Regression Models for the Analysis of Middle-censored Lifetime Data

Thesis submitted to the
Cochin University of Science and Technology
for the Award of Degree of

Doctor of Philosophy

under the Faculty of Science

by

PRASAD S.

DEPARTMENT OF STATISTICS
COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY
COCHIN-682 022

August 2017

CERTIFICATE

This is to certify that the thesis entitled **"Regression Models for the Analysis of Middle-censored Lifetime Data "** is a bona fide record of work done by Mr.Prasad S. under my guidance in the Department of Statistics, Cochin University of Science and Technology and that no part of it has been included anywhere previously for the award of any degree or title.

Kochi-22
August 2017

Dr. P. G. Sankaran
Professor,
Department of Statistics,
Cochin University of
Science and Technology.

CERTIFICATE

Certified that all the relevant corrections and modifications suggested by the audience during pre-submission seminar and recommended by the Doctoral committee of the candidate have been incorporated in the thesis.

Kochi- 22
August 2017

Dr. P. G. Sankaran
Professor,
Department of Statistics,
Cochin University of
Science and Technology.

DECLARATION

This thesis contains no material which has been accepted for the award of any other Degree or Diploma in any University and to the best of my knowledge and belief, it contains no material previously published by any other person, except where due references are made in the text of the thesis.

Kochi- 22
August 2017

Prasad S.

Acknowledgements

It is with all sincerity and high regards that I express my deep sense of gratitude to my supervising guide Dr. P. G. Sankaran, Professor and Head, Department of Statistics, Cochin University of Science and Technology, for his meticulous guidance, consistent encouragement and valuable suggestions throughout my research period. This feat was possible only because of the unconditional support provided by him.

I profoundly thank Dr. K. R. Muraleedharan Nair, Retired Professor, Dr. N. Balakrishna, Professor, Dr. K. C. James, Professor, Dr. Asha Gopalakrishnan, Professor and Dr. S. M. Sunoj, Professor, Department of Statistics, CUSAT for their valuable suggestions and help to complete this endeavour. I remember with deep gratefulness all my former teachers who gave me light in life through education. I extend my sincere thanks to all non-teaching staff of the Department of Statistics, CUSAT for their kind cooperation.

I thank my fellow research scholars, Department of Statistics, CUSAT for their help and support.

I take this opportunity to thank Directorate of Collegiate Education, Government of Kerala for permitting me to carry out this research work.

I am failing in words to express my feelings to my family members for their love, care and support. I owe everything to them.

Above all, I bow before the grace of the Almighty

Prasad S.

To my loving daughter Ameya and other members in family

Contents

List of Tables

List of Figures

Chapter 1

Preliminaries

1.1 Introduction

The statistical data that measure time to occurrence of some event is referred to as lifetime data, also called failure time data or survival data. Such data correspond to the time from a well-defined time origin until the occurrence of some particular event of interest or end-point, which marks the termination of the experiment. In medical research, the time origin may often corresponds to the recruitment of an individual into a c1inical trial to compare two or more treatments and the event of interest may be the time to the diagnosis of a particular condition, or the occurrence of some adverse event, or the time when a disease symptom subsided significantly, or the end-point which is the death of a patient. In an engineering study, one may be interested in observing the event of failure of an electronic component from the time when it was put on test. The branch of statistics that deals with modelling and analysis of lifetime data is referred to as survival analysis or reliability analysis. Survival analysis concerns with the models for the lifetime data from medical and biological studies, whereas reliability theory discusses the models for lifetime of components and systems in engineering studies. The origin of survival analysis might be attributed to the early work on mortality tables centuries ago. A major

advance in the field of survival analysis took place in the later of twentieth century. Over the past few decades, there has been considerable progress in the research for analyzing events observed over time, largely motivated by problems arising in the analysis of data from clinical trials in medical research. Survival analysis has found its applications in many areas including medicine, biology, public health, epidemiology and economics.

The definition of lifetime, as mentioned earlier, includes a time scale, time origin and the specification of the event that determines the lifetime. The lifetime or failure time is usually considered as a nonnegative real valued random variable. The analysis of lifetime data are not amenable to conventional statistical procedures because of its special features like censoring and truncation, which are discussed in a forthcoming section.

We now give some examples of lifetime data that arise in different practical situations.

Example 1.1. *One early example of the use of survival methods is found in the work by Turnbull et al. (1974). The study describes the survival experience and identification of risk factors associated with patients requiring heart transplants. The ultimate aim of a heart transplant programme is to restore the patient to the level of risk of his or her healthy contemporaries of the same age. In this heart transplant programme, patients are assessed for transplant and then, if suitable, have to await a donor heart. One consequence of this wait is that patients may die before a suitable donor has been found. For such patients, the waiting time from the date of assessment of suitability until death is considered as their lifetime. For those who receive a transplant their survival time is measured from the date of assessment*

of suitability and consists of their waiting time to transplant, plus their survival time from their transplant until death.

Example 1.2. *Nash et al. (1990) considered a group of patients with severe pain due to some disease. The patients were given transcutaneous electrical nerve stimulus (TENS) treatments for relieving pain. The study was intended to compare the value of high as opposed to low frequency TENS for the relief of pain in a randomised trial. They measured the time taken, from the date of randomisation for the patients to achieve a 50% reduction in pain levels as compared to those levels recorded at admission to the trial, where pain was measured using a visual analogue scale. This elapsed time is considered as the lifetime of the patient.*

Example 1.3. *A standard experiment in the investigation of carcinogenic substances is one in which laboratory animals are subjected to doses of the substance and then observed to see if they develop tumors. The main variable of interest is the time to appearance of a tumor, measured from when the dose is administered (Lawless (2003)).*

Example 1.4. *Suppose that we are interested in studying patients with systemic cancer who subsequently develop a brain metastasis, our ultimate goal is to prolong their lives by controlling the disease. A group of patients, treated with radio therapy, were followed from first day of their treatment until the recurrence of original tumor. In this study, lifetime is defined as the time to recurrence of tumor for each patient (Le (1997)).*

1.2 Basic Concepts

Let T be a nonnegative random variable representing the lifetime of an individual in a population. Let the distribution function of T be given by $F(t) = \mathrm{P}(T \leq t)$, $t \in \mathbb{R}$. Assume that $F(\cdot)$ is absolutely continuous with respect to the Lebesgue measure on the real line $\mathbb{R}$, and that $f(\cdot)$ is the probability density function of T. We now discuss some fundamental quantities related to the distribution of the lifetime variate T.

1.2.1 Survival Function

The basic quantity employed to analyze the lifetime data is the survival function, denoted by $S(t)$. It is defined by

$$S(t) = P(T > t) = \int_t^\infty f(u)du. \tag{1.1}$$

The function $S(t)$ measures the probability of an individual surviving beyond time t and $S(t) = 1 - F(t)$. Note that $S(t)$ is a non-increasing function with $S(0) = 1$ and $S(\infty) = \lim_{t\to\infty} S(t) = 0$. It may be noted that $f(t) = -\frac{dS(t)}{dt}$. In the context of reliability analysis, $S(t)$ is referred to as the reliability function.

1.2.2 Hazard Rate

One of the basic concepts associated with the lifetime distribution is the hazard rate $h(t)$, which is defined by

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t \,|T \geq t\,)}{\Delta t}. \tag{1.2}$$

The hazard rate specifies the instantaneous rate of failure or death of an individual at time t, given that the individual survives at least t units of time. Thus $h(t)\Delta t$ is the approximate probability of failure in $[t, t + \Delta t)$, given survival of at least t units of time. The hazard rate is sometimes referred to as hazard function or force of mortality.

Note that $h(t)$ is a nonnegative function and is related to the survival function and probability density function by the identity

$$h(t) = -\frac{\mathrm{d}\log(S(t))}{\mathrm{d}t} = \frac{f(t)}{S(t)}. \tag{1.3}$$

The cumulative hazard function, denoted by $H(t)$ is defined as

$$H(t) = \int_0^t h(u)\mathrm{d}u. \tag{1.4}$$

It is well known that $h(t)$, or equivalently $H(t)$, determines the distribution uniquely by the identity

$$S(t) = \exp(-H(t)) = \exp\Big(-\int_0^t h(u)\mathrm{d}u\Big). \tag{1.5}$$

1.3 Censoring

The main feature of survival data that renders standard statistical inference methods inappropriate is that lifetimes are frequently censored. The survival time of an individual is said to be censored when the information about his lifetime is incomplete. It may be due to certain unavoidable or uncontrollable circumstances. When censoring occurs, exact lifetimes are known only for a portion of the study subjects, and lifetimes for remaining individuals are known only to belong to a subset of the support of T. For example, in medical studies, patients in a clinical trial may withdraw from the study, or the study may need to be terminated at a prefixed time point. In such situations, only a partial information about the lifetimes are known for those censored cases. There are various categories of censoring that occur naturally in observation schemes. The well known censoring schemes are right censoring, left censoring, interval censoring, double censoring etc. We now give a short survey of these schemes of censoring.

1.3.1 Right Censoring

Assume that a group of individuals are put on statistical investigation and are continuously observed for a possible occurrence of a certain event of interest. The resulting set of observations often include some individuals who do not fail during their observation period and the data on these individuals are said to be right censored. In such situations the lower bounds of the lifetimes are available for some individuals. Thus right censoring may occur when the life test terminates before all

individuals fail, or when the individuals in a prospective study are lost to follow up because they move away from the region (Lawless (2003)).

Example 1.5. *Prentice (1973) discussed the data on 40 lung cancer patients taken from a study designed to compare the effect of two chemotherapy treatments in prolonging survival time. All patients in the study were received prior therapy and then randomly assigned to one of the two treatments termed standard and test. Survival times are measured in days from the start of the treatment for each patient. Observations corresponding to the patients who were still alive at the time of data collection are considered to be right censored.*

There are different forms of right censoring in practice. Some of them are Type I censoring, Type II censoring and independent random censoring schemes. We give a brief description of these schemes below.

(i) Type I Censoring: Suppose we fix a specific time point to terminate the experiment. In Type I censoring, the event of interest is observed only if it occurs prior to this pre-specified time. Let there be n individuals under investigation. Each individual has a fixed potential censoring time $C_i > 0$ such that the lifetime T_i is observed if $T_i \leq C_i$, otherwise we only know that $T_i > C_i$, $i = 1, 2, ..., n$. For example, consider a clinical trial concerning the duration of remission for patients with leukemia, which was planned to run for one year with patients entering the trial over that period. The lifetime variable T_i of the i'th patient is the duration of the remission measured from time of entry to the study, and C_i would be the time between the date of entry and the end of the study (Lawless (2003)).

(ii) Type II Censoring: Here, unlike the previous case, we don't fix the upper bound for the experiment duration, but we fix the number of individuals, say $r, (r \leq$

n) that should fail to mark the end of the experiment, where n is the total number of individuals that are put on test. That is, in Type II censoring, the study continues until the r smallest lifetimes in a random sample of size n are observed, and r is chosen before the experiment begins. A significant advantage of type II censoring is that, we know the number of observed lifetimes in advance. Such a censoring scheme is often found in a life testing experiment, where the experiment stops when the first r units under investigation fails. Thus, the r smallest lifetimes are observed and the rest are right censored.

(iii) Independent Random Censoring: In independent random censoring, each individual is assumed to have a lifetime T and a censoring time C, where T and C are independently distributed continuous random variables. This implies that the censoring time C is non-informative in analyzing the lifetime T. The observed vector will be (X, δ), where $X = \min(T, C)$ and the censoring indicator δ is defined in such a way that $\delta = 1$ if $T \leq C$, and $\delta = 0$ otherwise.

1.3.2 Left Censoring

Left censoring occurs when the event of interest has already occurred for certain individuals before they enter into the study. For others, the exact event time is observed. In such contexts, for those who are censored, the exact values of the lifetime are not observed, but only the upper bounds of the lifetimes are recorded.

Example 1.6. *If we follow individuals until they become HIV positive, we may record a failure when an individual first tests positive for the virus. However, we may not know exactly the time of first exposure to the virus, and therefore we do not know exactly when the failure occurred. Thus, the lifetime is left censored since*

the true lifetime, which ends at exposure, is shorter than the follow up time, which ends when the individual tests positive.

1.3.3 Interval Censoring

Interval censored data arise when the exact lifetime is not observable, but we know that it lies in an interval of time points obtained from a sequence of examination times. Here the study subject, or failure time processes of interest, is not under continuous observation. Consequently the event times are neither observed exactly nor right censored. Thus individuals are known to have experienced the event of interest within an interval of time, but observations are taken intermittently. More generally, one could define an interval censored observation as a union of several non-overlapping windows of time points (Turnbull (1976)).

Interval censored lifetime data occur in many areas including demographical, epidemiological, financial, medical, and sociological studies. A typical example of interval censored data occurs in medical or health studies that entail periodic follow-ups, and many clinical trials and longitudinal studies fall into this category. We report one such example below.

Example 1.7. *Odell et al. (1992) analyses the data set obtained from the Framingham Heart Study, where the ages at which individuals first developed coronary heart disease (CHD) are usually known exactly. However, the ages of first occurrence of the subcategory angina pectoris may be known only to be between two consecutive clinical examinations, approximately two years apart. Such observations would be interval censored.*

1.3.4 Double Censoring

Let us consider a survival study involving two related events E_1 and E_2, and let their times of occurrences be T_1 and T_2 respectively with $P(T_1 \leq T_2) = 1$. Assume that the survival time of interest is the inter-occurrence time, say $T = T_2 - T_1$. When the observations on both T_1 and T_2 are interval censored, we say that the survival time T is doubly censored or doubly interval censored (De Gruttola and Lagakos (1989), Sun (1995), Sun (2004)). This means that instead of observing T_1 and T_2 exactly, one only observes two intervals, say $(U_1, V_1]$ and $(U_2, V_2]$ such that $T_1 \in (U_1, V_1]$ and $T_2 \in (U_2, V_2]$, where $U_1 \leq V_1$ and $U_2 \leq V_2$ with probability 1, and $U_1 \leq U_2$ and $V_1 \leq V_2$ with probability 1. In other words, the observations on T are doubly censored. In biomedical studies, we often come across a special type of doubly censored data in which T_2 is only right censored. In this case, one has either $U_2 = V_2$ or $V_2 = \infty$. Doubly censored failure time data are often found in a disease progression study, where the two events may be thought of as an infection and the subsequent onset of a certain disease respectively. In the following, we report one such example.

Example 1.8. *Kim et al. (1993) analyzed AIDS data, where inter-occurrence time between two related events namely, HIV seroconversion and AIDS diagnosis times are of interest. Let T_1 and T_2 denote HIV seroconversion and AIDS diagnosis times respectively, and $T = T_2 - T_1$ be the AIDS incubation time or AIDS latency time. This incubation time provides information about HIV infection progression and plays an important role in predicting HIV prevalences. For most AIDS cohort studies, because HIV infection is usually determined through periodic blood tests, observations on it are not observed and are commonly interval censored. Also,*

observations on the diagnosis of AIDS could be, for example, right censored due to the end of the study, thus yielding doubly censored data on T.

Another type of doubly censored data that often arise in situations where both left censoring and right censoring occur together in a study. The survival time of interest is observed exactly if it is within a window of time points, and left or right censored if it is to the left or to the right of the window (Turnbull (1974), Chen and Zhou (2003), Cai and Cheng (2004)). Here, unlike the doubly interval censored data, some exact failure times are observed, but if not, they become case I interval censored data. The data can be represented by a pair of variables (X, δ), where $X = \max(\min(T, R), L)$; δ is 1 if X is an event time; 0 if T is right censored; and -1 if T is left censored. Here L is the time before which some individuals experience the event and R is the time after which some individuals experience the event. T will be known exactly if it is less than or equal to R and greater than or equal to L. The methods of data analysis for these two variants of double censoring are entirely different. We report a real life example for the latter case below.

Example 1.9. *FitzSimmons (1993) describes a biomedical study concerned with patients with cystic fibrosis. For these patients, the onset of Pseudomonas aeruginosa (PA) is monitored as an important landmark event for lung disease and this affects the survival rate adversely. Suppose we are interested in observing the age at which PA infection occurs, which is treated as the lifetime. The patients with cystic fibrosis enrolled in the study may either have developed PA infection before they enter the follow up study, or they may not develop PA infection by the end of follow up. The first situation renders left censoring and the second one causes right censoring. Also exact lifetimes are observable for those who develop the PA*

infection in between the entry time and end of study. This poses a double censoring scenario.

For an elaborate discussion on different schemes of censoring, one may refer to Lawless (2003), Klein and Moeschberger (2005), and Sun (2006).

1.3.5 Middle-censoring

Middle-censoring introduced by Jammalamadaka and Mangalam (2003) occurs in situations where a data point becomes unobservable if it falls inside a random censoring interval. In such situations, the exact lifetimes are available for some individuals, and for others random censoring intervals within which the true lifetimes belong are observed. To be more precise, let T be the random variable representing the lifetime of interest and let (U, V) be a bivariate random variable representing the censoring interval such that $P(U < V) = 1$. Under the middle-censoring set up, the exact lifetime T becomes unobservable if $T \in (U, V)$ and in such instances we only observe the censoring interval (U, V). On the other hand, exact observation on T will be available if $T \notin (U, V)$.

Jammalamadaka and Mangalam (2003) describe middle-censoring scheme as an important variation and generalization of censoring, since left censoring, right censoring are special cases of this middle-censorship by suitable extensions of the endpoints of the censoring interval. Middle-censoring, where a random middle part is missing, appears at a first glance as complementary to the idea of double censoring (Turnbull (1974)), where the middle part is what is actually observed. However, if we consider these two schemes carefully, along with the resulting data sets, they

turn out to be quite distinct ideas. Middle-censoring scheme is related to the mixed interval censoring scheme (Yu et al. (2001)), which includes interval censored data together with few exact lifetimes. Even though these two observation schemes are entirely different, they produce similar data sets under certain restrictions. A nice account of this is available in Shen (2011). Following are few situations where one can come across middle-censoring scheme.

Example 1.10. *In an early childhood learning center, interest often focuses upon testing children, to determine the age at which a child learns to accomplish certain skill. The age at which a child acquires the skill would be considered as the lifetime. Let us envisage a scenario where there are no late entries, which correspond to cases where the child already learned the skill before entering the study, and losses, which correspond to the cases where the child had not acquired the skill by the end of the study. Assume further that during a fixed time interval the observations were not possible, perhaps due to a sudden outbreak of war, or a natural calamity etc. If some children of varying ages develop the skill during this period of closure, we are unable to observe the exact lifetime T of skill development, rather we have only the information that the development occurred during a certain time interval (U, V).*

Example 1.11. *In a prognostic study, the patients under observation may be withdrawn from the study for a short period of time for some unforeseen reasons and may return to study with a changed status of event of interest. This causes the exact event time to become unobservable, but gives the information that it is contained in some censoring interval. For those who were not missing, the exact event times are recorded. Thus, the observed data consists of both exact observations and censoring intervals, leading to a middle-censored set up.*

Example 1.12. *In the case of reliability applications, suppose an experimenter observes several mechanical components of a large machine, which is subject to replacement of failed components. Suppose that the lifetime of interest is the exact time duration (in hours) of components at work. If the experimenter accidentally miss to observe a few of them for a short period of time and they happen to fail during that particular time period, he only knows that those lifetimes belong to the censoring interval. For other components, exact times of failure are observed. The resulting data set is also middle-censored.*

One important feature that can be noted in all these examples is that, in each case, although the unobserved time interval is a fixed one, it is indeed a random interval relative to the individual's lifetime of interest. Moreover, one can observe that the middle-censoring scheme is totally different from the scheme of interval censoring. This is because of the fact that under interval censoring, one never observes an exact lifetime and always observes the censoring events, as we see in Example 1.7. But under middle-censoring, since the individuals are under continuous observations, it is possible to observe exact lifetimes for some of the individuals. In Example 1.7, we stated that the ages of first occurrence of angina pectoris are known only to be between two consecutive clinical examinations, and these examination times are, in effect, serving as censoring events. Therefore exact age of first occurrence of angina pectoris remains totally unobservable. Beadle et al. (1984) report a retrospective study to compare the cosmetic effects of radiotherapy alone versus radiotherapy and adjuvant chemotherapy on women with early breast cancer. In this study, the patients were observed initially every 46 months, and when their recovery progressed, the interval between visits lengthened. Here the event of interest was the first appearance of moderate or severe breast retraction,

a cosmetic deterioration of the breast. For a patient, the exact time of retraction is known to fall only in the interval between visits, and the exact event time is never known. Thus the lifetimes of patients in this study are interval censored. In both these examples, the study subjects are observed intermittently, and hence only censoring events are observed. This feature distinguishes interval censoring scheme from middle-censoring scheme. Under middle-censoring scheme, since the subjects are continuously observed, except possibly within a random censoring interval, it is possible to observe exact lifetimes for some of the subjects, in which case we observe lifetime along with the censoring indicator, say, $\delta = 1$. For the rest, we observe censoring intervals along with $\delta = 0$. In view of these examples, it is apparent that the characteristics of middle-censoring and interval censoring schemes are entirely different, and they produce different types of data sets.

1.4 Truncation

Truncation is another feature in lifetime data observation schemes. Under truncation, an individual enters the study if and only if the corresponding lifetime exceeds some threshold value, say t_0. Therefore, if an individual enters the study, we observe the corresponding threshold value t_0 along with its lifetime or censoring time as the case may be. Individuals who do not enter the study are totally unobserved, and the experimenter doesn't know even their existence. Often truncation is observed along with different forms of censoring. The situation described here is known as left truncation. Other truncation schemes, like right truncation and double truncation, are also observed in practice. For a concise description and some practical examples, one could refer to Lawless (2003).

1.5 Inference Procedures

The inference procedures employed for the analysis of lifetime data can be broadly classified into three methods namely parametric, semiparametric, and nonparametric methods. In parametric approach, we assume that the lifetime variate T has a probability density function $f(t, \boldsymbol{\vartheta})$, where $f(t, \boldsymbol{\vartheta})$ has a specified known functional form, but the parameter vector $\boldsymbol{\vartheta}$ is unknown. Various parametric families of distributions are used in the analysis of lifetime data. In the continuous set up, exponential, gamma, Weibull, inverse Weibull, lognormal, log-logistic, Pareto and inverse Gaussian are the most widely used lifetime distributions. The estimation of the parameters is done by different procedures such as method of maximum likelihood, method of moments, Bayesian techniques etc. For a comprehensive review on parametric models, one may refer to Martz and Waller (1982), Sinha (1986), and Lawless (2003) among others.

In many practical situations, the lifetime data may not meet the assumptions of a parametric model. In such contexts, semiparametric or nonparametric methods are employed. Semiparametric methods do not make any assumption about the underlying form of the lifetime distribution, but make some postulations on the failure process. In contrast, nonparametric methods allow a completely distribution free approach.

1.5.1 Nonparametric Estimation Under Right Censoring

When we have a random sample of exact lifetimes, the empirical survival function given by

$$\hat{S}_{ESF}(t) = \frac{k}{n}, \tag{1.6}$$

where n is the total number of observations, k is the number of observations which are greater than or equal to t, $t \in \mathbb{R}$, provides a nonparametric maximum likelihood estimator (NPMLE) for the true survival function $S(t)$ of T. When the sample contains censored observations, the estimator (1.6) does not account the information provided by an individual whose lifetime is censored before time t. Thus, a modification to (1.6) becomes necessary. Accordingly, Kaplan and Meier (1958) suggested a nonparametric estimator for survival function under right censoring.

1.5.1.1 Kaplan-Meier Estimator

Let (X_i, δ_i) be a random sample of lifetimes under right censoring, where X_i is either an observed lifetime or a censoring time, and δ_i is the indicator function which equals to one if X_i is a lifetime, and 0 otherwise, $i = 1, ..., n$. Assume that there are $k(\leq n)$ distinct lifetimes $X'_1, ..., X'_k$. Let d_j represents the number of deaths at X'_j. Then the Kaplan-Meier estimator of $S(t)$ is defined as

$$\hat{S}(t) = \prod_{j:X'_j<t} \frac{n_j - d_j}{n_j}, \tag{1.7}$$

where n_j is the number of individuals at risk at X'_j, that is those individuals who are uncensored and alive just prior to X'_j. The estimator (1.7) is often referred

to as the product limit estimator. It can be noted that this estimator does not change its value at censoring times. It can be shown that $\hat{S}(t)$ is an NPMLE of $S(t)$ (see Lawless (2003)). When there are no censored observations, $\hat{S}(t)$ reduces to the estimator (1.6). The estimator for the variance of $\hat{S}(t)$ is given by

$$\hat{Var}(\hat{S}(t)) = \hat{S}(t)^2 \sum_{j:X'_j<t} \frac{d_j}{n_j(n_j - d_j)}.$$

1.5.1.2 Nelson-Aalen Estimator

Nelson (1969) proposed a nonparametric estimator of the cumulative hazard function (1.4). The estimator is given by

$$\hat{H}(t) = \sum_{j:X'_j \leq t} \frac{d_j}{n_j}, \tag{1.8}$$

where d_j and n_j are defined as in the case of Kaplan-Meier estimator. $\hat{H}(t)$ is sometimes called the empirical cumulative hazard function, but is more commonly known as the Nelson-Aalen (NA) estimator, as it was reinvented by Aalen in his doctoral work in 1972. Note that $\hat{H}(t)$ is an NPMLE of $H(t)$ (see Lawless (2003)). The estimator of the variance of $\hat{H}(t)$ is given by

$$\hat{Var}(\hat{H}(t)) = \sum_{j:X'_j \leq t} \frac{d_j(n_j - d_j)}{n_j^3}.$$

Using the identity given in (1.5), we can estimate $S(t)$ as

$$\tilde{S}(t) = \exp(-\hat{H}(t)).$$

Both Kaplan-Meier and Nelson-Aalen estimators possess desirable large sample properties like strong consistency and asymptotic normality. For more properties of these estimators, one may refer to Lawless (2003).

1.5.2 Bayesian Analysis

We now consider the Bayesian method of estimating the unknown parameter $\boldsymbol{\vartheta}$ in the probability density function $f(t;\boldsymbol{\vartheta})$ of the lifetime variate T. Unlike the frequentist approach of estimation, the Bayesian approach involves the terminology of subjective probability for representing the uncertainty about $\boldsymbol{\vartheta}$. This term is used to express our degree of belief about the truth of an event, here observing the true value of $\boldsymbol{\vartheta}$ in a range of its possible values, which is not based on the number of times it happens in a long series of experiments. Moreover, this subjective probability may change when new information is gathered as time evolves. Bayes' theorem gives us a formal rule for determining how our probabilities will change when we acquire new information. Since the parameter $\boldsymbol{\vartheta}$ is unknown, we regard it as a random variable and our knowledge about this parameter is represented by means of a probability distribution, called prior distribution, denoted by $\pi_0(\boldsymbol{\vartheta})$. Apart from the prior believes, we also have information about the unknown parameter contained in the observed sample $t_1, t_2, ..., t_n$. The term *likelihood* is used to represent the probability of the data outcome, given the prior belief, viewed as a function of the parameter. We usually denote the likelihood function as $L(t_1, t_2, ..., t_n|\boldsymbol{\vartheta})$. This function is particularly helpful in comparing the relative plausibility of different parameter values, given the observed data. The prior density $\pi_0(\boldsymbol{\vartheta})$ that we choose reflects our opinion about $\boldsymbol{\vartheta}$ before any data is observed. After the data has been

observed, more information may be obtained. We incorporate this updated information using a probability density function called posterior distribution, denoted by $\pi(\boldsymbol{\vartheta}|t_1, t_2, ..., t_n)$, which can be written as

$$\pi(\boldsymbol{\vartheta}|t_1, t_2, ..., t_n) = L(t_1, t_2, ..., t_n|\boldsymbol{\vartheta}) \cdot \pi_0(\boldsymbol{\vartheta}).$$

Now the inference on $\boldsymbol{\vartheta}$ is made from this posterior distribution, under a properly set loss function. In practice, such an analysis involves laborious computations which deter its advancement at an earlier stage. With the advent of sophisticated computational softwares, the area of Bayesian analysis recorded an immense growth. Several works on Bayesian analysis of survival data appeared in literature over the past few decades. Kalbfleisch (1978) carried out a seminal work on nonparametric treatment of Bayesian analysis with survival data. Some important works on Bayesian analysis of lifetime data are found in Sinha et al. (1999), Walker and Mallick (1999), Cheng et al. (1999) and Sinha et al. (2003) among others. Recently, Danish and Arshad (2017) discussed parametric proportional hazards model for randomly censored survival data. In the case of two sample censored data, a nonparametric Bayesian analysis is carried out by Shang and Reilly (2017). For double-censored durations, Dörre and Weißbach (2017) discussed a Bayesian estimation of a proportional hazards model. A comparative study of maximum likelihood and Bayes estimators for randomly censored discrete lifetime data is carried out by Krishna and Goel (2017). For progressive Type II censored data, maximum likelihood and Bayes estimators of the parameters are analyzed by Lee and Cho (2017).

1.6 Regression Models

In biomedical survival studies, the prediction of the future of a patient with respect to duration, course, and outcome of a disease, called prognosis, is of great importance. A medical history as well as information from pathologic, clinical and laboratory data are collated by a physician before he decides on the course of treatment needed. Therefore many medical charts contain a large number of patient characteristics, also called covariates or concomitant variables. In clinical trials, factors such as age, gender and general conditions of the patient can be considered as covariates. The covariates are mainly employed in survival studies to describe the heterogeneity in the population under consideration. In such contexts, the main objectives are to understand the relationship between the lifetime and the covariates, and then to exploit this relationship to the benefit of forecasting. To this end, we generally employ regression models in survival analysis. The primary aim in doing this is to study the relationship between the lifetime and given covariates. In some practical situations, the effect of covariates on lifetime variable may change over time and such covariates are referred to as time-dependent or time-varying covariates.

Parametric and semiparametric regression models are often employed to analyze lifetime data with covariates. The parametric regression analysis involves the specification of the distribution of the lifetime variate T, given a $p \times 1$ vector of covariates $\boldsymbol{z}$. A regression model can be developed by specifying a relationship between the model parameters and covariates. For example, consider the Weibull distribution with shape parameter α and scale parameter β. Conditional on $\boldsymbol{z}$, let

T assume Weibull distribution with parameters α and $\beta = \beta(\boldsymbol{z})$. Then the survival function of T, given covariate $\boldsymbol{z}$, is given by

$$S(t|\boldsymbol{z}) = \exp\left(-\left(\frac{t}{\beta(\boldsymbol{z})}\right)^{\alpha}\right).$$

A convenient specification of $\beta(\boldsymbol{z})$ is $\beta(\boldsymbol{z}) = \exp(\boldsymbol{z}^{\top}\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is the $p \times 1$ vector of regression parameters and $\boldsymbol{a}^{\top}$ represents the transpose of vector $\boldsymbol{a}$. In such situations, one would be interested in estimating the regression parameter $\boldsymbol{\theta}$ and testing whether it has a significant effect on lifetime variate.

The log-location scale family of distributions or the accelerated failure time models are the widely used types of parametric regression models. Parametric regression models have been studied in literature by Feigl and Zelen (1965), Zippin and Armitage (1966), Glasser (1967), and Prentice (1973) among others.

In semiparametric regression models, we do not make any assumption about the underlying form of the lifetime distribution, but some postulations on the relationship between covariates and the lifetime variable are made. The proportional hazards model introduced by Cox (1972) is the commonly employed semiparametric regression model in survival analysis.

1.6.1 Proportional Hazards Model

The proportional hazards model assumes that the covariates have a multiplicative effect on the hazard function of lifetimes of individuals. The proportional hazards model associates the covariates effect on lifetime variate T by specifying the hazard

function of T, given $\boldsymbol{z}$ as

$$h\left(t\,|\boldsymbol{z}\right) = h_0\left(t\right) r\left(\boldsymbol{z}, \boldsymbol{\theta}\right),$$

where $h_0\left(t\right)$ is the baseline hazard function and $r\left(\boldsymbol{z}, \boldsymbol{\theta}\right)$ is a positive real valued function. The proportional hazards model possesses the property that, any two individuals have hazard functions that are constant multiples of each other. The model involves a parameter $\boldsymbol{\theta}$. However, the baseline hazard function is usually treated nonparametrically and is left arbitrary.

A specification of the proportional hazards model proposed by Cox (1972) with $r\left(\boldsymbol{z}, \boldsymbol{\theta}\right) = \exp\left(\boldsymbol{\theta}^\top \boldsymbol{z}\right)$, is widely used in literature and is known as Cox proportional hazards model. Therefore the Cox proportional hazards model may be written as

$$h\left(t\,|\boldsymbol{z}\right) = h_0\left(t\right) \exp\left(\boldsymbol{z}^\top \boldsymbol{\theta}\right). \tag{1.9}$$

The model (1.9) assumes that covariates have multiplicative effect on the hazard function of the lifetime variable. The primary objective in this set up is to estimate the regression parameter and the baseline hazard function.

Suppose that the lifetime variable T is randomly right censored by the censoring variable C. We observe $(X, \delta, \boldsymbol{z})$ where $X = \min(T, C)$, $\delta = I(X = T)$ is the censoring indicator and $\boldsymbol{z}$ is a $p \times 1$ vector of covariates. The observed data consists of $(X_i, \delta_i, \boldsymbol{z}_i)$, $i = 1, 2, ..., n$, which are independent and identically distributed copies of $(X, \delta, \boldsymbol{z})$. To estimate $\boldsymbol{\theta}$, Cox (1972) proposed the partial likelihood method,

where the partial likelihood function is formulated as

$$L\left(\boldsymbol{\theta}\right)=\prod_{i=1}^{n}\left(\frac{\exp\left(\boldsymbol{\theta}^{\top}\boldsymbol{z}_{i}\right)}{\sum_{l=1}^{n}Y_{*l}\left(X_{i}\right)\exp\left(\boldsymbol{\theta}^{\top}\boldsymbol{z}_{l}\right)}\right)^{\delta_{i}},$$

where $Y_{*l}(t) = I(X_i \geq t)$. Maximum likelihood estimator (MLE) of $\boldsymbol{\theta}$ can be obtained by maximizing the partial likelihood function $L\left(\boldsymbol{\theta}\right)$. A nonparametric estimator of the baseline cumulative hazard function is then given by

$$\hat{H}_{0}\left(t\right)=\sum_{i:X_{i}\leq t}\left(\frac{\delta_{i}}{\sum_{l=1}^{n}Y_{*l}\left(X_{i}\right)\exp\left(\hat{\boldsymbol{\theta}}^{\top}\boldsymbol{z}_{l}\right)}\right).$$

Using the identity (1.5), $S_0(t)$ can be estimated as $\hat{S}_0(t) = \exp(-\hat{H}_0(t))$. From (1.9), the survival function of T, given $\boldsymbol{z}$, is given by

$$S(t|\boldsymbol{z})=S_{0}(t)^{\exp(\boldsymbol{z}^{\top}\boldsymbol{\theta})},$$

where $S_0(t)$ is the baseline survival function and the estimator of the survival function $S(t|\boldsymbol{z})$ is obtained as

$$\hat{S}(t|\boldsymbol{z})=\hat{S}_{0}(t)^{\exp(\boldsymbol{z}^{\top}\hat{\boldsymbol{\theta}})}.$$

For various properties of these estimators, one could refer to Lawless (2003).

1.6.2 Additive Hazards Model

The proportional hazards model provides a convenient way of summarizing covariate effects in terms of relative risks. However, there are occasions where a measure

of the additive effect of covariates is preferred over a multiplicative effect. In such situations, the additive hazards model, which relates the conditional hazard function of the lifetime linearly to the covariates is more suitable. In contrast to the proportional hazard models, the additive hazards model specifies that the hazard function, given a set of covariates, is the sum of the baseline hazard function and the regression function of the covariates.

Let $\boldsymbol{z}(t) = (z_1(t), ..., z_p(t))^\top$ be a $p \times 1$ vector of possibly time-dependent covariates. Then the additive hazards model is defined by

$$h(t|\boldsymbol{z}) = h_0(t) + \boldsymbol{z}(t)^\top \boldsymbol{\theta}(t), \tag{1.10}$$

where $h_0(t)$ is an arbitrary baseline hazard function and $\boldsymbol{\theta}(t) = (\theta_1(t), ..., \theta_p(t))^\top$ is the $p \times 1$ vector of regression parameters. The model (1.10) is due to Aalen (1989) and it allows $\boldsymbol{\theta}(t)$ to be a function whose values change over time. To estimate the cumulative regression functions, $B_k(t) = \int_0^t \theta_k(t)$ and the standard errors of these functions, Aalen (1989) used the least squares approach. These estimators can then be smoothed to obtain estimators of $\boldsymbol{\theta}_k(t)$. The least squares estimators and their variances were developed using the theory of counting process. Aalen (1993) discussed a method for the goodness of fit of this model. For more details on this topic, one could refer to Zahl and Tretli (1997), Borgan and Langholz (1997) and Klein and Moeschberger (2005).

Lin and Ying (1994, 1997) studied a simplified form of the additive hazards regression model by replacing the time-varying regression parameters in the Aalen model (1.10) with the time-independent regression parameters. The model is given

by

$$h(t|\boldsymbol{z}) = h_0(t) + \boldsymbol{z}(t)^\top \boldsymbol{\theta}, \tag{1.11}$$

where $h_0(t)$ is an arbitrary baseline hazard function and $\boldsymbol{\theta} = (\theta_1, ..., \theta_p)^\top$ is the $p \times 1$ vector of regression parameters.

