ali Abbas, Qamar Ali, Waqas Munir and Samama Zafar.

(Saad Waqas)

Abstract

Survival analysis is a set of statistical methods or tools for analyzing time-to-event data. Survival analysis is a very active field of inquiry that has applications in numerous areas of study, including engineering, physical, biological, and social sciences. The hazard function represents the rate of failure at a specific time t, considering the individual has survived up to that point. The Cox proportional hazards model is widely employed in survival analysis to explore the relationship between covariates and the hazard rate. Covariates can influence the hazard function, causing it to vary, which may lead to the occurrence of a change point. This study centers around the estimation of change point in the Cox proportional hazard model, involving an examination of three distinct hazard models. We utilize maximum likelihood method to estimate change points within these proposed hazard models. Through a thorough Monte Carlo simulation study, we assess the consistency and performance of our methods as sample size and censorship percentages vary. The chronic granulotamous disease (CGD) dataset is used as a practical application and results showed that the proposed models are effective in estimating the change point.

List of Abbreviations

AFT	Accelerated Failure Time
CGD	Chronic Granulotamous Disease
CDF	Cumulative Distribution Function
MERT	Maximum Efficiency Robust Test
ML	Maximum Likelihood
MLE	Maximum Likelihood Estimation
MSE	Mean Squared Error
ME	Moments Estimation
NHS	Nurses Health Study
O-U	Ornstein-Uhlenbeck
PWP	Prentice Williams and Peterson
PDF	Probability Density Function

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Survival analysis is a collection of statistical methods used to analyze and interpret the time-to-event data. An event can be the occurrence of a disease, the failure of a mechanical component, or the death of an individual (Kleinbaum and Klein, 2012). Survival analysis is also known as time-to-event analysis or event history analysis. Survival analysis takes into account censored data, which occurs when the event of interest has not yet occurred for some individuals by the end of the study period.

1.1 Censoring and its types

Censoring in survival analysis refers to the situation where complete information about the occurrence time of an event is not available for all individuals in a study. According to Grambsch and Therneau (2000) censoring occurs when an event time is known only to occur after a given time. It is a fundamental concept in survival analysis and must be appropriately handled during data analysis to avoid having biased results with traditional methods. Different types of censoring include

- **Left censoring** occurs when a data point is known to be below a certain value, but the specific magnitude of how much it is below that threshold remains unknown.
- **Right censoring** takes place when a data point is known to be above a certain value, but the exact magnitude above that threshold is unknown.
- **Interval censoring** happens when a data point falls within an interval between two values, and the exact value within that interval is unknown.
- **Type I censoring** occurs when an experiment is conducted for a predetermined duration or with a set sample size, and the study concludes at a fixed

time. Any individuals or subjects remaining at that time are considered right censored.

- **Type II censoring** arises when an experiment is conducted for a predetermined duration or with a set sample size, and the study concludes once a predetermined number of failures are observed. The remaining individuals or subjects are then right censored.

It is also important to consider random (or non-informative) censoring, where each subject's censoring time is statistically independent of their failure time.

1.2 Hazard function

The hazard function, also known as the hazard rate or instantaneous failure rate, is a fundamental concept in survival analysis. According to Klein and Moeschberger (2003) hazard function is defined as "the rate of occurrence of an event of interest, given that the individual has survived up to a particular time" It represents the instantaneous probability of experiencing the event at a specific time, conditional on the individual being at risk up to that time. The hazard function is the ratio of the probability density function and survival function. Mathematically, $\lambda(t) = f(t)/s(t)$.

1.2.1 Cox proportional hazard model

The Cox proportional hazards model, also known as the Cox regression, was developed by Sir David Cox in 1972. Cox (1972) is widely recognized as a significant breakthrough in the analysis of time-to-event data. The paper was greeted as a milestone in the analysis of time-to-event data and received feedback from statisticians worldwide, highlighting its impact and importance in the field (Kalbfleisch and Schaubel, 2023). The Cox proportional hazards model, as described by Kleinbaum and Klein (2012), is a widely used semi-parametric regression model for analysing survival data. It assumes that the hazard function for an individual is a function of time multiplied by an unknown baseline hazard function, which is independent of the covariates. The model relates the hazard function to the covariates through a linear relationship, allowing for the estimation of the effects of the predictor variables on the hazard rate.

There is one main type of Cox regression model, which is the standard Cox proportional hazards model. However, variations and extensions of the Cox model have been developed to handle specific scenarios or address certain limitations. Here are some notable types of Cox regression models.

- **Extended Cox Model:** The extended Cox model allows for the inclusion of time-varying covariates, which are predictor variables that can change their values over time. This model accounts for the dynamic nature of the covariates and their impact on the hazard rate over time (Husain et al., 2018)

- **Startified Cox Model:** The stratified Cox model allows for the incorporation of additional stratification variables that may affect the hazard function. It stratifies the data based on these variables and estimates separate baseline hazard functions within each stratum while assuming proportional hazards within each stratum (Kleinbaum and Klein, 2012)

- **Frailty Cox Model:** The frailty Cox model incorporates random effects or frailty terms to account for unobserved heterogeneity among individuals within a group. It assumes that the hazard function is a product of an individual-specific hazard and a common baseline hazard (Grambsch and Therneau, 2000)

Marginal Cox model, cure rate model, Prentice Williams and Peterson (PWP) model are also extensions of the Cox regression model provide additional flexibility and address specific complexities encountered in survival data analysis. The choice of the appropriate model depends on the research question, data characteristics, and underlying assumptions of the analysis.

1.3 Change Point

Survival analysis involves the examination of a group of individuals who experience an event of interest over a period of time, commonly referred to as the failure event. The duration until the occurrence of this event is known as the failure time. Failure times are observed in various fields, including biology, engineering, reliability theory, and medicine. In the medical domain, individuals with chronic degenerative diseases such as cancer, diabetes, and kidney diseases often exhibit changes within short time intervals due to treatment, which are considered covariates in hazard models. The specific point at which such changes manifest in a patient is commonly known as a change point (Arenas et al., 2021).

1.3.1 Estimation of change point in Cox proportional hazard model

The estimation of change points in a Cox proportional hazard model involves identifying the time points at which significant shifts occur in the hazard function.

These change points indicate potential changes in the underlying risk factors or covariates affecting the survival outcome. The study of change points involves two essential steps: the first step is to determine the presence of a change point, while the second step involves estimating its occurrence. Estimation methods for identifying the timing of change points are typically based on parametric models such as exponential, Weibull, normal, gamma, etc., as explored by (Chen and Gupta, 2012). Various techniques, including the maximum likelihood, stochastic processes, logistic regression, Bayesian tests, and non-parametric tests, among others, have been utilized for change point detection (Chen and Gupta, 2001; Goodman et al., 2006).To determine the change point within the Cox proportional hazard model, Liang et al. (1990) employed the Ornstein-Uhlenbeck (O-U) process. Their aim was to maximize the statistical test score, taking into account the variable nature of the change point over time. O'Sullivan (1993) employed a non-parametric linear regression approach with time-dependent covariates for the Cox model estimation. Pons (2003) studied the asymptotic behavior of the partial likelihood function to obtain estimates of the change point. Liu et al. (2008) proposed a Monte Carlo method for maximizing score functions to determine test thresholds, offering an alternative to the O-U process. Zhang et al. (2010) introduced a semi-parametric approach based on Splines for the maximum likelihood estimation. Li et al. (2013) focused on the parameter estimation problem for a piecewise hazard regression model with one change point, considering covariate effects and right censoring through maximum likelihood. Palmeros et al. (2018) recently proposed a method for calculating the change point in the Weibull regression model, with a specific emphasis on the log-likelihood function. Arenas et al. (2021) conducted an estimation of the change point in the Cox proportional hazard model using the maximum likelihood estimation and moments estimation (ME).

Because the change point problem arises in many practical situations, efficient change point estimation methods are needed. In this study, we offer two methods for the estimation of change point in the Cox proportional hazard model on three different hazard models. First method covers the estimation of the change point when other parameters of the Cox proportional hazard model like β and θ are not in closed form whereas in second method their MLEs are available in closed form. A Monte Carlo simulation study is performed for assessing the consistency of the proposed estimation methods in terms of the mean squared error (MSE).

1.4 Thesis objectives

The main objectives of the study are

1. To propose Cox proportional hazard model with different intensities in different change point situations.

2. To check the consistency of proposed estimation methods for change point estimation in the Cox proportional hazard model.

3. To access the performance of estimation methods under different levels of censorship.

The structure of this thesis is as follows: Chapter 2 provides a comprehensive review of the literature. Chapter 3 presents the methodology and results derived from the simulation study. The concluding remarks can be found in Chapter 4.

Chapter 2

Literature Review

Arenas et al. (2021) presented a study on the computational methods for estimating a change point in the Cox hazard model. In this work, a novel approach combines moment estimation (ME) with the maximum likelihood estimation to estimate the change point. It uses a numerical method to minimize an objective function provided by ME. A Monte Carlo simulation using various situations provides the mean squared error of the estimator.

Palmeros et al. (2018) computed estimates of a change point in the Weibull regression hazard model. Covariates and censored variables are used in this study to estimate a change point in the Weibull regression hazard model, an extension of the exponential model. A Monte Carlo simulation study indicates that this model can really be applied in practice.

Pons (2002) presented a study on estimation in a Cox regression model with a change point at an unknown time. The asymptotic properties of the maximum likelihood estimators of the parameters in a non-regular Cox model with a change-point in the regression on time-dependent covariates are examined in this study. Overall consistency is derived from the uniform convergence of the partial log-likelihood.

Liu et al. (2008) studied a Monte Carlo approach for change point detection in the Cox proportional hazards model. The maximal score tests were used for detecting change points in the Cox proportional hazards model with the censored data. The proposed method can be used to test a single change point in the Cox model with covariates and sample stratifications over a wide range of candidate regions, like discrete time-point sets or disjoint intervals. The proposed test statistics and Monte Carlo procedure are well suited to situations involving multiple change points.

Gandy et al. (2005) reported a study on a Cox model with a change point applied to an actuarial problem. They used survival analysis methods on an actuarial dataset to identify significant covariates for contract cancellation. The

well-known risk models suggested by the Cox (1972) and Aalen (1980) were used in their approach but the result suggests that the functional form of a covariate was misspecified. Then, a new variant of the Cox model with a change point at an unknown threshold was proposed that proved the consistency of the estimators.

Wang et al. (2021) worked on change point detection in the Cox proportional hazards mixture cure model, In which the covariate effects on the distribution of uncured subjects' failure time may jump when one of the covariates exceeds a change point. The semiparametric estimates are obtained using nonparametric maximum likelihood estimation. To implement the estimation a two-step computational procedure involving the expectation-maximization algorithm and finite sample performance is demonstrated using simulation studies and real data examples.

Dupuy (2006) studied estimation in a change point hazard regression model. The author take into account a parametric survival regression model with a change point in both the hazard and regression parameters. Estimators of the regression parameters, change point, and hazard are proposed and shown to be consistent.