In two sample set up, the additive hazards model addresses the risk difference, while the proportional hazards model concerns the risk ratio. The counting process approach together with the resultant martingale structure is used in the inference procedure of the parameters of the model (1.11). For modelling and analysis of lifetime data using the additive hazards model under different contexts, one could refer to Aalen (1989), McKeague and Utikal (1991), Lin and Ying (1994, 1997), Gupta et al. (1998), Sun et al. (2006), and Li and Ling (2012).

1.6.3 Quantile Regression Model

A fundamental quantity of interest in survival analysis is the quantile function. A lifetime distribution can be characterized by its quantile function. For the lifetime variate T, the quantile function $Q_T(\tau)$ is defined by

$$Q_T(\tau) = \inf\{t \in \mathbb{R} : F(t) \geq \tau\}, \quad 0 \leq \tau \leq 1. \tag{1.12}$$

In many instances the quantile function provides a better alternative to the distribution function in analyzing a distribution. Moreover, it has several useful features which are not shared by the distribution function or characteristic function. A rigorous treatment of quantile function and its properties is available in Gilchrist (2000), and for its applications in reliability theory, one may refer to Nair et al.

(2013).

In the presence of covariates, generally we analyze lifetime data using regression models like proportional hazards model and additive hazards model. One major drawback of hazard based regression approach is that it models the hazard function rather than the survival times directly. In literature, the accelerated failure time (AFT) model addresses this issue by regressing a monotone transformation of the lifetime variable, say, $\tilde{T} = \log T$ over the $p \times 1$ vector of recorded covariates $\boldsymbol{z}$. The model assumes the form

$$\tilde{T} = \boldsymbol{z}^\top \boldsymbol{b} + \epsilon, \tag{1.13}$$

where $\boldsymbol{b}$ is an unknown $p \times 1$ vector of regression parameters and ϵ is an error term. The inference procedures for the model (1.13) have been derived without specifying the actual distribution of ϵ, but generally require that the error term is independent of $\boldsymbol{z}$, as one can see in the works of Buckley and James (1979), Wei and Gail (1983), and Ritov (1990) among others. This assumption precludes data heteroscedasticity and entails a location-shift effect for each covariate, which sometimes mislead us when some covariate effects are nonconstant in nature over the support of T. An alternative and efficient approach called quantile regression was put forward by Koenker and Bassett (1978) by simply computing the regression quantiles for a transformation of the survival time by elegantly making use of the concept of conditional quantiles. Let $\boldsymbol{Z} = (1, \boldsymbol{z}^\top)^\top$ and $\tau \in [0, 1]$. We define the τ'th conditional quantile for $\tilde{T}$ as $Q_{\tilde{T}}(\tau|\boldsymbol{Z}) = \inf\{t : P(\tilde{T} \leq t|\boldsymbol{Z}) \geq \tau\}$. A quantile regression model may associate the conditional quantile $Q_{\tilde{T}}(\tau|\boldsymbol{Z})$ linearly to the covariate $\boldsymbol{Z}$ for each $0 < \tau < 1$ as

$$Q_{\tilde{T}}(\tau|\boldsymbol{Z}) = \boldsymbol{Z}^\top \boldsymbol{\beta}_0(\tau), \tag{1.14}$$

where $\boldsymbol{\beta}_0(\tau)$ is an unknown vector of regression parameters and represents the effects of covariates on the τ'th quantile of $\tilde{T}$ and may change with τ. Note that when $\boldsymbol{\beta}_0(\tau) = (Q_\epsilon(\tau), \boldsymbol{b}^\top)^\top$, where $Q_\epsilon(\tau)$ is the τ'th quantile of ϵ, the model (1.14) reduces to the model (1.13). The advantage of the model (1.14) as compared to (1.13) is that, it is more flexible in the sense that the effect of $\boldsymbol{Z}$ is not restricted to be constant across the range of values of τ. This approach was extended to the case of right censored survival data by Powell (1984, 1986) by introducing censored quantile regression, but with some restrictions on censoring variables. Later, Portnoy (2003) proposed a recursively re-weighted estimator as a generalization of the Kaplan-Meier estimator without imposing any stringent assumptions on censoring variable. Some modifications to this was made by Portnoy and Lin (2010). The analysis of right censored lifetime data using quantile regression model, when the data is subject to conditionally independent censoring was developed by Peng and Huang (2008). This method utilizes the martingale feature associated with the censored data, which is found to be helpful in developing inference procedures and in establishing the large sample properties of the estimators. Later, Ji et al. (2012) developed quantile regression model for doubly censored data. A much more generalized form of quantile regression model for T can be stated as

$$Q_T(\tau|\boldsymbol{Z}) = g(\boldsymbol{Z}^\top \boldsymbol{\beta}_0(\tau)), \quad 0 < \tau < 1, \tag{1.15}$$

where $Q_T(\tau|\boldsymbol{Z}) = \inf\{t : F(t|\boldsymbol{Z}) \geq \tau\}$ with $F(t|\boldsymbol{Z}) = P(T \leq t|\boldsymbol{Z})$, and $g(\cdot)$ is a known monotone link function. Here $\boldsymbol{\beta}_0(\tau)$ is a vector of unknown parameters representing true covariate effects on $Q_T(\tau|\boldsymbol{Z})$. A commonly used choice for $g(x)$

is e^x, in which case model (1.15) reduces to model (1.14). This specification of τ-varying coefficients can be effectively used to incorporate population heterogeneous covariate effects. That is, the effect of $\boldsymbol{z}$ is not restricted to have a constant impact for all values of τ. Our interest here is to draw inferences about the quantile process $\boldsymbol{\beta}_0(\tau)$, for $0 < \tau < 1$.

1.7 Motivation and Present Study

A large amount of literature has built up in the analysis of censored lifetime data, particularly on right censored, left censored, double censored and interval censored lifetime data. The development of new stochastic models for the analysis of middle-censored data in the presence of covariates is a topic of interest. However, only a limited number of works are reported in literature in this direction.

Jammalamadaka and Mangalam (2003) pointed out various applications of middle-censoring, and have developed an NPMLE of the distribution function of the lifetime variate. They have proved that NPMLE is always a Self Consistent Estimator (SCE) (see Tarpey and Flury (1996)). A variant of this SCE was proposed by Jammalamadaka and Iyer (2004) and its weak convergence was established. In the parametric context, Iyer et al. (2008) studied analysis of survival data under middle-censoring for the exponential distribution. Mangalam et al. (2008) developed a necessary and sufficient condition for the equivalence of SCE and NPMLE. Jammalamadaka and Mangalam (2009) have discussed middle-censored data for the von Mises model in the context of directional data. Shen (2010) proposed an inverse-probability-weighted estimator for the distribution function under middle-censoring

setup. Qin (2010) discussed a parametric estimation problem with exponential distribution under middle-censoring scheme. Shen (2011) has shown that NPMLE of the distribution function, under middle-censoring scheme, can be obtained by using Turnbull's EM algorithm or self-consistent estimating equation under certain conditions. Davarzani and Parsian (2011) developed an inference procedure for discrete middle-censored data. Davarzani et al. (2015) developed an inference procedure for middle-censored data under dependent setup. Jammalamadaka and Leong (2015) analyzed discrete middle-censored data in the presence of covariates. Wang (2016) developed a procedure for parametric estimation for middle-censored data with multiple causes of failure. Abuzaid et al. (2017) have discussed robustness of middle-censoring scheme in parametric survival models. Recently, Bennett et al. (2017) considered a parametric regression model under middle-censoring scheme. However, a systematic study on regression modelling and analysis of middle-censored lifetime data is not yet carried out. Motivated by this, we develop new stochastic models for the analysis of middle-censored lifetime data in the presence of covariates. The present work develops several parametric and semiparametric regression models for the analysis of middle-censored lifetime data arising from a continuous population, where the random censoring interval is assumed to have an absolutely continuous distribution function. We also consider the Bayesian analysis of regression models for middle-censored lifetime data.

The thesis is organized as follows. Chapter 1 gives a formal introduction to the subject area and covering a comprehensive review of literature on middle-censoring. In Chapter 2, we introduce and study a parametric proportional hazards model with the baseline distribution as Weibull. The parameters are estimated by the method

of maximum likelihood. The likelihood ratio test is proposed to test the significance of the regression parameter. We establish the asymptotic normality of the estimator. We carry out extensive simulation studies to assess the performance of the proposed estimator with finite sample sizes. The proposed inference procedure is illustrated using a real life data studied by Krall et al. (1975).

Chapter 3 discusses semiparametric proportional hazards model for the middle-censored lifetime data. We develop an iterative algorithm for finding the semiparametric MLE (SPMLE) of the regression parameter. The NPMLE of the baseline survival function is also derived. The consistency of the proposed estimators under certain regularity conditions is established. We also discuss a test concerning the significance of covariate effect. The finite sample performance of the proposed estimator is studied using simulation studies. The proposed method is illustrated with a real data set reported in Lee and Wang (2003).

Chapter 4 is devoted to a parametric additive hazards regression model for middle-censored lifetime data. We assume that the baseline distribution of lifetime variate is exponentiated exponential (Ahuja and Nash (1967)). The MLE for the parameters of the model are derived. We show that under certain regularity conditions, the estimator is asymptotically normally distributed. Extensive simulation studies are conducted to assess the finite sample performance of the estimator. The proposed method is applied to a real data set given in Karduan (1983).

Chapter 5 deals with an additive hazards regression model in a semiparametric setup for the middle-censored data. We derive estimators for the regression parameter and the baseline survival function using two different estimation procedures. The first one makes use of the martingale feature associated with the observed data and the other one is based on an iterative algorithm. We establish large sample

properties of the estimators under both estimation methods. Extensive simulation studies are carried out to assess the finite sample performance of the proposed estimators. The model is applied to a real life data studied by Ichida et al. (1993).

In Chapter 6, we develop a quantile regression model for the analysis of middle-censored lifetime data subject to conditionally independent censoring. We construct a stochastic integral estimating equation for estimating the regression quantile. The consistency and weak convergence of the estimators are established under certain regularity conditions. The model diagnostic method is developed using martingale residuals. Simulation studies are reported to assess the performance of proposed estimator with finite sample sizes. The proposed quantile regression method is applied to a real data set studied by Copelan et al. (1991).

Chapter 7 discusses Bayesian analysis of regression problem for lifetime data subject to middle-censoring in a parametric setup. We assume that the baseline distribution is Weibull. Simulation studies are carried out to assess the performance of estimators with finite sample sizes. The proposed method is applied to a real life data studied by Krall et al. (1975).

Chapter 8 summarizes the thesis, with major findings of the study. We also discuss future works to be carried out in this area.

Chapter 2

Parametric Proportional Hazards Regression Model

2.1 Introduction

In survival studies, we usually express a regression model either as a parametric model, where the distribution of underlying lifetime variate is assumed to be known, or as a semiparametric model, where the underlying distribution of the lifetime variable is left arbitrary. The advantage of using a parametric model as a lifetime model is that the inference procedure associated with it results in more precise estimators than with a semiparametric model. A well known probability distribution that is extensively used as a lifetime model is the Weibull distribution. It has wide variety of applications in biological, medical, and industrial fields. A key feature for this distribution is that it can accommodate monotone increasing, monotone decreasing, and constant hazard rates depending on its shape parameter. This flexibility can very well be exploited to the advantage of modelling lifetime distributions. Recently, Pradhan and Kundu (2014) used the Weibull distribution to model the lifetime data under the interval censoring scheme. Motivated by the wide

[1]The results in this chapter are published in *Journal of Statistics and Management Systems* (See Sankaran and Prasad (2014)).

range of applications of this distribution, in this chapter, we consider a parametric regression model with the baseline distribution of the lifetime variate as Weibull.

The rest of the chapter is organized as follows. In Section 2.2, we introduce a parametric regression model and derive the inference procedure for the parameters of the model. Section 2.3 describes the asymptotic properties of the proposed estimator. A simulation study to assess the finite sample performance of the estimator is described in Section 2.4. We apply the proposed method to a real life problem in Section 2.5. The chapter ends with major conclusions in Section 2.6.

2.2 The Model and Inference Procedure

Let T be the random variable representing the lifetime, $\boldsymbol{z}$ be the $p \times 1$ vector of covariates and $\boldsymbol{\theta}$ be the corresponding $p \times 1$ vector of regression parameters. The Cox proportional hazards model associates the covariate effects on the lifetime T by specifying

$$h\left(t \,|\boldsymbol{z}\right) = h_0\left(t\right) \exp\left(\boldsymbol{z}^\top \boldsymbol{\theta}\right). \tag{2.1}$$

We specify $h_0(t)$ in (2.1) as the hazard function of Weibull distribution with shape parameter α and scale parameter β, with probability density function $f_0(t)$ given by

$$f_0(t) = \alpha\beta^{-1}\left(\beta^{-1}t\right)^{\alpha-1} \exp(-\left(\beta^{-1}t\right)^{\alpha}), \quad t > 0, \alpha > 0, \beta > 0. \tag{2.2}$$

Thus the baseline hazard function is

$$h_0(t) = \alpha\beta^{-1}\left(\beta^{-1}t\right)^{\alpha-1}, \quad t > 0,$$

with the baseline survival function as

$$S_0(t) = \exp(-\left(\beta^{-1}t\right)^{\alpha}), \quad t > 0.$$

We now assume that T is middle-censored by the random censoring interval (U, V) with $P(U < V) = 1$, as defined in Chapter 1. Thus one can observe the vector $(X, \delta, \boldsymbol{z})$, where

$$X = \begin{cases} T & \text{if } \delta = 1 \\ (U, V) & \text{if } \delta = 0, \end{cases}$$

and $\delta = I(X = T)$ is the censoring indicator, with $I(\cdot)$ denoting the indicator function. Further, we assume that T is independent of (U, V) given $\boldsymbol{z}$, and that the censoring distribution is independent of $\boldsymbol{z}$.

Under the proportional hazards model assumption (2.1), the survival function of T given $\boldsymbol{z}$, is given by

$$S(t|\boldsymbol{z}) = (S_0(t))^{\exp(\boldsymbol{\theta}^\top \boldsymbol{z})} = \exp(-\left(\gamma^{-1}t\right)^{\alpha}), \tag{2.3}$$

where $\gamma = \beta e^{-\boldsymbol{\theta}^\top \boldsymbol{z}/\alpha}$. Thus the probability density function of T given on $\boldsymbol{z}$ is obtained from (2.3) as

$$f(t|\boldsymbol{z}) = \alpha\gamma^{-1}\left(\gamma^{-1}t\right)^{\alpha-1}\exp(-\left(\gamma^{-1}t\right)^{\alpha});\ t > 0, \alpha > 0, \gamma > 0. \tag{2.4}$$

The observed data consists of $(X_i, \delta_i, \boldsymbol{z}_i)$, $i = 1, 2, ..., n$, which are independent and identically distributed copies of $(X, \delta, \boldsymbol{z})$ corresponding to n subjects under investigation. The likelihood function corresponding to the observed data can be

written as

$$L(\boldsymbol{\psi}) \propto \prod_{i=1}^{n} f(t_i|\boldsymbol{z}_i)^{\delta_i} \left[S(u_i|\boldsymbol{z}_i) - S(v_i|\boldsymbol{z}_i)\right]^{1-\delta_i}, \tag{2.5}$$

where $\boldsymbol{\psi} = (\alpha, \beta, \boldsymbol{\theta}^\top)^\top$, and t_i and (u_i, v_i) are the realizations of X_i. Note that the likelihood formulation in (2.5) is quite distinct from that in the case of interval censored data (Groeneboom and Wellner (1992); Geskus and Groeneboom (1996)), since in the latter case one can only observe a censoring event and can note whether the lifetime of interest occurred before or after the occurrence of the censoring event, as we discussed in Chapter 1. Thus, the component specifying the exact event time is not present in the likelihood function for interval censored data consequent to this intermittent observation scheme. For a detailed discussion on this distinction, one may refer to Jammalamadaka and Mangalam (2003), Mangalam et al. (2008), Iyer et al. (2008), and Shen (2011).

Without loss of generality, we now arrange the observations in such a way that first n_1 observations are exact lifetimes and remaining n_2 are censored intervals, with $n_1 + n_2 = n$. We shall rewrite (2.5) as

$$L(\boldsymbol{\psi}) = \prod_{i=1}^{n_1} \alpha\gamma_i^{-1}(\gamma_i^{-1}t_i)^{\alpha-1}\exp(-(\gamma_i^{-1}t_i)^\alpha) \prod_{i=n_1+1}^{n} \left[\, \exp(-(\gamma_i^{-1}u_i)^\alpha) - \exp(-(\gamma_i^{-1}v_i)^\alpha)\right], \tag{2.6}$$

where $\gamma_i = \beta e^{-\boldsymbol{\theta}^\top \boldsymbol{z}_i/\alpha}$. Our objective is to estimate the regression parameter and the parameters of the Weibull distribution. To achieve this, we can make use of the Expectation-Maximization (EM) algorithm (Dempster et al. (1977)), which is a powerful tool in handling the incomplete data problems. It is an iterative method, where we repeatedly replace the missing data with estimated values and update

the parameters. To begin with the algorithm, a hypothetical complete-data likelihood function, say $L_C(\boldsymbol{\psi})$, is first constructed. Thus, it is straightforward to verify that the E-step requires the computation of the conditional expectation of the complete-data log-likelihood function, given the observed data and the current update of the parameter. The EM cycle is completed with M-step, where the MLE is computed with the complete-data log-likelihood function with the missing observations replaced by their conditional expectations. This cycle is repeated until the convergence up to a desired precision is met. Moreover, the convergence of this iterative procedure is guaranteed (see McLachlan and Krishnan (2007)). An elementary treatment of its computational aspect is given in Kundu and Basu (2004). Now, the complete-data likelihood function can be written as

$$L_C(\boldsymbol{\psi}) = \prod_{i=1}^{n_1} f\left(t_i|\boldsymbol{z}_i\right) \prod_{i=n_1+1}^{n} f\left(\check{t}_i|\boldsymbol{z}_i\right),$$

where $\check{t}_i$ are potentially observable, but unobserved realizations of T, such that $\check{t}_i \in (u_i, v_i)$, $i = n_1 + 1, ..., n$.

The complete-data log-likelihood function can be written as

$$l_c(\boldsymbol{\psi}) = \log L_C(\boldsymbol{\psi}) = \sum_{i=1}^{n_1} \log f\left(t_i|\boldsymbol{z}_i\right) + \sum_{i=n_1+1}^{n} \log f\left(\check{t}_i|\boldsymbol{z}_i\right). \tag{2.7}$$

Using the density function in (2.4), we can write (2.7) as

$$l_c(\boldsymbol{\psi}) = n\log\alpha - \sum_{i=1}^{n}\log\gamma_i + \sum_{i=1}^{n_1}\left((\alpha-1)\log(\gamma_i^{-1}t_i) - (\gamma_i^{-1}t_i)^{\alpha}\right) + \sum_{i=n_1+1}^{n}\left((\alpha-1)\log(\gamma_i^{-1}\check{t}_i) - (\gamma_i^{-1}\check{t}_i)^{\alpha}\right). \tag{2.8}$$

The E-step in EM algorithm is performed by taking the expectation of complete-data log-likelihood function given in (2.8), given the current update of $\boldsymbol{\psi}$ and the observed data. This expected value is given by

$$\phi(\boldsymbol{\psi}) = n\log\alpha - \alpha\sum_{i=1}^{n}\log\gamma_i + (\alpha-1)\sum_{i=1}^{n_1}\log t_i - \sum_{i=1}^{n_1}(\gamma_i^{-1}t_i)^{\alpha} + \sum_{i=n_1+1}^{n}\left((\alpha-1)\xi_{1i} - \gamma_i^{-\alpha}\xi_{2i}\right), \quad (2.9)$$

where the terms ξ_{1i} and ξ_{2i} are defined as follows.

$\xi_{1i} = E\left(\log T|\boldsymbol{\psi}, T\in(u_i,v_i), \boldsymbol{z}_i\right) = (S(u_i|\boldsymbol{\psi},\boldsymbol{z}_i) - S(v_i|\boldsymbol{\psi},\boldsymbol{z}_i))^{-1}\int_{u_i}^{v_i}\log t\, f(t|\boldsymbol{\psi},\boldsymbol{z}_i)\mathrm{d}t,$

and

$\xi_{2i} = E\left(T^{\alpha}|\boldsymbol{\psi}, T\in(u_i,v_i), \boldsymbol{z}_i\right) = (S(u_i|\boldsymbol{\psi},\boldsymbol{z}_i) - S(v_i|\boldsymbol{\psi},\boldsymbol{z}_i))^{-1}\int_{u_i}^{v_i} t^{\alpha}\, f(t|\boldsymbol{\psi},\boldsymbol{z}_i)\mathrm{d}t.$

We now give an algorithm to find the MLE $\hat{\boldsymbol{\psi}} = (\hat{\alpha}, \hat{\beta}, \hat{\boldsymbol{\theta}}^{\top})^{\top}$ of $\boldsymbol{\psi}$.

Algorithm 2.1

Step 1. Choose an initial value $\boldsymbol{\psi}^{(0)} = \left(\alpha^{(0)}, \beta^{(0)}, \boldsymbol{\theta}^{(0)\top}\right)^{\top}$.

Step 2. At the j'th iteration ($j \geq 1$), evaluate $\xi_1^{(j)}$ and $\xi_2^{(j)}$ using $\boldsymbol{\psi}^{(j-1)}$ and substitute in (2.9).

Step 3. With $\alpha^{(j-1)}$ and $\beta^{(j-1)}$ held fixed, express (2.9) as a function of $\boldsymbol{\theta}$ and solve $\frac{\partial\phi}{\partial\boldsymbol{\theta}} = 0$ to get $\boldsymbol{\theta} = \boldsymbol{\theta}^{(j)}$.

Step 4. With $\boldsymbol{\theta} = \boldsymbol{\theta}^{(j)}$, express (2.9) as a function of α and β and solve $\frac{\partial\phi}{\partial\alpha} = 0$ $\frac{\partial\phi}{\partial\beta} = 0$ to get $\alpha^{(j)}$ and $\beta^{(j)}$ and thus obtain $\boldsymbol{\psi}^{(j)} = \left(\alpha^{(j)}, \beta^{(j)}, \boldsymbol{\theta}^{(j)\top}\right)^{\top}$.

Step 5. Repeat *Step-2* to *Step-4* till convergence in $\boldsymbol{\psi}$ is met, say when $\|\boldsymbol{\psi}^{(k)} - \boldsymbol{\psi}^{(k+1)}\| < 0.0001$, for some finite positive integer k.

In regression analysis, often one would be interested to know whether the covariates have significant effect on the lifetime variate. This can be formulated as a

hypothesis $H_0 : \boldsymbol{\theta} = \mathbf{0}$, where $\mathbf{0}$ is the null vector of same order. This hypothesis can be tested against the alternative hypothesis $H_1 : \boldsymbol{\theta} \neq \mathbf{0}$ by means of likelihood ratio test. The test statistic is given by $\Lambda = 2 \log L\left(\hat{\boldsymbol{\theta}}, \hat{\alpha}, \hat{\beta}\right) - 2 \log L\left(\mathbf{0}, \tilde{\alpha}, \tilde{\beta}\right)$, where $\tilde{\alpha}$ and $\tilde{\beta}$ are the MLEs obtained when restricting $\boldsymbol{\theta}$ to be $\mathbf{0}$. Under the null hypothesis, Λ follows approximately chi-square distribution with p degrees of freedom for large samples and the test rejects the null hypothesis for small P-values.

2.2.1 Bootstrap Method

We now discuss a procedure for generating bootstrap samples from a middle-censored lifetime data for finding out the standard errors of estimates. Motivated by the re-sampling scheme for the right censored lifetime data discussed in Davison and Hinkley (1997), we now develop a conditional bootstrap procedure for a middle-censored lifetime data. Let the observed data consist of $\left(X_i, \delta_i, \boldsymbol{z}_i\right)$, $i = 1, 2, ..., n$. Let $\hat{G}_1(\cdot)$ be the empirical distribution function corresponding to the left censoring times U_i in the data, and let $\hat{G}_2(\cdot)$ be the empirical distribution function corresponding to the right censoring times V_i in the data. We designate the bootstrap observations and the associated estimates by attaching an asterisk mark as a superscript. In the following we provide an algorithm for the conditional bootstrap procedure for a middle-censored data, where the resample observations are generated with replacement.

Algorithm 2.2

For $i = 1, 2, ..., n$,

Step 1. Generate T_i^* from the fitted survival function $\left(\hat{S}_0(t)\right)^{\exp(\boldsymbol{z_i}^\top \hat{\boldsymbol{\theta}})}$, where $\hat{S}_0(t)$ and $\hat{\boldsymbol{\theta}}$ are the MLEs obtained using Algorithm 2.1.

Step 2. Generate K_i from a Bernoulli distribution with success probability 0.5, where $K_i = 0$ is used to denote the case when the unobserved censoring interval is assumed to fall before the occurrence of the observed event; and $K_i = 1$ is used to denote the case when the unobserved censoring interval is assumed to fall after the occurrence of the observed event.

Step 3. For $K_i = 0$, if $\delta_i = 0$, set $V_i^* = v_i$ and set $U_i^* = u_i$; and if $\delta_i = 1$, we first generate V_i^* from the conditional distribution $\hat{G}_2(v)/\hat{G}_2(t_i)$, for $v < t_i$, which is the estimated distribution of V_i conditional on $V_i < t_i$. Then we generate U_i^* from the conditional distribution $\hat{G}_1(u)/\hat{G}_1(v_i^*)$, for $u < v_i^*$, which is the estimated distribution of U_i conditional on $U_i < v_i^*$.

Step 4. For $K_i = 1$, if $\delta_i = 0$, set $U_i^* = u_i$ and set $V_i^* = v_i$; and if $\delta_i = 1$, we first generate U_i^* from the conditional distribution $\frac{\hat{G}_1(u)-\hat{G}_1(t_i)}{1-\hat{G}_1(t_i)}$, which is the estimated distribution of U_i conditional on $U_i > t_i$. We then generate V_i^* from the conditional distribution $\frac{\hat{G}_2(v)-\hat{G}_2(u_i^*)}{\hat{G}_2(u_i^*)}$, which is the estimated distribution of V_i conditional on $V_i > u_i^*$.

Step 5 If $T_i^* \notin (U_i^*, V_i^*)$, set $X_i^* = T_i^*$ with $\delta_i^* = 1$, otherwise set $X_i^* = (U_i^*, V_i^*)$ with $\delta_i^* = 0$.

In practice, if the largest observation in a data set happens to be a lifetime t_j, in the sense that there is no observed censoring interval which is a subset of (t_j, ∞), we fix $K_j = 0$ and then generate the censoring interval using *Step 3.* Similarly, if t_j happens to be a smallest observation in a data set, in the sense that there is no observed censoring interval which is a subset of $(0, t_j)$, we fix $K_j = 1$ and generate censoring intervals using *Step 4.*

Assume that we have B such bootstrap samples generated from the observed data set, and that $\hat{\eta}$ is an estimator of an unknown parameter η. Following Efron

(1982), we compute the bootstrap standard error (SE) of the estimator $\hat{\eta}$ as

$$\text{SE}(\hat{\eta}) = \Big[\frac{1}{B-1}\sum_{b=1}^{B}(\hat{\eta}_b^* - \bar{\eta}^*)^2\Big]^{1/2}, \tag{2.10}$$

where $\hat{\eta}_b^*$ is the estimate of η based on b'th bootstrap sample ($b = 1, 2, ..., B$) and $\bar{\eta}^* = \frac{1}{B}\sum_{b=1}^{B}\hat{\eta}_b^*$.

2.3 Asymptotic Properties

In this section, we discuss the asymptotic normality of the MLE $\hat{\boldsymbol{\psi}}$ of the parameter $\boldsymbol{\psi}$. We show that, under certain regularity conditions $\hat{\boldsymbol{\psi}}$ is asymptotically normally distributed with mean $\boldsymbol{\psi}$, the true mean value, and dispersion matrix $\mathscr{I}^{-1}(\boldsymbol{\psi})$, where $\mathscr{I}(\boldsymbol{\psi})$ is the Fisher information matrix given by $\mathscr{I}(\boldsymbol{\psi}) = -E\big(\frac{\partial^2 l(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}^\top}\big)$, where $l(\boldsymbol{\psi}) = \log L(\boldsymbol{\psi})$. We provide explicit expressions of the second order partial derivatives of $l(\boldsymbol{\psi})$ in equations (2.11) - (2.16) below. First we define the following terms for any collection of positive real numbers a_i and b_i, $i = 1, 2, ..., n$.

$D_1(a_i, b_i) = e^{-(\beta^{-1}a_i)^\alpha e^{\boldsymbol{\theta}^\top \boldsymbol{z}_i}} - e^{-(\beta^{-1}b_i)^\alpha e^{\boldsymbol{\theta}^\top \boldsymbol{z}_i}}$,

$K_1(a_i) = e^{-(\beta^{-1}a_i)^\alpha e^{\boldsymbol{\theta}^\top \boldsymbol{z}_i}}(\beta^{-1}a_i)^\alpha \cdot \log(\beta^{-1}a_i)$,

$M_1(a_i, b_i) = K_1(a_i) - K_1(b_i)$,

$K_2(a_i) = K_1(a_i)/\log(\beta^{-1}a_i)$,

$M_2(a_i, b_i) = K_2(a_i) - K_2(b_i)$ and

$A_1(a_i) = e^{\boldsymbol{\theta}^\top \boldsymbol{z}_i}(\beta^{-1}a_i)^\alpha$.

$$\frac{\partial^2 l(\boldsymbol{\psi})}{\partial \alpha^2} = -n_1\alpha^{-2} + \sum_{i=1}^{n_1} A_1(t_i)\log(t_i)\log(\beta^{-1}t_i) + \sum_{i=n_1+1}^{n} e^{\boldsymbol{\theta}\boldsymbol{z}_i} D_1(u_i,v_i)^{-1} \Big(- D_1(u_i,v_i)^{-1} e^{\boldsymbol{\theta}\boldsymbol{z}_i} M(u_i,v_i)^2 + \log(\beta^{-1}v_i)K_1(v_i)(1-A_1(v_i)) - \log(\beta^{-1}u_i)K_1(u_i)(1-A_1(u_i))\Big). \tag{2.11}$$

$$\frac{\partial^2 l(\boldsymbol{\psi})}{\partial \alpha \partial \theta_j} = \sum_{i=1}^{n_1} \boldsymbol{z}_{ij} A_1(t_i)\log(\beta t_i^{-1}) + \sum_{i=n_1+1}^{n} \boldsymbol{z}_{ij} e^{\boldsymbol{\theta}^\top \boldsymbol{z}_i} D_1(u_i,v_i)^{-1} \Big(e^{\boldsymbol{\theta}^\top \boldsymbol{z}_i} D_1(u_i,v_i)^{-1} M_2(u_i,v_i)M_1(u_i,v_i) + M_1(u_i,v_i) + e^{\boldsymbol{\theta}^\top \boldsymbol{z}_i}\left(K_1(u_i)(\beta^{-1}u_i)^\alpha - K_1(v_i)(\beta^{-1}v_i)^\alpha\right)\Big), \tag{2.12}$$

where z_{ij} is the j'th component of $\boldsymbol{z}_i$, $j = 1, 2, ..., p$.

$$\frac{\partial^2 l(\boldsymbol{\psi})}{\partial \alpha \partial \beta} = -n_1\beta^{-1} + \beta^{-1}\sum_{i=1}^{n_1} A_1(t_i)(1+\alpha\log(\beta^{-1}t_i)) + \sum_{i=n_1+1}^{n} e^{\boldsymbol{\theta}\boldsymbol{z}_i} D_1(u_i,v_i)^{-1} \Big(\beta^{-1}K_1(u_i) - \alpha\beta^{-1}e^{\boldsymbol{\theta}\boldsymbol{z}_i} D_1(u_i,v_i)^{-1} M_2(u_i,v_i)M_1(u_i,v_i) + \alpha\beta^{-1}K_1(v_i) (A_1(v_i)-1) - \beta^{-1}K_1(v_i) - \alpha\beta^{-1}K_1(u_i)(A_1(u_i)-1)\Big). \tag{2.13}$$

$$\frac{\partial^2 l(\boldsymbol{\psi})}{\partial \beta^2} = -n_1\alpha\beta^{-2} - \alpha(\alpha+1)\beta^{-2}\sum_{i=1}^{n_1} A_1(t_i) + \sum_{i=n_1+1}^{n} \alpha\beta^{-2} e^{\boldsymbol{\theta}^\top \boldsymbol{z}_i} D_1(u_i,v_i)^{-1} \Big(- M_2(u_i,v_i) - e^{\boldsymbol{\theta}^\top \boldsymbol{z}_i} D_1(u_i,v_i)^{-1} M_2(u_i,v_i)^2 + K_2(u_i)(A_1(u_i)-1) + K_2(v_i)(A_1(v_i)-1)\Big). \tag{2.14}$$

$$\frac{\partial^2 l(\boldsymbol{\psi})}{\partial\beta\partial\theta_j} = \alpha\beta^{-1}\sum_{i=1}^{n_1}\boldsymbol{z}_{ij}A_1(t_i) + \alpha\beta^{-1}\sum_{i=n_1+1}^{n}\boldsymbol{z}_{ij}D_1(u_i,v_i)^{-1}e^{\boldsymbol{\theta}^\top\boldsymbol{z}_i}\Big(M_2(u_i,v_i) + e^{\boldsymbol{\theta}^\top\boldsymbol{z}_i}D_1(u_i,v_i)^{-1}M_2(u_i,v_i)^2 - e^{\boldsymbol{\theta}^\top\boldsymbol{z}_i}\left(K_2(u_i)(\beta^{-1}u_i)^\alpha - K_2(v_i)(\beta^{-1}v_i)^\alpha\right)\Big), \tag{2.15}$$

for $j = 1, 2, ..., p$.

$$\frac{\partial^2 l(\boldsymbol{\psi})}{\partial\theta_j\partial\theta_k} = -\sum_{i=1}^{n_1}z_{ij}z_{ik}A_1(t_i) + \sum_{i=n_1+1}^{n}z_{ij}z_{ik}e^{\theta\boldsymbol{z}_i}D_1(u_i,v_i)^{-1}\Big(-e^{\boldsymbol{\theta}^\top\boldsymbol{z}_i}D_1(u_i,v_i)^{-1} M_2(u_i,v_i)^2 - M_2(u_i,v_i) + e^{\boldsymbol{\theta}^\top\boldsymbol{z}_i}\left(K_2(u_i)(\beta^{-1}u_i)^\alpha - K_2(v_i)(\beta^{-1}v_i)^\alpha\right)\Big), \tag{2.16}$$

for $j, k = 1, 2, ...p$.

For the evaluation of the Fisher information matrix, we make some distributional assumptions on the random censoring interval (U, V). We assume that U is exponentially distributed with mean λ_1^{-1}, $Y = V - U$ is exponentially distributed with mean λ_2^{-1}, and that T, U and Y are mutually independent. Now the probability density function of T, conditional on the event that $T \notin (U, V)$ and $\boldsymbol{z}$ is given by

$$\begin{aligned} f_{T|(T\notin(U,V),\boldsymbol{z})}(t) &= \lim_{\Delta t\to 0}\frac{1}{\Delta t}P(t \le T < t+\Delta t | T \notin (U,V), \boldsymbol{z}) \\ &= \lim_{\Delta t\to 0}\frac{1}{\Delta t}\frac{P(t \le T < t+\Delta t, T \notin (U,V)|\boldsymbol{z})}{P(T \notin (U,V)|\boldsymbol{z})} \\ &= \lim_{\Delta t\to 0}\frac{1}{\Delta t}\frac{1}{\acute{p}(\boldsymbol{\psi})}P(t \le T < t+\Delta t|\boldsymbol{z})P(T \notin (U,V)|t \le T < t+\Delta t, \boldsymbol{z}) \\ &= \frac{1}{\acute{p}(\boldsymbol{\psi})}f(t|\boldsymbol{z})P(T \notin (U,V)|T = t, \boldsymbol{z}), \end{aligned}$$

where $\acute{p}(\boldsymbol{\psi}) = P\left(T \notin (U,V)|\boldsymbol{z}\right) = \int f(t|\boldsymbol{z}) P\left(T \notin (U,V)|T=t,\boldsymbol{z}\right) \mathrm{d}t$.

Thus, for $\lambda_1 \neq \lambda_2$, we obtain

$$f_{T|(T\notin(U,V),\boldsymbol{z})}(t) = \frac{1}{\acute{p}(\boldsymbol{\psi})} \left(\alpha\gamma^{-1}(\gamma^{-1}t)^{\alpha-1}e^{-(\gamma^{-1}t)^{\alpha}}\right) \left(1 - \frac{\lambda_1}{\lambda_2-\lambda_1}e^{-\lambda_2 t}\left(e^{-(\lambda_1-\lambda_2)t}-1\right)\right), \tag{2.17}$$

and for $\lambda_1 = \lambda_2 = \lambda$ (say), we get

$$f_{T|(T\notin(U,V),\boldsymbol{z})}(t) = \frac{1}{\acute{p}(\boldsymbol{\psi})} \left(\alpha\gamma^{-1}(\gamma^{-1}t)^{\alpha-1}e^{-(\gamma^{-1}t)^{\alpha}}\right)(1-\lambda t e^{-\lambda t}). \tag{2.18}$$

In a similar way, we obtain the probability density function of U, conditional on the event that $T \in (U,V)$ and $\boldsymbol{z}$ as

$$f_{U|(T\in(U,V),\boldsymbol{z})}(u) = \frac{1}{1-\acute{p}(\boldsymbol{\psi})}\lambda_1\alpha\gamma^{-\alpha}e^{-u(\lambda_1+\lambda_2)}\int_{t=u}^{\infty} t^{\alpha-1}e^{-(\gamma^{-1}t)^{\alpha}}e^{-\lambda_2 t}dt, \tag{2.19}$$

and the probability density function of Y, conditional on the event that $T \in (U,V)$ and $\boldsymbol{z}$ is given by

$$f_{Y|(T\in(U,V),\boldsymbol{z})}(y) = \frac{1}{1-\acute{p}(\boldsymbol{\psi})}\lambda_2 e^{-\lambda_2 y}\int_{u=0}^{\infty}\lambda_1 e^{-\lambda_1 u}\left(e^{-(\gamma^{-1}u)^{\alpha}} - e^{-(\gamma^{-1}(u+y))^{\alpha}}\right)du. \tag{2.20}$$

Further, the joint distribution of (U,V) is obtained as

$$g_0(u,v) = \lambda_1\lambda_2 e^{-(\lambda_2-\lambda_1)u}e^{-\lambda_2 v}, \quad u,v \in \mathbb{R}^+, u<v. \tag{2.21}$$

Using (2.21), we obtain the density function of (U, V), conditional on $T \in (U, V)$ and $\boldsymbol{z}$ as

$$f_{UV|(T\in(U,V),\boldsymbol{z})}(u,v) = \frac{1}{1-\acute{p}(\boldsymbol{\psi})} g_0((u,v)) P\left(T \in (U,V) | (U,V) = (u,v), \boldsymbol{z}\right). \quad (2.22)$$

On simplification, (2.22) becomes

$$f_{UV|(T\in(U,V),\boldsymbol{z})}(u,v) = \frac{1}{1-\acute{p}(\boldsymbol{\psi})} \lambda_1\lambda_2 e^{-(\lambda_2-\lambda_1)u} e^{-\lambda_2 v}\left(e^{-(\gamma^{-1}u)^\alpha} - e^{-(\gamma^{-1}v)^\alpha}\right). \quad (2.23)$$

We use the density functions given in (2.17) - (2.23) to evaluate the expected values of the second order partial derivatives of the log-likelihood function given in (2.11) - (2.16), and hence we can evaluate the Fisher information matrix $\mathscr{I}(\boldsymbol{\psi})$. We now assume that the covariate space is bounded. Then the likelihood function (2.6) satisfies the standard regularity conditions (see Bain (1976)), and it follows that the asymptotic distribution of $\hat{\boldsymbol{\psi}}$ is $(p+2)$-variate normal with mean vector $\boldsymbol{\psi}$ and dispersion matrix $\mathscr{I}^{-1}(\boldsymbol{\psi})$. Thus, $\sqrt{n}\left(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}\right)^\top \to N_{p+2}\left(\mathbf{0}, n\mathscr{I}^{-1}(\boldsymbol{\psi})\right)$ in distribution.

2.4 Simulation Studies

Simulation studies are carried out to assess the finite sample performance of the estimators. We assume that the baseline distribution of T is Weibull with probability density function as given in (2.2). Under the proportional hazards assumption given in (2.1), the survival function of T given $\boldsymbol{z}$ is as given in (2.3). We consider a single covariate z in the present study which is generated from uniform distribution

over $[0, 15]$. Observations are generated from the distribution given in (2.4), for various values of α, β and θ. Now corresponding to each observation on T given z, a random censoring interval (U, V) is generated, where U and $Y = V - U$ are assumed to be independent exponential variates with means $\lambda_1^{-1} = 15$ and $\lambda_2^{-1} = 10$ respectively. If we find $T \notin (U, V)$, then T is selected as the sample observation, otherwise we choose the interval (U, V) as the observation. We consider different sample sizes and different censoring percentages for our study. Suppose that we require n_1 exact lifetimes and n_2 censoring intervals in a simulated sample of size $n = n_1 + n_2$, satisfying the required censoring percentage. To achieve this, we first generate a large number of observations as mentioned above. We then randomly select n_1 lifetimes from among the generated exact lifetimes and n_2 censoring intervals from among the generated censoring intervals. In this study, we consider three sample sizes, $n = 50$, $n = 100$, and $n = 200$. We consider three different censoring percentages viz. mild (10% censoring), moderate (20% censoring) and heavy (30% censoring) for analyzing the impact of censoring on the estimate of θ.

TABLE 2.1: Bias, MSE and CP of the estimator of θ under mild censoring.

			$n = 50$			$n = 100$			$n = 200$		
α	β	θ	Bias	MSE	CP	Bias	MSE	CP	Bias	MSE	CP
0.05	0.75	-0.50	0.049	0.0034	0.895	0.045	0.0011	0.898	0.016	0.0003	0.962
0.05	8.00	1.00	0.047	0.0033	0.904	0.044	0.0010	0.907	0.019	0.0007	0.971
0.05	15.00	0.75	0.038	0.0039	0.917	0.035	0.0014	0.921	0.022	0.0009	0.959
0.50	0.75	-1.00	0.037	0.0058	0.895	0.035	0.0016	0.899	0.020	0.0011	0.966
0.50	8.00	0.05	0.012	0.0039	0.934	0.011	0.0017	0.937	0.005	0.0008	0.977
0.50	15.00	0.80	0.020	0.0056	0.937	0.016	0.0017	0.942	0.009	0.0006	0.974
1.00	0.75	-0.05	0.026	0.0033	0.923	0.024	0.0009	0.927	0.013	0.0005	0.958
1.00	8.00	0.10	0.038	0.0042	0.929	0.035	0.0011	0.934	0.016	0.0005	0.959
1.00	15.00	0.04	0.023	0.0062	0.887	0.019	0.0018	0.890	0.007	0.0009	0.961
2.00	0.75	-0.10	0.006	0.0034	0.904	0.004	0.0020	0.906	0.001	0.0007	0.971
2.00	8.00	0.05	0.044	0.0045	0.905	0.039	0.0019	0.907	0.018	0.0006	0.965
2.00	15.00	0.80	0.006	0.0046	0.913	0.004	0.0012	0.915	0.002	0.0007	0.974

TABLE 2.2: Bias, MSE and CP of the estimator of θ under moderate censoring.

			$n = 50$			$n = 100$			$n = 200$		
α	β	θ	Bias	MSE	CP	Bias	MSE	CP	Bias	MSE	CP
0.05	0.75	-0.50	0.050	0.0045	0.892	0.046	0.0029	0.897	0.021	0.0007	0.955
0.05	8.00	1.00	0.048	0.0081	0.901	0.046	0.0022	0.905	0.033	0.0011	0.966
0.05	15.00	0.75	0.041	0.0061	0.913	0.036	0.0024	0.920	0.028	0.0012	0.957
0.50	0.75	-1.00	0.039	0.0074	0.891	0.036	0.0034	0.897	0.025	0.0017	0.962
0.50	8.00	0.05	0.016	0.0056	0.932	0.012	0.0027	0.936	0.008	0.0012	0.971
0.50	15.00	0.80	0.023	0.0068	0.935	0.018	0.0032	0.940	0.014	0.0011	0.970
1.00	0.75	-0.05	0.028	0.0047	0.921	0.026	0.0021	0.926	0.017	0.0008	0.955
1.00	8.00	0.10	0.040	0.0088	0.926	0.036	0.0027	0.933	0.024	0.0007	0.954
1.00	15.00	0.04	0.027	0.0102	0.884	0.021	0.0029	0.889	0.011	0.0012	0.960
2.00	0.75	-0.10	0.011	0.0049	0.899	0.006	0.0037	0.905	0.004	0.0015	0.965
2.00	8.00	0.05	0.046	0.0090	0.901	0.041	0.0032	0.905	0.020	0.0018	0.961
2.00	15.00	0.80	0.009	0.0064	0.909	0.006	0.0028	0.914	0.005	0.0013	0.968

TABLE 2.3: Bias, MSE and CP of the estimator of θ under heavy censoring.

			$n = 50$			$n = 100$			$n = 200$		
α	β	θ	Bias	MSE	CP	Bias	MSE	CP	Bias	MSE	CP
0.05	0.75	-0.50	0.054	0.0076	0.888	0.048	0.0042	0.895	0.028	0.0011	0.951
0.05	8.00	1.00	0.051	0.0101	0.898	0.048	0.0040	0.904	0.038	0.0014	0.960
0.05	15.00	0.75	0.046	0.0099	0.910	0.037	0.0039	0.918	0.034	0.0021	0.941
0.50	0.75	-1.00	0.040	0.0100	0.887	0.038	0.0047	0.896	0.029	0.0028	0.955
0.50	8.00	0.05	0.019	0.0104	0.930	0.013	0.0045	0.934	0.010	0.0024	0.957
0.50	15.00	0.80	0.025	0.0087	0.932	0.019	0.0042	0.939	0.017	0.0018	0.958
1.00	0.75	-0.05	0.031	0.0087	0.917	0.027	0.0036	0.924	0.025	0.0019	0.950
1.00	8.00	0.10	0.043	0.0129	0.923	0.038	0.0042	0.931	0.029	0.0011	0.950
1.00	15.00	0.04	0.029	0.0115	0.883	0.022	0.0040	0.888	0.018	0.0026	0.955
2.00	0.75	-0.10	0.013	0.0102	0.896	0.007	0.0055	0.903	0.005	0.0021	0.951
2.00	8.00	0.05	0.049	0.0129	0.898	0.042	0.0042	0.903	0.027	0.0029	0.958
2.00	15.00	0.80	0.013	0.0078	0.908	0.007	0.0040	0.912	0.006	0.0019	0.955

Further, different combinations of parameters are also considered. We now obtain the MLE of θ using Algorithm 2.1 given in Section 2.2. The initial value of θ is set as zero. The estimates obtained while fitting the data set consisting of only exact lifetimes at baseline level to the Weibull model given in (2.2) is used as initial values of distribution parameters α and β. We use 1000 iterations for each combination of parameters. Average absolute bias (Bias) and estimated mean squared error (MSE)

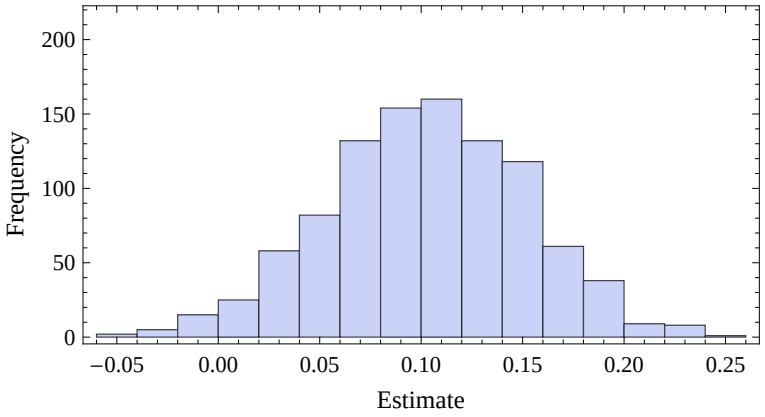


FIGURE 2.1: Histogram of estimates obtained for $\theta = 0.10$.

of $\hat{\theta}$ are computed and are reported in Tables 2.1 to 2.3. The coverage probabilities (CP) are computed using Wald 95% confidence intervals, which are also reported in these tables. It can be observed that the bias and MSE are less and they decrease as sample size increases, and the CP values are pretty high. Moreover, as the percentage of censoring increases, the bias and MSE increase, while the CP decreases. Figure 2.1 shows a histogram of the estimates of $\theta = 0.10$ under moderate censoring. It is evident that the estimator is roughly normally distributed with mean 0.10, the true parameter value, giving evidence in favour of asymptotic normality of the estimator $\hat{\theta}$. Similar results are obtained for other values of θ.

2.5 Data Analysis

In this section, we apply our model to a real life data. We consider the data on survival times in months for sixty five patients suffering from multiple myeloma studied by Krall et al. (1975). The complete data set is given in Lawless (2003), and we consider only the exact lifetimes for our study. We validate the distribution

assumption in two stages. We first select all exact lifetimes in this data set and the Weibull model given in (2.2) is fitted to this. The P-value is observed to be 0.8142 under the Kolmogorov-Smirnov test. This indicates that the Weibull distribution is a plausible lifetime model for the data. The estimates of distribution parameters obtained here are used later as initial values for Algorithm 2.1. At the second stage, motivated by the method adopted by Jammalamadaka and Mangalam (2003), and Iyer et al. (2008), we artificially middle-censored 20% of the lifetimes as described in previous section with $\lambda_1^{-1} = 15$ and $\lambda_2^{-1} = 8$. This resulted in a dataset consisting of both exact lifetimes and censoring intervals. This new data set is further checked for its suitability under the Weibull model assumption as described below. We first consider a likelihood function using baseline distribution, given by $L_0(\alpha, \beta) = \prod_{i=1}^{n_1} f_0(t_i) \prod_{i=n_1+1}^{n} (S_0(u_i) - S_0(v_i))$. The distribution parameters α and β are estimated via the EM algorithm. The estimators thus obtained are denoted by $\hat{\alpha}_0$ and $\hat{\beta}_0$.

We now give a Nelson-Aalen-type estimator for the cumulative hazard function when the data is subject to middle-censoring. For the i'th subject $(i = 1, 2, ..., n)$, we define the counting process $\{N_i(t);\, t \geq 0\}$ and the at-risk process $\{R_i(t);\, t \geq 0\}$ by $N_i(t) = I(\bar{X}_i \leq t, \delta_i = 1)$ and $R_i(t) = I(\bar{X}_i \geq t)$, where $\bar{X} = T$ if $\delta = 1$, and $\bar{X} = U$ if $\delta = 0$. The process $\{N_i(t);\, t \geq 0\}$ denotes the number of observed events up to time t for the i'th individual, and the at-risk process $R_i(t)$ is a 0-1 predictable process, where the value 1 indicates whether the i'th individual is at risk at time t, which means whether it is alive and uncensored just prior to t. Then the Nelson-Aalen-type estimator of baseline cumulative hazard function is defined as

$$\hat{H}_0(t) = \int_0^t \frac{\sum_{i=1}^n \mathrm{d}N_i(u)}{\sum_{i=1}^n R_i(u)}. \tag{2.24}$$

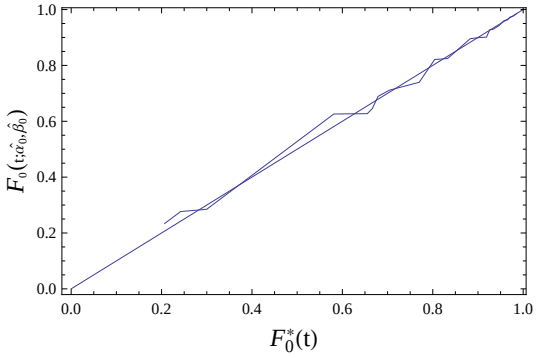


FIGURE 2.2: Plot of $F_0(t;\hat{\alpha}_0,\hat{\beta}_0)$ against $\breve{F}_0(t)$.

A probability-probability (PP)-plot is drawn with the distribution function of the Weibull distribution with parameters as $\hat{\alpha}_0$ and $\hat{\beta}_0$, say $F_0(t;\hat{\alpha}_0,\hat{\beta}_0)$ along the Y-axis and the distribution function $\breve{F}_0(t)$, which is the estimator of baseline distribution function obtained from (2.24) along the X-axis. If the model assumption holds, then a straight line passing through origin making an angle 45 degrees with the X-axis is expected. Figure 2.2 shows the PP-plot thus obtained. The graph seems to be close to the line, which validate the distribution assumption.

Although there are five covariates in the original data, we take only one covariate viz. logarithm of blood urea nitrogen measurement at diagnosis. We apply the model given in Section 2.2 and obtain the MLE of parameters using Algorithm 2.1, where the initial values of parameters are set as in previous section. The estimates are reported in Table 2.4. The standard errors (SE) of these estimates are obtained by the conditional bootstrap method and they are also reported in Table 2.4. It is evident that the covariate has an adverse effect on the lifetime and the SE's are small. To test the significance of the covariate effect, we consider the null hypothesis $H_0 : \theta = 0$ and we use the likelihood ratio test described in Section 2.2. The P-value is obtained as 0.0091 and we conclude that the covariate has significant effect on

TABLE 2.4: Estimates of parameters and their SE.

	MLE	SE
α	0.9134	0.1190
β	18.0546	1.8600
θ	0.1024	0.0314

lifetime.

We now check the overall fit of the model by using Cox-Snell residuals (Cox and Snell (1968)). Suppose that the model given in (2.1) is fitted to the data. If the model assumption is correct, then the probability integral transform of the true death time T assumes a uniform distribution over [0, 1] or equivalently the random variable $H(T|\boldsymbol{z})$, which is the true cumulative hazard function corresponding to (2.1), has an exponential distribution with hazard rate one.

Motivated by the definition of Cox-Snell residuals in the case of interval censored data given in Farrington (2000), we now describe the method of obtaining Cox-Snell residuals for the middle-censored data. For the observed data $\{(X_i, \delta_i, \boldsymbol{z}_i), 1 \leq i \leq n\}$, we assume that the first n_1 observations are exact lifetimes and the remaining n_2 are censoring intervals. Define the quantities $r_j^{(0)} = \hat{H}_0(t_j)e^{\boldsymbol{z}_j^\top \hat{\boldsymbol{\theta}}}$ for $j = 1, 2, ..., n_1$. For each $j = n_1+1, ..., n$, we define two quantities $r_j^{(1)} = \hat{H}_0(u_j)e^{\boldsymbol{z}_j^\top \hat{\boldsymbol{\theta}}}$ and $r_j^{(2)} = \hat{H}_0(v_j)e^{\boldsymbol{z}_j^\top \hat{\boldsymbol{\theta}}}$. Then we define the Cox-Snell residuals corresponding to the middle-censored data as $r_j = r_j^{(0)}$ if $\delta_j = 1$, and $r_j = (r_j^{(1)}, r_j^{(2)})$ if $\delta_j = 0$. If the model assumption is reasonable and the estimates of the parameters are close to the true values, then $\{(r_j, \delta_j), 1 \leq j \leq n\}$ should behave like a middle-censored sample from unit exponential distribution. To check this we compute the Nelson-Aalen-type estimator of the cumulative hazard function of the residuals using (2.24), where

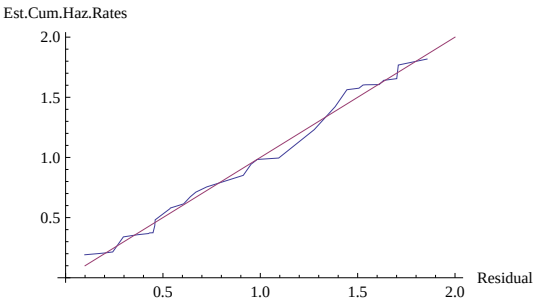


FIGURE 2.3: Plot of estimated cumulative hazard rates against r_j's.

the counting process and at-risk process are computed corresponding to the residuals. If the unit exponential distribution fits this data, then this estimator should approximately be equal to the cumulative hazard function of the unit exponential distribution. Thus, a plot of estimated cumulative hazard rates of r_j versus r_j's should be a straight line passing through origin and having a slope of one. Figure 2.3 shows the plot so obtained, which indicates that the data fits the model reasonably well.

2.6 Conclusion

In the present chapter, we considered a parametric regression model for the analysis of middle-censored lifetime data. The baseline distribution was assumed to be Weibull. Asymptotic normality of the estimator of regression parameter was established. Simulation studies indicated that the estimator perform satisfactorily under finite sample sizes. The model was applied to a real life data set studied by Krall et al. (1975). For finding the expected Fisher information matrix, we need to make some distribution assumptions for the random censoring interval (U, V). However, for numeric computation, we replaced the expected Fisher information

matrix with the observed information matrix, which does not require any such assumption. Moreover, this provides precise results since the estimator follows normal distribution asymptotically (see Efron and Hinkley (1978)).

Chapter 3

Semiparametric Proportional Hazards Regression Model

.

3.1 Introduction

In Chapter 2, we discussed a parametric proportional hazards model for a lifetime data subjected to middle-censoring. Such parametric methods are powerful if the underlying probability distribution of the lifetime variate is known. The inference procedures of parameters in such models are studied by applying standard likelihood techniques. However, in practice, the exact form of the underlying lifetime distribution is usually unknown and we may not be able to find an appropriate model. Therefore, the use of parametric methods in identifying significant prognostic factors is somewhat limited. In this chapter, we discuss the Cox (1972) proportional hazards model for middle-censored data in a semiparametric context, where the model does not require knowledge of the underlying lifetime distribution. Semiparametric models are flexible models for incorporating the covariate effect on lifetime variate. Over the past few decades, they became increasingly popular due

[1]The results in this chapter have been published in *Statistica* (see Jammalamadaka et al. (2016))

to their wide spectrum of applicability. For a review on this topic, one could refer to Lawless (2003).

The rest of the chapter is organized as follows. In Section 3.2, we develop an iterative method to derive the SPMLE of the regression parameter and the NPMLE of baseline survival function. Further, we discuss large sample properties of these estimators. Extensive simulation studies to assess the performance of these estimators under finite sample set up with different censoring rates are carried out in Section 3.3. We, in Section 3.4, apply the model to a real life data. Major findings and conclusions are given in Section 3.5.

3.2 The Model and Inference Procedure

Let T be a nonnegative random variable representing lifetime of a study subject with an unknown baseline distribution function $F_0(t)$. The only assumption we make about the underlying population is that $F_0(t)$ is absolutely continuous. In semiparametric context, we consider the proportional hazards model

$$h\left(t\,|\boldsymbol{z}\right) = h_0\left(t\right)\exp\left(\boldsymbol{z}^\top\boldsymbol{\theta}\right), \tag{3.1}$$

where the baseline hazard function $h_0(t)$ is arbitrary. Let (U, V) be the random vector representing the censoring interval. We assume that the distribution function $G_0(\cdot,\cdot)$ of (U, V) is absolutely continuous with $P\left(U < V\right) = 1$. Assume that (U, V) is independent of T, given $\boldsymbol{z}$, and that the censoring mechanism is independent of the covariates. When lifetime T is middle-censored by the random interval (U, V),

one can observe $(X, \delta, \boldsymbol{z})$, where

$$X = \begin{cases} T & \text{if } \delta = 1 \\ (U, V) & \text{if } \delta = 0, \end{cases}$$

where $\delta = I(X = T)$ represents the censoring indicator. Let us assume that there are n individuals under study. Then the observed data consists of $\left(X_i, \delta_i, \boldsymbol{z}_i\right)$, $i = 1, 2, ..., n$, which are independent and identically distributed copies of $\left(X, \delta, \boldsymbol{z}\right)$.

When we have incomplete data due to censoring, the idea of self-consistency plays a pivotal role in the estimation of the unknown baseline survival function $S_0(t)$. If we estimate $S_0(t)$ via the Expectation-Maximisation algorithm (Dempster et al. (1977)), as described by Tsai and Crowley (1985), the resulting estimating equation takes the form $\hat{S}_0(t) = E_{\hat{S}_0}[S_{ESF}(t)|\text{observed data}]$, where $\hat{S}_0(t)$ is the required estimate of $S_0(t)$ and $S_{ESF}(t)$ is the empirical survival function defined in (1.6). This equation is known as self-consistency equation and was first introduced by Efron (1967) to derive a class of estimators of $S_0(t)$ under right censoring. Jammalamadaka and Mangalam (2003) have shown that the NPMLE of $F_0(t)$ is always a Self Consistent Estimator (SCE) which takes the form

$$\hat{F}_0(t) = \frac{1}{n}\sum_{i=1}^{n}\Bigg\{\delta_i I(X_i \leq t) + (1-\delta_i) I(V_i \leq t) + (1-\delta_i) I(t \in (U_i, V_i)) \frac{\hat{F}_0(t) - \hat{F}_0(U_i)}{\hat{F}_0(V_i-) - \hat{F}_0(U_i)}\Bigg\}. \tag{3.2}$$

To estimate $S_0(t) = 1 - F_0(t)$ using (3.2), we first obtain its estimate at a number of time points t using (3.2). Then a cubic polynomial of the form $\hat{S}_0(t) = c_0 + c_1 t + c_2 t^2 + c_3 t^3$ is fitted using these estimated values, where c_0, c_1, c_2 and c_3 are the

coefficients of the fitted curve. This estimator is used in the forthcoming iterative algorithm.

Now, the survival function of T given $\boldsymbol{z}$ is given by

$$S\left(t|\boldsymbol{z}\right) = \left(S_0(t)\right)^{\exp(\boldsymbol{\theta}^\top \boldsymbol{z})}. \tag{3.3}$$

Differentiating (3.3) with respect to t, we get the probability density function of T given $\boldsymbol{z}$ as

$$f(t|\boldsymbol{z}) = f_0(t)\exp(\boldsymbol{\theta}^\top \boldsymbol{z})\left(S_0(t)\right)^{\exp(\boldsymbol{\theta}^\top \boldsymbol{z})-1},$$

where $f_0(t)$ is the baseline probability density function of T. Our objective is to estimate $\boldsymbol{\theta}$ and $S_0(t)$ under middle-censored observation scheme.

The likelihood function corresponding to the observed data is given by

$$L(\boldsymbol{\theta}) \propto \prod_{i=1}^{n} f(t_i|\boldsymbol{z}_i)^{\delta_i}\left[\left(S_0(u_i)\right)^{\exp(\boldsymbol{\theta}^\top \boldsymbol{z}_i)} - \left(S_0(v_i)\right)^{\exp(\boldsymbol{\theta}^\top \boldsymbol{z}_i)}\right]^{1-\delta_i}. \tag{3.4}$$

Without loss of generality, arrange the observations in such a way that the first n_1 observations are exact lifetimes, and the remaining n_2 are censored intervals, with $n_1 + n_2 = n$.

Now the likelihood function in (3.4), excluding the normalizing constant, is given by

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n_1} f(t_i|\boldsymbol{z}_i) \cdot \prod_{i=n_1+1}^{n}\left(\left(S_0(u_i)\right)^{\exp(\boldsymbol{\theta}^\top \boldsymbol{z}_i)} - \left(S_0(v_i)\right)^{\exp(\boldsymbol{\theta}^\top \boldsymbol{z}_i)}\right). \tag{3.5}$$

Then the log-likelihood function is given by

$$l(\boldsymbol{\theta}) = \sum_{i=1}^{n_1}[\log f_0(t_i) + \boldsymbol{\theta}^\top \boldsymbol{z}_i + \exp(\boldsymbol{\theta}^\top \boldsymbol{z}_i)\log S_0(t_i)] + \sum_{i=n_1+1}^{n} \log\left((S_0(u_i))^{\exp(\boldsymbol{\theta}^\top \boldsymbol{z}_i)} - (S_0(v_i))^{\exp(\boldsymbol{\theta}^\top \boldsymbol{z}_i)}\right). \tag{3.6}$$

The first order partial derivative of (3.6) with respect to θ_r, for $r = 1, 2, .., p$, is given by

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \theta_r} = \sum_{i=1}^{n_1}\left(z_{ir}(1 + \exp(\boldsymbol{\theta}^\top \boldsymbol{z}_i)\log S_0(t_i))\right) + \sum_{i=n_1+1}^{n}\left\{z_{ir}\exp(\boldsymbol{\theta}^\top \boldsymbol{z}_i)\left((S_0(u_i))^{\exp(\boldsymbol{\theta}^\top \boldsymbol{z}_i)} - (S_0(v_i))^{\exp(\boldsymbol{\theta}^\top \boldsymbol{z}_i)}\right)^{-1}\right.$$
$$\left.\left((S_0(u_i))^{\exp(\boldsymbol{\theta}^\top \boldsymbol{z}_i)}\log S_0(u_i) - (S_0(v_i))^{\exp(\boldsymbol{\theta}^\top \boldsymbol{z}_i)}\log S_0(v_i)\right)\right\}, \tag{3.7}$$

where z_{ir} is the r'th component of $\boldsymbol{z}_i$. We observe that (3.7) does not involve the baseline density function $f_0(t)$. We now give an algorithm for estimating the parameters $\boldsymbol{\theta}$ and $S_0(t))$.

Algorithm 3.1

Step 1. Set the vector $\boldsymbol{\theta} = \mathbf{0}$.

Step 2. At the first iteration, find NPMLE $S_0^{(1)}(t)$ of $S_0(t)$ using (3.2) and substitute this in (3.7) and solve $\partial l(\boldsymbol{\theta})/\partial\theta_r = 0, r = 1, 2, ..., p$ to get the estimate $\boldsymbol{\theta}^{(1)}$ of $\boldsymbol{\theta}$.

Step 3. Find $\tilde{t}_i^{(1)} = S_0^{(1)^{-1}}\left[S_0^{(1)}(t_i)^{\exp(\boldsymbol{\theta}^{(1)^\top}\boldsymbol{z}_i)}\right]$, $i = 1, 2, ..., n_1$, and similarly find $\tilde{u}_i^{(1)}$ and $\tilde{v}_i^{(1)}$, $i = n_1 + 1, ..., n$ as our updated observations at the first iteration.

Step 4. At the j'th iteration ($j > 1$), use $\tilde{t}_i^{(j-1)}, i = 1, 2, ..., n_1$, and $(\tilde{u}_i^{(j-1)}, \tilde{v}_i^{(j-1)}), i = n_1 + 1, ..., n$ as our data points in (3.2) and obtain $S_0^{(j)}(t)$. Substitute $S_0^{(j)}(t)$ in (3.7)

and solve $\partial l(\boldsymbol{\theta})/\partial\theta_r = 0, r = 1, 2, ..., p$ to obatain the j'th iterated update $\boldsymbol{\theta}^{(j)}$ of $\boldsymbol{\theta}$.

Step 5. Perform subsequent iterations by using Steps 3 and 4 until convergence is met, say when $\|\boldsymbol{\theta}^{(k)} - \boldsymbol{\theta}^{(k+1)}\| < 0.0001$ and $\sup_t \left\{|S_0^{(k)}(t) - S_0^{(k+1)}(t)|\right\} < 0.001$, for some positive integer k.

Note that *Step 3* in the algorithm is justified because if the model assumption (3.1) is true and if we write $a_i = (S_0^{(1)}(t_i))^{\exp(\boldsymbol{\theta}^{(1)^\top}\boldsymbol{z}_i)}$, then a_i 's follow uniform distribution over $[0, 1]$, being the survival function of a continuous type random variable. Therefore to scale these back to the baseline distribution, we need to find $\tilde{t}_i = \inf\{t : S_0^{(1)}(t) \leq a_i\}$. Thus, the correct choice is $\tilde{t}_i = S_0^{(1)^{-1}}(a_i) = S_0^{(1)^{-1}}\left(S_0^{(1)}(t_i)^{exp(\boldsymbol{\theta}^{(1)^\top}\boldsymbol{z}_i)}\right)$.

3.2.1 Asymptotic Properties

To derive asymptotic properties of the estimators, we now define the parameter space to be $(\boldsymbol{\Theta}, \Phi)$, where $\boldsymbol{\Theta}$, which is the parameter space for $\boldsymbol{\theta}$, is a bounded subset of $\mathbb{R}_p$, and Φ, which is the parameter space for $S_0(t)$, is defined as the class of all absolutely continuous survival functions. We denote the estimator of $\boldsymbol{\theta}$ as $\hat{\boldsymbol{\theta}}_{(n)}$ and that of $S_0(t)$ as $\hat{S}_{0(n)}(t)$.

We now state an important assumption regarding the identifiability of the distribution function $F_0(t)$. Let $[a, b]$, $a \leq b$ be any arbitrary interval in the support of T. Define, for each $r \in [a, b]$,

$$A_0(r) = G_0(r-, \infty) - G_0(r-, r) = P(U < r < V). \tag{3.8}$$

Consider a situation where $A_0(r) = 1$ for all $r \in [a, b]$ for which $F_0(b) > F_0(a-)$. That is, censoring occurs with probability 1 on this interval where $F_0(t)$ has a positive mass. Consequently, there will not be any exact observation in this interval, making it impossible to distinguish between two distributions which are identical outside $[a, b]$ but differing only within $[a, b]$. To overcome this issue, we make the following assumption.

A1: The probability defined in (3.8) is strictly less than one.

Besides this identifiability condition, the following conditions are also assumed to hold for establishing the consistency property.

A2: Conditional on $\boldsymbol{z}$, T is independent of (U, V) and the censoring mechanism is covariate independent.

A3: The joint distribution of $(U, V, \boldsymbol{z})$ does not depend on the true parameter $(\boldsymbol{\theta}^0, S_0^0(t))$.

A4: The covariate space is bounded. That is, there exist some finite $M_0 > 0$ such that $P\{\|\boldsymbol{z}\| \leq M_0\} = 1$, where $\|\cdot\|$ is the usual metric on $\mathbb{R}_p$.

A5: The distribution of $\boldsymbol{z}$ is not concentrated on any proper affine subspace of $\mathbb{R}_p$.

Theorem 3.1. Suppose that $\boldsymbol{\Theta} \in \mathbb{R}_p$ is bounded and assumptions (A1) to (A5) hold. Then the estimator $(\hat{\boldsymbol{\theta}}_{(n)}, \hat{S}_{0(n)}(t))$ is consistent for the true parameter $(\boldsymbol{\theta}^0, S_0^0(t))$ in the sense that if we define a metric $d_0 : \boldsymbol{\Theta} \times \Phi \to \mathbb{R}$ by

$$d_0((\boldsymbol{\theta}_1, S_{01}(t)), (\boldsymbol{\theta}_2, S_{02}(t))) = \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| + \int |S_{01}(t) - S_{02}(t)| dF_0(t) + \left[\int \left((S_{01}(u) - S_{02}(u))^2 + (S_{01}(v) - S_{02}(v))^2\right) dG_0(u, v)\right]^{\frac{1}{2}}, \tag{3.9}$$

where $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \boldsymbol{\Theta}$ and $S_{01}(t), S_{02}(t) \in \Phi$, then $d_0\left((\hat{\boldsymbol{\theta}}_{(n)}, \hat{S}_{0(n)}), (\boldsymbol{\theta}^0, S_0^0(t))\right) \to 0$ almost surely (a.s.).

Proof. In the following discussion we denote $D_i = (X_i, \delta_i)$. Let the probability function of D be given by

$$p(d; \boldsymbol{\theta}, S_0(t)) = \prod_{i=1}^{n} f(t_i|\boldsymbol{z}_i)^{\delta_i}[(S_0(u_i))^{\exp(\boldsymbol{\theta}^\top \boldsymbol{z}_i)} - (S_0(v_i))^{\exp(\boldsymbol{\theta}^\top \boldsymbol{z}_i)}]^{1-\delta_i} g_0(u_i, v_i) q(\boldsymbol{z}_i), \tag{3.10}$$

where $g_0(u, v)$ is the joint probability density function of (U, V), and q is the probability density function of $\boldsymbol{z}$.