Li et al. (2013) presented a study on estimation in a change point hazard regression model with long-term survivors. In the presence of right censoring and long-term survivors, they estimated the change point for a piecewise hazard regression model. The consistency of the maximum likelihood estimators of the change point and other parameters is demonstrated. The proposed method is demonstrated by analyzing data on kidney infection recurrence.

Yao (1986) worked on the maximum likelihood estimation in hazard rate models with a change point. The problem of parameter estimation in hazard rate models with a change point is explained. The likelihood function in this problem is unbounded, which is an interesting feature. A maximum likelihood change point estimator subject to a natural constraint is proposed and shown to be consistent. The limiting distributions are derived as well.

Fotopoulos and Jandhyala (2001) reported a study on the maximum likelihood estimation of a change point for the exponentially distributed random variables. The problem of estimating the unknown change point in the parameter of a sequence of independent and exponentially distributed random variables's considered. The analysis is based on the use of the Weiner-Hopf factorization identity, which involves the distribution of ascending and descending ladder heights, as well as the renewal measure in random walks.

Fu and Curnow (1990) presented a study on the maximum likelihood estimation of multiple change points. If there are known lower bounds on the lengths of the sub-sequences between the change points, the maximum likelihood estimation of the locations of changes in sequences of independent categorical random variables is examined. A method is developed which finds the maximum likelihood solution.

The method also allows the boundary distributions of the changed segments to differ from the distribution of the changed segments' central region.

Matthews and Farewell (1982) studied testing for a constant hazard against a change point alternative. They constructed a test of a constant failure rate against the alternative of a failure rate involving a single change point. For the given alternative, a likelihood ratio test is derived and simulated. Tests based on alternatives in the log gamma family are also taken into consideration since they function effectively when the change-point model is accurate.

Lee et al. (2020) reported a study on testing for change point in the covariate effects based on the Cox regression model. A Cox model with a change point in covariate is considered, and the pattern of the change point effects can be specified in a variety of ways. Three statistical tests, the maximal score, the maximal normalized score, and the maximal Wald tests are suggested to examine the possibility of change-point effects. Monte Carlo approaches to simulate the critical values are suggested.

Liang et al. (1990) studied an epidemiologic application using the Cox proportional hazards model with change point. The model possesses two features, different relative risk parameters are allowed for early versus late onset of the disease of interest, and an additional parameter is introduced so that specification is not required for the time (age) at which a change in the magnitude of the relative risks takes place. They used a set of data on a group of white male medical students who attended The Johns Hopkins Medical School between 1948 and 1964 to apply the model.

Zucker et al. (2013) presented a study on testing for a change point in the Cox survival regression model. They examined testing for a threshold effect in the case when a potential threshold value is unknown. They considered a maximum efficiency robust test (MERT) of linear combination form and supremum type tests. They discussed the relevant theory, conduct a simulation study comparing the power of various test statistics, and demonstrate the use of the tests on data from the Nurses Health Study (NHS) relating to the relationship between chronic exposure to particulate matter with a diameter of 10 m or less (PM10) and fatal myocardial infarction.

Husain et al. (2018) worked on the application of the extended Cox proportional hazard method for estimating the survival time of breast cancer. The main objective was to model the various factors influencing both the survival time and the rate of cure in breast cancer patients. They utilized the extended Cox model, which is a modified version of the proportional hazard Cox model designed to account for situations where the proportional hazard assumptions are not satisfied. The researchers employed the maximum likelihood estimation approach to estimate the

model's parameters.

Zhang et al. (2010) suggested a semiparametric maximum likelihood method based on splines to analyse the Cox model using interval-censored data. They introduced a straightforward approach to consistently estimate the standard error of the regression parameter, making it easier to perform inference for the Cox model with interval-censored data. Their proposed method facilitates accurate estimation and inference procedures for this specific type of data.

Goodman et al. (2006) introduced the inclusion of multiple change points. They put forward a model selection technique utilizing sequential testing, which encompasses two types of models: the piecewise constant hazard model and the piecewise linear hazard model. These approaches are driven by the data and enable estimation of both the overall trend in the hazard function and the specific locations where changes in the trend occur.

Chapter 3

Changepoint Estimation

The background of the Cox proportional hazard model with its different extensions is discussed in section 1.2.1. Cox (1972) introduced a diverse range of survival models that specifically target the hazard function. Among these models, the most basic one is known as the proportional hazard model. where the hazard at time t for an individual with covariates vector $\mathbf{z}$ is assumed to be

$$\lambda(t|z) = \lambda_0(t)\exp(\beta' z)$$

In this model, $\lambda_0(t)$ is a baseline hazard function that describes the risk for individuals with $z = 0$, and $\exp(\beta' z)$ is a proportionate increase or decrease in risk, associated with the set of covariates z. Arenas et al. (2021) studied the hazard model as follows

$$\lambda(t; \mathbf{Z}) = \lambda_0(t)\exp[\{\beta + \theta \mathbf{I}(t \leq \tau)\}\mathbf{Z}],$$

where $\lambda_0(t)$ is a unspecified baseline hazard function, $\tau > 0$ is the change point parameter, $\beta \in \mathbf{R}^k$ is the vector of regression coefficients, $\theta \in \mathbf{R}^k$ denotes the effect of change due to treatment, and $\mathbf{Z} \in \mathbf{R}^k$ is the covariate vector.

We use the setting of Kleinbaum and Klein (2012) who considered a clinical trial with a total of n individuals each of whom is assigned a random failure time represented by $\tilde{T}_i$ and their censorship time C_i. For the i-th individual, the observations on their failure time consist of two components T_i and δ_i, $T_i = \min(\tilde{T}_i, C_i)$ and $\delta_i = I(\tilde{T}_i \leq C_i)$, with $\delta_i = 1$ if the failure event has been observed, and $\delta_i = 0$.

To derive the probability density function (PDF), cumulative distribution function (CDF), and survival function from proposed hazard functions, we use the following mathematical relationships

Let $h(t)$ denote the hazard function, to find the cumulative hazard function we can integrate the hazard function over time. The cumulative hazard function,

denoted by $H(t)$, represents the cumulative risk or cumulative failure rate up to time t.

$$H(t) = \int_0^t h(u)du, t > 0$$

The survivor function provides information about the probability that the specific event has not happened up to a given time point t. If we consider T as the time until failure, S(t) represents the probability of surviving beyond time t.

$$S(t) = \exp(-H(t))$$

then the CDF is expressed as

$$F(t) = 1 - S(t)$$

To obtain PDF we use this relationship

$$f(t) = h(t)S(t)$$

3.1 Model 1

In this section, we present certain well-known properties associated with the Cox proportional hazard model. The first hazard model under consideration is presented as follows

$$\lambda(t; \mathbf{Z}) = \lambda_0(t) \exp[\{\beta + \theta^2 \mathbf{I}(t \leq \tau)\}\mathbf{Z}] \tag{3.1}$$

The expression for the hazard function in equation 3.1 can be stated as follows

$$\lambda(t; \mathbf{Z}) = \lambda_0(t) \exp\{(\beta + \theta^2)\mathbf{Z}\}\mathbf{I}(t \leq \tau) + \lambda_0(t) \exp\{\beta \mathbf{Z}\}\mathbf{I}(t > \tau) \tag{3.2}$$

where the corresponding survival function is

$$\begin{aligned} S(t; \mathbf{Z}) = &\exp\Big(-t\lambda_0(t) \exp\{(\beta + \theta^2)\mathbf{Z}\}\Big)\mathbf{I}(t \leq \tau) + \\ &\exp\Big(-\tau\lambda_0(t) \exp\{(\beta + \theta^2)\mathbf{Z}\} - (t - \tau)\lambda_0(t) \exp\{\beta \mathbf{Z}\}\Big)\mathbf{I}(t > \tau) \end{aligned} \tag{3.3}$$

From equations 3.2 and 3.3, the PDF is

$$\begin{aligned} f(t; \mathbf{Z}) = &\Big(\lambda_0(t) \exp\{(\beta + \theta^2)\mathbf{Z}\} \exp(-t\lambda_0(t) \exp\{\beta + \theta^2)\mathbf{Z}\}\Big)\mathbf{I}(t \leq \tau) + \\ &\Big(\lambda_0(t) \exp\{\beta \mathbf{Z}\} \exp(-\tau\lambda_0(t) \exp\{\beta + \theta^2)\mathbf{Z}\} - \\ &(t - \tau)\lambda_0(t) \exp\{\beta \mathbf{Z}\})\Big)\mathbf{I}(t > \tau) \end{aligned} \tag{3.4}$$

Taking into account the assumption of independent censoring for the data, the log-likelihood function associated with equation 3.4 can be expressed as follows

$$\begin{aligned} l(\theta,\beta) = &\sum_{i}^{n}\Big[\delta_i \log(\lambda_0(t_i)) + \delta_i(\beta+\theta^2)\mathbf{Z}_i - t_i\lambda_0(t_i)\exp\{(\beta+\theta^2)\mathbf{Z}_i\}\Big]\mathbf{I}(t_i \leq \tau) + \\ &\sum_{i}^{n}\Big[\delta_i \log\lambda_0(t_i) + \delta_i\beta\mathbf{Z}_i - \tau\lambda_0(t_i)\exp\{(\beta+\theta^2)\mathbf{Z}_i\} - \\ &(t_i-\tau)\lambda_0(t_i)\exp\{\beta\mathbf{Z}_i\}\Big]\mathbf{I}(t_i > \tau) \end{aligned} \tag{3.5}$$

To simplify the model and reduce computational time, we have assumed that Z is a dichotomous variable that takes the values 0 or 1. This simplification allows for easier implementation of the proposed method. The log-likelihood function given in equation 3.5 is reduced for $z_i = 1$

$$\begin{aligned} l(\theta,\beta) = &\sum_{i}^{n}\Big[\delta_i \log\lambda_0(t_i) + \delta_i(\beta+\theta^2) - t_i\lambda_0(t_i)\exp(\beta+\theta^2)\Big]\mathbf{I}(t_i \leq \tau) + \\ &\sum_{i}^{n}\Big[\delta_i \log\lambda_0(t_i) + \delta_i\beta - \tau\lambda_0(t_i)\exp(\beta+\theta^2) - \\ &(t_i-\tau)\lambda_0(t_i)\exp(\beta)\Big]\mathbf{I}(t_i > \tau) \end{aligned} \tag{3.6}$$

3.2 Model 2

In this section, we will describe the properties of our second proposed hazard model. The second hazard model that we are considering is

$$\lambda(t;\mathbf{Z}) = \lambda_0(t)\exp[\{\beta\theta\mathbf{I}(t \leq \tau)\}\mathbf{Z}] \tag{3.7}$$

The hazard function in equation 3.7 can be formulated as follows.