Using $(A2)$ and $(A3)$, the log-likelihood function scaled by $1/n$ for the sample $(d_i, \boldsymbol{z}_i), i = 1, 2, ..., n$, up to terms not depending on $(\boldsymbol{\theta}^0, S_0^0(t))$, is

$$l(\boldsymbol{\theta}, S_0(t)) = \frac{1}{n} \sum_{i=1}^{n} \left\{ \delta_i \log f(t_i|\boldsymbol{z}_i) + (1 - \delta_i) \log [(S_0(u_i))^{\exp(\boldsymbol{\theta}^\top \boldsymbol{z}_i)} - (S_0(v_i))^{\exp(\boldsymbol{\theta}^\top \boldsymbol{z}_i)}] \right\}. \tag{3.11}$$

We write $p_n(d) = p(d; \hat{\boldsymbol{\theta}}_{(n)}, \hat{S}_{0(n)}(t))$ and $p_0(d) = p(d; \boldsymbol{\theta}^0, S_0^0(t))$ where $(\hat{\boldsymbol{\theta}}_{(n)}, \hat{S}_{0(n)}(t))$ is the MLE that maximizes the likelihood function over $\boldsymbol{\Theta} \times \boldsymbol{\Phi}$ and $(\boldsymbol{\theta}^0, S_0^0(t)) \in \boldsymbol{\Theta} \times \boldsymbol{\Phi}$. Therefore

$$\sum_{i=1}^{n} \log p_n(D_i) \geq \sum_{i=1}^{n} \log p_0(D_i),$$

and hence

$$\sum_{i=1}^{n} \log \frac{p_n(D_i)}{p_0(D_i)} \geq 0.$$

By the concavity of the function $x \mapsto \log x$, for any real α, $0 < \alpha < 1$,

$$\frac{1}{n} \sum_{i=1}^{n} \log \left((1 - \alpha) + \alpha \frac{p_n(D_i)}{p_0(D_i)} \right) \geq 0. \tag{3.12}$$

The left hand side can be written as

$$\int \log\left((1-\alpha)+\alpha\frac{p_n(D_i)}{p_0(D_i)}\right)\mathrm{d}(\mathbb{P}_n-\mathbb{P})(D)+\int \log\left((1-\alpha)+\alpha\frac{p_n(D_i)}{p_0(D_i)}\right)\mathrm{d}\mathbb{P}(D), \tag{3.13}$$

where $\mathbb{P}_n$ is the empirical measure of D and $\mathbb{P}$ is the joint probability measure of D. We assume that the sample space Ω consists of all infinite sequences $\{D_1, D_2, ...\}$, along with the usual sigma field generated by the product topology on $\prod_1^\infty(\mathbb{R}^3 \times \{0,1\})$ and the product measure $\mathbf{P}$. For p defined in (3.10), let us define a class of functions $\mathcal{P} = \left\{p(d,\boldsymbol{\theta},S_0(t)),(\boldsymbol{\theta},S_0(t)) \in (\boldsymbol{\Theta}\times\boldsymbol{\Phi})\right\}$ and a class of functions $\mathcal{H} = \left\{\log(1-\alpha+\alpha p/p_0) : p \in \mathcal{P}\right\}$, where $p_0 = p(d;\boldsymbol{\theta}^0,S_0^0(t))$. Then it follows from Huang and Wellner (1995) that $\mathcal{H}$ is a Donsker class. Further, from Glivenko-Cantelli theorem, there exists a set $\Omega_0 \subset \Omega$ with $\mathbf{P}(\Omega_0) = 1$ such that for every $\omega \in \Omega_0$, the first term of (3.13) converges to zero. Now fix a point $\omega \in \Omega_0$ and write $\hat{\boldsymbol{\theta}}_{(n)} = \hat{\boldsymbol{\theta}}_{(n)}(\omega)$ and $\hat{S}_{0(n)}(\cdot) = \hat{S}_{0(n)}(\cdot,\omega)$. By our assumption $\boldsymbol{\Theta}$ is bounded, and hence for any subsequence of $\hat{\boldsymbol{\theta}}_{(n)}$, we can find a subsequence converging to $\boldsymbol{\theta}_* \in \boldsymbol{\Theta}^{\mathcal{C}}$, the closure of $\boldsymbol{\Theta}$. Also by Helly's selection theorem, for any subsequence of $\hat{S}_{0(n)}(t)$, we can find a further subsequence converging to some decreasing function $S_{0*}(t)$. Choose the convergent subsequence of $\hat{\boldsymbol{\theta}}_{(n)}$ and the convergent subsequence of $\hat{S}_{0(n)}(t)$ so that they have the same indices, and without loss of generality, assume that $\hat{\boldsymbol{\theta}}_{(n)}$ converges to $\boldsymbol{\theta}_*$ and that $\hat{S}_{0(n)}(t)$ converges to $S_{0*}(t)$. Let $p_*(d) = p(d;\boldsymbol{\theta}_*,S_{0*}(t))$. By the bounded convergence theorem, the second term of (3.13) converges to

$$A_* = \int \log\left((1-\alpha)+\alpha\frac{p_*(d)}{p_0(d)}\right)d\mathbb{P}(d), \tag{3.14}$$

which is nonnegative by (3.12). However, by Jensen's inequality, (3.14) must be non-positive. Therefore (3.14) must be zero and it follows that

$$p_*(d) = p_0(d) \quad \mathbb{P} - \text{almost surely},$$

which implies

$$S_{0*}(t) = S_0^0(t) \quad F_0 - \text{almost surely}.$$

Therefore by bounded convergence theorem,

$$\int |\hat{S}_{0(n)}(t) - S_0^0(t)| dF_0(t) \to 0 \tag{3.15}$$

and also

$$(S_{0*}(u))^{\exp(\boldsymbol{\theta}_*^\top \boldsymbol{z})} = (S_0^0(u))^{\exp({\boldsymbol{\theta}^0}^\top \boldsymbol{z})} \quad \mathbb{P} - \text{almost surely},$$

and

$$(S_{0*}(v))^{\exp(\boldsymbol{\theta}_*^\top \boldsymbol{z})} = (S_0^0(v))^{\exp({\boldsymbol{\theta}^0}^\top \boldsymbol{z})} \quad \mathbb{P} - \text{almost surely}.$$

This together with (A5) imply that there exist $\boldsymbol{z}_1 \neq \boldsymbol{z}_2$ such that for some point c interior to the support of T,

$$(S_{0*}(c))^{\exp(\boldsymbol{\theta}_*^\top \boldsymbol{z}_1)} = (S_0^0(c))^{\exp({\boldsymbol{\theta}^0}^\top \boldsymbol{z}_1)} \quad \text{and} \quad (S_{0*}(c))^{\exp(\boldsymbol{\theta}_*^\top \boldsymbol{z}_2)} = (S_0^0(c))^{\exp({\boldsymbol{\theta}^0}^\top \boldsymbol{z}_2)}.$$

Since $S_{0*}(c) > 0$ and $S_0^0(c) > 0$, this implies $(\boldsymbol{\theta}_* - \boldsymbol{\theta}^0)^\top (\boldsymbol{z}_1 - \boldsymbol{z}_2) = 0$. Again by (A5), the collection of such $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ has positive probability and there exist at least b such pairs that constitute a full rank $b \times b$ matrix, it follows that $\boldsymbol{\theta}_* = \boldsymbol{\theta}^0$,

This in turn implies that

$$S_{0*}(u) = S_0^0(u) \quad \text{and} \quad S_{0*}(v) = S_0^0(v) \quad G_0 - \text{almost surely.}$$

Therefore by bounded convergence theorem,

$$\int \left((\hat{S}_{0(n)}(u) - S_0^0(u))^2 + (\hat{S}_{0(n)}(v) - S_0^0(v))^2 \right) dG_0(u,v) \to 0. \tag{3.16}$$

The equations (3.15) and (3.16) together with $\boldsymbol{\theta}_* = \boldsymbol{\theta}^0$ hold for all $\omega \in \Omega_0$ with $\mathbf{P}(\Omega_0) = 1$. This completes the proof. □

3.2.2 Nonparametric Bootstrap Method

We discuss nonparametric bootstrap method for middle-censored data and some inference procedures based on it. The method of conditional bootstrapping discussed in Chapter 2 was based on parametric simulations. In this chapter we use a nonparametric counterpart of it. We obtain bootstrap samples using Algorithm 2.2 with the changes that instead of the fitted parametric model of the baseline survival function, we use the estimate $\hat{S}_{0(n)}(t)$, and instead of $\hat{\boldsymbol{\theta}}$ we use $\hat{\boldsymbol{\theta}}_{(n)}$ for generating T_i^*. For an arbitrary parameter of interest say η, the standard error (SE) of its estimate $\hat{\eta}$ can be computed using (2.10).

The bootstrap method can be applied to derive a confidence interval for a parameter η. To achieve this we adopt the method of studentized bootstrap confidence interval (Davison and Hinkley (1997)) of confidence coefficient $(1-\alpha)$, $0 < \alpha < 1$. We define the studentized bootstrap statistic as $W^* = (\hat{\eta}^* - \hat{\eta})/{M^*}^{1/2}$, where $\hat{\eta}^*$ is the

estimate of η, and $M^{*^{1/2}}$ is an estimate of variance of $\hat{\eta}^*$, both based on a simulated random sample $(X_i^*, \delta_i^*, \boldsymbol{z}_i), 1 \leq i \leq n$ using Algorithm 2.2. Here $M^{*^{1/2}}$ can be obtained by double bootstrapping, that is, bootstrapping within each bootstrap sample. Then the required confidence interval is given by $(\hat{\eta} - m^{1/2} w^*_{(1-\frac{\alpha}{2})}, \hat{\eta} - m^{1/2} w^*_{(\frac{\alpha}{2})})$, where m is an estimate of $\text{Var}(\hat{\eta})$ which can be obtained by using bootstrap method as before, and $w^*_{(\frac{\alpha}{2})}$, $w^*_{(1-\frac{\alpha}{2})}$ are respectively the $\alpha/2$'th and $(1-\alpha/2)$'th quantiles of the studentized bootstrap statistic values $w_1^*, w_2^*, ..., w_B^*$.

Suppose now that we want to test the significance of the regression parameter $\boldsymbol{\theta}$. For this we first consider a general hypothesis $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$, where $\boldsymbol{\theta}_0$ is a specific value of $\boldsymbol{\theta}$, and can be tested against the alternative hypothesis $H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. We then let $\boldsymbol{\theta}_0 = \mathbf{0}$, where $\mathbf{0}$ is the null vector of same order. We follow the studentized bootstrap test procedure outlined in Section 4.4 of Davison and Hinkley (1997) and accordingly we define the studentized test statistic as $\rho = (\hat{\boldsymbol{\theta}}_{(n)} - \boldsymbol{\theta}_0)^\top \Gamma_0^{-1} (\hat{\boldsymbol{\theta}}_{(n)} - \boldsymbol{\theta}_0)$, where Γ_0 is an estimated variance of $\hat{\boldsymbol{\theta}}_{(n)}$ under the null hypothesis, which may be obtained by means of bootstrap method. Let ρ_0 be the observed value of ρ corresponding to the data. We now generate B bootstrap samples under the null hypothesis and for the b'th sample $(b = 1, 2, ..., B)$ we compute the studentized bootstrap statistic $\rho_b^* = (\hat{\boldsymbol{\theta}}^*_{(n)b} - \hat{\boldsymbol{\theta}}_{(n)})^\top \Gamma^{*^{-1}}_{0b} (\hat{\boldsymbol{\theta}}_{(n)b} - \hat{\boldsymbol{\theta}}_{(n)})$, where Γ^*_{0b}, which is the estimated variance of the estimator based on b'th bootstrap sample can be obtained by the method of double bootstrap. Then the P-value of the test can be approximated by $p = \frac{1 + \sum_{b=1}^{B} I(\rho_b^{*2} \geq \rho_0^2)}{B+1}$, where the quantities $\rho_1^*, \rho_2^*, ..., \rho_B^*$ are obtained from the B studentized bootstrap statistic values generated under the null hypothesis.

TABLE 3.1: Bias, MSE and BCP of the estimator of θ under mild censoring.

		$n = 50$			$n = 100$		
λ	θ	Bias	MSE	BCP	Bias	MSE	BCP
0.10	-0.50	0.028	0.0044	0.926	0.025	0.0009	0.929
0.10	1.00	0.048	0.0037	0.913	0.044	0.0009	0.914
1.00	0.05	0.032	0.0042	0.933	0.029	0.0015	0.937
1.00	0.80	0.051	0.0042	0.922	0.047	0.0008	0.924
3.00	-0.05	0.006	0.0044	0.947	0.002	0.0010	0.950
3.00	0.10	0.008	0.0038	0.914	0.003	0.0012	0.916
8.00	-0.50	0.048	0.0056	0.919	0.045	0.0009	0.921
8.00	0.07	0.038	0.0046	0.948	0.036	0.0015	0.953

TABLE 3.2: Bias, MSE and BCP of the estimator of θ under heavy censoring.

		$n = 50$			$n = 100$		
λ	θ	Bias	MSE	BCP	Bias	MSE	BCP
0.10	-0.50	0.035	0.0109	0.924	0.028	0.0035	0.925
0.10	1.00	0.054	0.0099	0.907	0.048	0.0036	0.911
1.00	0.05	0.039	0.0129	0.925	0.032	0.0043	0.933
1.00	0.80	0.059	0.0109	0.919	0.050	0.0045	0.921
3.00	-0.05	0.012	0.0107	0.941	0.004	0.0034	0.947
3.00	0.10	0.014	0.0109	0.911	0.007	0.0039	0.913
8.00	-0.50	0.051	0.0097	0.914	0.047	0.0038	0.917
8.00	0.07	0.044	0.0115	0.939	0.039	0.0053	0.951

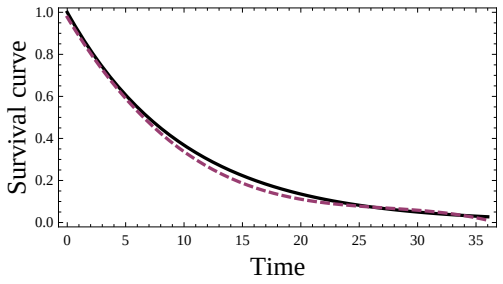
3.3 Simulation Studies

Simulation studies are carried out to assess the finite sample performance of the estimators. We consider exponential distribution with mean λ^{-1} as the baseline distribution of lifetime variable T. We choose exponential distribution with mean $\lambda_1^{-1} = 8$ as the distribution for U and exponential distribution with mean $\lambda_2^{-1} = 5$ as the distribution for $Y = V - U$. We assume that T, U and Y are mutually independent. We consider a single covariate z in the present study, which is generated from uniform distribution over $[0, 10]$ and let θ be the corresponding regression parameter. Under the proportional hazards assumption (3.1), the survival function

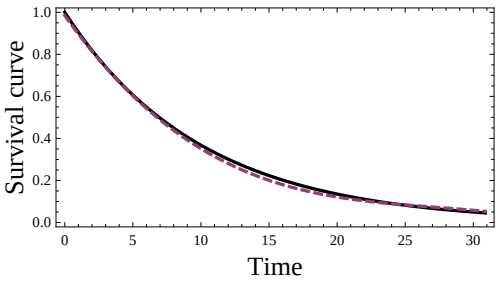
of T given $\boldsymbol{z}$ is given by

$$S(t|\boldsymbol{z}) = \exp\big(-\lambda \exp(\theta \boldsymbol{z}) t\big). \tag{3.17}$$

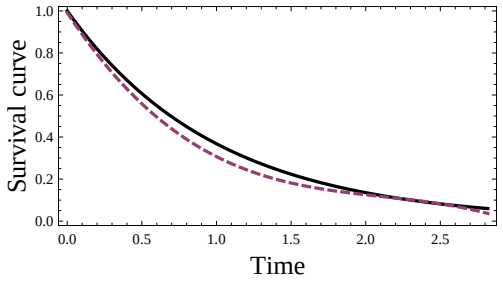
We generate sample of sizes $n = 50$ and 100 from (3.17) for several combinations of λ and θ. Now corresponding to each generated lifetime, a random censoring interval is formed with (U, V). If we find $T \notin (U, V)$, then T is selected as a sample observation, otherwise we choose the interval (U, V) as a sample observation. We consider two different censoring rates: 10% (mild censoring) and 30% (heavy censoring) for our inference. The average absolute bias (Bias) and estimated mean squared error (MSE) are computed using 1000 iterations and are given in Tables 3.1 - 3.2. In each case, a 95% symmetric studentized bootstrap confidence interval for regression parameter is computed with $B = 1000$. The proportion of times the true parameter value lies in such intervals is called bootstrap coverage probabilities (BCP). They are also reported in Tables 3.1 - 3.2. It is evident that both bias and MSE are small in each case and they decrease as the sample size increases. The bootstrap coverage probabilities are found fairly large, close to 0.95. Further, as the censoring rate increases the bias and MSE increase, while the BCP decreases. For each combination of parameter values of (λ, θ) considered, we found out a cubic polynomial estimate of $S_0(t)$ by running the iterative algorithm, where each coefficient in the estimate $\hat{S}_{0(n)}(t)$ is obtained by taking the average of corresponding coefficients computed for all the 1000 iterations. These estimates are compared in Figures 3.1 - 3.4. We see that the estimated survival functions are close to the true survival functions and the estimates become closer to the true survival functions as the sample size increases. Moreover, as the censoring rate increases, the estimated



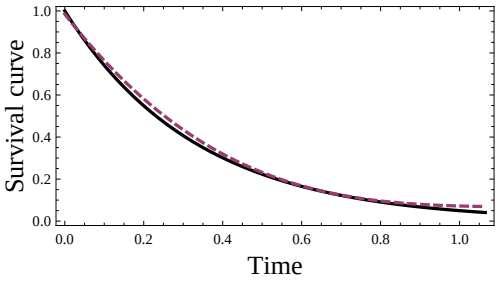
(a) $\lambda = 0.1, \theta = -0.50; n = 50$



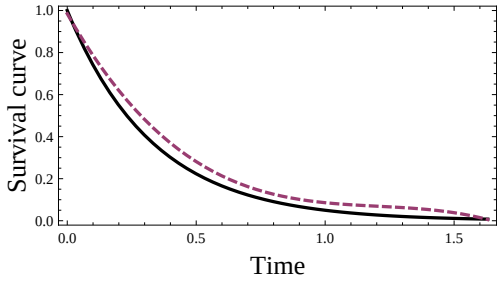
(b) $\lambda = 0.1, \theta = -0.50; n = 100$



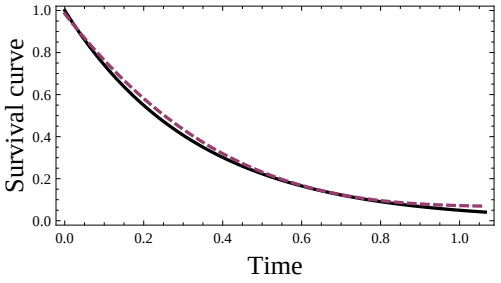
(c) $\lambda = 1, \theta = 0.05; n = 50$



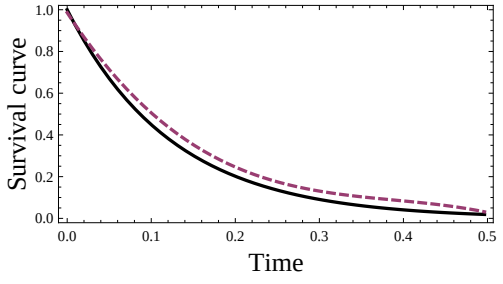
(d) $\lambda = 1, \theta = 0.05; n = 100$



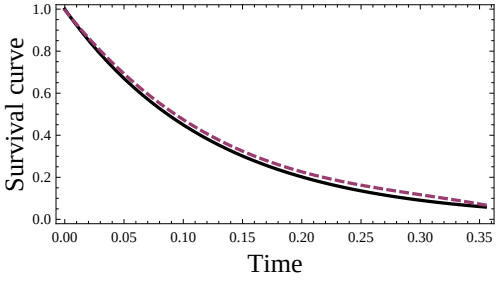
(e) $\lambda = 3, \theta = -0.05; n = 50$



(f) $\lambda = 3, \theta = -0.05; n = 100$

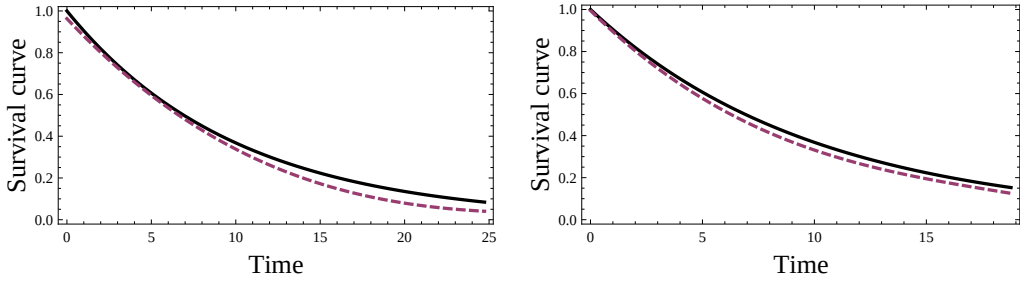


(g) $\lambda = 8, \theta = -0.50; n = 50$



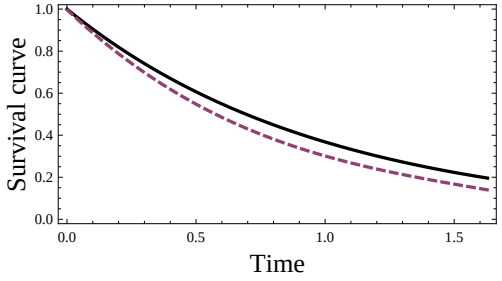
(h) $\lambda = 8, \theta = -0.50; n = 100$

FIGURE 3.1: Plots of baseline survival function (continuous curve) and its estimate (dashed curve) under mild censoring.

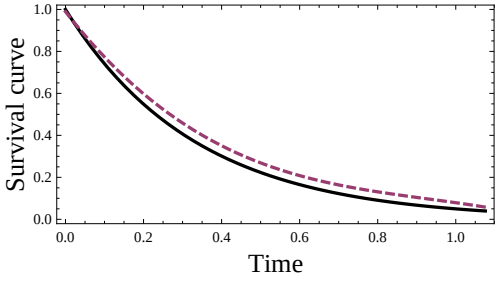


(a) $\lambda = 0.1, \theta = -0.50; n = 50$

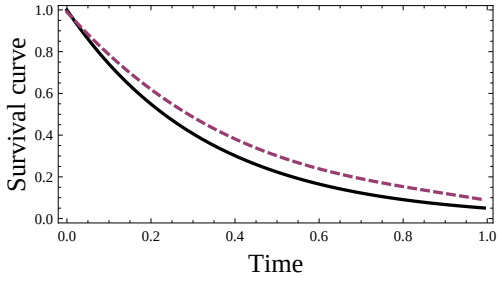
(b) $\lambda = 0.1, \theta = -0.50; n = 100$



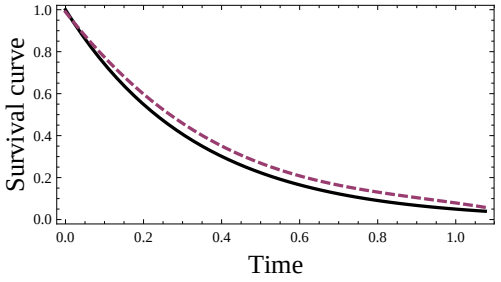
(c) $\lambda = 1, \theta = 0.05; n = 50$



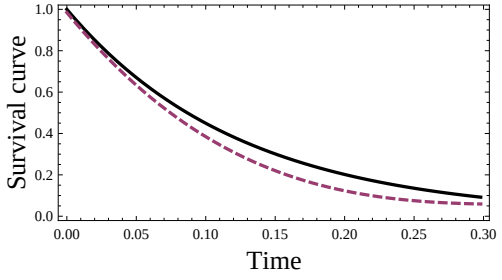
(d) $\lambda = 1, \theta = 0.05; n = 100$



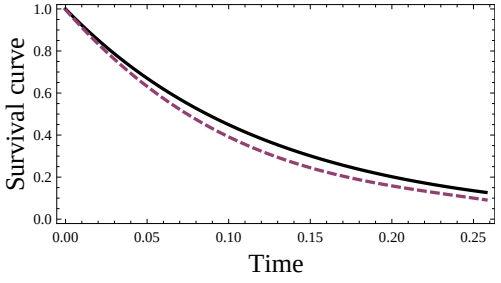
(e) $\lambda = 3, \theta = -0.05; n = 50$



(f) $\lambda = 3, \theta = -0.05; n = 100$

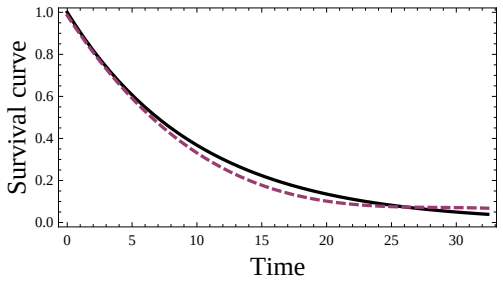


(g) $\lambda = 8, \theta = -0.50; n = 50$

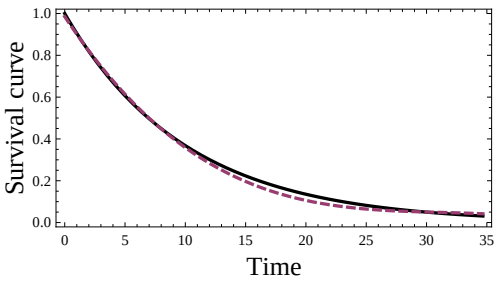


(h) $\lambda = 8, \theta = -0.50; n = 100$

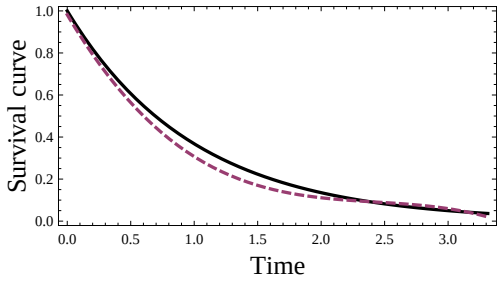
FIGURE 3.2: Plots of baseline survival function (continuous curve) and its estimate (dashed curve) under heavy censoring.



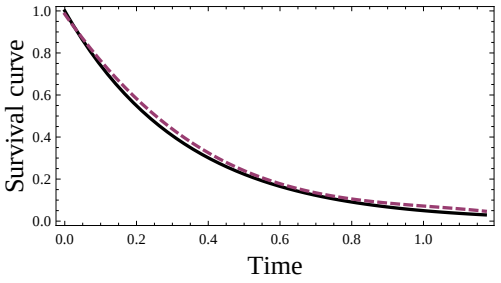
(a) $\lambda = 0.1, \theta = 1.00; n = 50$



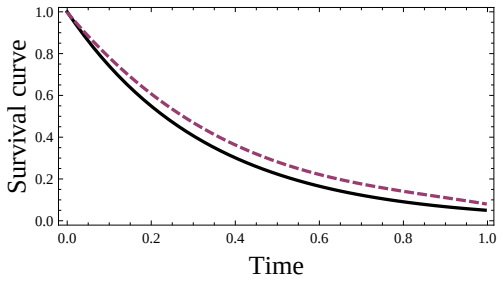
(b) $\lambda = 0.1, \theta = 1.00; n = 100$



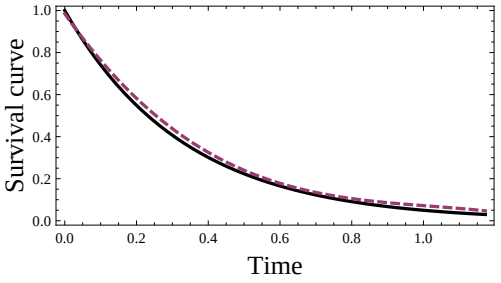
(c) $\lambda = 1, \theta = 0.80; n = 50$



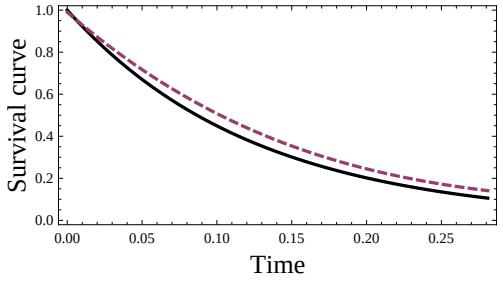
(d) $\lambda = 1, \theta = 0.80; n = 100$



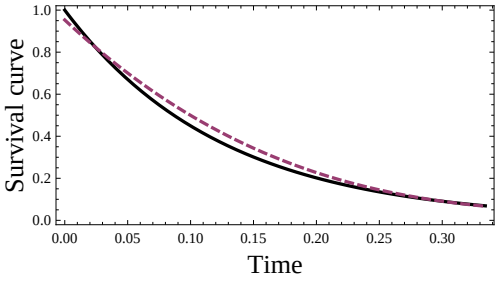
(e) $\lambda = 3, \theta = 0.10; n = 50$



(f) $\lambda = 3, \theta = 0.10; n = 100$

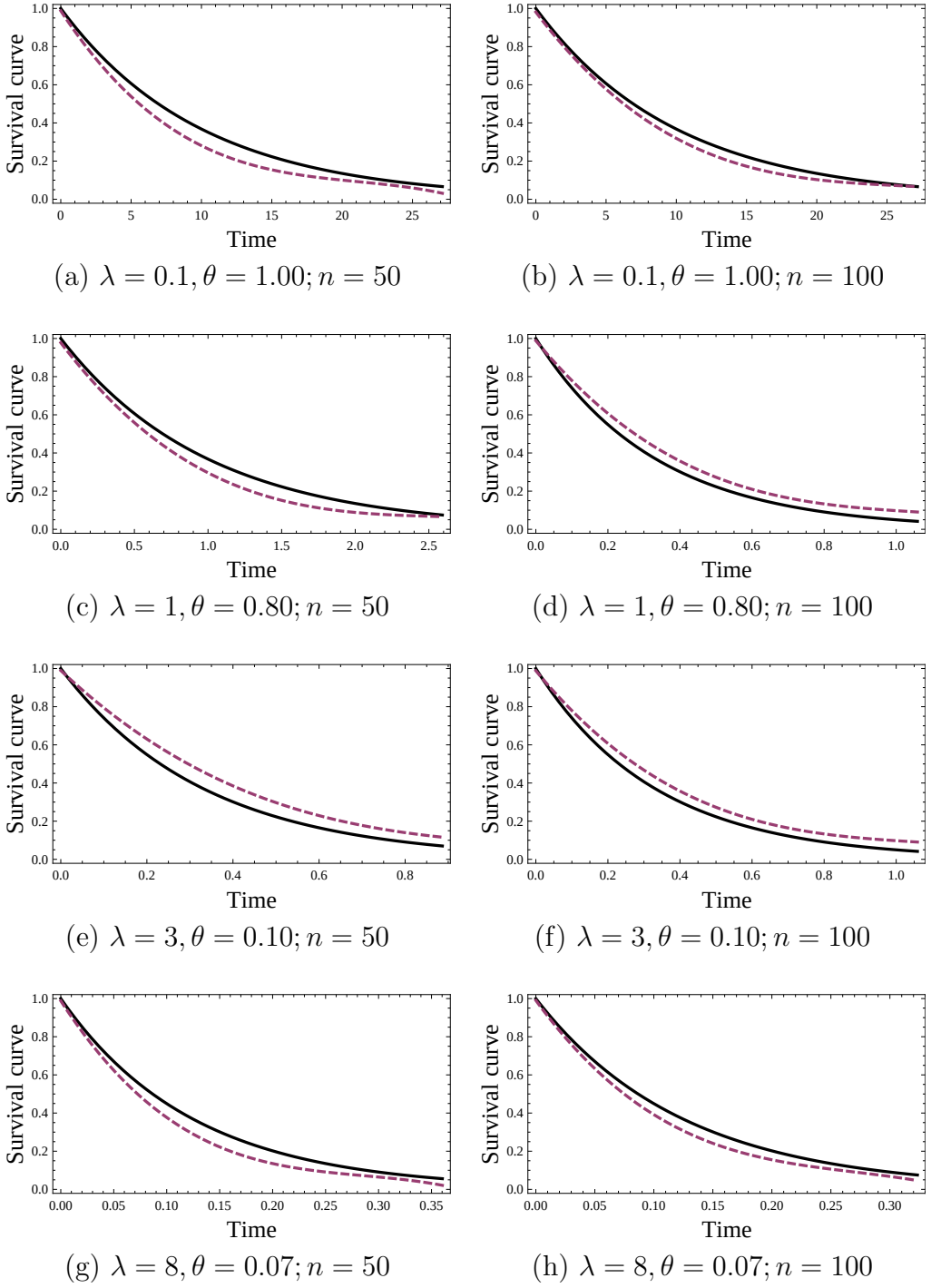


(g) $\lambda = 8, \theta = 0.07; n = 50$



(h) $\lambda = 8, \theta = 0.07; n = 100$

FIGURE 3.3: Plots of baseline survival function (continuous curve) and its estimate (dashed curve) under mild censoring.



(a) $\lambda = 0.1, \theta = 1.00; n = 50$ (b) $\lambda = 0.1, \theta = 1.00; n = 100$

(c) $\lambda = 1, \theta = 0.80; n = 50$ (d) $\lambda = 1, \theta = 0.80; n = 100$

(e) $\lambda = 3, \theta = 0.10; n = 50$ (f) $\lambda = 3, \theta = 0.10; n = 100$

(g) $\lambda = 8, \theta = 0.07; n = 50$ (h) $\lambda = 8, \theta = 0.07; n = 100$

FIGURE 3.4: Plots of baseline survival function (continuous curve) and its estimate (dashed curve) under heavy censoring.

TABLE 3.3: Estimates of parameters and their SE.

	Estimate	SE
θ_1	0.0044	0.0081
θ_2	0.1096	0.0703
c_0	0.8806	0.1166
c_1	-0.2218	0.0863
c_2	-0.0023	0.0092
c_3	7.8860×10^{-7}	1.9481×10^{-8}

survival functions depart more from their respective true survival functions.

3.4 An Application

In this section, we apply our model to a real life data set. We consider a data set on survival times (in years) for 149 diabetic patients reported in Lee and Wang (2003). We first consider all exact lifetimes for our study and then we impose artificial middle-censoring as we did in Chapter 2. For illustrative purpose we take two covariates from the data set, namely age denoted by z_1 and coronary heart disease (CHD) denoted by z_2, with respective regression parameters θ_1 and θ_2. The data is middle-censoring by the following method. A random censoring interval (U, V), where U and $Y = V - U$ are independent exponential variates with means $\lambda_1^{-1} = 20$ and $\lambda_2^{-1} = 12.5$ respectively is generated first. An individual with exact lifetime is selected at random and if that lifetime happens to fall in the generated censoring interval, it is assumed to be middle-censored and the corresponding censoring interval is considered as the observation. Otherwise the lifetime is taken. This process is repeated until around 25% of the observations are censored. We apply the model given in Section 3.2 to this new data set consisting of both exact lifetimes and censoring intervals and obtained the estimates of regression parameters as well

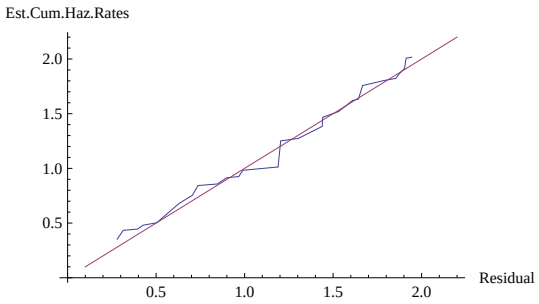


FIGURE 3.5: Plot of estimated cumulative hazard rates against r_j's.

as the coefficients of cubic polynomial form of baseline survival function and are reported in Table 3.3. We also report the standard errors of these estimates based on nonparametric bootstrap method with B=1000. The estimates of the regression parameters indicate that the covariates have adverse effect on the lifetime. Further the SE values are small. We now test the significance of regression parameter $\boldsymbol{\theta}$ by studentized bootstrap test. The test shows a P-value of 0.0271, indicating that the covariates are significant at 5% level of significance.

We now check the overall fit of the model by using Cox-Snell residuals (Cox and Snell (1968)). The method is similar to that given in Chapter 2, except that for defining residuals we use $\hat{\boldsymbol{\theta}}_{(n)}$ and the baseline cumulative hazard function obtained from $\hat{S}_{0(n)}(t)$. Figure 3.5 shows the Cox-Snell residuals plot obtained. The curve is close to the straight line indicating data fits the model reasonably well.

3.5 Conclusion

In this chapter, we discussed a semiparametric proportional hazards regression problem for the analysis of middle-censored data. We obtained the SPMLE of regression parameter as well as the NPMLE of baseline survival function by using an iterative

algorithm. Moreover the consistency of these estimators were established. Simulation studies indicated that the estimators are performing well in terms of bias, MSE, and BCP. The model was applied to a real data set. Asymptotic normality of $\hat{\boldsymbol{\theta}}_{(n)}$ and weak convergence of $\hat{S}_{0(n)}(t)$ do not appear to be easy to establish, although one can perhaps extend the ideas used in Huang and Wellner (1995).

Chapter 4

Parametric Additive Hazards Regression Model

4.1 Introduction

In the last two chapters, we examined regression models for survival data based on proportional hazards model, where the effect of the covariates act multiplicatively on the baseline hazard rate. There are situations where the proportional hazards model is not suitable to associate covariate effect on lifetime. An alternative model is the additive hazards model. In Chapter 1, we have discussed advantages of additive hazards model over proportional hazards model. For the two sample situation, the additive hazards model addresses the risk difference, while the proportional hazards model concerns the risk ratio. In tumorigenicity experiments that investigate the dose effect on tumor risk, an additive hazards model may be more reasonable since the excess risk is often the quantity of interest (see Breslow and Day (1980)). Motivated by this, in this chapter, we introduce and study additive hazards model for lifetime data subject to middle-censoring in parametric context.

[1]The results in this chapter are accepted for publication in *Communications in Statistics - Simulation and Computation* (See Sankaran and Prasad (2017a)).

The rest of the chapter is organized as follows. In Section 4.2, we discuss additive hazards model in a parametric context, where the baseline distribution of the lifetime variate is assumed to be exponentiated exponential. We also provide inference procedures, where we estimate the unknown parameters using Expectation-Maximization algorithm (Dempster et al. (1977)). In Section 4.3, we present various conditional distributions and establish asymptotic normality property of the estimator. Section 4.4 discusses simulation studies for assessing the finite sample performance of the proposed estimator. In Section 4.5, the proposed method is illustrated to show the utility in practical situations. Finally, Section 4.6 gives major conclusions of this study.

4.2 The Model and Inference Procedure

Let T denote the lifetime of interest. We consider the additive hazards model defined by

$$h(t|\boldsymbol{z}) = h_0(t) + \boldsymbol{z}^\top \boldsymbol{\theta}, \tag{4.1}$$

where $h_0(t)$ is the baseline survival function, $\boldsymbol{z}$ is the $p \times 1$ vector of covariates and $\boldsymbol{\theta}$ is the corresponding $p \times 1$ vector of regression parameters. We assume that the baseline distribution of T is exponentiated exponential (EE) with scale parameter λ and shape parameter α with distribution function given by

$$F_0(t) = (1 - \exp(-\lambda t))^\alpha, \quad \lambda > 0,\ \alpha > 0. \tag{4.2}$$

Thus $h_0(t)$ in the model (4.1) takes the form

$$h_0(t) = \frac{\alpha\lambda \exp(-\lambda t)\left(1-\exp(-\lambda t)\right)^{\alpha-1}}{1-\left(1-\exp(-\lambda t)\right)^{\alpha}}. \tag{4.3}$$

The distribution defined in (4.2) was introduced by Ahuja and Nash (1967) and further studied by Gupta and Kundu (1999). This family of distributions can be well used as an alternative to gamma and Weibull families of distributions in analyzing lifetime data (Gupta and Kundu (2001)). Further, this distribution has an advantage over Weibull distribution in modelling lifetime data as the model has non-monotonic hazard rates which often suits a regular maintenance environment. Using (4.3), the model formulation given in (4.1) can be written as

$$h(t|\boldsymbol{z}) = \frac{\alpha\lambda \exp(-\lambda t)\left(1-\exp(-\lambda t)\right)^{\alpha-1}}{1-\left(1-\exp(-\lambda t)\right)^{\alpha}} + \boldsymbol{z}^{\top}\boldsymbol{\theta}. \tag{4.4}$$

Therefore the survival function of T given $\boldsymbol{z}$, is given by

$$S\left(t|\boldsymbol{z}\right) = \exp\left(-H_0(t) - \boldsymbol{z}^{\top}\boldsymbol{\theta}t\right), \tag{4.5}$$

where $H_0(t) = \int_0^t h_0(a)da$. We thus obtain the probability density function of T given on $\boldsymbol{z}$, as

$$\begin{aligned} f(t|\boldsymbol{z}) &= -\frac{d}{dt}S(t|\boldsymbol{z}) \\ &= \left(h_0(t) + \boldsymbol{z}^{\top}\boldsymbol{\theta}\right)\exp\{-H_0(t) - \boldsymbol{z}^{\top}\boldsymbol{\theta}t\}. \end{aligned} \tag{4.6}$$

We now assume that the lifetime T is middle-censored by the random censoring interval (U, V), and that given $\boldsymbol{z}$, (U, V) is independent of T, with $P\left(U < V\right) = 1$.