$$\lambda(t;\mathbf{Z}) = \lambda_0(t)\exp\{(\beta\theta\mathbf{Z}\}\mathbf{I}(t \leq \tau) + \lambda_0(t)\exp\{\beta\mathbf{Z}\}\mathbf{I}(t > \tau) \tag{3.8}$$

where the associated survival function is

$$\begin{aligned} S(t;\mathbf{Z}) = &\exp\Big(-t\lambda_0(t)\exp\{\beta\theta\mathbf{Z}\}\Big)\mathbf{I}(t \leq \tau) + \\ &\exp\Big(-\tau\lambda_0(t)\exp\{\beta\theta\mathbf{Z}\} - (t-\tau)\lambda_0(t)\exp\{\beta\mathbf{Z}\}\Big)\mathbf{I}(t > \tau) \end{aligned} \tag{3.9}$$

From equations 3.8 and 3.9, the PDF is

$$\begin{aligned} f(t;\mathbf{Z}) =& \Big(\lambda_0(t)\exp\{\beta\theta\mathbf{Z}\}\exp(-t\lambda_0(t)\exp\{\beta\theta\mathbf{Z}\}\Big)\mathbf{I}(t\leq\tau)+ \\ & \Big(\lambda_0(t)\exp\{\beta\mathbf{Z}\}\exp(-\tau\lambda_0(t)\exp\{\beta\theta\mathbf{Z}\}- \\ & (t-\tau)\lambda_0(t)\exp\{\beta\mathbf{Z}\})\Big)\mathbf{I}(t>\tau) \end{aligned} \tag{3.10}$$

Considering the assumption of independent censoring for the data, the log-likelihood function related to equation 3.10 can be stated as

$$\begin{aligned} l(\theta,\beta) =& \sum_i^n \Big[\delta_i\log(\lambda_0(t_i)) + \delta_i\beta\theta\mathbf{Z}_i - t_i\lambda_0(t_i)\exp\{\beta\theta\mathbf{Z}_i\}\Big]\mathbf{I}(t_i\leq\tau)+ \\ & \sum_i^n \Big[\delta_i\log\lambda_0(t_i) + \delta_i\beta\mathbf{Z}_i - \tau\lambda_0(t_i)\exp\{\beta\theta\mathbf{Z}_i\}- \\ & (t_i-\tau)\lambda_0(t_i)\exp\{\beta\mathbf{Z}_i\}\Big]\mathbf{I}(t_i>\tau) \end{aligned} \tag{3.11}$$

As we assumed that Z is a dichotomous variable that takes the values 0 or 1, the log-likelihood function given in equation 3.11 is reduced for $z_i = 1$

$$\begin{aligned} l(\theta,\beta) =& \sum_i^n \Big[\delta_i\log\lambda_0(t_i) + \delta_i\beta\theta - t_i\lambda_0(t_i)\exp(\beta\theta)\Big]\mathbf{I}(t_i\leq\tau)+ \\ & \sum_i^n \Big[\delta_i\log\lambda_0(t_i) + \delta_i\beta - \tau\lambda_0(t_i)\exp(\beta\theta)- \\ & (t_i-\tau)\lambda_0(t_i)\exp(\beta)\Big]\mathbf{I}(t_i>\tau) \end{aligned} \tag{3.12}$$

3.3 Model 3

Matthews and Farewell (1982) introduced a simple model where the hazard rate function piecewise with a single change point that is

$$\lambda(t) = \beta + \theta\mathbf{I}(t\geq\tau) \tag{3.13}$$

Here, we considered the following extension of above mentioned hazard model

$$\lambda(t;\mathbf{Z}) = \lambda_0(t)\exp(\beta\mathbf{Z})\mathbf{I}(t\leq\tau) + \lambda_0(t)\exp\{(\beta+\theta)\mathbf{Z}\}\mathbf{I}(t>\tau) \tag{3.14}$$

The corresponding survival function is given as

$$\begin{aligned} S(t;\mathbf{Z}) =& \exp\Big(-t\lambda_0(t)\exp(\beta\mathbf{Z})\Big)\mathbf{I}(t\leq\tau)+ \\ & \exp\Big(-\tau\lambda_0(t)\exp(\beta\mathbf{Z}) - (t-\tau)\lambda_0(t)\exp\{(\beta+\theta)\mathbf{Z}\}\Big)\mathbf{I}(t>\tau) \end{aligned} \tag{3.15}$$

From equations 3.14 and 3.15, the PDF becomes

$$\begin{aligned} f(t;\mathbf{Z}) =& \Big(\lambda_0(t)\exp(\beta\mathbf{Z})\exp(-t\lambda_0(t)\exp(\beta\mathbf{Z})\Big)\mathbf{I}(t\leq\tau)+ \\ & \Big(\lambda_0(t)\exp\{(\beta+\theta)\mathbf{Z}\}\exp(-\tau\lambda_0(t)\exp(\beta\mathbf{Z})- \\ & (t-\tau)\lambda_0(t)\exp\{(\beta+\theta)\mathbf{Z}\})\Big)\mathbf{I}(t>\tau) \end{aligned} \tag{3.16}$$

Given the assumption of independent censoring for the data, we can express the log-likelihood function for equation 3.16 in the following manner

$$\begin{aligned} l(\theta,\beta) =& \sum_i^n \Big[\delta_i\log(\lambda_0(t_i)) + \delta_i\beta\mathbf{Z}_i - t_i\lambda_0(t_i)\exp(\beta\mathbf{Z}_i)\Big]\mathbf{I}(t_i\leq\tau)+ \\ & \sum_i^n \Big[\delta_i\log\lambda_0(t_i) + \delta_i(\beta+\theta)\mathbf{Z}_i - \tau\lambda_0(t_i)\exp(\beta\mathbf{Z}_i)- \\ & (t_i-\tau)\lambda_0(t_i)\exp\{(\beta+\theta)\mathbf{Z}_i\}\Big]\mathbf{I}(t_i>\tau) \end{aligned} \tag{3.17}$$

From log-likelihood function in equation 3.17, the following score equations are obtained

$$\frac{\partial l(\theta,\beta)}{\partial\theta} = \sum_i^n \Big[\delta_i\mathbf{Z}_i - (t_i-\tau)\lambda_0(t_i)\exp\{(\beta+\theta)\mathbf{Z}_i\}\Big]\mathbf{I}(t_i>\tau) = 0 \tag{3.18}$$

$$\begin{aligned} \frac{\partial l(\theta,\beta)}{\partial\beta} =& \sum_i^n \Big[\delta_i\mathbf{Z}_i - t_i\lambda_0(t_i)\exp(\beta\mathbf{Z}_i)\Big]\mathbf{I}(t_i\leq\tau)+ \\ & \sum_i^n \Big[\delta_i\mathbf{Z}_i - \tau\lambda_0(t_i)\exp(\beta\mathbf{Z}_i)- \\ & (t_i-\tau)\lambda_0(t_i)\exp\{(\beta+\theta)\mathbf{Z}_i\}\Big]\mathbf{I}(t_i>\tau) = 0 \end{aligned} \tag{3.19}$$

Also, the score equations given in 3.18 and 3.19 are reduced to the following equations for $Z_i = 1$, because it was assumed that Z is a dichotomous variable that takes the values 0 or 1.

$$\frac{\partial l(\theta,\beta)}{\partial\theta} = \sum_i^n \Big[\delta_i - (t_i-\tau)\lambda_0(t_i)\exp(\beta+\theta)\Big]\mathbf{I}(t_i>\tau) = 0 \tag{3.20}$$

$$\begin{aligned} \frac{\partial l(\theta,\beta)}{\partial\beta} =& \sum_i^n \Big[\delta_i - t_i\lambda_0(t_i)\exp(\beta)\Big]\mathbf{I}(t_i\leq\tau)+ \\ & \sum_i^n \Big[\delta_i - \tau\lambda_0(t_i)\exp(\beta) - (t_i-\tau)\lambda_0(t_i)\exp(\beta+\theta)\Big]\mathbf{I}(t_i>\tau) = 0 \end{aligned} \tag{3.21}$$

In this scenario, it is possible to solve equations 3.20 and 3.21 simultaneously to find the maximum likelihood estimators for β and θ

$$\hat{\beta} = \log\left[\frac{\sum_i^n \delta_i \mathbf{I}(t_i \leq \tau)}{\lambda_0(t_i)\sum_i^n(t_i)\mathbf{I}(t_i \leq \tau) + \tau\lambda_0(t_i)\sum_i^n \mathbf{I}(t_i > \tau)}\right] \tag{3.22}$$

$$\hat{\theta} = \log\left[\frac{\Big(\lambda_0(t_i)\sum_i^n(t_i)\mathbf{I}(t_i \leq \tau) + \tau\lambda_0(t_i)\sum_i^n \mathbf{I}(t_i > \tau)\Big)\Big(\sum_i^n \delta_i \mathbf{I}(t_i > \tau)\Big)}{\lambda_0(t_i)\Big(\sum_i^n(t_i - \tau)\mathbf{I}(t_i > \tau)\Big)\Big(\sum_i^n \delta_i \mathbf{I}(t_i \leq \tau)\Big)}\right] \tag{3.23}$$

3.4 Simulation study

We use the Monte Carlo simulation method to study the proposed methodology. Three sets of values are used for the parameters θ, β, and τ to generate failure times for three proposed hazard models. Furthermore, we randomly assigned censorship times and covariate values to the failure times from a binomial distribution. Four percentages of censorship were considered 0%, 20%, 50% and 70% along with four sample sizes $n = 50, 200, 500, 1000$. To obtain $\hat{\beta}$, $\hat{\theta}$ and $\hat{\tau}$, 5000 iterations are performed for each sample size at each level of censorship. Function $\lambda_0(t)$ was taken to be a constant equal to 0.5 as used in Liu et al. (2008). A characteristic of the Cox proportional hazard model is that in estimating regression coefficients, it is not necessary to know the baseline hazard function $\lambda_0(t)$.

3.5 Results

In this section, the estimated values of $\hat{\beta}$, $\hat{\theta}$, and $\hat{\tau}$ along with their MSE are shown in Tables 3.1-3.15 under different settings. The model parameters are estimated by the maximum likelihood estimation method. For Model 1 and Model 2, we maximize the log-likelihood functions given in the equation 3.6 and 3.12, respectively, by using the 'optim' function of R Core Team (2016). For Model 3, first we obtained the MLEs of β and θ from equation 3.22 and 3.23. Then $\hat{\tau}$ is found by using these values of $\hat{\beta}$ and $\hat{\theta}$, we maximize log-likelihood functions given in equation 3.17 for $z_i = 1$.