We also assume that the censoring mechanism is independent of covariates. Thus one can observe the vector $(X, \delta, \boldsymbol{z})$, where $X = T$ if $\delta = 1$, and $X = (U, V)$ if $\delta = 0$, with $\delta = I(X = T)$ as the censoring indicator. Let $(X_i, \delta_i, \boldsymbol{z}_i)$, $i = 1, 2, ..., n$ be independent and identically distributed copies of $(X, \delta, \boldsymbol{z})$ corresponding to n individuals under investigation. The likelihood function corresponding to the observed data is given by

$$L(\boldsymbol{\psi}) \propto \prod_{i=1}^{n} f(t_i|\boldsymbol{z}_i)^{\delta_i} \left[S(u_i|\boldsymbol{z}_i) - S(v_i|\boldsymbol{z}_i)\right]^{1-\delta_i}, \tag{4.7}$$

where $\boldsymbol{\psi} = (\alpha, \lambda, \boldsymbol{\theta}^\top)^\top$ and t_i or (u_i, v_i) is the realization corresponding to the i'th individual, $i = 1, 2, ..., n$. Without loss of generality, arrange these observations in such a way that the first n_1 observations are exact lifetimes and remaining n_2 are censored intervals, with $n_1 + n_2 = n$. We can now rewrite the likelihood function in (4.7) as

$$\begin{aligned} L(\boldsymbol{\psi}) &= \prod_{i=1}^{n_1} f(t_i|\boldsymbol{z}_i) \prod_{i=n_1+1}^{n} \left[S(u_i|\boldsymbol{z}_i) - S(v_i|\boldsymbol{z}_i)\right] \\ &= \prod_{i=1}^{n_1} \big(h_0(t_i) + \boldsymbol{z}_i^\top \boldsymbol{\theta}\big) \exp\{-H_0(t_i) - \boldsymbol{z}_i^\top \boldsymbol{\theta} t_i\} \times \\ &\qquad \prod_{i=n_1+1}^{n} \left[\exp\{-H_0(u_i) - \boldsymbol{z}_i^\top \boldsymbol{\theta} u_i\} - \exp\{-H_0(v_i) - \boldsymbol{z}_i^\top \boldsymbol{\theta} v_i\}\right]. \end{aligned} \tag{4.8}$$

Now we estimate $\boldsymbol{\psi}$ via the maximum likelihood method. To achieve this we adopt the Expectation-Maximization(EM) algorithm. The complete-data log-likelihood function corresponding to the censored data likelihood function in (4.7), excluding

the normalizing constant, is given by

$$l_c(\boldsymbol{\psi}) = \sum_{i=1}^{n_1} \log f\left(t_i|\boldsymbol{z}_i\right) + \sum_{i=n_1+1}^{n} \log f\left(\check{t}_i|\boldsymbol{z}_i\right), \tag{4.9}$$

where $\check{t}_i$'s are potentially observable, but unobserved realizations of T, such that $\check{t}_i \in (u_i, v_i)$, $i = n_1 + 1, ..., n$.

Using (4.6), we can rewrite (4.9) as

$$l_c(\boldsymbol{\psi}) = \sum_{i=1}^{n_1}\Bigg(- (H_0(t_i) + \boldsymbol{z}_i^\top\boldsymbol{\theta}t_i) + \log(H_0(t_i) + \boldsymbol{z}_i^\top\boldsymbol{\theta}t_i)\Bigg) + \sum_{i=n_1+1}^{n}\Bigg(- (H_0(\check{t}_i) + \boldsymbol{z}_i^\top\boldsymbol{\theta}\check{t}_i) + \log(H_0(\check{t}_i) + \boldsymbol{z}_i^\top\boldsymbol{\theta}\check{t}_i)\Bigg). \tag{4.10}$$

For the E-step we consider the expectation of the complete-data log-likelihood function in (4.10) given by

$$\phi(\boldsymbol{\psi}) = \sum_{i=1}^{n_1}\Bigg(- (H_0(t_i) + \boldsymbol{z}_i^\top\boldsymbol{\theta}t_i) + \log(H_0(t_i) + \boldsymbol{z}_i^\top\boldsymbol{\theta}t_i)\Bigg) + \sum_{i=n_1+1}^{n}(\zeta_{1i} - \zeta_{2i}), \tag{4.11}$$

where ζ_{1i} and ζ_{2i} are respectively the expected values of $\log(H_0(\check{t}_i) + \boldsymbol{z}_i^\top\boldsymbol{\theta}\check{t}_i)$ and $(H_0(\check{t}_i) + \boldsymbol{z}_i^\top\boldsymbol{\theta}\check{t}_i)$, conditional on the current update of the parameter and observed data and are given by

ζ_{1i}= $(S(u_i|\boldsymbol{\psi}, \boldsymbol{z}_i) - S(v_i|\boldsymbol{\psi}, \boldsymbol{z}_i))^{-1}\int_{u_i}^{v_i}\log(H_0(t) + \boldsymbol{z}_i^\top\boldsymbol{\theta}t)f(t|\boldsymbol{\psi}, \boldsymbol{z}_i)\mathrm{d}t$ and

ζ_{2i}= $(S(u_i|\boldsymbol{\psi}, \boldsymbol{z}_i) - S(v_i|\boldsymbol{\psi}, \boldsymbol{z}_i))^{-1}\int_{u_i}^{v_i}(H_0(t) + \boldsymbol{z}_i^\top\boldsymbol{\theta}t)f(t|\boldsymbol{\psi}, \boldsymbol{z}_i)\mathrm{d}t$.

Closed form expressions for these quantities are not available. For the M-step, we provide the following algorithm to find the MLE $\hat{\boldsymbol{\psi}} = (\hat{\alpha}, \hat{\lambda}, \hat{\boldsymbol{\theta}}^\top)^\top$ of $\boldsymbol{\psi}$.

Algorithm 4.1

Step 1. Choose an initial value $\boldsymbol{\psi}^{(0)} = \left(\alpha^{(0)}, \lambda^{(0)}, \boldsymbol{\theta}^{(0)^\top}\right)^\top$.

Step 2. At the j'th iteration, for $(j \geq 1)$, evaluate $\zeta_1^{(j)}$ and $\zeta_2^{(j)}$ using $\boldsymbol{\psi}^{(j-1)}$ and substitute in (4.11).

Step 3. With $\alpha^{(j-1)}$ and $\lambda^{(j-1)}$ held fixed, express (4.11) as a function of $\boldsymbol{\theta}$ and solve $\frac{\partial \phi}{\partial \boldsymbol{\theta}} = 0$ to get $\boldsymbol{\theta} = \boldsymbol{\theta}^{(j)}$.

Step 4. With $\boldsymbol{\theta} = \boldsymbol{\theta}^{(j)}$, express (4.11) as a function of α and λ and solve $\frac{\partial \phi}{\partial \alpha} = 0$ $\frac{\partial \phi}{\partial \lambda} = 0$ to get $\alpha^{(j)}$ and $\lambda^{(j)}$ and thus obtain $\boldsymbol{\psi}^{(j)} = \left(\alpha^{(j)}, \lambda^{(j)}, \boldsymbol{\theta}^{(j)^\top}\right)^\top$.

Step 5. Repeat *Step 2* to *Step 4* till convergence in $\boldsymbol{\psi}$ is met, say when $\|\boldsymbol{\psi}^{(k)} - \boldsymbol{\psi}^{(k+1)}\| < 0.0001$, for some finite positive integer k.

In survival analysis, one would often be interested to know whether the covariates have significant effect on the lifetime variate. We use the likelihood ratio test to know such a significance. We set the null hypothesis $H_0 : \boldsymbol{\theta} = \mathbf{0}$, and test this against the alternative hypothesis $H_1 : \boldsymbol{\theta} \neq \mathbf{0}$. The test statistic is given by $\Lambda = 2 \log L\left(\hat{\boldsymbol{\theta}}, \hat{\alpha}, \hat{\lambda}\right) - 2 \log L\left(\mathbf{0}, \tilde{\alpha}, \tilde{\lambda}\right)$, where $\tilde{\alpha}$ and $\tilde{\lambda}$ are the MLEs obtained under the null hypothesis. For large samples, the null distribution of the test statistic Λ is approximately chi-square with p degrees of freedom. The test rejects the null hypothesis for small P-values.

4.3 Asymptotic Properties

In this section, we discuss the asymptotic normality of the MLE $\hat{\boldsymbol{\psi}}$ of the parameter $\boldsymbol{\psi}$. We show that, under certain regularity conditions, the MLE is asymptotically normally distributed with mean $\boldsymbol{\psi}$, the true mean value, and dispersion

matrix $\mathscr{I}^{-1}(\boldsymbol{\psi})$, where $\mathscr{I}(\boldsymbol{\psi})$ is the Fisher information matrix given by $\mathscr{I}(\boldsymbol{\psi}) = -E\big(\frac{\partial^2 l(\boldsymbol{\psi})}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}^\top}\big)$. We provide explicit expressions of the second order derivatives of the log-likelihood function in equations (4.12) to (4.17) below. First we define the following terms. Let ξ, ξ_1 and ξ_2 be arbitrary positive real numbers.

$b(\xi) = 1 - \exp(-\lambda\xi), c(\xi) = \alpha\lambda \log\left(b(\xi)\right),$

$D_1(u, v) = S(u|\boldsymbol{z}) - S(v|\boldsymbol{z}),$

$K_1(\xi) = (\exp(\lambda\xi) - 1))^{-1}(S(\xi|\boldsymbol{z})F_0(\xi)\xi\alpha((F_0(\xi) - 1),$

$K_2(\xi) = ((F_0(\xi) - 1))^{-1}(S(\xi|\boldsymbol{z})F_0(\xi)\log b(\xi)),$

$K_3(\xi) = ((F_0(\xi)-1))^{-1}(S(\xi|\boldsymbol{z})b(\xi)^{\alpha-1}\xi\alpha\log b(\xi))\Big(1-\frac{\xi}{\alpha\log b(\xi)}-\frac{F_0(\xi)}{F_0(\xi)-1}+\frac{F_0(\xi)b(\xi)}{(F_0(\xi)-1)(-b(\xi))}\Big),$

$K_4(\xi) = ((F_0(\xi) - 1))^{-1}(\exp(\lambda\xi) - 1)^{-2}(S(\xi|\boldsymbol{z})F_0(\xi)^{\alpha}\xi\alpha),$

$K_5(\xi) = (F_0(\xi)-1)^{-2}(\exp(\lambda\xi)-1)^{-2}S(\xi|\boldsymbol{z})F_0(\xi)^{2\alpha}\xi^2\alpha^2\Big(1-\exp(\lambda\xi)b(\xi)^{-1}(-b(\xi))+$
$(1 - b(\xi))b(\xi)^{-\alpha-1}(F_0(\xi) - 1)^{-1}\Big).$

$$
\begin{aligned}
\frac{\partial^2 l(\boldsymbol{\psi})}{\partial\alpha^2} = \sum_{i=1}^{n_1}\Bigg(& - \boldsymbol{z}_i^\top\boldsymbol{\theta}t_i - H_0(t_i) - h(t_i|\boldsymbol{z}_i)^{-2}\big\{\alpha^{-1}h_0(t_i) + (1 - b(t_i))b(t_i)^{2\alpha-1}c(t_i) \\
& S_0(t_i)^{-2} + (1 - b(t_i))b(t_i)^{\alpha-1}c(t_i)S_0(t_i)^{-1}\big\} + h(t_i|\boldsymbol{z}_i)^{-1}(1 - b(t_i))b(t_i)^{\alpha-1}c(t_i) \\
& S_0(t_i)^{-2}\big\{2b(t_i)^{\alpha}c(t_i)\alpha^{-1} + 2c(t_i)\alpha^{-1}S_0(t_i)^{-1} + 4b(t_i)^{\alpha} + 6b(t_i)^{\alpha} + 2S_0(t_i)^{-1}\big\}\Bigg) \\
& + \sum_{i=n_1+1}^{n}\Bigg(- D_1(u_i, v_i)^{-2}\big[\frac{S(u_i|\boldsymbol{z}_i)F_0(u_i)\log b(u_i)}{F_0(u_i) - 1} - \frac{S(v_i|\boldsymbol{z}_i)F_0(v_i)\log b(v_i)}{F_0(v_i) - 1}\big] \\
& + 2D_1(u_i, v_i)^{-1}\big[\frac{S(u_i|\boldsymbol{z}_i)F_0(u_i)\log b(u_i)}{F_0(u_i) - 1} - \frac{S(v_i|\boldsymbol{z}_i)F_0(v_i)\log b(v_i)}{F_0(v_i) - 1}\big]\Bigg).
\end{aligned}
\tag{4.12}
$$

For $j, k = 1, 2, ..., p$,

$$\frac{\partial^2 l(\boldsymbol{\psi})}{\partial\theta_j \partial\theta_k} = \sum_{i=1}^{n_1} \Bigg(- \boldsymbol{z}_i^\top \boldsymbol{\theta} t_i - H_0(t_i) - h(t_i|\boldsymbol{z}_i)^{-2} \boldsymbol{z}_{ij}^2 \Bigg) + \sum_{i=n_1+1}^{n} \Bigg(- D_1(u_i, v_i)^{-2} \boldsymbol{z}_{ij} (S(u_i|\boldsymbol{z}_i)u_i - S(v_i|\boldsymbol{z}_i)v_i) + D_1(u_i, v_i)^{-1} \boldsymbol{z}_{ij}^2 (S(u_i|\boldsymbol{z}_i)u_i^2 - S(v_i|\boldsymbol{z}_i)v_i^2) \Bigg). \tag{4.13}$$

$$\begin{aligned}
\frac{\partial^2 l(\boldsymbol{\psi})}{\partial\alpha\partial\lambda} = \sum_{i=1}^{n_1} \Bigg(& - \boldsymbol{z}_i^\top \boldsymbol{\theta} t_i - H_0(t_i) - h(t_i|\boldsymbol{z}_i))^{-2} \Big\{ (1 - b(t_i)) b(t_i)^{\alpha-1} \alpha\lambda S_0(t_i)^{-1})^2 \\
& [\lambda^{-1} - t_i + (1 - b(t_i)) b(t_i)^{-1} (\alpha - 1) + (1 - b(t_i)) b(t_i)^{\alpha-1} \alpha t_i S_0(t_i)^{-1})] \\
& [\alpha^{-1} + b(t_i)^\alpha \log b(t_i) S_0(t_i)^{-1}) + \log b(t_i)] \Big\} + h(t_i|\boldsymbol{z}_i)^{-1} \Big\{ (1 - b(t_i)) b(t_i)^{\alpha-1} \\
& S_0(t_i)^{-1}) \left[1 - \lambda t_i + (1 - b(t_i) b(t_i)^{-1} t_i \lambda(\alpha - 1)\right] 2(1 - b(t_i) b(t_i)^{\alpha-1} S_0(t_i)^{-1} + \\
& (1 - b(t_i) b(t_i)^{-1} t_i \alpha\lambda + b(t_i)^\alpha c(t_i) \lambda^{-1} S_0(t_i)^{-1} + c(t_i) \lambda^{-1} b(t_i)^\alpha c(t_i) t_i S_0(t_i)^{-1} \\
& - c(t_i) t_i S_0(t_i)^{-1} + (1 - b(t_i)) b(t_i)^{-1} c(t_i) t_i (\alpha - 1) + 2(1 - b(t_i)) b(t_i)^{2\alpha-1} \\
& c(t_i) t_i \alpha S_0(t_i)^{-2}) 2(1 - b(t_i)) b(t_i)^{\alpha-1} c(t_i) t_i (2\alpha - 1) S_0(t_i)^{-1}) \Big\} \Bigg) + \\
& \sum_{i=n_1+1}^{n} \Bigg(- D_1(u_i, v_i)^{-2} (K_1(u_i) - K_1(v_i))(K_2(u_i) - K_2(v_i)) + D_1(u_i, v_i)^{-1} \\
& (K_3(u_i) - K_3(v_i)) \Bigg).
\end{aligned} \tag{4.14}$$

$$\frac{\partial^2 l(\boldsymbol{\psi})}{\partial\theta_j\partial\lambda} = \sum_{i=1}^{n_1}\Bigg(- \boldsymbol{z}_i^\top\boldsymbol{\theta}t_i - H_0(t_i) - h(t_i|\boldsymbol{z}_i)^{-2}z_{ij}\Bigg((1-b(t_i))b(t_i)^{\alpha-1}\alpha\lambda S_0(t_i)^{-1}$$
$$\Big[\lambda^{-1} - t_i + (1-b(t_i))b(t_i)^{-1}t_i(\alpha-1) + (1-b(t_i))b(t_i)^{\alpha-1}t_i\alpha S_0(t_i)^{-1}\Big]\Bigg)\Bigg)$$
$$+ \sum_{i=n_1+1}^{n}\Bigg(- D_1(u_i,v_i)^{-2}z_{ij}(S(v_i|\boldsymbol{z}_i)v_i - S(u_i|\boldsymbol{z}_i)u_i)(K_4(u_i) - K_4(v_i)) +$$
$$D_1(u_i,v_i)^{-1}z_{ij}(v_iK_4(v_i) - u_iK_4(u_i))\Bigg), \tag{4.15}$$

for $j = 1, 2, ..., p$.

$$\frac{\partial^2 l(\boldsymbol{\psi})}{\partial\lambda^2} = \sum_{i=1}^{n_1}\Bigg(- \boldsymbol{z}_i^\top\boldsymbol{\theta}t_i - h(t_i|\boldsymbol{z}_i)^{-2}\Bigg((1-b(t_i))b(t_i)^{\alpha-1}\alpha\lambda S_0(t_i)^{-1}\Big[\lambda^{-1} - t_i +$$
$$(1-b(t_i))b(t_i)^{-1}t_i(\alpha-1) + (1-b(t_i))b(t_i)^{\alpha-1}t_i\alpha S_0(t_i)^{-1}\Big]^2\Bigg) + h(t_i|\boldsymbol{z}_i)^{-2}$$
$$\Bigg(\alpha t_i S_0(t_i)^{-1}(1-b(t_i))b(t_i)^{\alpha-1}\Big[-2 + 2(1-b(t_i))b(t_i)^{\alpha-1} + t_i\lambda - 3(1-b(t_i))$$
$$b(t_i)^{-1}\alpha(\alpha-1)\lambda t_i^{-1} + (1-b(t_i))^2b(t_i)^{-2}(\alpha-1)(\alpha-2)\lambda t_i\Big] + 2\alpha^3\lambda t_i S_0(t_i)^{-3}$$
$$(1-b(t_i))^3b(t_i)^{3\alpha-3} + (1-b(t_i))^2b(t_i)^{2\alpha-2}\alpha^2 t_i S_0(t_i)^{-2}\Big[2 - 3\lambda t_i + (1-b(t_i))$$
$$b(t_i)^{-1}t_i\lambda(2\alpha-2)\Big]\Bigg)\Bigg) + \sum_{i=n_1+1}^{n}\Bigg(- D_1(u_i,v_i)^{-2}(K_1(u_i) - K_1(v_i))^2 +$$
$$D_1(u_i,v_i)^{-1}((K_5(u_i) - K_5(v_i) - (K_4(u_i) - K_4(v_i))\Bigg). \tag{4.16}$$

$$\frac{\partial^2 l(\boldsymbol{\psi})}{\partial\alpha\partial\theta_j} = \sum_{i=1}^{n_1}\Bigg(- \boldsymbol{z}^\top\boldsymbol{\theta}t_i - H_0(t_i) - \boldsymbol{z}_{ij}(1-b(t_i))b(t_i)^{\alpha-1}S_0(t_i)^{-2}h(t_i|\boldsymbol{z}_i)^{-2}$$
$$\big(\lambda S_0(t_i) + (b(t_i)^\alpha + S_0(t_i))c(t_0)\big)\Bigg) + \sum_{i=n_1+1}^{n}\Bigg(- D_1(u_i,v_i)^{-2}(S(v_i|\boldsymbol{z}_i)v_i\boldsymbol{z}_{ij} -$$
$$S(u_i|\boldsymbol{z}_i)u_i\boldsymbol{z}_{ij})\Big(\frac{S(u_i|\boldsymbol{z}_i)F_0(u_i)\log b(u_i)}{F_0(u_i)-1} - \frac{S(v_i|\boldsymbol{z}_i)F_0(v_i)\log b(v_i)}{F_0(v_i)-1}\Big) + D_1(u_i,v_i)^{-1}$$
$$\Big(\frac{S(v_i|\boldsymbol{z}_i)F_0(v_i)\log b(v_i)}{F_0(v_i)-1} - \frac{S(u_i|\boldsymbol{z}_i)F_0(u_i)\log b(u_i)}{F_0(u_i)-1}\Big)\Bigg), \tag{4.17}$$

for $j = 1, 2, ..., p$. For the evaluation of the Fisher information matrix, we need some distributional assumptions on the random censoring interval (U, V). We assume that U is exponentially distributed with mean λ_1^{-1} and $Y = V - U$ is exponentially distributed with mean λ_2^{-1} and that T, U and Y are mutually independent. Now the probability density function of T, conditional on the event that $T \notin (U, V)$ and $\boldsymbol{z}$ is given by

$$\begin{aligned} f_{T|(T\notin(U,V),\boldsymbol{z})}(t) &= \lim_{\Delta t\to 0}\frac{1}{\Delta t}P(t \le T < t+\Delta t|T \notin (U,V),\boldsymbol{z}) \\ &= \lim_{\Delta t\to 0}\frac{1}{\Delta t}\frac{P(t \le T < t+\Delta t, T \notin (U,V)|\boldsymbol{z})}{P(T \notin (U,V)|\boldsymbol{z})} \\ &= \lim_{\Delta t\to 0}\frac{1}{\Delta t}\frac{1}{\acute{p}(\boldsymbol{\psi})}P(t \le T < t+\Delta t|\boldsymbol{z})P(T \notin (U,V)|t \le T < t+\Delta t,\boldsymbol{z}) \\ &= \frac{1}{\acute{p}(\boldsymbol{\psi})}f(t|\boldsymbol{z})P(T \notin (U,V)|T=t,\boldsymbol{z}). \end{aligned}$$

where $\acute{p}(\boldsymbol{\psi}) = P\left(T \notin (U,V)|\boldsymbol{z}\right)$. Thus, for $\lambda_1 \neq \lambda_2$, we reach at

$$f_{T|(T\notin(U,V),\boldsymbol{z})}(t) = \frac{1}{\acute{p}(\boldsymbol{\psi})}\big(\big(h_0(t) + \boldsymbol{z}^\top\boldsymbol{\theta}\big)\exp\big(- H_0(t) - \boldsymbol{z}^\top\boldsymbol{\theta}t\big)\big)$$
$$\big(1 - \frac{\lambda_1}{\lambda_2-\lambda_1}e^{-\lambda_2 t}\big(e^{-(\lambda_1-\lambda_2)t} - 1\big)\big), \tag{4.18}$$

and for $\lambda_1 = \lambda_2 = \lambda$ (say),

$$f_{T|(T\notin(U,V),\boldsymbol{z})}(t) = \frac{1}{\acute{p}(\boldsymbol{\psi})}\big(\big(h_0(t) + \boldsymbol{z}^\top\boldsymbol{\theta}\big)\exp\big(-H_0(t) - \boldsymbol{z}^\top\boldsymbol{\theta}t\big)\big)(1-\lambda te^{-\lambda t}). \quad (4.19)$$

In a similar way we obtain the probability density function of U, conditional on the event that $T \in (U,V)$ and $\boldsymbol{z}$ as

$$\begin{aligned} f_{U|(T\in(U,V),\boldsymbol{z})}(u) =& \frac{1}{1-\acute{p}(\boldsymbol{\psi})}\lambda_1\exp(-\lambda_1 u)\int_{y=0}^{\infty}\Big(\exp\big(-H_0(u) - \boldsymbol{z}^\top\boldsymbol{\theta}u\big) - \\ & \exp\big(-H_0(u+y) - \boldsymbol{z}^\top\boldsymbol{\theta}(u+y)\big)\Big)\lambda_2\exp(-\lambda_2 y)\mathrm{d}y, \end{aligned} \quad (4.20)$$

and the probability density function of Y, conditional on the event that $T \in (U,V)$ and $\boldsymbol{z}$ is

$$\begin{aligned} f_{Y|(T\in(U,V),\boldsymbol{z})}(y) =& \frac{1}{1-\acute{p}(\boldsymbol{\psi})}\lambda_2\exp(-\lambda_2 y)\int_{u=0}^{\infty}\Big(\exp\big(-H_0(u) - \boldsymbol{z}^\top\boldsymbol{\theta}u\big) - \\ & \exp\big(-H_0(u+y) - \boldsymbol{z}^\top\boldsymbol{\theta}(u+y)\big)\Big)\lambda_1\exp(-\lambda_1 u)\mathrm{d}u. \end{aligned} \quad (4.21)$$

Further, the joint distribution of (U,V) is obtained as

$$g_0(u,v) = \lambda_1\lambda_2 e^{-(\lambda_2-\lambda_1)u}e^{-\lambda_2 v}, \quad u,v \in \mathbb{R}^+, u < v,$$

and using this we obtain the distribution of (U,V), conditional on $T \in (U,V)$ and $\boldsymbol{z}$ as

$$f_{UV|(T\in(U,V),\boldsymbol{z})}(u,v) = \frac{1}{1-\acute{p}(\boldsymbol{\psi})}g_0((u,v))P\left(T\in(U,V)|(U,V)=(u,v),\boldsymbol{z}\right),$$

which on simplification becomes

$$f_{UV|(T\in(U,V),\boldsymbol{z})}(u,v) = \frac{1}{1-\acute{p}(\boldsymbol{\psi})}\lambda_1\lambda_2 e^{-(\lambda_2-\lambda_1)u}e^{-\lambda_2 v}\{S\left(u|\boldsymbol{z}\right) - S\left(v|\boldsymbol{z}\right)\}. \tag{4.22}$$

The conditional density functions given in (4.18) - (4.22) can be used to evaluate the expected values of the second order partial derivatives given in (4.12) - (4.17), and hence we can obtain the Fisher information matrix $\mathscr{I}(\boldsymbol{\psi})$. We now assume that the covariate space is bounded. Then the likelihood function (4.8) satisfies the standard regularity conditions (see Bain (1976)), and it follows that the asymptotic distribution of $\hat{\boldsymbol{\psi}}$ is $(p+2)$-variate normal with mean vector $\boldsymbol{\psi}$ and dispersion matrix $\mathscr{I}^{-1}(\boldsymbol{\psi})$.

Thus, $\sqrt{n}\big(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}\big)^\top \to N_{p+2}\big(\mathbf{0}, n\mathscr{I}^{-1}(\boldsymbol{\psi})\big)$ in distribution.

4.4 Simulation Studies

Simulation studies are carried out to assess the finite sample performance of the estimator. We assume that the baseline distribution of the lifetime variable T is exponentiated exponential with shape parameter α and scale parameter λ. We consider a single covariate, say z, in the present study which is generated from uniform distribution over $[0, 10]$. We generate observations on T, conditional on z as follows. First we fix values for α, λ, θ and z. We generate lifetimes using (4.5) for given values of α, λ, θ and z. A random interval is generated with (U, V) where U and $Y = V - U$ are assumed to be independent exponential variates with means $\lambda_1^{-1} = 15$ and $\lambda_2^{-1} = 10$ respectively. If we find $T \notin (U, V)$ then T is selected as the sample observation, otherwise we choose the interval (U, V) as the observation.

In the present study, we consider two sample sizes viz., $n = 50$ and $n = 100$. We consider three different censoring percentages viz., mild (10% censoring), moderate (20% censoring) and heavy (30% censoring) for analyzing the impact of censoring. Further, different combinations of parameters are also considered. For any value of scale parameter λ, the hazard rate given in (4.3) is increasing for $\alpha > 1$, decreasing for $\alpha < 1$, both up to λ, and is a constant ($= \lambda$) for $\alpha = 1$. These three cases are considered for our simulation study. We compute the MLE of parameters using Algorithm 4.1, where the initial values of the parameters are set in a similar manner as in Chapter 2. The average absolute bias (Bias) and estimated mean squared error (MSE) of the estimate of θ along with coverage probabilities (CP) are computed using 1000 iterations. The results are reported in Tables 4.1 - 4.3. It is evident that as the sample size n increases, the bias and MSE decrease. When censoring percentage increases the bias and MSE increase and CP decreases. Figure 4.1 shows

TABLE 4.1: Bias, MSE and CP of the estimator of θ under mild censoring.

			$n = 50$			$n = 100$		
α	λ	θ	Bias	MSE	CP	Bias	MSE	CP
0.75	4	0.50	0.0079	0.0013	0.931	0.0015	0.0062	0.947
	10	-0.10	0.0092	0.0011	0.931	0.0067	0.0004	0.953
	14	-0.75	0.0098	0.0027	0.931	0.0041	0.0006	0.947
0.25	1.5	1	0.0122	0.0018	0.934	0.0063	0.0005	0.952
	7	0.75	0.0088	0.0018	0.928	0.0037	0.0007	0.943
	12	- 1.00	0.0086	0.0093	0.927	0.0026	0.0008	0.935
1.5	0.5	0.50	0.0069	0.0081	0.940	0.0035	0.0060	0.943
	1	-0.75	0.0179	0.0022	0.910	0.0105	0.0005	0.928
	2.5	-1.00	0.0178	0.0102	0.918	0.0046	0.0020	0.950
7	0.25	-0.25	0.0077	0.0060	0.917	0.0025	0.0006	0.924
	1	1.00	0.0031	0.0036	0.920	0.0009	0.0015	0.932
	3	-0.75	0.0069	0.0103	0.909	0.0024	0.0031	0.951
1.0	0.2	0.75	0.0206	0.0016	0.930	0.0071	0.0004	0.948
	1	-0.50	0.0265	0.0174	0.896	0.0036	0.0011	0.944
	4	1.00	0.0453	0.0127	0.950	0.0214	0.0071	0.955

TABLE 4.2: Bias, MSE and CP of the estimator of θ under moderate censoring.

			$n = 50$			$n = 100$		
α	λ	θ	Bias	MSE	CP	Bias	MSE	CP
0.75	4	0.50	0.0196	0.0099	0.912	0.0053	0.0084	0.924
	10	-0.10	0.0152	0.0048	0.918	0.0082	0.0011	0.941
	14	-0.75	0.0107	0.0028	0.932	0.0078	0.0019	0.941
0.25	1.5	1.00	0.0195	0.0064	0.911	0.0082	0.0009	0.929
	7	0.75	0.0088	0.0066	0.910	0.0064	0.0009	0.940
	12	-1.00	0.0087	0.0244	0.904	0.0048	0.0078	0.931
1.5	0.5	0.50	0.0107	0.0189	0.901	0.0051	0.0102	0.921
	1	-0.75	0.0264	0.0029	0.900	0.0132	0.0021	0.922
	2.5	-1.00	0.0222	0.0168	0.909	0.0129	0.0067	0.929
7	0.25	-0.25	0.0111	0.0084	0.900	0.0085	0.0009	0.908
	1	1.00	0.0247	0.0051	0.890	0.0010	0.0020	0.915
	3	-0.75	0.0138	0.0107	0.893	0.0077	0.0062	0.933
1.0	0.2	0.75	0.0249	0.0019	0.930	0.0124	0.0009	0.930
	1	-0.50	0.0373	0.0192	0.891	0.0064	0.0027	0.925
	4	1.00	0.0475	0.0193	0.908	0.0275	0.0092	0.939

TABLE 4.3: Bias, MSE and CP of the estimator of θ under heavy censoring.

			$n = 50$			$n = 100$		
α	λ	θ	Bias	MSE	CP	Bias	MSE	CP
0.75	4	0.50	0.0205	0.0250	0.894	0.0152	0.0178	0.919
	10	-0.10	0.0382	0.0074	0.888	0.0213	0.0019	0.913
	14	-0.75	0.0463	0.0051	0.900	0.0178	0.0028	0.927
0.25	1.5	1.00	0.0202	0.0277	0.881	0.0171	0.0139	0.900
	7	0.75	0.0126	0.0109	0.900	0.0087	0.0029	0.928
	12	-1.00	0.0281	0.0350	0.890	0.0150	0.0218	0.903
1.5	0.5	0.50	0.0291	0.0205	0.887	0.0169	0.0164	0.894
	1	-0.75	0.0297	0.0259	0.881	0.0168	0.0147	0.897
	2.5	-1.00	0.0270	0.0183	0.890	0.0158	0.0119	0.905
7	0.25	-0.25	0.0197	0.0145	0.883	0.0111	0.0054	0.892
	1	1.00	0.0292	0.0268	0.880	0.0219	0.0149	0.891
	3	-0.75	0.0268	0.0233	0.884	0.0199	0.0163	0.904
1.0	0.2	0.75	0.0487	0.0281	0.887	0.0172	0.0129	0.917
	1	-0.50	0.0478	0.0295	0.888	0.0287	0.0029	0.904
	4	1.00	0.0632	0.0396	0.882	0.0341	0.0147	0.915

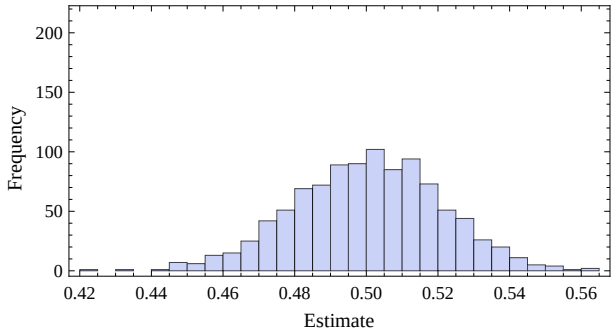


FIGURE 4.1: Histogram of estimates obtained for $\theta = 0.50$.

a histogram of the estimates of $\theta = 0.50$ obtained for 1000 iterations with moderate censoring. It is evident that the estimator is roughly normally distributed with mean 0.50, giving evidence in favour of asymptotic normality of the estimator $\hat{\theta}$. The other values of θ also show similar patterns.

4.5 Data Analysis

In this section, we apply our model to a real life data set. We consider the data corresponding to larynx cancer patients studied by Karduan (1983). The data set is described in Klein and Moeschberger (2005) and consists of time in years from first treatment until death or end of study. Three covariates are recorded namely patient's age, year of diagnosis and the stage of patient's cancer, which is grouped into classes 1, 2, 3 and 4 according as early stage to worst stage. Following Jammalamadaka and Mangalam (2003), and Iyer et al. (2008), we first selected all 50 exact lifetimes and fitted these uncensored lifetimes with the EE distribution. The Kolmogorov-Smirnov goodness of fit statistic shows a P-value of 0.35. For the same data set, we also considered a Weibull model with shape parameter a and scale

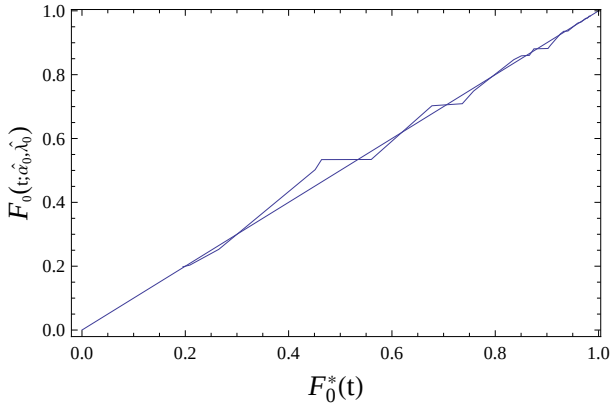


FIGURE 4.2: Plot of $F_0(t;\hat{\alpha}_0,\hat{\lambda}_0)$ against $\breve{F}_0(t)$.

parameter b for a comparative study under the model (4.1). The Weibull model assumption gives the P-value of 0.41 under the Kolmogorov-Smirnov goodness of fit test. Therefore both these distribution assumptions cannot be ruled out. The estimates of the shape and scale parameters of these distributions while fitting are later used as their respective initial values in Algorithm 4.1. Motivated by the method adopted by Jammalamadaka and Mangalam (2003), and Iyer et al. (2008), we then artificially middle-censored 20% of the lifetimes using the method described in previous section with $\lambda_1^{-1} = 3$ and $\lambda_2^{-1} = 1.5$. This resulted in 10 censored intervals. This new data set consisting of exact lifetimes as well as censored intervals is further checked for its suitability under the EE model assumption as described below. First we consider a likelihood function at baseline level, given by $L_0(\alpha,\lambda) = \prod_{i=1}^{n_1} f_0(t_i) \prod_{i=n_1+1}^{n} (S_0(u_i) - S_0(v_i))$, where $f_0(t)$ is the probability density function corresponding to $F_0(t)$. The distribution parameters α and λ are estimated via the EM algorithm, and denote the estimators so obtained as $\hat{\alpha}_0$ and $\hat{\lambda}_0$. Next we obtain the Nelson-Aalen-type estimator of the baseline cumulative hazard function under middle-censoring setup, as we described in Chapter 2. Let $\breve{F}_0(t)$ be the distribution function corresponding to the Nelson-Aalen-type estimator thus

TABLE 4.4: Estimates of parameters and their SE.