Table 3.1: Estimated MSE for Model 1 with different percentages of censorship and different values of β, θ, τ for $n = 50$

C	β	θ	τ	$\hat{\beta}$	MSE($\hat{\beta}$)	$\hat{\theta}$	MSE($\hat{\theta}$)	$\hat{\tau}$	MSE($\hat{\tau}$)
0%	-1	1	1.3862	-1.0089	0.0410	1.0149	0.0200	1.3667	0.0174
	-1	1	0.5099	-1.0014	0.0255	1.0341	0.0375	0.4935	0.0110
	-1	1	0.1876	-1.0046	0.0326	1.1126	0.0580	0.1722	0.0067
	-1	2	1.3862	-0.9889	0.0043	1.9989	0.0007	1.4996	0.0160
	-1	2	0.5099	-0.9635	0.4128	1.9895	0.0314	0.5977	0.0160
	-1	2	0.1876	-0.9816	0.1196	2.0018	0.0085	0.1798	0.0004
	-1	3	1.3862	-0.9832	0.0088	2.9977	0.0003	1.5513	0.0336
	-1	3	0.5099	-0.9832	0.0090	2.9978	0.0003	0.6738	0.0330
	-1	3	0.1876	-0.9834	0.0068	2.9986	0.0003	0.1829	0.0368
20%	-1	1	1.3862	-1.2529	0.1137	1.0185	0.0354	1.3841	0.0269
	-1	1	0.5099	-1.2343	0.0885	1.0367	0.0570	0.5005	0.0198
	-1	1	0.1876	-1.2371	0.0906	1.1437	0.0773	0.1675	0.0111
	-1	2	1.3862	-1.0004	0.0018	1.9502	0.0042	1.5414	0.0260
	-1	2	0.5099	-0.9783	0.4492	1.9550	0.0939	0.6344	0.0252
	-1	2	0.1876	-1.1206	0.2048	1.9973	0.0143	0.1775	0.0006
	-1	3	1.3862	-0.9984	0.0014	2.9615	0.0020	1.5747	0.0486
	-1	3	0.5099	-0.9983	0.0014	2.9612	0.0020	0.6982	0.0486
	-1	3	0.1876	-0.9984	0.0012	2.9610	0.0020	0.1930	0.0484
50%	-1	1	1.3862	-1.7530	0.6718	1.0320	0.1640	1.5098	0.1189
	-1	1	0.5099	-1.7259	0.5984	1.0387	0.1078	0.5961	0.0756
	-1	1	0.1876	-1.8498	0.6460	1.2598	0.1128	0.1656	0.0308
	-1	2	1.3862	-0.9787	0.0084	1.7978	0.0415	1.6239	0.0575
	-1	2	0.5099	-0.9924	0.1519	1.8299	0.1744	0.6895	0.0488
	-1	2	0.1876	-1.5832	0.7106	1.9706	0.0546	0.1724	0.0015
	-1	3	1.3862	-1.0040	0.0080	2.8850	0.0130	1.6539	0.0943
	-1	3	0.5099	-1.0042	0.0083	2.8850	0.0131	0.7779	0.0950
	-1	3	0.1876	-1.0042	0.0081	2.8813	0.0143	0.2204	0.0952
70%	-1	1	1.3862	-2.3169	2.0048	1.0475	0.7123	1.6575	0.3404
	-1	1	0.5099	-2.2749	1.7876	1.0458	0.2078	0.7879	0.2068
	-1	1	0.1876	-2.4500	1.9508	1.2656	0.2095	0.2167	0.1139
	-1	2	1.3862	-0.9041	0.0113	1.6427	0.1395	1.6797	0.0957
	-1	2	0.5099	-0.9317	0.1364	1.6898	0.2787	0.7874	0.0821
	-1	2	0.1876	-2.1881	1.6570	2.0105	0.2272	0.1643	0.0031
	-1	3	1.3862	-1.0963	0.0123	2.8115	0.0436	1.7631	0.1332
	-1	3	0.5099	-1.0966	0.0121	2.8094	0.0435	0.8868	0.1339
	-1	3	0.1876	-1.0962	0.0119	2.8091	0.0442	0.2644	0.2328

Table 3.2: Estimated MSE for Model 1 with different percentages of censorship, θ and τ for $n = 50$

C	β	θ	τ	$\hat{\beta}$	MSE($\hat{\beta}$)	$\hat{\theta}$	MSE($\hat{\theta}$)	$\hat{\tau}$	MSE($\hat{\tau}$)
0%	-1	1	1.3862	-1.0089	0.0410	1.0149	0.0200	1.3667	0.0174
	-1	2	0.5099	-0.9635	0.4128	1.9895	0.0314	0.5977	0.0160
	-1	3	0.1876	-0.9834	0.0068	2.9986	0.0003	0.1829	0.0368
20%	-1	1	1.3862	-1.2529	0.1137	1.0185	0.0354	1.3841	0.0269
	-1	2	0.5099	-0.9783	0.4492	1.9550	0.0939	0.6344	0.0252
	-1	3	0.1876	-0.9984	0.0012	2.9610	0.0020	0.1930	0.0484
50%	-1	1	1.3862	-1.7530	0.6718	1.0320	0.1640	1.5098	0.1189
	-1	2	0.5099	-0.9924	0.1519	1.8299	0.1744	0.6895	0.0488
	-1	3	0.1876	-1.0042	0.0081	2.8813	0.0143	0.2204	0.0952
70%	-1	1	1.3862	-2.3169	2.0048	1.0475	0.7123	1.6575	0.3404
	-1	2	0.5099	-0.9317	0.1364	1.6898	0.2787	0.7874	0.0821
	-1	3	0.1876	-1.0962	0.0119	2.8091	0.0442	0.2644	0.2328

Table 3.3: Estimated MSE for Model 1 with different percentages of censorship, θ and τ for $n = 200$

C	β	θ	τ	$\hat{\beta}$	MSE($\hat{\beta}$)	$\hat{\theta}$	MSE($\hat{\theta}$)	$\hat{\tau}$	MSE($\hat{\tau}$)
0%	-1	1	1.3862	-1.0045	0.0099	1.0043	0.0049	1.3790	0.0059
	-1	2	0.5099	-0.9699	0.3652	1.9966	0.0296	0.4887	0.0077
	-1	3	0.1876	-0.9896	0.0012	2.9989	0.0001	0.1863	0.0351
20%	-1	1	1.3862	-1.2318	0.0666	1.0052	0.0067	1.3922	0.0124
	-1	2	0.5099	-1.0006	0.3767	1.9532	0.0604	0.4839	0.0130
	-1	3	0.1876	-0.9983	0.0008	2.9608	0.0016	0.2024	0.0472
50%	-1	1	1.3862	-1.7154	0.5374	0.9978	0.0164	1.5643	0.0746
	-1	2	0.5099	-1.0113	0.4739	1.8428	0.2802	0.4671	0.0229
	-1	3	0.1876	-1.0003	0.0062	2.8819	0.0140	0.2323	0.0811
70%	-1	1	1.3862	-2.2411	1.6033	0.9814	0.1464	1.8165	0.2414
	-1	2	0.5099	-1.0260	0.3770	1.7209	0.3781	0.4505	0.0369
	-1	3	0.1876	-1.0999	0.0112	2.8092	0.0381	0.2644	0.1447

Table 3.4: Estimated MSE for Model 1 with different percentages of censorship, θ and τ for $n = 500$

C	β	θ	τ	$\hat{\beta}$	MSE($\hat{\beta}$)	$\hat{\theta}$	MSE($\hat{\theta}$)	$\hat{\tau}$	MSE($\hat{\tau}$)
0%	-1	1	1.3862	-1.0005	0.0039	1.0019	0.0018	1.3821	0.0021
	-1	2	0.5099	-0.9742	0.2604	1.9939	0.0226	0.4924	0.0025
	-1	3	0.1876	-0.9909	0.0005	2.9989	0.0000	0.1891	0.0350
20%	-1	1	1.3862	-1.2280	0.0569	1.0024	0.0026	1.3854	0.0053
	-1	2	0.5099	-1.0074	0.3505	1.9497	0.0525	0.4881	0.0038
	-1	3	0.1876	-0.9980	0.0006	2.9607	0.0015	0.1967	0.0449
50%	-1	1	1.3862	-1.7066	0.5085	0.9922	0.0053	1.5264	0.0608
	-1	2	0.5099	-1.4099	0.6564	1.9291	0.1882	0.4772	0.0064
	-1	3	0.1876	-1.0000	0.0035	2.8835	0.0134	0.2282	0.0670
70%	-1	1	1.3862	-2.2217	1.5160	0.9547	0.0216	1.8614	0.2223
	-1	2	0.5099	-1.8905	0.9194	1.9204	0.3276	0.4591	0.0104
	-1	3	0.1876	-1.1001	0.0106	2.8115	0.0366	0.2636	0.1347

Table 3.5: Estimated MSE for Model 1 with different percentages of censorship, θ and τ for $n = 1000$

C	β	θ	τ	$\hat{\beta}$	MSE($\hat{\beta}$)	$\hat{\theta}$	MSE($\hat{\theta}$)	$\hat{\tau}$	MSE($\hat{\tau}$)
0%	-1	1	1.3862	-1.0002	0.0019	1.0015	0.0010	1.3841	0.0007
	-1	2	0.5099	-0.9849	0.1186	1.9959	0.0082	0.4999	0.0007
	-1	3	0.1876	-0.9919	0.0003	2.9989	0.0000	0.1867	0.0346
20%	-1	1	1.3862	-1.2262	0.0531	1.0018	0.0013	1.3850	0.0022
	-1	2	0.5099	-1.0174	0.2054	1.9491	0.0160	0.4977	0.0010
	-1	3	0.1876	-0.9945	0.0004	2.9607	0.0015	0.1968	0.0436
50%	-1	1	1.3862	-1.7010	0.4950	0.9896	0.0032	1.4900	0.0514
	-1	2	0.5099	-1.5742	0.7294	1.9700	0.0723	0.4914	0.0023
	-1	3	0.1876	-0.9999	0.0016	2.8849	0.0130	0.2277	0.0579
70%	-1	1	1.3862	-2.2186	1.4940	0.9492	0.0071	1.8672	0.2183
	-1	2	0.5099	-2.1073	1.4128	1.9726	0.1707	0.4803	0.0039
	-1	3	0.1876	-1.1005	0.0107	2.8093	0.0363	0.2646	0.1312

Table 3.6: Estimated MSE for Model 2 with different percentages of censorship, β, θ and τ for $n = 50$