	EE		Weibull	
	Estimate	SE	Estimate	SE
θ_1	0.084	0.0108	0.071	0.0190
θ_2	0.106	0.0541	0.122	0.0607
shape	0.95	0.2003	1.41	0.6530
scale	0.46	0.0863	2.82	0.7721

obtained. Now a PP-plot is drawn with the distribution function of EE distribution with parameters as $\hat{\alpha}_0$ and $\hat{\lambda}_0$, say $F_0(t; \hat{\alpha}_0, \hat{\lambda}_0)$ along the Y-axis and the estimator $\breve{F}_0(t)$ along the X-axis. If the model assumption holds, then a straight line passing through origin making an angle 45 degrees with the X-axis is expected. Figure 4.2 shows the graph thus obtained for the censored data set considered. The graph seems to be close to the line, which supports the distribution assumption.

We consider two covariates viz., patient's age (z_1) and the disease stage (z_2) for our illustration. Let θ_1 and θ_2 respectively be their unknown regression parameters. We apply the model given in Section 4.2 for both EE and Weibull distributions. The estimates of shape, scale and regression parameters are found out using Algorithm 4.1 with initial values of parameters selected in a similar way as in Chapter 2. These estimates are reported in Table 4.4. Moreover, the standard errors (SE) of these estimates are computed using bootstrap method as discussed in Chapter 2 with $B = 1000$, where the lifetimes T_i^* are generated from the fitted value of the survival function given in (4.5). These are also given in Table 4.4. It can be observed that the covariates have adverse effects on the lifetime under both model assumptions and SE's are small.

We now check the overall fit of the model by using Cox-Snell residuals (Cox and Snell (1968)). The method is exactly the same as that defined in Chapter

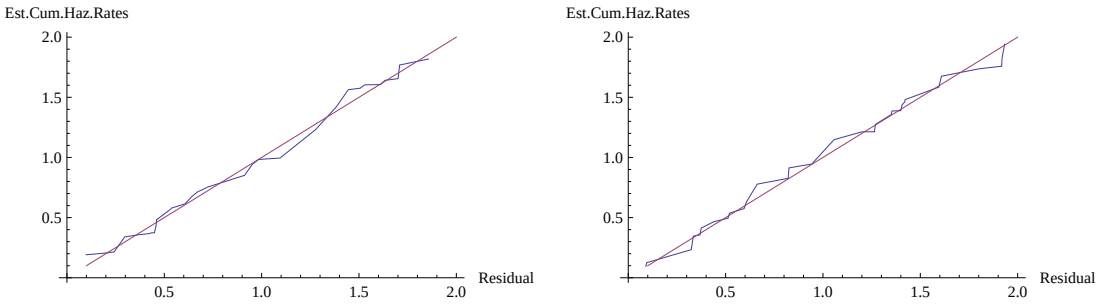


FIGURE 4.3: Plots of estimated cumulative hazard rates against r_j's for (i) EE distribution, (ii) Weibull distribution.

2, except that under model (4.1), the Cox-Snell residuals are defined to be the fitted cumulative hazard function values given by $r_j^{(0)} = \hat{H}_0(t_j) + \boldsymbol{z}_j^\top \hat{\boldsymbol{\theta}}\, t_j$ for $j = 1, 2, ..., n_1$, and $r_j^{(1)} = \hat{H}_0(u_j) + \boldsymbol{z}_j^\top \hat{\boldsymbol{\theta}}\, u_j$ and $r_j^{(2)} = \hat{H}_0(v_j) + \boldsymbol{z}_j^\top \hat{\boldsymbol{\theta}}\, v_j$ for $j = n_1+1, ..., n$. Figure 4.3 shows the residuals plots obtained for the two distributions. The curves are close to the straight line indicating that data fits both the models reasonably well. It can be observed that both these model assumptions are justifiable. We now employ the Akaike information criterion (Akaike (1974)), for identifying the model that fits the data better. We obtain Akaike information measures as 235 and 287 respectively for EE distribution and Weibull distribution. Thus we conclude that the EE distribution provides a better model for the data considered. We also carried out the likelihood ratio test to test the hypothesis $H_0 : \boldsymbol{\theta} = \mathbf{0}$ to assess whether the covariate effect is significant or not. The P-value is obtained to be 0.018 and we infer that the covariate effect is significant.

4.6 Conclusion

In this chapter, we considered the parametric additive hazards regression problem for the middle-censored survival data. We developed inference procedures for finding the MLE of the parameters and for testing the significance of the regression parameters in Section 4.2. In Section 4.3, we established asymptotic normality of the proposed estimator, and derived conditional probability density functions required for the computation of Fisher information matrix. We then carried out extensive simulation studies in Section 4.4, which indicated that the estimators perform satisfactorily. We presented an illustration of the proposed method with a real life data in Section 4.5.

Chapter 5

Semiparametric Additive Hazards Regression Model

5.1 Introduction

In Chapter 4, we assumed that the lifetime follows a parametric model viz. exponentiated exponential distribution. We then estimated the parameters of the model using the method of maximum likelihood. However, in practice, the exact form of the underlying lifetime distribution is usually unknown and we may not be able to find an appropriate model, as we discussed in Chapter 3. Thus, the parametric methods in identifying significant prognostic factors may not be adequate in practice. In this chapter, we focus on the semiparametric additive hazards regression model for the analysis of middle-censored survival data. The model is given by

$$h(t|\boldsymbol{z}) = h_0(t) + \boldsymbol{z}^\top \boldsymbol{\theta}, \tag{5.1}$$

where $h_0(t)$ is an arbitrary unspecified baseline hazard function.

For a comprehensive review on properties and inference procedures of model (5.1) under right censoring, one may refer to Aranda-Ordaz (1983), Cox and Oakes (1984),

[1]The results in this chapter are accepted for publication in *Statistics in Transition* (see Sankaran and Prasad (2017b))

Thomas (1986), Breslow and Day (1980), and Lin and Ying (1994). The semiparametric treatment of the model (5.1) under right censoring is available in Aalen (1989).

In the present work, we consider the model (5.1) with an unknown baseline survival function $S_0(t)$ of a continuous type lifetime variate T, which is subject to middle-censoring. We aim at estimating the regression parameters and baseline survival function under model (5.1).

The rest of the chapter is organized as follows. We, in Section 5.2, propose two different methods of estimation of parameters of the model (5.1). The asymptotic properties of the estimators are also presented. Simulation studies to assess the performance of these estimators under both methods are carried out and the results are compared in Section 5.3. The utility of the methods are illustrated in Section 5.4. Finally, Section 5.5 provides important conclusions of the study.

5.2 Inference Procedure

Let the lifetime variate T admit an absolutely continuous distribution function $F_0(t)$. Assume that T is is middle-censored by the random censoring interval (U, V) having absolutely continuous distribution function given by $G_0(\cdot, \cdot)$. We further assume that under model (5.1), T is independent of (U, V), given the covariate $\boldsymbol{z}$, and that the censoring mechanism is independent of covariates. Thus we observe the vector $(X, \delta, \boldsymbol{z})$, where

$$X = \begin{cases} T & \text{if } \delta = 1 \\ (U, V) & \text{if } \delta = 0, \end{cases}$$

and $\delta = I(X = T)$ is the censoring indicator. We now state an important assumption regarding the identifiability of $F_0(t)$. Let $[a, b]$, $a \leq b$ be any arbitrary interval in the support of T. Define, for each $r \in [a, b]$,

$$A_0(r) = G_0(r-, \infty) - G_0(r-, r) = P(U < r < V). \tag{5.2}$$

Now consider a situation where $A_0(r) = 1$ for $r \in [a, b]$ for which $F_0(b) > F_0(a-)$. That is, censoring occurs with probability one on this interval where $F_0(t)$ has a positive mass. Consequently there will not be any exact observation in this interval, making it impossible to distinguish between two distributions which are identical outside $[a, b]$ but differing only within $[a, b]$. To overcome this issue, we make the following assumption.

A1: The probability defined in (5.2) is strictly less than one.

In the following, we describe two different estimation methods; one make use of the classic martingale theory and the other by using an iterative method.

5.2.1 Martingale Method

We provide an inference procedure to estimate the baseline cumulative hazard function and the regression parameter by mimicking the inference based on partial likelihood method for the proportional hazards model (see Kalbfleisch and Prentice (2011)). We assume that there are n individuals under investigation. The observed data consists of n independent and identically distributed replicates $(X_i, \delta_i, \boldsymbol{z}_i)$ of $(X, \delta, \boldsymbol{z}), 1 \leq i \leq n$. In Chapter 2 we defined the counting process $\{N_i(t);\, t \geq 0\}$ and the at-risk process $\{R_i(t);\, t \geq 0\}$ for the middle-censored

data as $N_i(t) = I(\bar{X}_i \leq t, \delta_i = 1)$ and $R_i(t) = I(\bar{X}_i \geq t)$, where $\bar{X} = T$ if $\delta = 1$, and $\bar{X} = U$ if $\delta = 0$ for $i = 1, 2, ..., n$. For these processes we denote the filtration $\sigma\{N_i(u), R_i(u+), \boldsymbol{z}_i : i = 1, 2, ..., n;\ 0 \leq u \leq t\}$ by $\mathcal{F}_t$. Under model (5.1), the cumulative hazard function for the i'th individual given $\boldsymbol{z}_i$ is given by $H(t|\boldsymbol{z}_i) = H_0(t) + \boldsymbol{z}_i^\top \boldsymbol{\theta} t$, where $H_0(t) = \int_0^t h_0(a)\mathrm{d}a$ is the baseline cumulative hazard function. The model (5.1) assumes that

$$E[N_i(t)|\mathcal{F}_{t-}] = (h_0(t) + \boldsymbol{\theta}^\top \boldsymbol{z}_i)R_i(t)dt,$$

approximately and the intensity function corresponding to the counting process $N_i(t)$ can thus be written as $R_i(t)\mathrm{d}H(t|\boldsymbol{z}_i) = R_i(t)\{dH_0(t) + \boldsymbol{z}_i^\top \boldsymbol{\theta}\mathrm{d}t\}$. With this, the counting process can be uniquely decomposed, so that for every $i = 1, 2, ..., n$, and t,

$$N_i(t) = M_i(t) + \int_0^t R_i(a)\,\mathrm{d}H(a|\boldsymbol{z}_i), \tag{5.3}$$

where $M_i(\cdot)$ is a local square integrable martingale (Andersen and Gill (1982)) approximately. From (5.3), we have the following relation approximately for $i = 1, 2, ..., n$,

$$dN_i(t) = dM_i(t) + R_i(t)\mathrm{d}H(t|\boldsymbol{z}_i). \tag{5.4}$$

Therefore we have

$$\sum_{i=1}^{n} dM_i(t) = \sum_{i=1}^{n}[dN_i(t) - R_i(t)(dH_0(t) + \boldsymbol{\theta}^\top \boldsymbol{z}_i dt)] = 0, \tag{5.5}$$

approximately. Thus from (5.3), a Breslow type estimator (Breslow (1972)) for the cumulative hazard function $H_0(t)$ is obtained as

$$\hat{H}_0(t,\hat{\boldsymbol{\theta}}) = \int_0^t \frac{\sum_{i=1}^n \{\mathrm{d}N_i(a) - R_i(a)\boldsymbol{z}_i^\top \hat{\boldsymbol{\theta}}\mathrm{d}a\}}{\sum_{i=1}^n R_i(a)}, \tag{5.6}$$

where $\hat{\boldsymbol{\theta}}$ is a consistent estimator of $\boldsymbol{\theta}$. Motivated by Lin and Ying (1994), we propose the following estimating equation, which mimics the partial likelihood score function in the case of proportional hazards model discussed in Cox (1975).

$$\mathscr{U}(\boldsymbol{\theta}) = \sum_{i=1}^n \int_0^\infty \boldsymbol{z}_i\{\mathrm{d}N_i(t) - R_i(t)d\hat{H}_0(\boldsymbol{\theta},t) - R_i(t)\boldsymbol{z}_i^\top\boldsymbol{\theta}\mathrm{d}t\}. \tag{5.7}$$

Using (5.6) the middle term on the right hand side of (5.7) can be written as

$$\begin{aligned}\sum_{i=1}^n \int_0^\infty \boldsymbol{z}_i R_i(t)d\hat{H}_0(\boldsymbol{\theta},t) = &\sum_{i=1}^n \int_0^\infty \boldsymbol{z}_i R_i(t)(\sum_{j=1}^n R_j(t))^{-1}\sum_{k=1}^n \mathrm{d}N_k(t) - \\ &\sum_{i=1}^n \int_0^\infty \boldsymbol{z}_i R_i(t)(\sum_{j=1}^n R_j(t))^{-1}\sum_{k=1}^n R_k(t)\boldsymbol{z}_k^\top\boldsymbol{\theta}\mathrm{d}t. \end{aligned}\tag{5.8}$$

On interchanging the summation and integration operations on the right hand side of (5.8), we obtain

$$\begin{aligned}\sum_{i=1}^n \int_0^\infty \boldsymbol{z}_i R_i(t)d\hat{H}_0(\boldsymbol{\theta},t) &= \int_0^\infty \bar{\boldsymbol{z}}\sum_{k=1}^n \mathrm{d}N_k(t) - \int_0^\infty \bar{\boldsymbol{z}}\sum_{k=1}^n R_k(t)\boldsymbol{z}_k^\top\boldsymbol{\theta}\mathrm{d}t \\ &= \sum_{i=1}^n \int_0^\infty \bar{\boldsymbol{z}}\mathrm{d}N_i(t) - \sum_{i=1}^n \int_0^\infty \bar{\boldsymbol{z}}R_i(t)\boldsymbol{z}_i^\top\boldsymbol{\theta}\mathrm{d}t, \end{aligned}\tag{5.9}$$

where $\bar{\boldsymbol{z}} = \sum_{i=1}^{n} \boldsymbol{z}_i R_i(t) / \sum_{i=1}^{n} R_i(t)$ with the convention that $\frac{0}{0} = 0$. Using (5.9), the score function in (5.7) takes the form

$$\mathscr{U}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \int_0^{\infty} \boldsymbol{z}_i \mathrm{d}N_i(t) - \sum_{i=1}^{n} \int_0^{\infty} \bar{\boldsymbol{z}} \mathrm{d}N_i(t) + \sum_{i=1}^{n} \int_0^{\infty} \bar{\boldsymbol{z}} R_i(t) \boldsymbol{z}_i^{\top} \boldsymbol{\theta} \mathrm{d}t - \sum_{i=1}^{n} \int_0^{\infty} \boldsymbol{z}_i R_i(t) \boldsymbol{z}_i^{\top} \boldsymbol{\theta} \mathrm{d}t. \tag{5.10}$$

On rearrangement of terms of (5.10), we get

$$\mathscr{U}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \int_0^{\infty} \{\boldsymbol{z}_i - \bar{\boldsymbol{z}}\} \{\mathrm{d}N_i(t) - R_i(t) \boldsymbol{z}_i^{\top} \boldsymbol{\theta} \mathrm{d}t\}. \tag{5.11}$$

The identity (5.11) is based on a simple fact that when $\boldsymbol{\theta}_0$ is the true parameter value, $\mathscr{U}(\boldsymbol{\theta}_0)$ has mean approximately zero. Note that (5.11) is linear in $\boldsymbol{\theta}$ and the resulting estimator takes an explicit form given by

$$\hat{\boldsymbol{\theta}} = \Big[\sum_{i=1}^{n} \int_0^{\infty} [\boldsymbol{z}_i - \bar{\boldsymbol{z}}]^{\otimes 2} R_i(t) \mathrm{d}t \Big]^{-1} \sum_{i=1}^{n} \int_0^{\infty} [\boldsymbol{z}_i - \bar{\boldsymbol{z}}] \mathrm{d}N_i(t), \tag{5.12}$$

where $\boldsymbol{a}^{\otimes 2} = \boldsymbol{a}\boldsymbol{a}^{\top}$. This naturally leads to the following estimator of the survival function $S(t|\boldsymbol{z})$,

$$\hat{S}(t|\boldsymbol{z}) = \exp\{-\hat{H}_0(t, \hat{\boldsymbol{\theta}}) - \boldsymbol{z}^{\top} \hat{\boldsymbol{\theta}} t\}. \tag{5.13}$$

To prove asymptotic properties of $\hat{\boldsymbol{\theta}}$, an algebraic manipulation of (5.4) yields

$$\mathscr{U}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \int_0^{\infty} (\boldsymbol{z}_i - \bar{\boldsymbol{z}}) dM_i(t). \tag{5.14}$$

It follows from standard counting process theory (Andersen and Gill (1982)) that $n^{-1/2}\mathscr{U}(\boldsymbol{\theta}_0)$ converges weakly to p-variate normal distribution with mean zero and

a covariance matrix which can be estimated consistently by

$$A_0 = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\infty}(\boldsymbol{z}_i - \bar{\boldsymbol{z}})^{\otimes 2} dN_i(t). \tag{5.15}$$

Thus the random vector $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ converges weakly to p-variate normal variate with mean zero and a covariance matrix which can be consistently estimated by $B_0^{-1}A_0B_0^{-1}$, where

$$B_0 = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\infty} R_i(t)(\boldsymbol{z}_i - \bar{\boldsymbol{z}})^{\otimes 2} dt. \tag{5.16}$$

Specifically, $(B_0^{-1}A_0B_0^{-1})^{-\frac{1}{2}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ converges in distribution to $N(0, \mathcal{I}_p)$, where $\mathcal{I}_p$ is the identity matrix of order p. It can be observed that neither A_0 nor B_0 involves the regression parameters. The estimator (5.6) provides the basis for estimating survival probabilities. Using standard counting process techniques, it follows that the process $\sqrt{n}(\hat{H}_0(t, \hat{\boldsymbol{\theta}}) - H_0(t))$ converges weakly to a zero-mean Gaussian process whose covariance function at $(t, s), t \geq s$ can be estimated consistently by

$$\int_0^s \frac{n\sum_{i=1}^n dN_i(a)}{(\sum_1^n R_i(a))^2} + C_0'(t)B_0^{-1}A_0B_0^{-1}C_0(s) - C_0'(t)B_0^{-1}D_0(s) - C_0'(s)B_0^{-1}D_0(t),$$

where $C_0(t) = \bar{\boldsymbol{z}}t$, $D_0(t) = \int_0^t \frac{\sum_1^n(\boldsymbol{z}_i - \bar{\boldsymbol{z}})dN_i(a)}{\sum_1^n R_i(a)}$, and for any given function $k(a)$, $k'(a)$ represents $\frac{\mathrm{d}k(a)}{\mathrm{d}a}$.

Using functional delta method (Andersen et al. (2012)), it follows that the process $\sqrt{n}(\hat{S}(t|\boldsymbol{z}) - S(t|\boldsymbol{z}))$ converges weakly to a zero-mean Gaussian process whose covariance function at $(t, s), t \geq s$ can be estimated consistently by

$$\hat{S}(t|\boldsymbol{z})\hat{S}(s|\boldsymbol{z})\Big(\int_0^s \frac{n\sum_{i=1}^n dN_i(a)}{(\sum_1^n R_i(a))^2} + W_0'(t, \boldsymbol{z})B_0^{-1}A_0B_0^{-1}W_0(s, \boldsymbol{z}) + W_0'(t, \boldsymbol{z})B_0^{-1}D_0(s) + W_0'(s, \boldsymbol{z})B_0^{-1}D_0(t)\Big),$$

where $W_0(t, \boldsymbol{z}) = (\boldsymbol{z} - \bar{\boldsymbol{z}})t$.

5.2.2 The Iterative Method

In this section, an iterative method is proposed for estimating the unknown baseline survival function $S_0(t)$ of the lifetime variate T and the regression parameter $\boldsymbol{\theta}$ under model (5.1). Let the observed data be as before. For convenience, we arrange the observations in such a way that the first n_1 observations are exact lifetimes, and the remaining n_2 are censored intervals, with $n_1 + n_2 = n$. Now the likelihood function corresponding to the observed data, excluding the normalizing constant, can be written as

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n_1} f(t_i|\boldsymbol{z}_i) \cdot \prod_{i=n_1+1}^{n} \left(S(u_i|\boldsymbol{z}_i) - S(v_i|\boldsymbol{z}_i)\right). \tag{5.17}$$

Under the model assumption given in (5.1), the survival function of T given $\boldsymbol{z}$ is obtained as

$$S(t|z) = S_0(t)\exp(-\boldsymbol{\theta}^\top \boldsymbol{z}t), \tag{5.18}$$

where $S_0(t) = \exp(-H_0(t))$. Thus the probability density function of T given $\boldsymbol{z}$ is given by

$$f(t|\boldsymbol{z}) = \exp(-\boldsymbol{\theta}^\top \boldsymbol{z}t)\big(\boldsymbol{\theta}^\top \boldsymbol{z}S_0(t) - S_0'(t)\big). \tag{5.19}$$

Therefore (5.17) becomes

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n_1} \exp(-\boldsymbol{\theta}^\top \boldsymbol{z}_i t_i)\big(\boldsymbol{\theta}^\top \boldsymbol{z}_i S_0(t_i) - S_0'(t_i)\big) \prod_{i=n_1+1}^{n} \big(S_0(u_i)\exp(-\boldsymbol{\theta}^\top \boldsymbol{z}_i u_i) - S_0(v_i)\exp(-\boldsymbol{\theta}^\top \boldsymbol{z}_i v_i)\big)\,. \tag{5.20}$$

The log-likelihood function corresponding to (5.20) is given by

$$l(\boldsymbol{\theta}) = \sum_{i=1}^{n_1} \big(-\boldsymbol{\theta}^\top \boldsymbol{z}_i t_i + \log(\boldsymbol{\theta}^\top \boldsymbol{z}_i S_0(t_i) - S_0'(t_i))\big) + \sum_{i=n_1+1}^{n} \log\big(S_0(u_i)\exp(-\boldsymbol{\theta}^\top \boldsymbol{z}_i u_i) - S_0(v_i)\exp(-\boldsymbol{\theta}^\top \boldsymbol{z}_i v_i)\big). \tag{5.21}$$

To find the SPMLE of $\boldsymbol{\theta}$, we need to find out the derivative of (5.21) with respect to $\boldsymbol{\theta}$ and equate it with null vector. We carry out this process component-wise, keeping $S_0(t)$ fixed in the forthcoming maximization algorithm. The derivative of (5.21) with respect to θ_r, for $r = 1, 2, ..., p$ is given by

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \theta_r} = \sum_{i=1}^{n_1} z_{ir}(t_i + (\boldsymbol{\theta}^\top \boldsymbol{z}_i S_0(t_i) - S_0'(t_i))^{-1} S_0(t_i)) + \sum_{i=n_1+1}^{n} z_{ir}\Big(S_0(u_i)\exp(-\boldsymbol{\theta}^\top \boldsymbol{z}_i u_i) - S_0(v_i)\exp(-\boldsymbol{\theta}^\top \boldsymbol{z}_i v_i)\Big)^{-1} \Big(v_i S_0(v_i)\exp(-\boldsymbol{\theta}^\top \boldsymbol{z}_i v_i) - u_i S_0(u_i)\exp(-\boldsymbol{\theta}^\top \boldsymbol{z}_i u_i)\Big), \tag{5.22}$$

where z_{ir} is the r'th component of the covariate vector $\boldsymbol{z_i}$. Note that (5.22) involves both unknown quantities $\boldsymbol{\theta}$ and $S_0(t)$. Thus an explicit solution for $\boldsymbol{\theta}$ can not be obtained directly from it. We provide an iterative algorithm to estimate the MLE of these two quantities, where at each iteration a better update is obtained. To begin with the algorithm, we consider the NPMLE of the baseline survival function

as an initial approximation to the true baseline survival function.

In the case of middle-censored data, as mentioned in Chapter 3, Jammalamadaka and Mangalam (2003) have shown that the NPMLE of $F_0(t)$ is always a SCE (Tarpey and Flury (1996)) which satisfies

$$\hat{F}_0(t) = \frac{1}{n}\sum_{i=1}^{n}\left\{\delta_i I(X_i \le t) + (1-\delta_i)I(V_i \le t) + (1-\delta_i)I(t \in (U_i, V_i))\frac{\hat{F}_0(t) - \hat{F}_0(U_i)}{\hat{F}_0(V_i-) - \hat{F}_0(U_i)}\right\}. \tag{5.23}$$

We use the cubic polynomial estimate for the baseline survivor function $S_0(t) = 1 - F_0(t)$ using (5.23) as we did in Chapter 3. The algorithm for the estimating $\boldsymbol{\theta}$ and $S_0(t)$ is given below.

Algorithm 5.1

Step 1. Set the vector $\boldsymbol{\theta} = \mathbf{0}$.

Step 2. At the first iteration, find the NPMLE $S_0^{(1)}(t)$ of $S_0(t)$ using (5.23) and substitute this in (5.22) and solve $\partial l(\boldsymbol{\theta})/\partial\theta_r = 0, r = 1, 2, ..., p$ to get the estimator $\boldsymbol{\theta}^{(1)}$ of $\boldsymbol{\theta}$.

Step 3. Find $\tilde{t}_i^{(1)} = S_0^{(1)^{-1}}\left(S_0^{(1)}(t_i)\exp(-\boldsymbol{\theta}^{(1)^\top}\boldsymbol{z}_i t_i)\right)$, $i = 1, ..., n_1$ and similarly find $\tilde{u}_i^{(1)}$ and $\tilde{v}_i^{(1)}$, $i = n_1 + 1, ..., n$ as our updated observations at first iteration.

Step 4. At the j'th iteration ($j > 1$), use $\tilde{t}_i^{(j-1)}, i = 1, 2, ..., n_1$ and $(\tilde{u}_i^{(j-1)}, \tilde{v}_i^{(j-1)}), i = n_1 + 1, ..., n$ as our data points in (5.23) and obtain $S_0^{(j)}(t)$. Substitute $S_0^{(j)}(t)$ in (5.22) and solve $\partial l(\boldsymbol{\theta})/\partial\theta_r = 0, r = 1, 2, ..., p$ to obtain the j'th iterated update $\boldsymbol{\theta}^{(j)}$ of $\boldsymbol{\theta}$.

Step 5. Repeat Steps 3 and 4 until convergence is met, say when $\|\boldsymbol{\theta}^{(m)} - \boldsymbol{\theta}^{(m+1)}\| < 0.0001$ and $\sup_t\left\{\left|S_0^{(m)}(t) - S_0^{(m+1)}(t)\right|\right\} < 0.001$, for some finite positive integer m.

Note that *Step 3* in the algorithm is justified because, if $a_i = S_0^{(1)}(t_i)\exp(-\boldsymbol{\theta}^{(1)^\top}\boldsymbol{z}_i t_i)$, then the a_i 's have a uniform distribution over $[0,1]$, since under the model assumption (5.1), a_i is the survival function of continuous type random variable T given $\boldsymbol{z}_i$. Therefore to scale these back to the baseline distribution we need to find $\tilde{t_i} = \inf\{t : S_0^{(1)}(t) \leq a_i\}$. Thus the correct choice is $\tilde{t_i} = S_0^{(1)^{-1}}(a_i) = S_0^{(1)^{-1}}\left(S_0^{(1)}(t_i)\exp(-\boldsymbol{\theta}^{(1)^\top}\boldsymbol{z}_i t_i)\right)$.

5.2.2.1 Asymptotic Properties

To derive the asymptotic properties, we define the parameter space to be $(\boldsymbol{\Theta}, \Phi)$, where $\boldsymbol{\Theta}$, which is the parameter space for $\boldsymbol{\theta}$, is a bounded subset of $\mathbb{R}_p$, and Φ, which is the parameter space for $S_0(t)$, is defined as the class of all absolutely continuous survival functions. Let us name the estimator obtained for $\boldsymbol{\theta}$ as $\hat{\boldsymbol{\theta}}_{(n)}$ and that for $S_0(t)$ as $\hat{S}_{0(n)}(t)$. Besides the identifiability condition A1, the following conditions are also assumed to hold for establishing the consistency property.

A2: Conditional on $\boldsymbol{z}$, T is independent of (U, V) and the censoring distribution is independent of covariates.

A3: The joint distribution of $(U, V, \boldsymbol{z})$ does not depend on the true parameter value $(\boldsymbol{\theta}^0, S_0^0(t))$.

A4: The covariate space is bounded. That is, there exist some finite $M_0 > 0$ such that $P\{\|\boldsymbol{z}\| \leq M_0\} = 1$, where $\|\cdot\|$ is the usual metric on $\mathbb{R}_p$.

A5: The distribution of $\boldsymbol{z}$ is not concentrated on any proper affine subspace of $\mathbb{R}_p$.

Theorem 5.1. Suppose that $\boldsymbol{\Theta} \in \mathbb{R}_p$ is bounded and assumptions (A1) to (A5) hold. Then the estimator $(\hat{\boldsymbol{\theta}}_{(n)}, \hat{S}_{0(n)}(t))$ is consistent for the true parameter $(\boldsymbol{\theta}^0, S_0^0(t))$

in the sense that if we define a metric $d_0 : \boldsymbol{\Theta} \times \Phi \to \mathbb{R}$ by

$$d_0\big((\boldsymbol{\theta}_1, S_{01}(t)),(\boldsymbol{\theta}_2, S_{02}(t))\big) = \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| + \int |S_{01}(t) - S_{02}(t)| dF_0(t) + \left[\int \left((S_{01}(u) - S_{02}(u))^2 + (S_{01}(v) - S_{02}(v))^2\right) dG_0(u,v)\right]^{\frac{1}{2}}, \quad (5.24)$$

where $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \boldsymbol{\Theta}$ and $S_{01}(t), S_{02}(t) \in \Phi$, then $d_0\left((\hat{\boldsymbol{\theta}}_{(n)}, \hat{S}_{0(n)}(t)), (\boldsymbol{\theta}^0, S_0^0(t))\right) \to 0$ almost surely (a.s.).

Proof. In the following discussion, we denote $D_i = (X_i, \delta_i)$. Let the probability function of D be given by

$$p(d; \boldsymbol{\theta}, S_0(t)) = \prod_{i=1}^{n} f(t_i|\boldsymbol{z}_i)^{\delta_i}[S_0(u_i)\exp(-\boldsymbol{\theta}^\top \boldsymbol{z}_i u_i) - S_0(v_i)\exp(-\boldsymbol{\theta}^\top \boldsymbol{z}_i v_i)]^{1-\delta_i} g_0(u_i, v_i) q(\boldsymbol{z}_i), \quad (5.25)$$

where $g_0(u,v)$ is the joint density of (U,V) and $q(\boldsymbol{z})$ is the density of $\boldsymbol{z}$. Using $(A2)$ and $(A3)$, the log-likelihood function scaled by $1/n$ for the sample $(d_i, \boldsymbol{z}_i), i = 1, 2, ..., n$, up to terms not depending on $(\boldsymbol{\theta}^0, S_0^0(t))$, is

$$l(\boldsymbol{\theta}, S_0(t)) = \frac{1}{n}\sum_{i=1}^{n} \big\{\delta_i \log f(t_i|\boldsymbol{z}_i) + (1-\delta_i)\log\left[S_0(u_i)\exp(-\boldsymbol{\theta}^\top \boldsymbol{z}_i u_i) - S_0(v_i)\exp(-\boldsymbol{\theta}^\top \boldsymbol{z}_i v_i)\right]\big\}. \quad (5.26)$$

We write $p_n(d) = p(d; \hat{\boldsymbol{\theta}}_{(n)}, \hat{S}_{0(n)}(t))$ and $p_0(d) = p(d; \boldsymbol{\theta}^0, S_0^0(t))$, where $(\hat{\boldsymbol{\theta}}_{(n)}, \hat{S}_{0(n)}(t))$ is the MLE that maximizes the likelihood function over $\boldsymbol{\Theta} \times \Phi$ and $(\boldsymbol{\theta}^0, S_0^0(t)) \in \boldsymbol{\Theta} \times \Phi$. Therefore

$$\sum_{i=1}^{n} \log p_n(D_i) \geq \sum_{i=1}^{n} \log p_0(D_i),$$

and hence

$$\sum_{i=1}^{n} \log \frac{p_n(D_i)}{p_0(D_i)} \geq 0.$$

By the concavity of the function $x \mapsto \log x$, for any real number α $(0 < \alpha < 1)$,

$$\frac{1}{n} \sum_{i=1}^{n} \log \left((1-\alpha) + \alpha \frac{p_n(D_i)}{p_0(D_i)} \right) \geq 0. \tag{5.27}$$

The left hand side can be written as

$$\int \log \left((1-\alpha) + \alpha \frac{p_n(D_i)}{p_0(D_i)} \right) \mathrm{d}(\mathbb{P}_n - \mathbb{P})(D) + \int \log \left((1-\alpha) + \alpha \frac{p_n(D_i)}{p_0(D_i)} \right) \mathrm{d}\mathbb{P}(D), \tag{5.28}$$

where $\mathbb{P}_n$ is the empirical measure of D and $\mathbb{P}$ is the joint probability measure of D. Let us assume that the sample space Ω consists of all infinite sequences $\{D_1, D_2, ...\}$, along with the usual sigma field generated by the product topology on $\prod_1^\infty (\mathbb{R}^3 \times \{0,1\})$ and the product measure $\mathbf{P}$. For p defined in (5.25), we define a class of functions $\mathcal{P} = \left\{ p(d; \boldsymbol{\theta}, S_0(t)) : (\boldsymbol{\theta}, S_0(t)) \in (\boldsymbol{\Theta} \times \Phi) \right\}$ and a class of functions $\mathcal{H} = \left\{ \log(1 - \alpha + \alpha p/p_0) : p \in \mathcal{P} \right\}$, where $p_0 = p(d; \boldsymbol{\theta}^0, S_0^0(t))$. Then it follows from Huang and Wellner (1995) that $\mathcal{H}$ is a Donsker class. Further, from Glivenko-Cantelli theorem, there exists a set $\Omega_0 \subset \Omega$ with $\mathbf{P}(\Omega_0) = 1$ such that for every $\omega \in \Omega_0$, the first term of (5.28) converges to zero. Now fix a point $\omega \in \Omega_0$ and write $\hat{\boldsymbol{\theta}}_{(n)} = \hat{\boldsymbol{\theta}}_{(n)}(\omega)$ and $\hat{S}_{0(n)}(t) = \hat{S}_{0(n)}(t, \omega)$. By our assumption $\boldsymbol{\Theta}$ is bounded, and hence for any subsequence of $\hat{\theta}_{(n)}$, we can find a subsequence converging to $\boldsymbol{\theta}_* \in \boldsymbol{\Theta}^C$, the closure of $\boldsymbol{\Theta}$. Also by Helly's selection theorem, for any subsequence of $\hat{S}_{0(n)}(t)$, we can find a further subsequence converging to some nonincreasing function $S_{0*}(t)$. Choose the convergent subsequence of $\hat{\boldsymbol{\theta}}_{(n)}$ and a convergent subsequence of $\hat{S}_{0(n)}(t)$

so that they have the same indices, and without loss of generality, assume that $\hat{\boldsymbol{\theta}}_{(n)}$ converges to $\boldsymbol{\theta}_*$ and that $\hat{S}_{0(n)}(t)$ converges to $S_{0*}(t)$. Let $p_*(d) = p(d; \boldsymbol{\theta}_*, S_{0*}(t))$. By the bounded convergence theorem, the second term of (5.28) converges to

$$A_* = \int \log\left((1-\alpha) + \alpha \frac{p_*(d)}{p_0(d)}\right) d\mathbb{P}(d), \tag{5.29}$$

which is nonnegative by (5.27). However, by Jensen's inequality, (5.29) must be nonpositive. Therefore (5.29) must be zero and it follows that

$$p_*(d) = p_0(d) \quad \mathbb{P} - \text{almost surely},$$

which implies that

$$S_{0*}(t) = S_0^0(t) \quad F_0 - \text{almost surely}.$$

Therefore by bounded convergence theorem,

$$\int |\hat{S}_{0(n)}(t) - S_0^0(t)| \mathrm{d}F_0(t) \to 0. \tag{5.30}$$

Also

$$S_{0*}(u)\exp(-\boldsymbol{\theta}_*^\top \boldsymbol{z}u) = S_0^0(u)\exp(-\boldsymbol{\theta}^{0^\top} \boldsymbol{z}u) \quad \mathbb{P} - \text{almost surely},$$

and

$$S_{0*}(v)\exp(-\boldsymbol{\theta}_*^\top \boldsymbol{z}v) = S_0^0(v)\exp(-\boldsymbol{\theta}^{0^\top} \boldsymbol{z}v) \quad \mathbb{P} - \text{almost surely}.$$

This together with (A5) imply that there exist $\boldsymbol{z}_1 \neq \boldsymbol{z}_2$ such that for some interoir point $c > 0$ in the support of T,

$$S_{0*}(c)\exp(-\boldsymbol{\theta}_*^\top \boldsymbol{z}_1 c) = S_0^0(c)\exp(-\boldsymbol{\theta}^{0^\top} \boldsymbol{z}_1 c),$$

and

$$S_{0*}(c)\exp(-\boldsymbol{\theta}_*^\top \boldsymbol{z}_2 c) = S_0^0(c)\exp(-\boldsymbol{\theta}^{0^\top} \boldsymbol{z}_2 c).$$

Since $S_{0*}(c) > 0$ and $S_0^0(c) > 0$, this implies $(\boldsymbol{\theta}_* - \boldsymbol{\theta}^{0^\top})(\boldsymbol{z}_1 - \boldsymbol{z}_2) = 0$. Again by (A5), the collection of such $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ has positive probability and there exist at least b such pairs that constitute a full rank $b \times b$ matrix, it follows that $\boldsymbol{\theta}_* = \boldsymbol{\theta}^0$, This in turn implies that

$$S_{0*}(u) = S_0^0(u) \quad \text{and} \quad S_{0*}(v) = S_0^0(v) \quad G_0 - \text{almost surely.}$$

Therefore by bounded convergence theorem,

$$\int \left((\hat{S}_{0(n)}(u) - S_0^0(u))^2 + (\hat{S}_{0(n)}(v) - S_0^0(v))^2 \right) \mathrm{d}G_0(u, v) \to 0. \tag{5.31}$$

Equations (5.30) and (5.31) together with $\boldsymbol{\theta}_* = \boldsymbol{\theta}^0$ hold for all $\omega \in \Omega_0$ with $\mathbf{P}(\Omega_0) = 1$. This completes the proof. □

We can derive the studentized bootstrap confidence interval and can carry out studentized bootstrap test for testing the significance of regression parameters as we did in Chapter 3.