C	β	θ	τ	$\hat{\beta}$	MSE($\hat{\beta}$)	$\hat{\theta}$	MSE($\hat{\theta}$)	$\hat{\tau}$	MSE($\hat{\tau}$)
0%	-1	1	1.3862	-0.9873	0.0315	1.1326	0.3028	1.4966	0.0732
	-1	1	0.5099	-0.9923	0.0246	1.2188	0.5915	0.6098	0.0491
	-1	1	0.1876	-0.9933	0.0221	0.8434	0.6737	0.2801	0.0497
	-1	2	1.3862	-0.9716	0.0230	2.2221	0.5727	1.4788	0.0545
	-1	2	0.5099	-0.9727	0.0218	2.2472	0.4427	0.6140	0.0574
	-1	2	0.1876	-0.9670	0.0226	2.0995	0.2324	0.2772	0.0640
	-1	3	1.3862	-0.9762	0.0206	3.1247	0.4949	1.4592	0.0485
	-1	3	0.5099	-0.9687	0.0221	2.8852	0.3334	0.5669	0.0535
	-1	3	0.1876	-0.9671	0.0225	3.1787	0.4238	0.1610	0.0509
20%	-1	1	1.3862	-1.2011	0.0821	1.1402	0.2327	1.5695	0.1089
	-1	1	0.5099	-1.2048	0.0740	1.2528	0.4414	0.6715	0.0840
	-1	1	0.1876	-1.2052	0.0723	1.3449	0.4930	0.3644	0.0860
	-1	2	1.3862	-1.1824	0.0599	2.0983	0.3189	1.5415	0.0934
	-1	2	0.5099	-1.1843	0.0599	2.1205	0.2692	0.6664	0.0907
	-1	2	0.1876	-1.1804	0.0586	2.0452	0.2738	0.3526	0.0894
	-1	3	1.3862	-1.1808	0.0580	2.8649	0.3778	1.5026	0.0767
	-1	3	0.5099	-1.1811	0.0583	2.6223	0.5994	0.6116	0.0817
	-1	3	0.1876	-1.1810	0.0601	3.1194	0.6676	0.2319	0.0781
50%	-1	1	1.3862	-1.6563	0.5250	1.1593	0.1906	1.7610	0.2541
	-1	1	0.5099	-1.6617	0.4988	1.2928	0.2906	0.7923	0.2182
	-1	1	0.1876	-1.6606	0.4922	1.3829	0.3161	0.4819	0.2115
	-1	2	1.3862	-1.6261	0.4481	1.8515	0.2643	1.6408	0.2403
	-1	2	0.5099	-1.6859	0.4706	1.9086	0.3345	0.7881	0.2260
	-1	2	0.1876	-1.6279	0.4409	1.9161	0.4544	0.4771	0.2344
	-1	3	1.3862	-1.6240	0.4426	2.3544	0.8892	1.6555	0.2097
	-1	3	0.5099	-1.6287	0.4449	2.4061	1.1143	0.8076	0.2146
	-1	3	0.1876	-1.6324	0.4487	3.1566	1.2095	0.4604	0.2078
70%	-1	1	1.3862	-2.1541	1.5366	1.1919	0.1945	1.8666	0.6575
	-1	1	0.5099	-2.1438	1.4441	1.3278	0.3047	0.9274	0.4978
	-1	1	0.1876	-1.3407	0.1516	1.3514	0.3818	0.3976	0.1180
	-1	2	1.3862	-2.1061	1.3531	1.6466	0.4151	1.9041	0.6903
	-1	2	0.5099	-2.2142	1.4675	1.6955	0.6032	0.9708	0.6971
	-1	2	0.1876	-2.1112	1.3685	1.5921	0.7597	0.7619	0.7473
	-1	3	1.3862	-2.1069	1.3254	2.1534	1.4317	1.8248	0.5368
	-1	3	0.5099	-2.1263	1.3539	2.8858	1.5631	0.9778	0.5638
	-1	3	0.1876	-2.1208	1.4429	3.1398	1.6925	0.6543	0.5523

Table 3.7: Estimated MSE for Model 2 with different percentages of censorship, θ and τ for $n = 50$

C	β	θ	τ	$\hat{\beta}$	MSE($\hat{\beta}$)	$\hat{\theta}$	MSE($\hat{\theta}$)	$\hat{\tau}$	MSE($\hat{\tau}$)
0%	-1	1	1.3862	-0.9873	0.0315	1.1326	0.3028	1.4966	0.0732
	-1	2	0.5099	-0.9727	0.0218	2.2472	0.4427	0.6140	0.0574
	-1	3	0.1876	-0.9671	0.0225	3.1787	0.4238	0.1610	0.0509
20%	-1	1	1.3862	-1.2011	0.0821	1.1402	0.2327	1.5695	0.1089
	-1	2	0.5099	-1.1843	0.0599	2.1205	0.2692	0.6664	0.0907
	-1	3	0.1876	-1.1810	0.0601	3.1194	0.6676	0.2319	0.0781
50%	-1	1	1.3862	-1.6563	0.5250	1.1593	0.1906	1.7610	0.2541
	-1	2	0.5099	-1.6859	0.4706	1.9086	0.3345	0.7881	0.2260
	-1	3	0.1876	-1.6324	0.4487	3.1566	1.2095	0.4604	0.2078
70%	-1	1	1.3862	-2.1541	1.5366	1.1919	0.1945	1.8666	0.6575
	-1	2	0.5099	-2.2142	1.4675	1.6955	0.6032	0.9708	0.6971
	-1	3	0.1876	-2.1208	1.4429	3.1398	1.6925	0.6543	0.5523

Table 3.8: Estimated MSE for Model 2 with different percentages of censorship, θ and τ for $n = 200$

C	β	θ	τ	$\hat{\beta}$	MSE($\hat{\beta}$)	$\hat{\theta}$	MSE($\hat{\theta}$)	$\hat{\tau}$	MSE($\hat{\tau}$)
0%	-1	1	1.3862	-0.9975	0.0076	1.0432	0.0508	1.4949	0.0457
	-1	2	0.5099	-0.9931	0.0050	2.1399	0.1612	0.5278	0.0150
	-1	3	0.1876	-0.9948	0.0050	3.0914	0.3767	0.1960	0.0053
20%	-1	1	1.3862	-1.2184	0.0578	1.0546	0.0391	1.5591	0.0627
	-1	2	0.5099	-1.2074	0.0497	2.0453	0.1784	0.5500	0.0354
	-1	3	0.1876	-1.2136	0.0520	3.0109	0.4479	0.2009	0.0121
50%	-1	1	1.3862	-1.6841	0.4919	1.0508	0.0324	1.7042	0.2059
	-1	2	0.5099	-1.6779	0.4615	1.7552	0.2382	0.6576	0.1163
	-1	3	0.1876	-1.6766	0.4686	3.2734	0.8115	0.2224	0.0480
70%	-1	1	1.3862	-2.1785	1.4565	1.0729	0.0386	1.8865	0.5657
	-1	2	0.5099	-2.1979	1.4107	1.7406	0.4907	0.9100	0.6038
	-1	3	0.1876	-2.0745	1.4551	3.3069	1.3644	0.3202	0.1649

Table 3.9: Estimated MSE for Model 2 with different percentages of censorship, θ and τ for $n = 500$

C	β	θ	τ	$\hat{\beta}$	MSE($\hat{\beta}$)	$\hat{\theta}$	MSE($\hat{\theta}$)	$\hat{\tau}$	MSE($\hat{\tau}$)
0%	-1	1	1.3862	-0.9991	0.0030	1.0213	0.0192	1.4821	0.0363
	-1	2	0.5099	-0.9965	0.0020	2.0700	0.0835	0.5160	0.0050
	-1	3	0.1876	-0.9991	0.0020	3.0740	0.2506	0.1927	0.0009
20%	-1	1	1.3862	-1.2221	0.0530	1.0266	0.0141	1.5577	0.0520
	-1	2	0.5099	-1.2136	0.0486	1.9730	0.0825	0.5209	0.0180
	-1	3	0.1876	-1.2182	0.0503	3.1039	0.2799	0.1925	0.0026
50%	-1	1	1.3862	-1.6901	0.4852	1.0175	0.0114	1.6820	0.1835
	-1	2	0.5099	1.6655	0.4549	1.7121	0.2106	0.5365	0.0871
	-1	3	0.1876	-1.6903	0.4792	3.6376	0.7605	0.1913	0.0147
70%	-1	1	1.3862	-2.1960	1.4511	1.0275	0.0138	1.7779	0.4756
	-1	2	0.5099	-2.1830	1.3806	1.8256	0.4168	0.8205	0.5505
	-1	3	0.1876	-2.0597	1.4116	3.6133	1.0150	0.1972	0.0734

Table 3.10: Estimated MSE for Model 2 with different percentages of censorship, θ and τ for $n = 1000$

C	β	θ	τ	$\hat{\beta}$	MSE($\hat{\beta}$)	$\hat{\theta}$	MSE($\hat{\theta}$)	$\hat{\tau}$	MSE($\hat{\tau}$)
0%	-1	1	1.3862	-0.9994	0.0015	1.0155	0.0091	1.4791	0.0329
	-1	2	0.5099	-0.9977	0.0010	2.0467	0.0361	0.5131	0.0016
	-1	3	0.1876	-0.9993	0.0010	3.0462	0.2023	0.1902	0.0003
20%	-1	1	1.3862	-1.2222	0.0511	1.0143	0.0068	1.5456	0.0437
	-1	2	0.5099	-1.2158	0.0482	1.9247	0.0532	0.5126	0.0082
	-1	3	0.1876	-1.2212	0.0501	2.9506	0.2544	0.1901	0.0008
50%	-1	1	1.3862	-1.6918	0.4827	1.0070	0.0051	1.6428	0.1636
	-1	2	0.5099	-1.6338	0.4504	1.7058	0.1882	0.5107	0.0643
	-1	3	0.1876	-1.6942	0.4825	3.6265	0.6135	0.1785	0.0067
70%	-1	1	1.3862	-2.1987	1.4508	1.0125	0.0065	1.6283	0.3945
	-1	2	0.5099	-2.1631	1.3621	1.8602	0.3632	0.7666	0.4827
	-1	3	0.1876	-2.0208	1.3618	3.6828	0.7594	0.1574	0.0278

Table 3.11: Estimated MSE for Model 3 with different percentages of censorship, β, θ and τ for $n = 50$

C	β	θ	τ	$\hat{\beta}$	MSE($\hat{\beta}$)	$\hat{\theta}$	MSE($\hat{\theta}$)	$\hat{\tau}$	MSE($\hat{\tau}$)
0%	-1	1	1.3862	-1.0600	0.1161	1.1056	0.1520	1.4230	0.0123
	-1	1	0.5099	-1.1278	0.3390	1.1701	0.3741	0.5412	0.0071
	-1	1	0.1876	-1.1796	0.3030	1.2324	0.3389	0.2161	0.0063
	-1	2	1.3862	-1.0320	0.1133	2.0868	0.1438	1.3981	0.0021
	-1	2	0.5099	-1.1100	0.3403	2.1274	0.2725	0.5204	0.0016
	-1	2	0.1876	-0.9263	0.2870	2.0545	0.3071	0.1976	0.0015
	-1	3	1.3862	-1.0269	0.1049	3.0710	0.1336	1.3904	0.0002
	-1	3	0.5099	-1.1046	0.2917	3.0966	0.3204	0.5136	0.0001
	-1	3	0.1876	-0.8717	0.2907	2.9738	0.3093	0.1910	0.0001
20%	-1	1	1.3862	-1.3044	0.2573	1.1442	0.1967	1.4366	0.0172
	-1	1	0.5099	-1.3896	0.5121	1.1939	0.4147	0.5499	0.0105
	-1	1	0.1876	-1.4644	0.3673	1.2055	0.3616	0.2248	0.0100
	-1	2	1.3862	-1.2551	0.2286	2.0909	0.1821	1.4009	0.0030
	-1	2	0.5099	-1.3695	0.4935	2.1339	0.3035	0.5230	0.0022
	-1	2	0.1876	-1.4176	0.2609	2.0974	0.2579	0.1999	0.0021
	-1	3	1.3862	-1.2505	0.2226	3.0878	0.1768	1.3920	0.0003
	-1	3	0.5099	-1.3755	0.4643	3.1287	0.3811	0.5145	0.0002
	-1	3	0.1876	-1.3970	0.4720	3.0572	0.2870	0.1917	0.0002
50%	-1	1	1.3862	-1.8033	0.9780	1.1835	0.3436	1.4570	0.0283
	-1	1	0.5099	-1.8427	1.0686	1.2401	0.4367	0.5728	0.0254
	-1	1	0.1876	-1.5691	0.5708	0.9678	0.5603	0.2390	0.0205
	-1	2	1.3862	-1.7523	0.9024	2.1217	0.3215	1.4094	0.0058
	-1	2	0.5099	-1.8064	0.9203	2.1209	0.3587	0.5304	0.0044
	-1	2	0.1876	-1.5267	0.2925	1.8312	0.3113	0.2075	0.0045
	-1	3	1.3862	-1.7674	0.8770	3.1216	0.3003	1.3952	0.0007
	-1	3	0.5099	-1.7955	0.9072	3.0777	0.3747	0.5173	0.0005
	-1	3	0.1876	-1.5207	1.2581	2.7638	0.3551	0.1948	0.0006
70%	-1	1	1.3862	-2.3619	2.3418	1.2556	0.5434	1.4884	0.0674
	-1	1	0.5099	-2.4410	1.7825	1.2196	0.4867	0.5977	0.0500
	-1	1	0.1876	-1.6831	0.8038	0.6328	1.2235	0.2466	0.0348
	-1	2	1.3862	-2.3208	2.1627	2.1890	0.4440	1.4252	0.0142
	-1	2	0.5099	-2.2274	2.4256	2.0443	0.4050	0.5442	0.0110
	-1	2	0.1876	-1.5770	0.4129	1.5151	0.6038	0.2204	0.0099
	-1	3	1.3862	-2.3095	2.1225	3.1761	0.4535	1.4020	0.0021
	-1	3	0.5099	-2.2525	1.3747	3.0042	0.3106	0.5222	0.0015
	-1	3	0.1876	-1.5393	1.2856	2.3707	0.7289	0.1998	0.0019

Table 3.12: Estimated MSE for Model 3 with different percentages of censorship, θ and τ for $n = 50$

C	β	θ	τ	$\hat{\beta}$	MSE($\hat{\beta}$)	$\hat{\theta}$	MSE($\hat{\theta}$)	$\hat{\tau}$	MSE($\hat{\tau}$)
0%	-1	1	1.3862	-1.0600	0.1161	1.1056	0.1520	1.4230	0.0123
	-1	2	0.5099	-1.1100	0.3403	2.1274	0.2725	0.5204	0.0016
	-1	3	0.1876	-0.8717	0.2907	2.9738	0.3093	0.1910	0.0001
20%	-1	1	1.3862	-1.3044	0.2573	1.1442	0.1967	1.4366	0.0172
	-1	2	0.5099	-1.3695	0.4935	2.1339	0.3035	0.5230	0.0022
	-1	3	0.1876	-1.3970	0.4720	3.0572	0.2870	0.1917	0.0002
50%	-1	1	1.3862	-1.8033	0.9780	1.1835	0.3436	1.4570	0.0283
	-1	2	0.5099	-1.8064	0.9203	2.1209	0.3587	0.5304	0.0044
	-1	3	0.1876	-1.5207	1.2581	2.7638	0.3551	0.1948	0.0006
70%	-1	1	1.3862	-2.3619	2.3418	1.2556	0.5434	1.4884	0.0674
	-1	2	0.5099	-2.2274	2.4256	2.0443	0.4050	0.5442	0.0110
	-1	3	0.1876	-1.5393	1.2856	2.3707	0.7289	0.1998	0.0019

Table 3.13: Estimated MSE for Model 3 with different percentages of censorship, θ and τ for $n = 200$

C	β	θ	τ	$\hat{\beta}$	MSE($\hat{\beta}$)	$\hat{\theta}$	MSE($\hat{\theta}$)	$\hat{\tau}$	MSE($\hat{\tau}$)
0%	-1	1	1.3862	-1.0197	0.0246	1.0349	0.0320	1.3965	0.0036
	-1	2	0.5099	-1.0295	0.0666	2.0403	0.0726	0.5120	0.0001
	-1	3	0.1876	-1.0626	0.2129	3.0601	0.2190	0.1884	6.6E-06
20%	-1	1	1.3862	-1.2499	0.0974	1.0451	0.0410	1.3990	0.0042
	-1	2	0.5099	-1.2546	0.1617	2.0435	0.0917	0.5126	0.0002
	-1	3	0.1876	-1.3070	0.3730	3.0769	0.2735	0.1887	1.1E-05
50%	-1	1	1.3862	-1.7221	0.5953	1.0562	0.0649	1.4093	0.0080
	-1	2	0.5099	-1.7445	0.7649	2.0680	0.1651	0.5148	0.0005
	-1	3	0.1876	-1.8159	1.0321	3.0866	0.3581	0.1893	2.9E-05
70%	-1	1	1.3862	-2.2484	1.7227	1.0774	0.1165	1.4230	0.0178
	-1	2	0.5099	-2.2870	2.0699	2.1039	0.2976	0.5174	0.0012
	-1	3	0.1876	-2.2335	1.8800	3.0342	0.3746	0.1904	9.2E-05

Table 3.14: Estimated MSE for Model 3 with different percentages of censorship, θ and τ for $n = 500$

C	β	θ	τ	$\hat{\beta}$	MSE($\hat{\beta}$)	$\hat{\theta}$	MSE($\hat{\theta}$)	$\hat{\tau}$	MSE($\hat{\tau}$)
0%	-1	1	1.3862	-1.0073	0.0094	1.0149	0.0121	1.3894	0.0010
	-1	2	0.5099	-1.0138	0.0237	2.0172	0.0261	0.5108	2.0E-05
	-1	3	0.1876	-1.0128	0.0690	3.0239	0.0710	0.1879	1.1E-06
20%	-1	1	1.3862	-1.2327	0.0679	1.0196	0.0151	1.3908	0.0014
	-1	2	0.5099	-1.2425	0.0936	2.0226	0.0337	0.5110	3.2E-05
	-1	3	0.1876	-1.2645	0.1693	3.0370	0.0958	0.1880	1.9E-06
50%	-1	1	1.3862	-1.7086	0.5310	1.0254	0.0246	1.3935	0.0030
	-1	2	0.5099	-1.7171	0.5871	2.0325	0.0544	0.5117	0.0001
	-1	3	0.1876	-1.7501	0.7707	3.0498	0.1662	0.1883	4.7E-06
70%	-1	1	1.3862	-2.2272	1.5650	1.0408	0.0417	1.4005	0.0069
	-1	2	0.5099	-2.2431	1.7028	2.0470	0.1017	0.5128	0.0002
	-1	3	0.1876	-2.2404	2.0961	3.0808	0.3015	0.1887	1.4E-05

Table 3.15: Estimated MSE for Model 3 with different percentages of censorship, θ and τ for $n = 1000$

C	β	θ	τ	$\hat{\beta}$	MSE($\hat{\beta}$)	$\hat{\theta}$	MSE($\hat{\theta}$)	$\hat{\tau}$	MSE($\hat{\tau}$)
0%	-1	1	1.3862	-1.0052	0.0047	-1.0052	0.0060	1.3876	0.0003
	-1	2	0.5099	-1.0042	0.0117	2.0068	0.0130	0.5103	4.8E-06
	-1	3	0.1876	-1.0130	0.0305	3.0139	0.0316	0.1878	3.0E-07
20%	-1	1	1.3862	-1.2312	0.0590	1.0105	0.0074	1.3882	0.0005
	-1	2	0.5099	-1.2363	0.0732	2.0169	0.0166	0.5105	7.6E-06
	-1	3	0.1876	-1.2336	0.1044	3.0206	0.0425	0.1878	4.9E-07
50%	-1	1	1.3862	-1.7048	0.5080	1.0174	0.0124	1.3890	0.0010
	-1	2	0.5099	-1.7015	0.5308	2.0103	0.0262	0.5108	2.1E-05
	-1	3	0.1876	-1.7147	0.6131	3.0293	0.0716	0.1879	1.1E-06
70%	-1	1	1.3862	-2.2174	1.5090	1.0211	0.0202	1.3912	0.0026
	-1	2	0.5099	-2.2340	1.5822	2.0294	0.0465	0.5114	0.0001
	-1	3	0.1876	-2.2313	1.7622	3.0409	0.1352	0.1882	3.1E-06

3.6 Discussion

Table 3.1 shows the results for Model 1 under different percentages of censorship when the sample size is 50. In this table, the estimated results are obtained by using different values of τ when β and θ are kept fixed. At each level of censorship, we observe that the $MSE(\hat{\theta})$ increases as the value of τ decreased. It is also seen that the $MSE(\hat{\tau})$ decreases as the censoring level decreased. Tables 3.2-3.5 shows the ML estimated results for Model 1 under different percentages of censorship with different values of θ and τ for sample sizes 50, 200, 500, and 1000, respectively. It is observed that for each sample size $MSE(\hat{\tau})$ decreases as the censorship decreased. From these tables, we also notice that at each level of censorship the $MSE(\hat{\tau})$ decreases as n increased.

For Model 2, Table 3.6 presents the ML estimates for the parameters β, θ, and τ along with their MSE. The results are obtained by using four different levels of censorship with a fixed sample size $n = 50$. It can be seen that when the censorship percentage increases, there is greater variability, therefore $MSE(\hat{\tau})$ increased, although here we do not see any increase or decrease in the $MSE(\hat{\theta})$ due to τ. Tables 3.7-3.10 show the ML estimated results for Model 2 under different percentages of censorship with different values of θ and τ for sample sizes 50, 200, 500, and 1000 respectively. These tables provide evidence that when the sample size 'n' remains fixed and the censorship percentage increases, greater variability occurs, leading to an increase in the MSE. Conversely, by increasing the sample size while maintaining a fixed censorship percentage there is greater accuracy and the MSE decreased.

Table 3.11 displays the results for Model 3 with varying percentages of censorship while keeping the sample size fixed at 50. The estimated results in this table are obtained using different values of τ, while β and θ remain constant. The table indicates that the inclusion of τ does not result in any observable change in the $MSE(\hat{\theta})$. However, an increase in the percentage of censorship levels causes an increase in MSE. Tables 3.12-3.15 list the estimated results for Model 3 having sample sizes 50, 200, 500, and 1000, respectively. It can be seen that when increasing sample size n, and fixing the censorship percentage, $MSE(\hat{\tau})$ decreased. When the sample size n is fixed and the censorship percentage increases, the $MSE(\hat{\tau})$ increased.