5.3 Simulation Studies

Simulation studies are conducted to assess the finite sample performance of the estimators. We consider the exponential distribution with mean λ^{-1} as the baseline distribution of lifetime variate T. We choose exponential distribution with mean

TABLE 5.1: Bias, MSE and BCP of the estimator of θ for Method-1 and Method-2 under mild censoring.

			$n = 30$			$n = 75$			$n = 125$		
λ	θ	Method	Bias	MSE	BCP	Bias	MSE	BCP	Bias	MSE	BCP
0.1	0.25	1	0.0091	0.0067	0.898	0.0033	0.0008	0.903	0.0017	0.0004	0.958
		2	0.0396	0.0073	0.934	0.0347	0.0011	0.940	0.0025	0.0009	0.961
1.0	0.50	1	0.0163	0.0069	0.889	0.0104	0.0009	0.895	0.0022	0.0004	0.963
		2	0.0454	0.0078	0.920	0.0373	0.0018	0.928	0.0029	0.0007	0.970
2.5	-0.50	1	0.0108	0.0073	0.915	0.0077	0.0019	0.921	0.0034	0.0009	0.960
		2	0.0307	0.0067	0.921	0.0247	0.0012	0.926	0.0041	0.0005	0.958
4.0	-0.01	1	0.0410	0.0055	0.918	0.0336	0.0017	0.924	0.0027	0.0008	0.956
		2	0.0507	0.0106	0.929	0.0448	0.0013	0.934	0.0033	0.0009	0.963

TABLE 5.2: Bias, MSE and BCP of the estimator of θ for Method-1 and Method-2 under heavy censoring.

			$n = 30$			$n = 75$			$n = 125$		
λ	θ	Method	Bias	MSE	BCP	Bias	MSE	BCP	Bias	MSE	BCP
0.1	0.25	1	0.0151	0.0129	0.889	0.0057	0.0042	0.900	0.0020	0.0015	0.952
		2	0.0441	0.0147	0.925	0.0384	0.0031	0.937	0.0114	0.0018	0.955
1.0	0.50	1	0.0222	0.0143	0.882	0.0141	0.0041	0.892	0.0088	0.0017	0.950
		2	0.0539	0.0174	0.916	0.0405	0.0044	0.925	0.0144	0.0020	0.957
2.5	-0.50	1	0.0170	0.0151	0.904	0.0104	0.0050	0.918	0.0087	0.0028	0.954
		2	0.0372	0.0156	0.915	0.0283	0.0038	0.923	0.0091	0.0020	0.949
4.0	-0.01	1	0.0462	0.0101	0.913	0.0361	0.0056	0.921	0.0127	0.0022	0.951
		2	0.0561	0.0131	0.921	0.0484	0.0044	0.931	0.0166	0.0019	0.955

λ_1^{-1} as the distribution of U and exponential distribution with mean λ_2^{-1} as the distribution of $Y = V - U$. Further, the variates T, U and Y are assumed to be mutually independent. We consider a single covariate z in the present study which is generated from uniform distribution over $[0, 10]$ and let θ be the corresponding regression parameter. Under the model assumption in (5.1), the survival function of T given z may be written as

$$S(t|z) = S_0(t) \exp(-\theta z t), \tag{5.32}$$

where $S_0(t) = \exp(-\lambda t)$. It can be observed that (5.32) is the survival function corresponding to an exponential random variable with mean $(\lambda + \theta z)^{-1}$. Samples of sizes $n = 30$, 75 and 125 are generated from (5.32) for fixed values of λ and θ. Now corresponding to each observation on T, a random censoring interval (U, V) is generated, where the distribution parameters are fixed as $\lambda_1^{-1} = 15$ and $\lambda_2^{-1} = 10$. If $T \notin (U, V)$, then T is selected in the sample, otherwise we choose the interval as the observation. We consider two different censoring rates viz., 10% (mild censoring) and 30% (heavy censoring) for our inference.

The martingale based inference procedure, denoted as Method-1 and iterative inference procedure, denoted as Method-2 are employed to obtain the estimates of $S_0(t)$ and θ. We used 1000 iterations for various choices of λ and θ. The average absolute bias (Bias) and estimated mean squared error (MSE) are computed and are given in Tables 5.1 - 5.2. In each case, a 95% symmetric studentized bootstrap confidence interval for regression parameter is computed. The proportion of times the true parameter value lies in such intervals is called bootstrap coverage probabilities (BCP). They are also reported in Tables 5.1 - 5.2. From the tables we observe that both bias and MSE are small and they decrease as the sample size increases. The bootstrap coverage probabilities are found fairly large, close to 0.95. Further, as the censoring rate increases, the bias and MSE increase, while the BCP decreases. For each combination of parameter values of (λ, θ) considered, we found out a cubic polynomial estimate of the form $\hat{S}_0(t) = c_0 + c_1 t + c_2 t^2 + c_3 t^3$ with each of its coefficients being taken as the average of corresponding coefficient estimates obtained for 1000 iterations. The estimated survival curves are compared in Figures 5.1 - 5.4, where continuous curve represents the true baseline survival function, dashed curve represents corresponding estimated survival function under

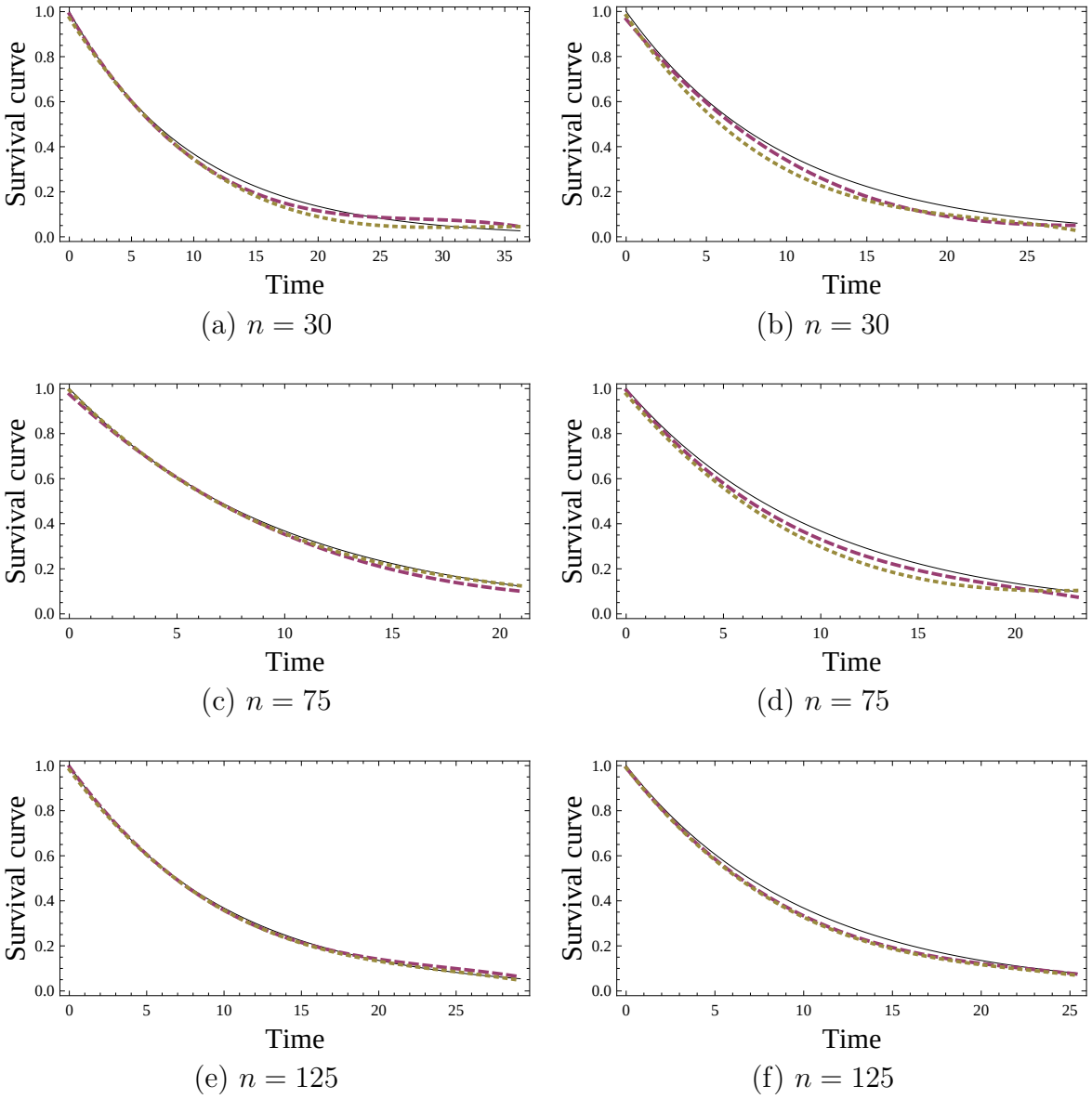


FIGURE 5.1: Plots of baseline survival function (continuous curve) and its estimates for Method-1 (dashed curve) and Method-2 (dotted curve) for $\lambda = 0.1, \theta = 0.25$.

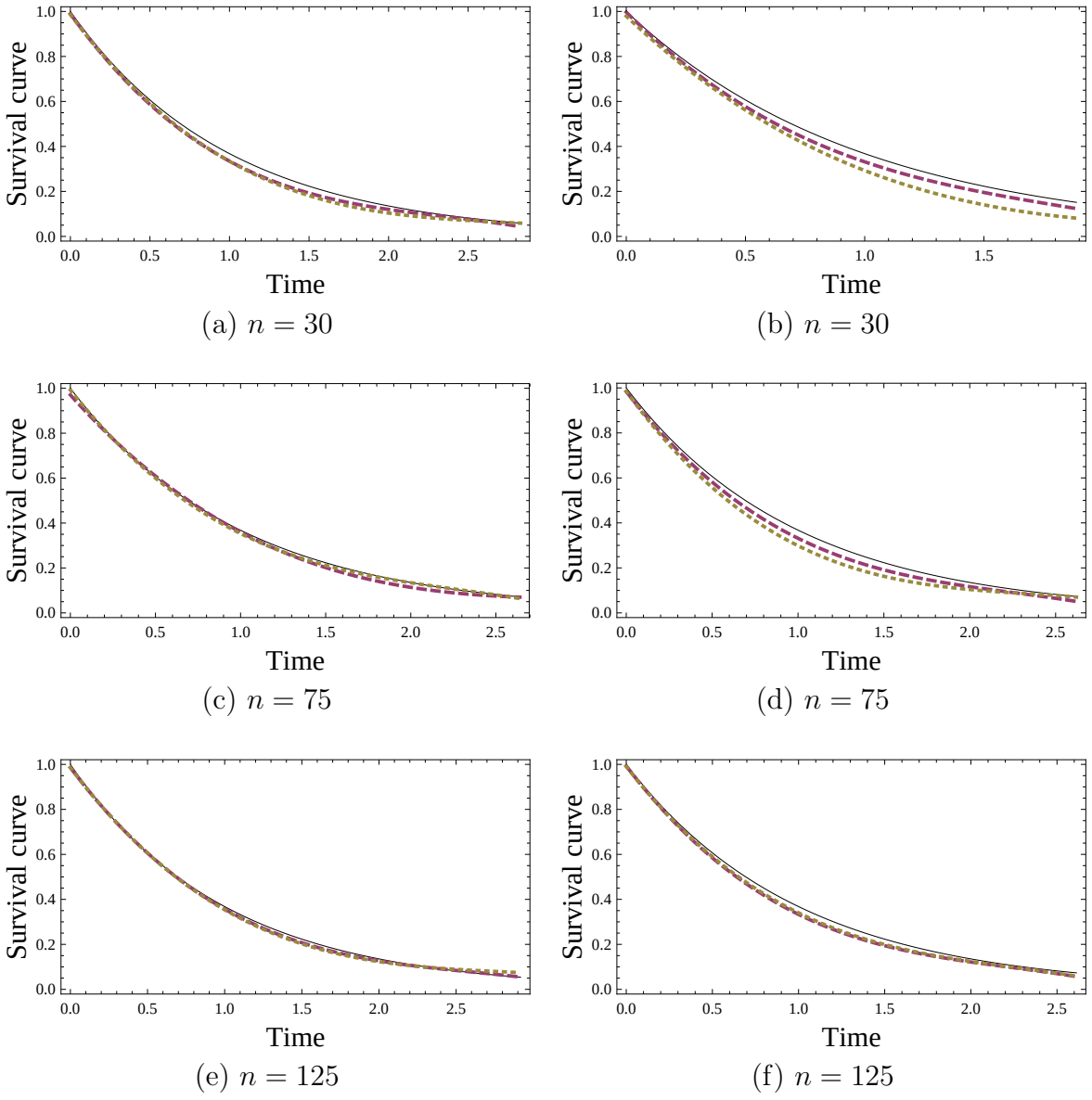


FIGURE 5.2: Plots of baseline survival function (continuous curve) and its estimates for Method-1 (dashed curve) and Method-2 (dotted curve) for $\lambda = 1, \theta = 0.50$.

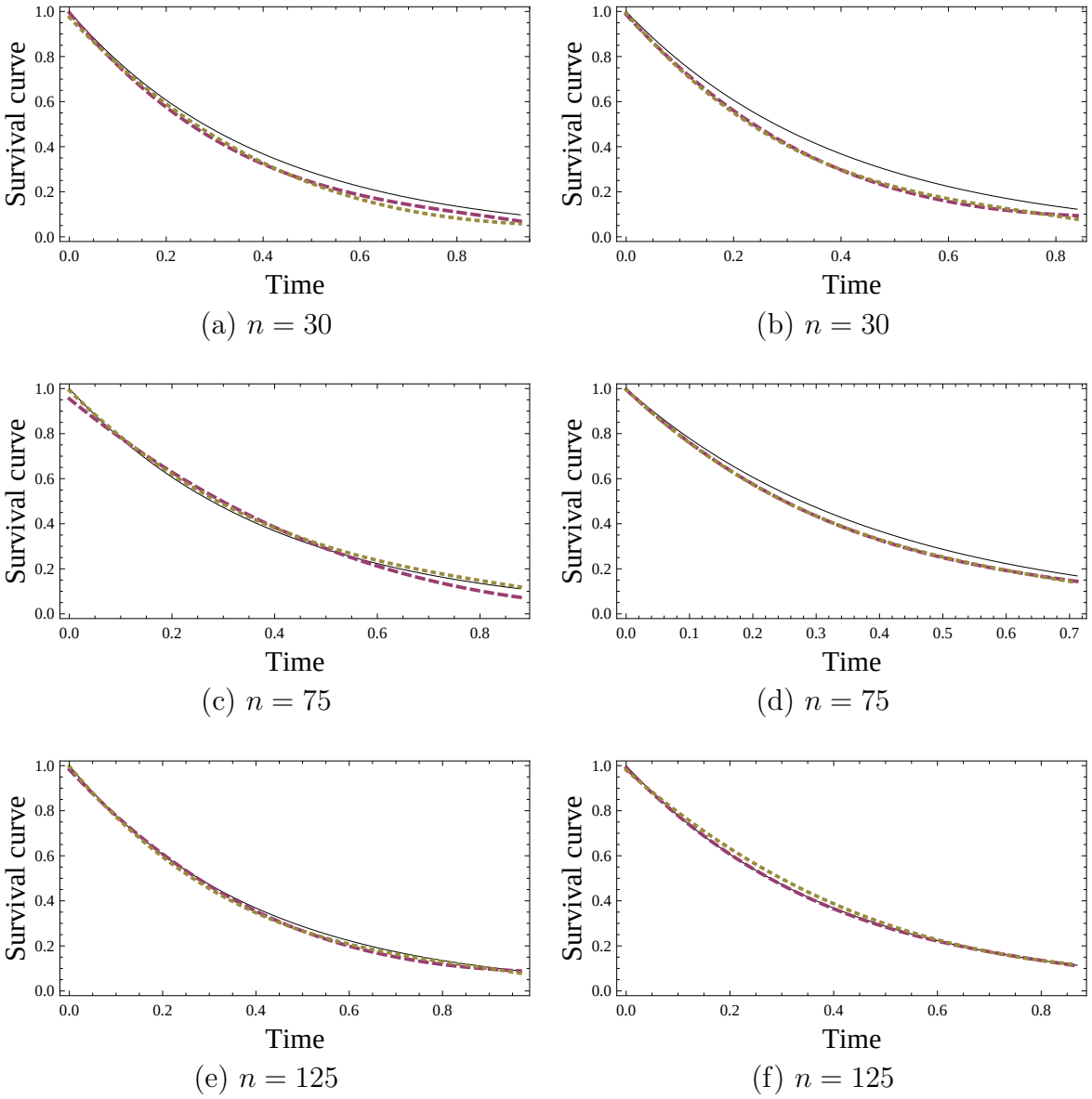


FIGURE 5.3: Plots of baseline survival function (continuous curve) and its estimates for Method-1 (dashed curve) and Method-2 (dotted curve) for $\lambda = 2.5, \theta = -0.50$.

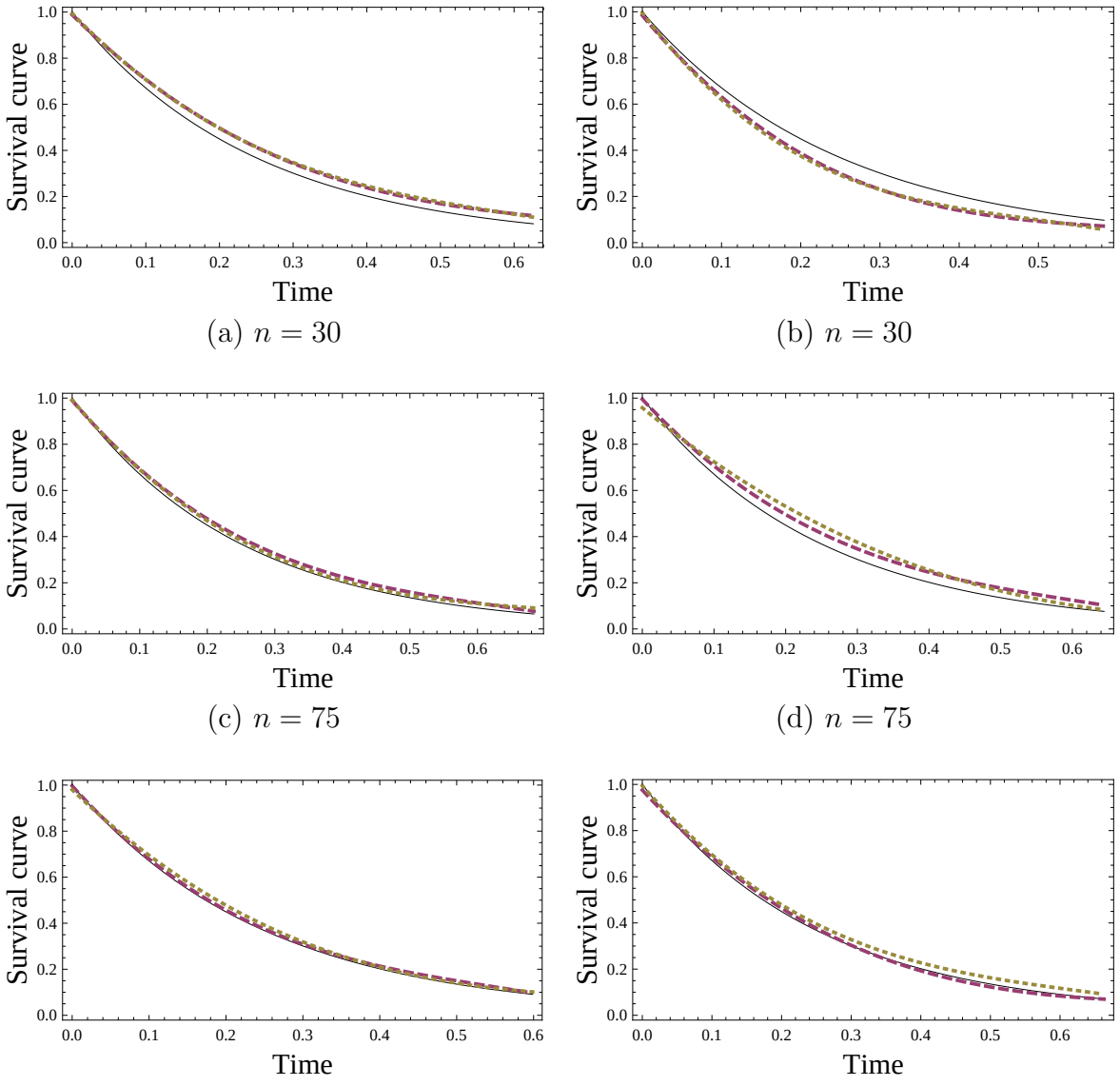


FIGURE 5.4: Plots of baseline survival function (continuous curve) and its estimates for Method-1 (dashed curve) and Method-2 (dotted curve) for $\lambda = 4, \theta = -0.01$.

TABLE 5.3: Estimates of parameters and their SE.

	Method - 1		Method - 2	
	Estimate	SE	Estimate	SE
θ_1	0.0112	0.0095	0.00895	0.0072
θ_2	0.1006	0.0181	0.1760	0.0105
c_0	0.9367	0.1011	0.9657	0.0914
c_1	-0.0487	0.0560	-0.0599	0.0504
c_2	0.0006	0.00029	0.00121	7.6824×10^{-5}
c_3	-9.2230×10^{-6}	6.9851×10^{-7}	-9.2560×10^{-6}	1.4597×10^{-7}

Method-1 and dotted curve represents corresponding estimated survival function under Method-2. We observe that both the estimated survival functions become close to the true survival function when the sample size increases. Moreover, as the censoring rate increases, they tend to depart more from the true survival curve.

5.4 Illustrative Data Analysis

In this section, the proposed methods are applied to a real life data studied by Ichida et al. (1993). The data deal with an evaluation of a protocol change in disinfectant practices in a medical center, where patients are suffering from burn wounds. The major concern in burn management is the control of infection and the study aims at comparing two different controlling methods namely routine bathing care method and body cleansing method. The time (in days) until staphylococcus infection is recorded and we consider it as the lifetime. We consider all exact lifetimes for our study. We then artificially middle-censor a portion of these lifetimes. Though the original study involves several covariates, for the illustration purpose we consider two of them namely treatment (z_1), which is coded as 1 for routine bathing and 2 for body cleansing, and percentage of total surface area burned (z_2). Let θ_1 and

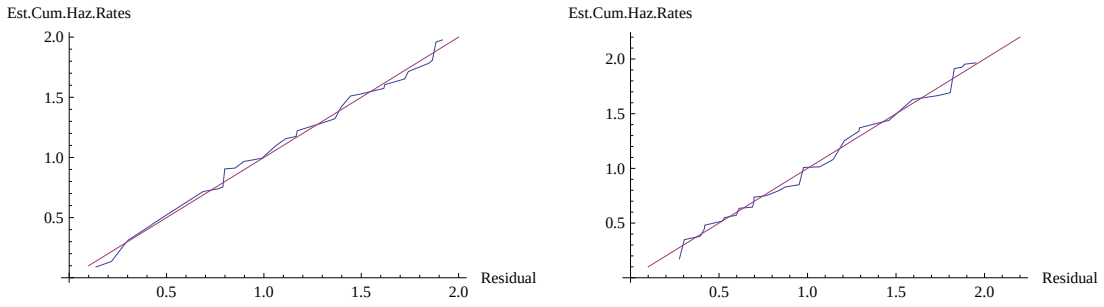


FIGURE 5.5: Plot of estimated cumulative hazard rates against r_j's for (i) Method-1 and (ii) Method-2.

θ_2 respectively be the corresponding regression parameters. A random censoring interval (U, V), where U and $Y = V - U$ are independent exponential variates with respective means $\lambda_1^{-1} = 20$ and $\lambda_2^{-1} = 10$ is generated first. Then, if the lifetime of a patient happens to fall in the generated censoring interval, that lifetime is assumed to be middle-censored and the interval is considered as the observation. Otherwise the exact lifetime is considered. This process is repeated until 25% of the observations are censored. The resulted data set includes both exact observations and censoring intervals. We apply the two methods of estimation given in Section 5.2 to this new data set consisting of both exact and censored observations and we obtained the estimates of the baseline survival function and the regression parameter $\boldsymbol{\theta}$. The estimated values of the coefficients of survival curves as well as regression parameters are given in Table 5.3. The bootstrap based standard errors (SE) of these estimates are computed with $B = 1000$ and they are also reported in Table 5.3. It can be observed that both the covariates have adverse effect on lifetime and that SE's are less. To test the significance of the regression parameters, we apply the studentized bootstrap test. For the null hypothesis $H_0 : \boldsymbol{\theta} = \mathbf{0}$ the test gives a P-value of 0.0037, indicating that the covariate effects are significant.

We now check the overall fit of the model using Cox-Snell residuals. The method of defining the residuals is similar to that described in Chapter 4, except that instead of fitted parametric model for the baseline cumulative hazards function, we use its nonparametric counterpart obtained from $\hat{S}_{0(n)}$. Figure 5.5 shows the plots so obtained under both methods of estimation. The curves are close to the straight line indicating that the model assumption given in (5.1) is reasonable.

5.5 Conclusion

The present chapter discussed the semiparametric additive hazards regression problem for middle-censored lifetime data. We have considered two different methods of estimation of the regression parameter and baseline survival function; one uses martingale based theory and other based on an iterative method for which a maximization procedure for finding the MLE is developed. Large sample properties including consistency and weak convergence of the estimators were established under the martingale based method. Consistency of the estimators was proved under the iterative method, whereas the weak convergence do not appear easy to establish, although one can perhaps extend the ideas used in Huang and Wellner (1995). Simulation studies were carried out in Section 5.3, which indicated that the inference procedures are performing satisfactorily. An application of the proposed inference procedures was illustrated using a real life data set in Section 5.4.

Chapter 6

The Quantile Regression Model

6.1 Introduction

In previous four chapters, we developed hazard based regression models for analyzing middle-censored liftime data. As mentioned in Chapter 1, one may be interested to associate the covariate effect directly on the survival time, rather than on its hazard function. Such methods would be practically useful if the covariates have non-constant effect across the support of lifetime variate. A quantile regression model defined in (1.15) is more adequate in such situations. Motivated by this, we develop a quantile regression model for the analysis of middle-censored lifetime data subject to conditionally independent censoring. We exploit the martingale feature associated with observed data to develop the inference procedure and to establish the asymptotic properties of the estimator.

The rest of the chapter is organized in the following way. We define the data structure and quantile regression model in Section 6.2. The estimation procedure and the asymptotic properties of the estimator are presented. The testing of hypothesis as well as model diagnostics are also discussed therein. Section 6.3 reports

[1]The results in this chapter have been communicated as entitled "Quantile regression model for the analysis of middle-censored lifetime data"(see Sankaran and Prasad (2017c))

an extensive simulation study to assess the finite sample performance of the proposed estimator. In Section 6.4, we apply the proposed model to a real data set on survival times of acute leukemia patients undergoing bone marrow transplantation studied by Copelan et al. (1991). The chapter ends with major conclusions in Section 6.5.

6.2 The Model and Inference Procedure

Let T be the lifetime of interest for an individual under investigation. We assume that T is middle-censored by (U, V) as before. Let $\boldsymbol{z}$ be a $p \times 1$ vector of recorded covariates and let $\boldsymbol{Z} = (1, \boldsymbol{z}^\top)^\top$. We denote the censoring indicator by $\delta = I(T \notin (U, V))$ and assume that given $\boldsymbol{Z}$, T is independent of (U, V). The observed vector is $(X, \boldsymbol{Z}, \delta)$, where $X = T$ if $\delta = 1$, and $X = (U, V)$ if $\delta = 0$. We consider a quantile regression model which takes the form

$$Q_T(\tau|\boldsymbol{Z}) = g(\boldsymbol{Z}^\top \boldsymbol{\beta}_0(\tau)), \tag{6.1}$$

for each $0 < \tau < 1$, where $g(\cdot)$ is a known monotone link function and $\boldsymbol{\beta}_0(\tau)$ is a vector of unknown parameters representing true covariate effects on $Q_T(\tau|\boldsymbol{Z})$. Common choices for $g(\cdot)$ include $g(x) = e^x$ and the identity mapping. Our objective is to estimate the regression quantile $\boldsymbol{\beta}_0(\tau)$ when $g(\cdot)$ is known. The observed data consists of n independent and identically distributed replicates $(X_i, \boldsymbol{Z}_i, \delta_i)$ of $(X, \boldsymbol{Z}, \delta), 1 \leq i \leq n$. In Chapter 2 we defined the counting process $\{N_i(t); t \geq 0\}$ and the at-risk process $\{R_i(t); t \geq 0\}$ for the i'th subject $(i = 1, 2, ..., n)$ under middle-censoring scheme by $N_i(t) = I(\bar{X}_i \leq t, \delta_i = 1)$ and $R_i(t) = I(\bar{X}_i \geq t)$, where

$\bar{X} = T$ if $\delta = 1$, and $\bar{X} = U$ if $\delta = 0$. Note that $\{N(t);\, t \geq 0\}$ and $\{R(t);\, t \geq 0\}$ reduce respectively to the counting process and at-risk process of randomly right censored data when the right end point V of the censoring interval is extended to infinity, and those of randomly left censored data when the left end point U of the censoring interval is limited to zero.

We now construct a stochastic integral estimating equation for estimating the unknown parameter $\boldsymbol{\beta}_0(\tau)$. We define $M_i(t) = N_i(t) - \int_0^t R_i(u)\mathrm{d}\Lambda_T(u|\boldsymbol{Z}_i)$, where $\Lambda_T(u|\boldsymbol{Z}_i)$ denotes the cumulative hazard function of T given $\boldsymbol{Z}_i$. Denote the filtration $\sigma\{N_i(u), R_i(u+), \boldsymbol{Z}_i : i = 1, 2, ..., n;\, 0 \leq u \leq t\}$ by $\mathcal{F}_t$. We then obtain the following relation approximately.

$$\begin{aligned} E[\mathrm{d}N_i(t)|\mathcal{F}_{t-}] &= P[t \leq T_i < t + dt, R_i(t) = 1|\mathcal{F}_{t-}] \\ &= R_i(t)P[t \leq T_i < t + dt, |T_i \geq t, U_i \geq t, V_i \leq t, \boldsymbol{Z}_i] \\ &= R_i(t)\mathrm{d}\Lambda_T(t|\boldsymbol{Z}_i). \end{aligned}$$

This shows that $M_i(t)$ is a martingale (Andersen and Gill (1982)) approximately, and hence we have

$$E\{M_i(t)|\boldsymbol{Z}_i\} = 0, \forall\, t \geq 0, \tag{6.2}$$

approximately. Under the model (6.1), we have

$$\begin{aligned} \Lambda_T(g(\boldsymbol{Z}_i^\top \boldsymbol{\beta}_0(\tau))|\boldsymbol{Z}_i) &= -\log[1 - F_T(g(\boldsymbol{Z}_i^\top \boldsymbol{\beta}_0(\tau))|\boldsymbol{Z}_i)] \\ &= -\log(1 - \tau). \end{aligned}$$

Thus, it follows from the use of variable transformation within the integral, that

$$M_i[g(\boldsymbol{Z}_i^\top \boldsymbol{\beta}_0(\tau))] = N_i[g(\boldsymbol{Z}_i^\top \boldsymbol{\beta}_0(\tau))] - \int_0^\tau I[\bar{X}_i \geq g(\boldsymbol{Z}_i^\top \boldsymbol{\beta}_0(\nu))] \mathrm{d}\mathscr{H}(\nu), \tag{6.3}$$

approximately, where $\mathscr{H}(x) = -\log(1-x)$. We now define

$$\boldsymbol{S}_n(\boldsymbol{\beta}, \tau) = \frac{1}{n}\sum_{i=1}^n \boldsymbol{Z}_i \Big(N_i[g(\boldsymbol{Z}_i^\top \boldsymbol{\beta}(\tau))] - \int_0^\tau I[\bar{X}_i \geq g(\boldsymbol{Z}_i^\top \boldsymbol{\beta}(\nu))] \mathrm{d}\mathscr{H}(\nu) \Big). \tag{6.4}$$

It is easy to see that $E\{\boldsymbol{S}_n(\boldsymbol{\beta}_0, \tau)\} = 0$ approximately by using (6.2) and (6.3). Therefore we propose to estimate $\boldsymbol{\beta}_0(\cdot)$ by using the estimating equation

$$n^{1/2}\boldsymbol{S}_n(\boldsymbol{\beta}, \tau) = 0. \tag{6.5}$$

A careful examination reveals that when the upper bounds V_i of the censoring intervals become infinity, equation (6.5) reduces to the estimating equation for randomly right censored data discussed by Peng and Huang (2008), and when the lower bounds U_i of the censoring intervals become zero, equation (6.5) gives an estimating equation for a randomly left censored data discussed by Ji et al. (2012).

The stochastic integral representation of $\boldsymbol{S}_n(\boldsymbol{\beta}, \tau)$ suggests a grid-based estimation procedure to obtain an estimator of $\boldsymbol{\beta}_0(\tau)$, denoted by $\hat{\boldsymbol{\beta}}(\tau)$ based on equation (6.5) as a right continuous piecewise-constant function that jumps only on a pre-specified grid $\mathscr{S}_{L_n} = \{0 = \tau_0 < \tau_1 < ... < \tau_{L_n} = \tau^\circ < 1\}$, where τ° is a pre-specified constant subject to certain theoretical constraints discussed in Section 6.2.1. We define the norm function for this grid by $\|\mathscr{S}_{L_n}\|$ =max $\{|\tau_j - \tau_{j-1}|, j = 1, 2, ..., L_n\}$. Since by definition $g(\boldsymbol{Z}^\top \boldsymbol{\beta}_0(0)) = 0$, we always set $g(\boldsymbol{Z}^\top \hat{\boldsymbol{\beta}}(0)) = 0$.

We shall obtain $\hat{\boldsymbol{\beta}}(\tau_j)$ for $j = 1, 2, ..., L_n$ by sequentially solving the estimating

equation for $\boldsymbol{\beta}(\tau_j)$ given by

$$n^{-1/2}\sum_{i=1}^{n}\boldsymbol{Z}_i\Big(N_i[g(\boldsymbol{Z}_i^\top\boldsymbol{\beta}(\tau_j))]-\sum_{k=0}^{j-1}I[\bar{X}_i\geq g(\boldsymbol{Z}_i^\top\hat{\boldsymbol{\beta}}(\tau_k))]\{\mathscr{H}(\tau_{k+1})-\mathscr{H}(\tau_k)\}\Big)=0. \tag{6.6}$$

It can be observed that the estimating equation (6.6) is a monotone estimating equation (Fygenson and Ritov (1994)), which is not continuous and therefore an exact solution may not exist. At the same time its monotonic nature can be well exploited to the benefit of computation, that all its generalized solutions belong to a convex set and the left hand side of (6.6) equals $2^{-1}n^{-1/2}$ times the gradient of the following L_1-type convex function.

$$\begin{aligned} l_j(\boldsymbol{h}) = &\sum_{i=1}^{n}\left|I(\delta_i=1)g^{-1}(\bar{X}_i)-\boldsymbol{h}^\top I(\delta_i=1)\boldsymbol{Z}_i\right|+\left|K-\boldsymbol{h}^\top\sum_{l=1}^{n}-I(\delta_i=1)\boldsymbol{Z}_i\right|+ \\ &\left|K-\boldsymbol{h}^\top\sum_{r=1}^{n}2\boldsymbol{Z}_r\sum_{k=0}^{j-1}I[\bar{X}_r\geq g(\boldsymbol{Z}_r^\top\hat{\boldsymbol{\beta}}(\tau_k))]\{\mathscr{H}(\tau_{k+1})-\mathscr{H}(\tau_k)\}\right|, \end{aligned} \tag{6.7}$$

where K is a very large number chosen in such a way that it can bound the quantities $\left|\boldsymbol{h}^\top\sum_{l=1}^{n}-I(\delta_i=1)\boldsymbol{Z}_i\right|$ and $\left|\boldsymbol{h}^\top\sum_{r=1}^{n}2\boldsymbol{Z}_r\sum_{k=0}^{j-1}I[\bar{X}_r\geq g(\boldsymbol{Z}_r^\top\hat{\boldsymbol{\beta}}(\tau_k))]\ \{\mathscr{H}(\tau_{k+1})-\mathscr{H}(\tau_k)\}\right|$ for all values of the vector $\boldsymbol{h}$ in the compact parameter space for $\boldsymbol{\beta}_0(\tau)$ and $j=1,2,...,L_n$. The solutions of (6.6) turn out to be the minimizing values (6.7). The equation (6.7) can be formulated as a linear programming problem and can easily be solved by using simplex method which leads to $\hat{\boldsymbol{\beta}}(\tau)$.