3.7 Real Data Analysis

In this section, we explore clinical data from a placebo-controlled randomized trial of gamma interferon in chronic granulotamous disease (CGD) (Fleming and

Harrington, 2013) to show the practical use of the proposed estimation process. The CGD is an unusual inherited disorder of the immune system. The specialized cells (neutrophils and monocytes) that are meant to eliminate harmful bacteria and fungi, do not function as they should in patients with CGD. As a result, those affected are more susceptible to serious and life-threatening bacterial and fungal infections, even though they can fight viruses normally. Additionally, they carry out chronic inflammatory signs, frequently of a granulomatous type. Although the exact amount of CGD cases is unknown. Song et al. (2011) estimated that it affects 8.5 out of every million newborns in the United Kingdom and Ireland.

The CGD data set is available in the R package "survival". The response variable is the time to the first infection (days). The covariate we consider here is gamma interferon from treatment/control. According to Group* (1991), there is reason to believe that gamma interferon has an important function in the treatment of CGD patients. In the dataset variable, Status is a censoring indicator that holds a value of 1 for uncensored and 0 for censored.

The proposed models were fitted to the treatment group by using the proposed estimation methods. Figure 3.1 presents the Kaplan-Meier estimate of the survival probability along with their confidence interval, the vertical lines on the Kaplan-Meier curve indicate censored observations. Arenas et al. (2021) detected that a change point occur on the 274th day. Figures 3.2, 3.3, and 3.4 present fitted survival functions for Model 1, Model 2, and Model 3, respectively, all proposed models are effective in estimating the change point. Table 3.16 shows the parameter estimates for the proposed models. The correlation between survival probabilities of the Kaplan-Meier method and Model 3 is higher (0.99241) than Model 1 (0.99067) and Model 2 (0.97882), which indicates better fitting.

Table 3.16: Parameter estimates for CGD data

	$\hat{\beta}$	$\hat{\theta}$	$\hat{\tau}$
Model 1	-6.1802	0.2104	274
Model 2	-6.7760	0.9020	274
Model 3	-6.1664	0.0572	274

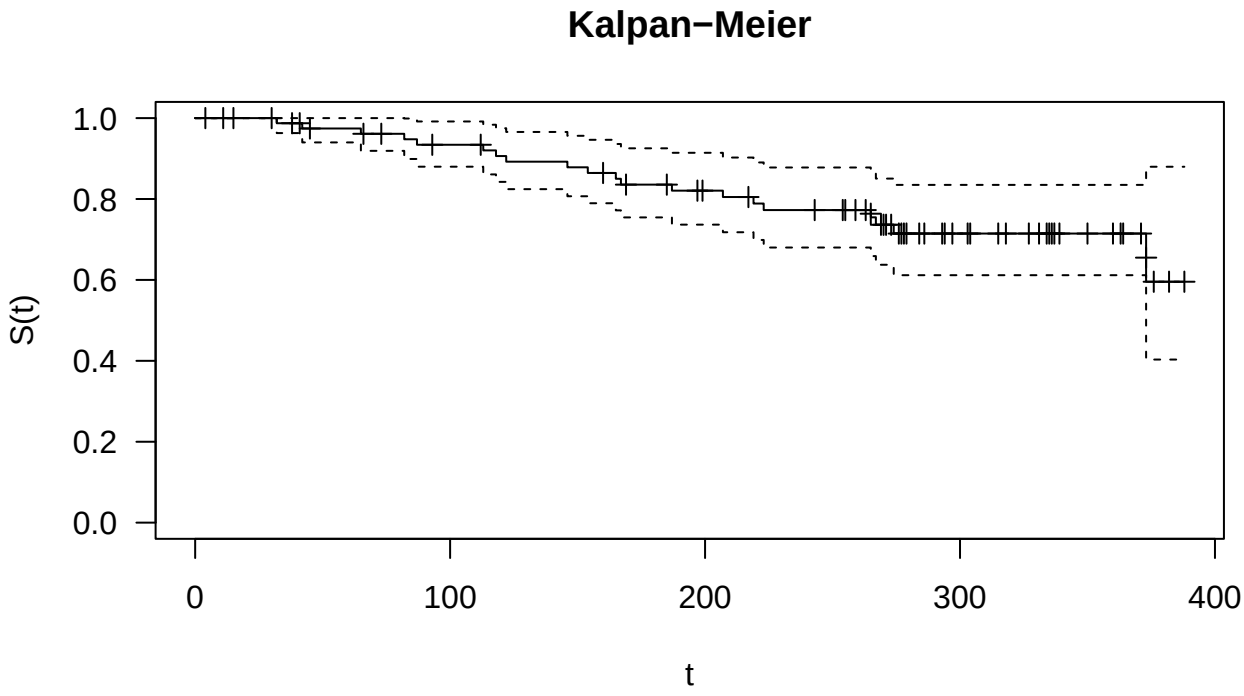

Figure 3.1: Kaplan-Meier estimated survival functions for treatment group

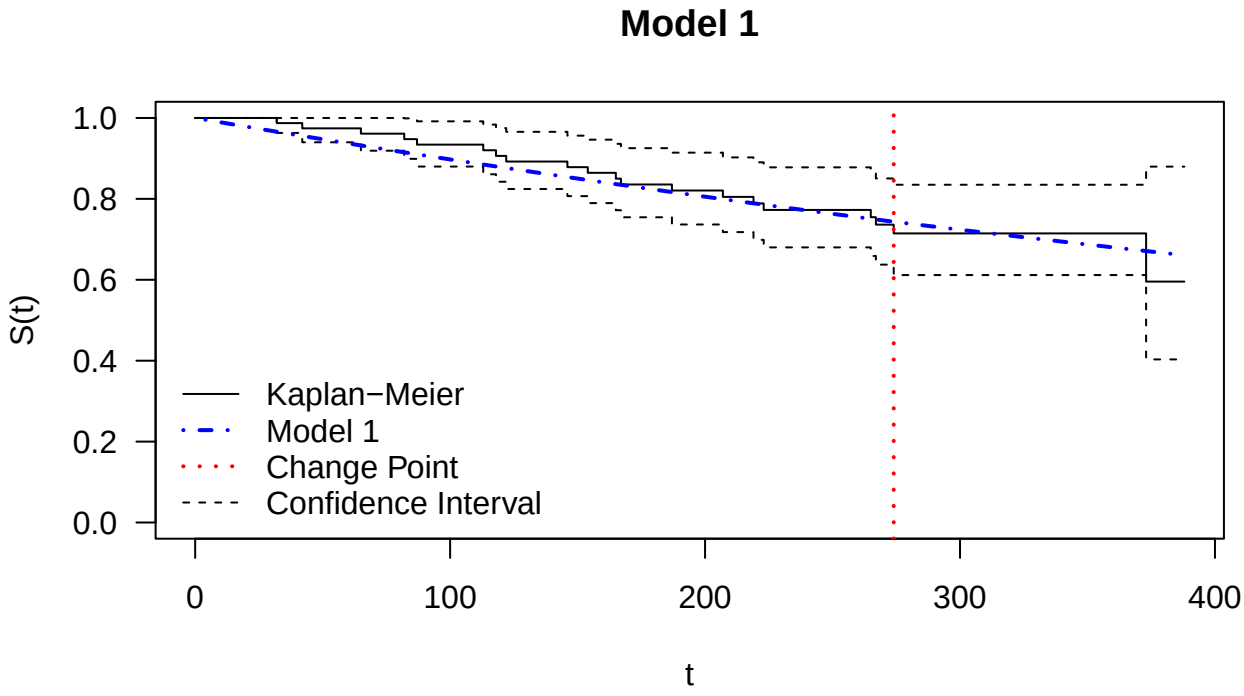

Figure 3.2: Model 1 estimated survival functions for treatment group and estimated change point

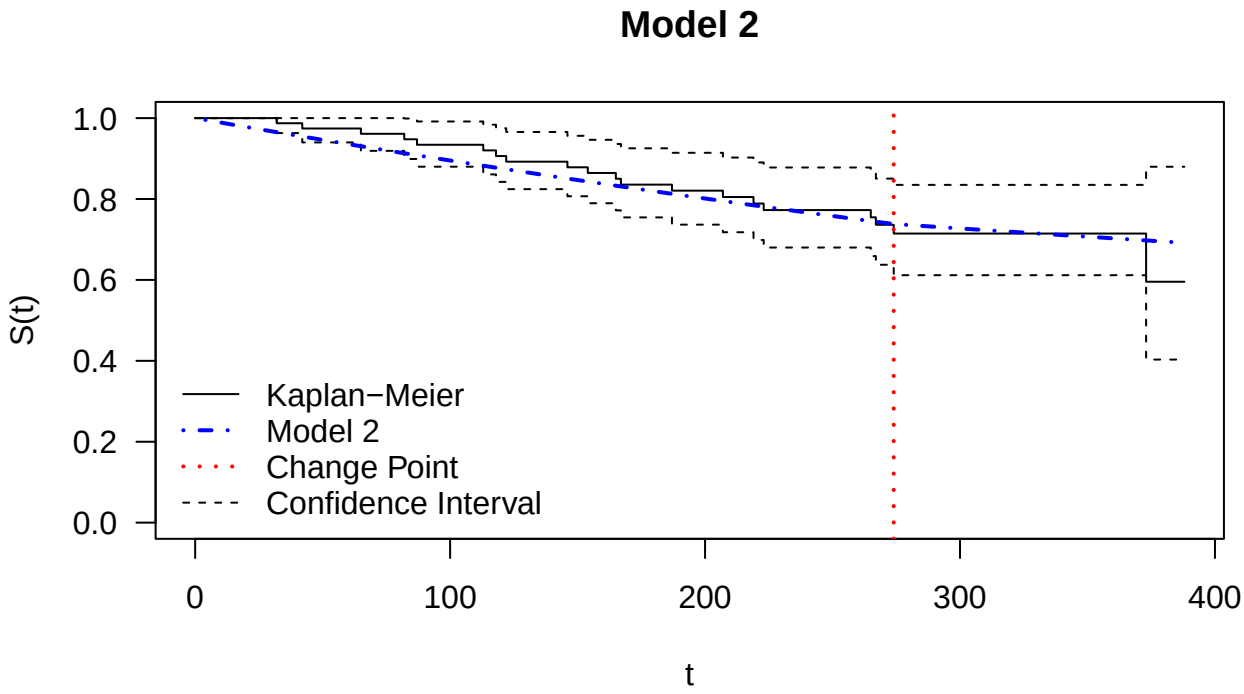

Figure 3.3: Model 2 estimated survival functions for treatment group and estimated change point

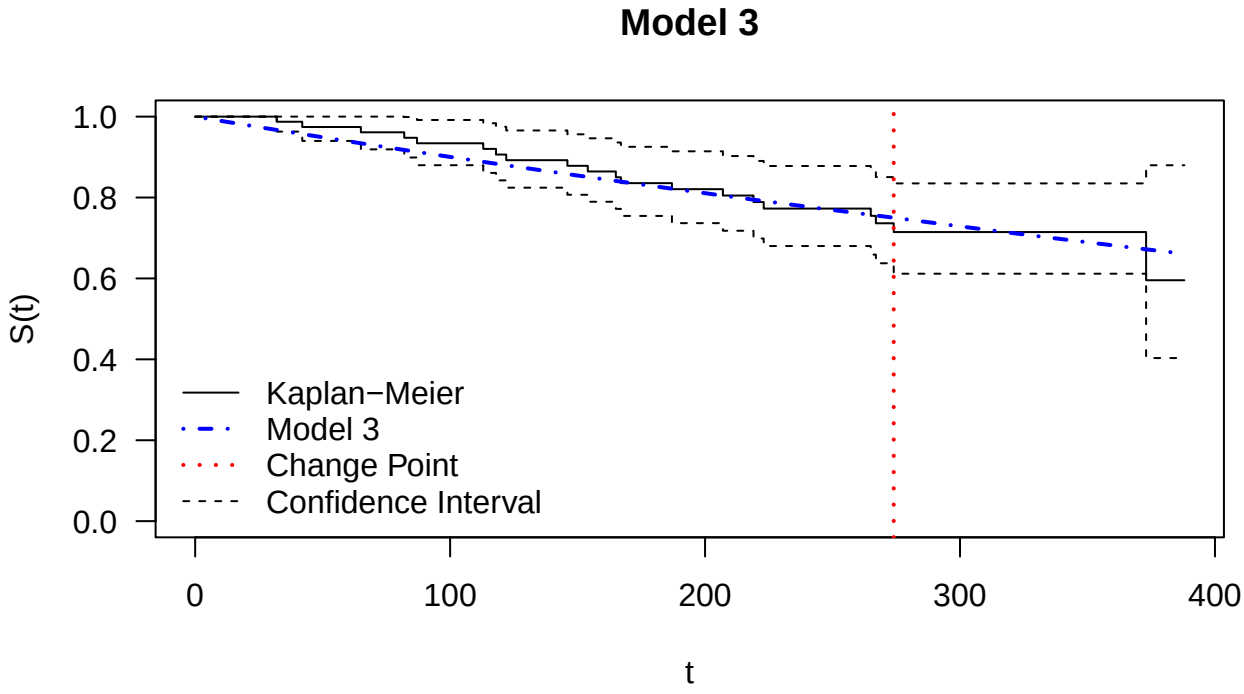

Figure 3.4: Model 3 estimated survival functions for treatment group and estimated change point

Chapter 4

Conclusion and Recommendations

Hazard functions in survival analysis measure the risk of failure at a given time. A change point in the failure rate is essential for identifying important population subgroups. For multiple change point scenarios, dividing the data into many subgroups may result in limited observations and increased variability in estimated results. Such multi-group results could also be overfitted. Therefore, a single change point, or equivalently two subgroups analysis, is still valuable for many scientific applications.

This study discussed the maximum likelihood method to estimate the change point in the Cox proportional hazard model. Three hazard models were analyzed, with the first two models having non-closed mathematical forms for the regression parameters β and θ. In contrast, the third model provides closed-form MLEs for these parameters. The simulation study provides evidence supporting the consistency of the estimator $\hat{\tau}$ for the change point τ. As the sample size n increases, the MSE of the estimator $\hat{\tau}$ decreases across all hazard models. Moreover, higher censorship percentages lead to an increase in the $MSE(\hat{\tau})$, which is attributed to the loss of information by censoring.

The CGD dataset is utilized as an authentic real-life instance of data. We apply the proposed models to study the group that received gamma interferon treatment and conclude that all proposed models are effective in estimating the change point. However, the survival function of Model 3 closely resembles the Kaplan-Meier survival function.

4.1 Open Studies

The research work has several limitations. Firstly, the study focuses on estimating a single change point, but there is potential for extension to explore two or more change points. Additionally, the covariate being studied is currently limited to a

dichotomous variable, but it could be expanded to include continuous or multinomial variables. Moreover, there is scope for further exploration by incorporating other hazard models such as the accelerated failure time (AFT) model and Frailty models. Also, other estimation methods can be studied for the discussed models.

References

Aalen, O. (1980). A model for nonparametric regression analysis of counting processes. In *Mathematical Statistics and Probability Theory: Proceedings, Sixth International Conference, Wisła (Poland), 1978*, pages 1–25. Springer.

Arenas, G., Villaseñor, J. A., Palmeros, O., and Tajonar, F. (2021). A computational method for estimating a change point in the cox hazard model. *Computational Statistics*, pages 1–16.

Chen, J. and Gupta, A. K. (2001). On change point detection and estimation. *Communications in statistics-simulation and computation*, 30(3):665–697.

Chen, J. and Gupta, A. K. (2012). Parametric statistical change point analysis: with applications to genetics, medicine, and finance.

Core Team, R. (2016). A language and environment for statistical computing (vienna, austria: R foundation for statistical computing) retrieved from www. *R-project. org*.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202.

Dupuy, J.-F. (2006). Estimation in a change-point hazard regression model. *Statistics & probability letters*, 76(2):182–190.

Fleming, T. R. and Harrington, D. P. (2013). *Counting processes and survival analysis*, volume 625. John Wiley & Sons.

Fotopoulos, S. and Jandhyala, V. (2001). Maximum likelihood estimation of a change-point for exponentially distributed random variables. *Statistics & probability letters*, 51(4):423–429.

Fu, Y.-X. and Curnow, R. (1990). Maximum likelihood estimation of multiple change points. *Biometrika*, 77(3):563–573.

Gandy, A., Jensen, U., and Lütkebohmert, C. (2005). A cox model with a change-point applied to an actuarial problem. *Brazilian Journal of Probability and Statistics*, pages 93–109.

Goodman, M. S., Li, Y., and Tiwari, R. C. (2006). Survival analysis with change point hazard functions.

Grambsch, P. and Therneau, T. (2000). Modeling survival data: extending the cox model.

Group*, I. C. G. D. C. S. (1991). A controlled trial of interferon gamma to prevent infection in chronic granulomatous disease. *New England Journal of Medicine*, 324(8):509–516.

Husain, H., Thamrin, S. A., Tahir, S., Mukhlisin, A., and Apriani, M. M. (2018). The application of extended cox proportional hazard method for estimating survival time of breast cancer. In *Journal of physics: Conference series*, volume 979, page 012087. IOP Publishing.

Kalbfleisch, J. D. and Schaubel, D. E. (2023). Fifty years of the cox model. *Annual Review of Statistics and Its Application*, 10.

Klein, J. P. and Moeschberger, M. L. (2003). *Survival analysis: techniques for censored and truncated data*, volume 1230. Springer.

Kleinbaum, D. and Klein, M. (2012). Survival analysis: A self-learning text. 3rd springer. *New York*.

Lee, C. Y., Chen, X., and Lam, K. F. (2020). Testing for change-point in the covariate effects based on the cox regression model. *Statistics in Medicine*, 39(10):1473–1488.

Li, Y., Qian, L., and Zhang, W. (2013). Estimation in a change-point hazard regression model with long-term survivors. *Statistics & Probability Letters*, 83(7):1683–1691.

Liang, K.-Y., Self, S. G., and Liu, X. (1990). The cox proportional hazards model with change point: An epidemiologic application. *Biometrics*, pages 783–793.

Liu, M., Lu, W., and Shao, Y. (2008). A monte carlo approach for change-point detection in the cox proportional hazards model. *Statistics in medicine*, 27(19):3894–3909.

Matthews, D. E. and Farewell, V. T. (1982). On testing for a constant hazard against a change-point alternative. *Biometrics*, pages 463–468.

O'Sullivan, F. (1993). Nonparametric estimation in the cox model. *The Annals of Statistics*, pages 124–145.

Palmeros, O., Villaseñor, J. A., and González, E. (2018). On computing estimates of a change-point in the weibull regression hazard model. *Journal of Applied Statistics*, 45(4):642–648.

Pons, O. (2002). Estimation in a cox regression model with a change-point at an unknown time. *Statistics*, 36(2):101–124.

Pons, O. (2003). Estimation in a cox regression model with a change-point according to a threshold in a covariate. *The Annals of Statistics*, 31(2):442–463.

Song, E., Jaishankar, G. B., Saleh, H., Jithpratuck, W., Sahni, R., and Krishnaswamy, G. (2011). Chronic granulomatous disease: a review of the infectious and inflammatory complications. *Clinical and Molecular Allergy*, 9(1):1–14.

Wang, B., Li, J., and Wang, X. (2021). Change point detection in cox proportional hazards mixture cure model. *Statistical Methods in Medical Research*, 30(2):440–457.

Yao, Y.-C. (1986). Maximum likelihood estimation in hazard rate models with a change-point. *Communications in Statistics-Theory and Methods*, 15(8):2455–2466.

Zhang, Y., Hua, L., and Huang, J. (2010). A spline-based semiparametric maximum likelihood estimation method for the cox model with interval-censored data. *Scandinavian Journal of Statistics*, 37(2):338–354.

Zucker, D. M., Agami, S., and Spiegelman, D. (2013). Testing for a changepoint in the cox survival regression model. *Journal of Statistical Theory and Practice*, 7:360–380.

Saad Waqas MPhil Statistics Thesis

ORIGINALITY REPORT

17%	12%	20%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	G. Y. Arenas, J. A. Villaseñor, O. Palmeros, F. Tajonar. "A computational method for estimating a change point in the Cox hazard model", Computational Statistics, 2021 Publication	4%
2	www.tandfonline.com Internet Source	1%
3	Submitted to BENEMERITA UNIVERSIDAD AUTONOMA DE PUEBLA BIBLIOTECA Student Paper	1%
4	fau.digital.flvc.org Internet Source	1%
5	cris.technion.ac.il Internet Source	1%
6	Emílio A. Coelho-Barros, Josmar Mazucheli, Jorge A. Achcar, Kelly Vanessa Parede Barco, José Rafael Tovar Cuevas. "The inverse power Lindley distribution in the presence of left-censored data", Journal of Applied Statistics, 2017 Publication	<1%

Associate Professor
Department of Statistics
Quaid-i-Azam University
Islamabad
2023

7	pt.scribd.com Internet Source	<1 %
8	Oscar Palmeros, Jose A. Villaseñor, Elizabeth González. "On computing estimates of a change-point in the Weibull regression hazard model", Journal of Applied Statistics, 2017 Publication	<1 %
9	Tingyao Xiong, Hung-Ping Tsao, Jonathan I. Hall. "General Eulerian Numbers and Eulerian Polynomials", Journal of Mathematics, 2013 Publication	<1 %
10	epdf.pub Internet Source	<1 %
11	Bing Wang, Jialiang Li, Xiaoguang Wang. "Change point detection in Cox proportional hazards mixture cure model", Statistical Methods in Medical Research, 2020 Publication	<1 %
12	Submitted to Cardiff University Student Paper	<1 %
13	Submitted to Koc University Student Paper	<1 %
14	Ryoto Ozaki, Yoshiyuki Ninomiya. "Information criteria for detecting change - points in the cox proportional hazards model", Biometrics, 2023 Publication	<1 %