6.2.1 Asymptotic Results

We establish asymptotic properties of the proposed estimator whose derivation is greatly facilitated by the stochastic integral representation of the estimating function (6.7). Strong consistency and weak convergence of the estimator are established under certain regularity conditions. We first give following notations.

Define $\tilde{F}_{X,\delta}(t|\boldsymbol{Z}) = P\{\bar{X} \le t, \delta = 1|\boldsymbol{Z}\}$. Let $F_\Psi(t|\boldsymbol{Z})$ and $\bar{F}_\Psi(t|\boldsymbol{Z})$ be the cumulative distribution function and survival function respectively for Ψ given $\boldsymbol{Z}$, where $\Psi = T$ if $\delta = 1$, and $\Psi = V$ if $\delta = 0$. We denote $\bar{F}_0(t|\boldsymbol{Z}) = P\{U < t < V|\boldsymbol{Z}\}$. Also let $f_T(\cdot|\boldsymbol{Z}), f_\Psi(\cdot|\boldsymbol{Z}), \bar{f}_\Psi(\cdot|\boldsymbol{Z}), \tilde{f}_{X,\delta}(\cdot|\boldsymbol{Z}), \bar{f}_0(\cdot|\boldsymbol{Z})$ and $g'(\cdot)$ denote the first order derivatives of $F_T(\cdot|\boldsymbol{Z}), F_{\Psi|\boldsymbol{Z}}(\cdot|\boldsymbol{Z}), \bar{F}_{\Psi|\boldsymbol{Z}}(\cdot|\boldsymbol{Z}), \tilde{F}_{X,\delta}(\cdot|\boldsymbol{Z}), \bar{F}_0(\cdot|\boldsymbol{Z})$ and $g(\cdot)$ respectively. We define the following quantities first.

$\boldsymbol{\mu}(\boldsymbol{b}) = E\big(\boldsymbol{Z}N[g(\boldsymbol{Z}^\top\boldsymbol{b})]\big)$,

$\boldsymbol{B}(\boldsymbol{b}) = E\big[\boldsymbol{Z}^{\otimes 2}\tilde{f}_{X,\delta}\big(g(\boldsymbol{Z}^\top\boldsymbol{b})|\boldsymbol{Z}\big)g'(\boldsymbol{Z}^\top\boldsymbol{b}))\big]$,

$\boldsymbol{\upsilon}_n(\boldsymbol{b}) = n^{-1}\sum_{i=1}^n \boldsymbol{Z}_i N_i[g(\boldsymbol{Z}_i^\top\boldsymbol{b})|\boldsymbol{Z}] - \boldsymbol{\mu}(\boldsymbol{b})$,

$\tilde{\boldsymbol{\mu}}(\boldsymbol{b}) = E[\boldsymbol{Z}I(\bar{X} \ge g(\boldsymbol{Z}^\top\boldsymbol{b})]$,

$\boldsymbol{J}(\boldsymbol{b}) = E\Big[\boldsymbol{Z}^{\otimes 2}\big(\bar{f}_\Psi(g(\boldsymbol{Z}^\top\boldsymbol{b})|\boldsymbol{Z}) - (1-\delta)\bar{f}_0(g(\boldsymbol{Z}^\top\boldsymbol{b})|\boldsymbol{Z})\big)g'(\boldsymbol{Z}^\top\boldsymbol{b})\Big]$ and

$\bar{\boldsymbol{\upsilon}}_n(\boldsymbol{b}) = n^{-1}\sum_{i=1}^n \boldsymbol{Z}_i I(\bar{X}_i \ge g(\boldsymbol{Z}_i^\top\boldsymbol{b}) - \tilde{\boldsymbol{\mu}}(\boldsymbol{b})$.

We now establish asymptotic properties under the following regularity conditions.

C1. The covariate space $\boldsymbol{\mathcal{Z}}$ is bounded, i.e., $\sup_i\|\boldsymbol{\mathcal{Z}}_i\| < \infty$.

C2. (i) Each component of $E\big(\boldsymbol{Z}N[g(\boldsymbol{Z}^\top\boldsymbol{\beta}_0(\tau))]\big)$ is a Lipschitz function of τ;

(ii) $\tilde{f}_{X,\delta}(t|\boldsymbol{Z})$ and $f_\Psi(t|\boldsymbol{Z})$ are uniformly bounded in t and $\boldsymbol{Z}$.

C3. (i) $\tilde{f}_{X,\delta}(g(\boldsymbol{Z}^\top\boldsymbol{b})|\boldsymbol{Z}) > 0$ for all $\boldsymbol{b} \in \mathscr{B}(d_0)$;

(ii) $E(\boldsymbol{Z}^{\otimes 2}) > 0$;

(iii) each component of $\boldsymbol{J}(\boldsymbol{b})\boldsymbol{B}(\boldsymbol{b})^{-1}$ is uniformly bounded in $\boldsymbol{b} \in \mathscr{B}(d_0)$, where

$\mathscr{B}(d_0)$ is a neighborhood containing $\{\boldsymbol{\beta}_0(\tau), \tau \in (0, \tau^\circ)\}$ defined as $\mathscr{B}(d_0) = \{\boldsymbol{b} \in \mathbb{R}^p : \inf_{\tau\in(0,\tau^\circ)} \|\boldsymbol{\mu}(\boldsymbol{b}) - \boldsymbol{\mu}(\boldsymbol{\beta}_0(\tau))\| \leq d_0\}$.

C4. $\inf_{\tau\in(h,\tau^\circ)} eigmin\boldsymbol{B}\{\boldsymbol{\beta}_0(\tau)\} > 0$ for any $h \in (0, \tau^\circ)$, where $eigmin(\cdot)$ denotes the minimal eigen value of a matrix.

Condition C1 specifies that the covariates are bounded. C2 (i) requires the smoothness of the quantile process $\boldsymbol{\beta}_0(\cdot)$ and C2(ii) ensures boundedness of $\tilde{f}_{X,\delta}(t|\boldsymbol{Z})$ and $f_\Psi(t|\boldsymbol{Z})$, while C3 (i) and (ii) respectively ensures the positive density and positive definiteness of $E(\boldsymbol{Z}^{\otimes 2})$. Condition C4 is a crucial assumption required for the identifiability of the quantile process. Note that $\tau^\circ; 0 < \tau^\circ < 1$ is a deterministic constant subject to identifiability constraints. Motivated by the identifiability condition proposed by Chang and Yang (1987), we set $\tau^\circ \leq F_T(T^+)$, where T^+ is the upper bound of support of T. These assumptions guarantee that C3(iii) is satisfied with $\boldsymbol{b} = \boldsymbol{\beta}_0(\tau)$ and ensure the continuity of $J(\boldsymbol{b})B(\boldsymbol{b})^{-1}$. With these assumptions, it is reasonable to expect C3(iii) holds in $\mathscr{B}(d_0)$. It can be noted that the distribution law of the random censoring interval (U, V) also plays an important role in the foregoing assumptions. We now establish the strong consistency and weak convergence of our proposed estimator $\hat{\boldsymbol{\beta}}(\tau)$ stated in the following theorems.

Theorem 6.1. *Under the conditions C1 -C4,* $sup_{\tau\in[h,\tau^\circ]} \|\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\| \to_P 0$, *where* $0 < h < \tau^\circ$, *provided* $lim_{n\to\infty}\|\mathscr{S}_{L_n}\| = 0$.

Proof. Define $\mathscr{A}(d) = \{\boldsymbol{\mu}(\boldsymbol{b}) : \boldsymbol{b} \in \mathscr{B}(d)\}$ and denote $\boldsymbol{\alpha}_0(\tau) = \boldsymbol{\mu}\{\boldsymbol{\beta}_0(\tau)\}$ and $\hat{\boldsymbol{\alpha}}_0(\tau) = \boldsymbol{\mu}\{\hat{\boldsymbol{\beta}}(\tau)\}$. It follows from Peng and Huang (2008) that the mapping $\boldsymbol{\mu}(\cdot)$ is invertible from $\mathscr{A}(d_0)$ to $\mathscr{B}(d_0)$ with the inverse mapping being denoted by $\boldsymbol{\kappa}$. Taylor series expansion of $\boldsymbol{\kappa}\{\hat{\boldsymbol{\alpha}}(\tau)\}$ around $\boldsymbol{\alpha}_0(\tau)$ for $\tau \in [h, \tau^\circ]$ along with the

assumptions C1-C4 lead to the required result which follows directly from Peng and Huang (2008). □

Theorem 6.2. *Under the conditions C1 -C4, the process* $n^{1/2}\{\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\}$ *converges weakly to a Gaussian process, for* $\tau \in [h, \tau^\circ]$, *where* $0 < h < \tau^\circ$, *provided* $lim_{n\to\infty} n^{1/2}\|\mathscr{S}_{L_n}\| = 0$.

Proof. Following the proof of Lemma B.1. given in Peng and Huang (2008), we can show that

$$(i) \quad \sup_{\tau\in(0,\tau^\circ]}\|n^{-1/2}\sum_{i=1}^{n}\boldsymbol{Z}_i\big[N_i(g(\boldsymbol{Z}_i^\top\hat{\boldsymbol{\beta}}(\tau))) - N_i(g(\boldsymbol{Z}_i^\top\boldsymbol{\beta_0}(\tau)))\big] - n^{-1/2}\big[\boldsymbol{\mu}(\hat{\boldsymbol{\beta}}(\tau)) - \boldsymbol{\mu}(\boldsymbol{\beta_0}(\tau))\big]\| \quad \text{converges in probability to zero,}$$

and

$$(ii) \quad \sup_{\tau\in(0,\tau^\circ]}\|n^{-1/2}\sum_{i=1}^{n}\boldsymbol{Z}_i\big[I[\bar{X}_i \geq g(\boldsymbol{Z}_i^\top\hat{\boldsymbol{\beta}}(\tau))] - I[\bar{X}_i \geq g(\boldsymbol{Z}_i^\top\boldsymbol{\beta_0}(\tau))]\big] - n^{-1/2}\big(\tilde{\boldsymbol{\mu}}(\hat{\boldsymbol{\beta}}(\tau)) - \tilde{\boldsymbol{\mu}}(\boldsymbol{\beta_0}(\tau))\big)\| \quad \text{converges in probability to zero.}$$

These two convergence properties along with the fact that $\boldsymbol{\mu}(\hat{\boldsymbol{\beta}}(\tau))$ converges uniformly to $\boldsymbol{\mu}(\boldsymbol{\beta_0}(\tau))$ for $\tau \in (0, \tau^\circ]$ enables us to write $n^{1/2}[\boldsymbol{\mu}(\hat{\boldsymbol{\beta}}(\tau)) - \boldsymbol{\mu}(\boldsymbol{\beta_0}(\tau))] = \boldsymbol{\phi}\{-n^{1/2}\boldsymbol{S}_n(\boldsymbol{\beta_0}, \tau)\} + o_{(0,\tau^\circ]}(1)$, where $\boldsymbol{\phi}$ is a linear operator as defined in equation B.3 of Peng and Huang (2008) and $o_I(a_n)$ denote a term that converges uniformly to zero in probability where $\tau \in I$, after being divided by a_n. The required result follows by an application of Donsker theorem to this class whose derivation is similar to that given in Peng and Huang (2008). □

6.2.2 Resampling and Hypothesis Testing

One important issue that occurs in inferences of the regression quantile process is that the covariance matrix of the limiting process of $n^{1/2}\{\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\}$ involves the unknown probability density functions $f_\Psi(t|\boldsymbol{Z})$ and $\tilde{f}_{X,\delta}(t|\boldsymbol{Z})$. Estimation of these quantities by conventional methods may be time consuming and tedious or even unstable with samples of small sizes. To overcome this difficulty, Jin et al. (2001) introduced a simple resampling method by perturbing the objective function which is to be optimized repeatedly, and thereby facilitating the inferences of the parameters based on a large collection of the resulting optimizers. This minimand perturbing technique has been generalized by Peng and Huang (2008) with functional estimands. To adopt this method, we choose independent observations $\zeta_1, \zeta_2, ..., \zeta_n$ drawn from a known non-negative distribution with unit mean and unit variance, for example *exponential* with mean unity. Using these variates, we perturb the objective function given in (6.7) and the resulting objective function is given by

$$\tilde{l}_j(\boldsymbol{h}) = \sum_{i=1}^{n} \left|\zeta_i I(\delta_i = 1) g^{-1}(\bar{X}_i) - \boldsymbol{h}^\top \zeta_i I(\delta_i = 1)\boldsymbol{Z}_i\right| + \left|K - \boldsymbol{h}^\top \sum_{i=1}^{n} -\zeta_i I(\delta_i = 1)\boldsymbol{Z}_i\right| + \left|K - \boldsymbol{h}^\top \sum_{r=1}^{n} 2\zeta_r \boldsymbol{Z}_r \sum_{k=0}^{j-1} I(\bar{X}_r \geq g(\boldsymbol{Z}_r^\top \boldsymbol{\beta}^*(\tau_k)))\{\mathscr{H}(\tau_{k+1}) - \mathscr{H}(\tau_k)\}\right|, \quad (6.8)$$

for $j = 1, 2, ..., L_n$, where $\boldsymbol{\beta}^*(\tau_j)$ is defined as the minimizer of (6.8) and this can be obtained sequentially with the same procedure employed for (6.7). For a fixed quantile point τ, the variance of the quantile process estimator at τ can be approximated by repeatedly generating variates set $\{\zeta_1, \zeta_2, ..., \zeta_n\}$ for a large number of times, say B, and calculating the variance of the resulting sequence $\{\boldsymbol{\beta}_k^*(\tau), k = 1, 2, ..., B\}$. A normal approximation for the process at this specific quantile point is valid since B

is taken to be large and therefore we can find out a $100(1-\alpha)\%$ confidence interval for $\boldsymbol{\beta}_0(\tau_*)$, for a given level of significance α. Following the lines of Ji et al. (2012), a $100(1-\alpha)\%$ confidence band for $\{\boldsymbol{\beta}_0(\tau) : \tau \in [l_1, l_2],\, 0 < l_1 < l_2 < \tau^\circ\}$ can be constructed as $\{\hat{\boldsymbol{\beta}}(\tau) \pm \rho_{1-\alpha}\hat{\sigma}(\tau), \tau \in [l_1, l_2]\}$, where $\rho_{1-\alpha}$ is the $100(1-\alpha)\%$ empirical percentile of $\sup_{\tau\in[l_1,l_2]}|\boldsymbol{\beta}^*(\tau) - \hat{\boldsymbol{\beta}}(\tau)|/\hat{\sigma}(\tau)$ with $\hat{\sigma}(\tau)$ being the empirical standard deviation of $\boldsymbol{\beta}^*(\tau)$ obtained from resampling.

It is often of interest in a regression analysis to assess whether the covariate, say $Z^{(r)}$, the r'th component of covariate vector $\boldsymbol{Z}$ has significant effect over $\tau \in [l_1, l_2]$ for $2 \le r \le p+1$, where $l_1 < l_2 \in (0, \tau^\circ)$. This can be posed as a hypothesis testing problem by considering the general hypothesis $H_0 : \Phi\{\boldsymbol{\beta}_0(\tau)\} = \gamma_0(\tau)$, $\tau \in [l_1, l_2]$, where Φ is a known function and $\gamma_0(\tau)$ is a hypothesized value of $\Phi\{\boldsymbol{\beta}_0(\tau)\}$. It follows that testing the covariate significance at $\tau \in [l_1, l_2]$ is equivalent to setting $\Phi(\boldsymbol{s}) = s^{(r)}$, which is the r'th component of vector $\boldsymbol{s}$ and $\gamma_0(\tau) = 0$. One natural test is given by $\Gamma = n^{1/2}\int_{l_1}^{l_2}\{\Phi\{\hat{\boldsymbol{\beta}}(a)\} - \gamma_0(a)\}\Theta(a)\mathrm{d}a$, where $\Theta(a)$ is a nonnegative weight function and the distribution of Γ, being unknown, can be approximated by the empirical distribution of $\Gamma^* = n^{1/2}\int_{l_1}^{l_2}\{\Phi\{\boldsymbol{\beta}^*(a)\} - \Phi\{\hat{\boldsymbol{\beta}}(a)\}\}\Theta(a)\mathrm{d}a$, given the observed data. Further, it can be shown that Γ is a consistent test under certain conditions (Peng and Huang (2008)). In practice, the weight function $\Theta(a)$ may be appropriately chosen to accentuate the deviation from H_0 by meeting the desired power for an observed sample size. Under H_0, the conditional distribution of Γ^*, given the observed data, is equivalent to the unconditional distribution of Γ and for a given size α, reject H_0 if either Γ is greater than the $(1-\alpha/2)$'th percentile of, or less than the $\alpha/2$'th percentile of the empirical distribution of Γ^*.

Another important hypothesis of interest is to verify whether the effect of a

covariate is constant over $\tau \in [l_1, l_2]$. This can be formulated by first considering a general hypothesis $\tilde{H}_0 : \tilde{\Phi}\{\boldsymbol{\beta}_0(\tau)\} = \eta_0$, for $\tau \in [l_1, l_2]$, where $\tilde{\Phi}(\cdot)$ is a known function and η_0 is an unspecified constant, and then by assigning $\tilde{\Phi}(\boldsymbol{s}) = s^{(r)}$, the r'th component of the vector $\boldsymbol{s}$. This hypothesis can be tested by using the test statistic $\tilde{\Gamma} = n^{1/2} \int_{l_1}^{l_2} [\tilde{\Phi}\{\hat{\boldsymbol{\beta}}(a)\} - \hat{\rho}]\tilde{\Theta}(a)\mathrm{d}a$, where $\tilde{\Theta}(\cdot)$ is a nonconstant weight function and $\hat{\rho} = (l_2 - l_1)^{-1} \int_{l_1}^{l_2} \tilde{\Phi}\{\hat{\boldsymbol{\beta}}(a)\}\mathrm{d}(a)$. For a given level α, one may reject $\tilde{H}_0$ when $\tilde{\Gamma}$ is greater than the $(1-\alpha/2)$'th percentile or less than the $(\alpha/2)$'th percentile of empirical distribution $\tilde{\Gamma}^* = n^{1/2} \int_{l_1}^{l_2} ([\tilde{\Phi}\{\boldsymbol{\beta}^*(a)\} - \tilde{\Phi}\{\hat{\boldsymbol{\beta}}(a)\}] - (\rho^* - \hat{\rho}))\tilde{\Theta}(a)\mathrm{d}a$.

6.2.3 Model Diagnostics

Model diagnostics is an important concern for all model-based inference procedures. As our inference procedure largely exploits the martingale structure of the underlying quantile process, it is natural to employ martingale residuals or their transformations, as discussed by Lin et al. (1993), for model diagnostics. We consider a simple class of stochastic processes given by

$$K(\tau) = n^{-1/2} \sum_{i=1}^{n} q_0(\boldsymbol{Z}_i) M_i(\tau; \hat{\boldsymbol{\beta}}),$$

where $q_0(\cdot)$ is a known bounded function and

$$M_i(\tau; \boldsymbol{\beta}) = N_i[g(\boldsymbol{Z}_i^\top \boldsymbol{\beta}(\tau))] - \int_0^\tau I(\bar{X}_i \geq g(\boldsymbol{Z}_i^\top \boldsymbol{\beta}(\nu)))\mathrm{d}\mathscr{H}(\nu).$$

It follows from Peng and Huang (2008) that if the model (6.1) is specified correctly, then the process $K(\tau)$ converges weakly to a zero-mean Gaussian process, and its

distribution can be approximated by that of

$$K^*(\tau) = n^{-1/2}\sum_{i=1}^{n} q_0(\boldsymbol{Z}_i)M_i(\tau;\hat{\boldsymbol{\beta}})(1-\zeta_i) + n^{-1/2}\sum_{i=1}^{n} q_0(\boldsymbol{Z}_i)\{M_i(\tau;\boldsymbol{\beta}^*) - M_i(\tau;\hat{\boldsymbol{\beta}})\},$$

where ζ_i's are defined as earlier and we consider the observed data as fixed for evaluating $K^*(\cdot)$. Also the null distribution of $K(\cdot)$ can be approximated by simulating $K^*(\cdot)$ by repeatedly generating $\{\zeta_1, \zeta_2, ..., \zeta_n\}$. A numerical measure for the lack of fit may be taken to be the supremum statistic $\sup_{\tau\in[l_1,l_2]}|K(\tau)|$, which is reasonable since under model (6.1), $K(\tau)$ is expected to fluctuate around zero. It follows from Lin et al. (1993) that for a properly chosen $q_0(\cdot)$, this test is consistent against the general alternative that the model assumption (6.1) is violated.

6.3 Simulation Studies

In this section, the finite sample performance of the proposed estimator is assessed by means of Monte Carlo simulation studies. We first consider an accelerated failure time (AFT) model with two independent covariates given by

$$\tilde{T} = b_1 z_1 + b_2 z_2 + \epsilon, \tag{6.9}$$

where $\tilde{T} = \log T$, b_1 and b_2 are regression parameters, and ϵ is a random term assumed to follow standard normal distribution. It then follows that the quantile function of T given $\boldsymbol{Z} = (z_0, z_1, z_2)^\top$, where $z_0 = 1$, is $Q_T(\tau) = g(z_0\beta_0(\tau)+z_1\beta_1(\tau)+z_2\beta_2(\tau))$, where $g(x) = \mathrm{e}^x$, $\beta_0(\tau) = Q_\epsilon(\tau)$, $\beta_1(\tau) = b_1$, and $\beta_2(\tau) = b_2$. Thus

with $\boldsymbol{\beta}_0(\tau) = (\beta_0(\tau), \beta_1(\tau), \beta_2(\tau))^\top$, the model (6.9) is a special case of the model specified in (6.1).

The covariate z_1 is generated from uniform distribution over the interval $(0, 1)$ and z_2 is generated from the Bernoulli distribution with success probability 0.5. The observations are generated using the model (6.9). We generate random censoring interval (U, V) in such a way that $\log U$ follows uniform(c_1, c_2) distribution and $\log Y = \log(V - U)$ follows uniform distribution over $(c_3 \times z_2, c_4)$, where c_1, c_2, c_3 and c_4 are appropriately chosen nonnegative constants. Clearly the censoring interval is covariate dependent in this setting. We now give step-by-step procedure for generating observations of required size.

(i) Randomly generate covariates z_1 and z_2, and using them generate $\log T$ from the AFT model given in (6.9).

(ii) With z_2 held fixed, generate U and V as described above.

(iii) Choose $X = T$ as an observed value if $T \notin (U, V)$, otherwise select (U, V) as the observation.

We consider two different censoring rates viz., 15% (mild censoring) and 30% (heavy censoring), for comparing the impact of censoring on estimators. We choose two different sample sizes viz., 50 and 100 and we fix $b_1 = 0$ and $b_2 = -1$ along with $c_1 = 0.1$, $c_2 = 4.5$, $c_3 = 0.5$ and $c_4 = 2.6$ for generating mild censored data. The heavy censored data is obtained using same values for c_1, c_2 and c_3, but with $c_4 = 3.5$. Data sets are replicated in 1000 iterations and the resampling method is carried out for $B = 250$ with $\{\xi_1, \xi_2, ..., \xi_B\}$ generated from exponential distribution with unit mean. An equally spaced grid with subintervals of length 0.01 is adopted for $\tau \in (0.1, 0.7)$ for the estimation of $\boldsymbol{\beta}_0(\tau)$. Testing of hypotheses concerning the overall significance of the covariates as well as the constancy of their effects

are also performed using weight functions $\Theta(k) = 1$ and $\Theta(k) = I[k \geq (l_1 + l_2)/2]$ respectively. One can also use a weight function putting unit weight on the first half of the interval (l_1, l_2).

TABLE 6.1: Bias, AERSD, ESD and CP for the AFT model under mild censoring.

		$n = 50$				$n = 100$			
τ		Bias	AERSD	ESD	CP	Bias	AERSD	ESD	CP
0.1	$\hat{\beta}_0$	0.048	0.755	0.711	0.926	0.025	0.685	0.626	0.930
	$\hat{\beta}_1$	0.024	0.178	0.180	0.901	0.010	0.100	0.170	0.904
	$\hat{\beta}_2$	0.084	0.341	0.361	0.930	0.041	0.315	0.320	0.934
0.3	$\hat{\beta}_0$	0.074	0.790	0.807	0.895	0.031	0.724	0.726	0.900
	$\hat{\beta}_1$	0.086	0.899	0.890	0.919	0.021	0.888	0.849	0.924
	$\hat{\beta}_2$	0.045	0.533	0.553	0.899	0.040	0.469	0.490	0.901
0.5	$\hat{\beta}_0$	0.063	0.942	0.941	0.947	0.041	0.725	0.708	0.951
	$\hat{\beta}_1$	0.045	0.557	0.514	0.894	0.044	0.311	0.329	0.898
	$\hat{\beta}_2$	0.034	0.231	0.200	0.900	0.032	0.169	0.183	0.902
0.7	$\hat{\beta}_0$	0.063	0.414	0.438	0.920	0.030	0.198	0.211	0.923
	$\hat{\beta}_1$	0.039	0.305	0.326	0.924	0.037	0.222	0.248	0.928
	$\hat{\beta}_2$	0.077	0.479	0.500	0.946	0.034	0.309	0.299	0.950

TABLE 6.2: Bias, AERSD, ESD and CP for the AFT model under heavy censoring.

		$n = 50$				$n = 100$			
τ		Bias	AERSD	ESD	CP	Bias	AERSD	ESD	CP
0.1	$\hat{\beta}_0$	0.054	0.905	0.821	0.919	0.043	0.885	0.802	0.928
	$\hat{\beta}_1$	0.028	0.279	0.261	0.895	0.027	0.165	0.190	0.900
	$\hat{\beta}_2$	0.091	0.453	0.441	0.925	0.086	0.379	0.381	0.928
0.3	$\hat{\beta}_0$	0.079	0.908	0.921	0.886	0.078	0.833	0.851	0.890
	$\hat{\beta}_1$	0.088	0.916	0.901	0.915	0.058	0.900	0.885	0.916
	$\hat{\beta}_2$	0.055	0.811	0.832	0.894	0.049	0.622	0.610	0.895
0.5	$\hat{\beta}_0$	0.070	0.980	0.963	0.941	0.057	0.743	0.800	0.946
	$\hat{\beta}_1$	0.057	0.714	0.710	0.888	0.051	0.500	0.488	0.894
	$\hat{\beta}_2$	0.040	0.411	0.418	0.894	0.037	0.305	0.290	0.895
0.7	$\hat{\beta}_0$	0.071	0.633	0.645	0.913	0.057	0.418	0.470	0.915
	$\hat{\beta}_1$	0.047	0.609	0.584	0.915	0.045	0.411	0.444	0.919
	$\hat{\beta}_2$	0.083	0.606	0.591	0.941	0.059	0.439	0.473	0.945

TABLE 6.3: Bias, AERSD, ESD and CP for the heteroscedastic model under mild censoring.

		$n = 50$				$n = 100$			
τ		Bias	AERSD	ESD	CP	Bias	AERSD	ESD	CP
0.1	$\hat{\beta}_0$	0.016	0.231	0.259	0.910	0.013	0.184	0.171	0.915
	$\hat{\beta}_1$	0.018	0.705	0.699	0.918	0.017	0.588	0.600	0.920
	$\hat{\beta}_2$	0.037	0.276	0.290	0.922	0.035	0.217	0.283	0.925
0.3	$\hat{\beta}_0$	0.079	0.169	0.200	0.943	0.047	0.122	0.154	0.945
	$\hat{\beta}_1$	0.063	0.408	0.444	0.921	0.060	0.277	0.222	0.930
	$\hat{\beta}_2$	0.064	0.355	0.311	0.942	0.040	0.189	0.200	0.944
0.5	$\hat{\beta}_0$	0.025	0.215	0.188	0.902	0.022	0.116	0.122	0.906
	$\hat{\beta}_1$	0.043	0.515	0.500	0.944	0.040	0.368	0.390	0.948
	$\hat{\beta}_2$	0.049	0.723	0.791	0.956	0.046	0.515	0.498	0.958
0.7	$\hat{\beta}_0$	0.065	0.444	0.410	0.930	0.063	0.188	0.201	0.950
	$\hat{\beta}_1$	0.068	0.283	0.300	0.940	0.064	0.184	0.201	0.944
	$\hat{\beta}_2$	0.030	0.614	0.621	0.925	0.028	0.583	0.600	0.929

TABLE 6.4: Bias, AERSD, ESD and CP for the heteroscedastic model under heavy censoring.

		$n = 50$				$n = 100$			
τ		Bias	AERSD	ESD	CP	Bias	AERSD	ESD	CP
0.1	$\hat{\beta}_0$	0.019	0.301	0.321	0.901	0.017	0.212	0.242	0.904
	$\hat{\beta}_1$	0.026	0.884	0.854	0.915	0.022	0.669	0.687	0.918
	$\hat{\beta}_2$	0.044	0.317	0.298	0.919	0.041	0.276	0.286	0.920
0.3	$\hat{\beta}_0$	0.084	0.433	0.410	0.930	0.082	0.400	0.386	0.932
	$\hat{\beta}_1$	0.071	0.631	0.624	0.890	0.067	0.400	0.391	0.901
	$\hat{\beta}_2$	0.070	0.704	0.691	0.888	0.069	0.299	0.312	0.891
0.5	$\hat{\beta}_0$	0.035	0.408	0.395	0.895	0.031	0.209	0.230	0.899
	$\hat{\beta}_1$	0.062	0.714	0.705	0.900	0.054	0.485	0.468	0.908
	$\hat{\beta}_2$	0.066	0.803	0.860	0.911	0.059	0.719	0.720	0.912
0.7	$\hat{\beta}_0$	0.070	0.619	0.640	0.915	0.068	0.303	0.290	0.920
	$\hat{\beta}_1$	0.082	0.505	0.489	0.931	0.080	0.308	0.300	0.938
	$\hat{\beta}_2$	0.051	0.880	0.854	0.904	0.043	0.714	0.738	0.913

These weight functions seem to be providing good powers during the simulation study and therefore it is reasonable to consider them for our data analysis. We set $l_1 = 0.1$ and $l_2 = 0.7$. Tables 6.1 - 6.2 present the results obtained by estimation of the parameters of the AFT model (6.9). We give average absolute values of biases (Bias), average estimated resampling-based standard deviations (AERSD), empirical standard deviations (ESD) of $\hat{\boldsymbol{\beta}}(\tau)$ along with the coverage probabilities (CP) of 95% confidence intervals of $\boldsymbol{\beta}_0(\tau)$ based on normal approximation for $\tau = 0.1, 0.3, 0.5, 0.7$. It can be observed that biases are small, the empirical standard deviations and resampling based standard deviations are small and are close to each other. The coverage probabilities are pretty high. We can observe that the bias, AERSD and ESD decrease when the sample size increases. Moreover, when the censoring rate increases they increase, while the CP decreases. Tables 6.5 - 6.6 show the results of testing of hypotheses concerning significance of regression quantile and constancy of covariate effects. The empirical rejection rates (ERR) for both tests at significance level 0.05 are given. We also present estimated average effects (EAE), empirical standard deviations (ESD) and average resampling-based standard deviation (ARSD) of the average effects. It can be observed that the probability of type-I error is close to nominal value 0.05, the estimated average covariate effects of z_1 and z_2 are close to the true values and the average resampling-based standard deviation and empirical standard deviations agree well with each other. It is evident that the ERR shows better results with increase in sample size, but it worsens with increase in censoring rate. The average estimated effects obtained from the proposed method (in dashed lines) along with the true coefficients (in continuous lines) are displayed in Figures 6.1 - 6.2. It can be observed that the true values and their mean estimates are close to each other.

TABLE 6.5: Hypothesis testing results for AFT model under mild censoring.

	$H_0 : \beta(\tau) = 0,\ l_1 \leq \tau \leq l_2$								$H_0 : \beta(\tau) = \eta_0,\ l_1 \leq \tau \leq l_2$	
	$n_1 = 50$				$n_2 = 100$				$n_1 = 50$	$n_2 = 100$
	ERR	EAE	ESD	ARSD	ERR	EAE	ESD	ARSD	ERR	ERR
$\hat{\beta}_0$	0.72	-0.265	0.50	0.47	0.81	-0.275	0.47	0.43	0.80	0.93
$\hat{\beta}_1$	0.06	0.005	0.02	0.05	0.04	0.003	0.01	0.04	0.08	0.06
$\hat{\beta}_2$	0.89	-0.968	0.03	0.05	0.91	-0.981	0.01	0.04	0.07	0.05

TABLE 6.6: Hypothesis testing results for AFT model under heavy censoring.

	$H_0 : \beta(\tau) = 0,\ l_1 \leq \tau \leq l_2$								$H_0 : \beta(\tau) = \eta_0,\ l_1 \leq \tau \leq l_2$	
	$n_1 = 50$				$n_2 = 100$				$n_1 = 50$	$n_2 = 100$
	ERR	EAE	ESD	ARSD	ERR	EAE	ESD	ARSD	ERR	ERR
$\hat{\beta}_0$	0.61	-0.255	0.53	0.56	0.72	-0.272	0.52	0.52	0.65	0.70
$\hat{\beta}_1$	0.09	0.010	0.05	0.08	0.07	0.007	0.04	0.05	0.09	0.08
$\hat{\beta}_2$	0.69	-0.944	0.06	0.09	0.78	-0.968	0.02	0.05	0.08	0.06

TABLE 6.7: Hypothesis testing results for heteroscedastic model under mild censoring.

	$H_0 : \beta(\tau) = 0,\ l_1 \leq \tau \leq l_2$								$H_0 : \beta(\tau) = \eta_0,\ l_1 \leq \tau \leq l_2$	
	$n_1 = 50$				$n_2 = 100$				$n_1 = 50$	$n_2 = 100$
	ERR	EAE	ESD	ARSD	ERR	EAE	ESD	ARSD	ERR	ERR
$\hat{\beta}_0$	0.69	-0.303	0.53	0.49	0.70	-0.312	0.49	0.45	0.80	0.91
$\hat{\beta}_1$	0.07	0.005	0.08	0.09	0.06	0.002	0.06	0.06	0.08	0.06
$\hat{\beta}_2$	0.75	-1.471	0.39	0.39	0.85	-1.476	0.38	0.36	0.85	0.89

TABLE 6.8: Hypothesis testing results for heteroscedastic model under heavy censoring.

	$H_0 : \beta(\tau) = 0,\ l_1 \leq \tau \leq l_2$								$H_0 : \beta(\tau) = \eta_0,\ l_1 \leq \tau \leq l_2$	
	$n_1 = 50$				$n_2 = 100$				$n_1 = 50$	$n_2 = 100$
	ERR	EAE	ESD	ARSD	ERR	EAE	ESD	ARSD	ERR	ERR
$\hat{\beta}_0$	0.65	-0.270	0.60	0.53	0.67	-0.262	0.55	0.48	0.71	0.80
$\hat{\beta}_1$	0.09	0.008	0.10	0.14	0.08	0.005	0.05	0.06	0.10	0.08
$\hat{\beta}_2$	0.72	-1.458	0.42	0.41	0.74	-1.468	0.40	0.38	0.80	0.85

As we mentioned earlier, the ability to catch the τ- varying effect of the regressors is a great advantage of the quantile regression model. To assess this, we now consider a log-linear model with heteroscedastic errors. We generate lifetimes from the model

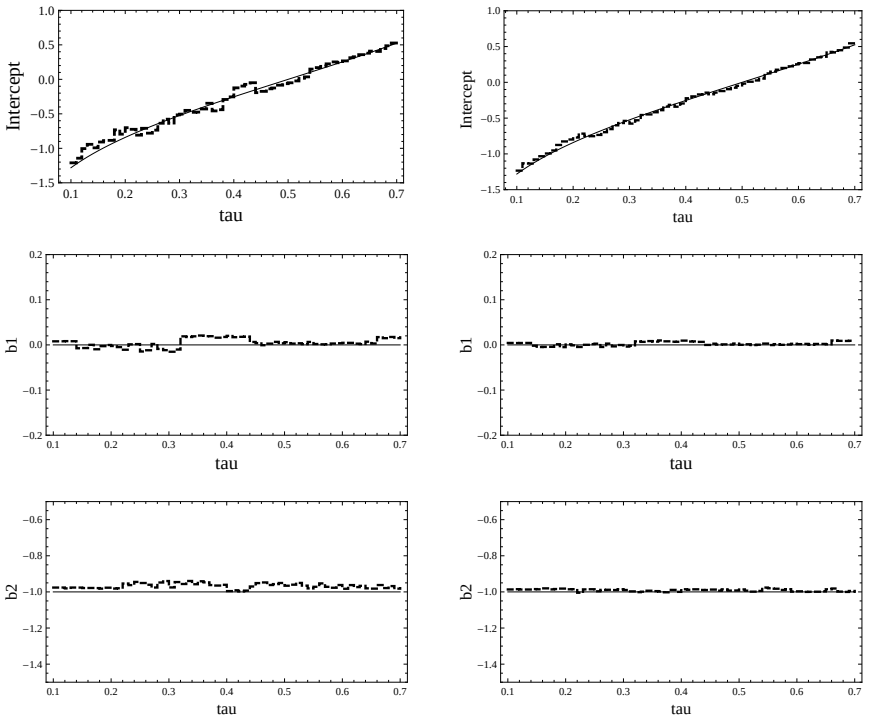
$$\tilde{T} = b_1 z_1 + b_2 z_2 \xi + \epsilon, \tag{6.10}$$

where z_1 and z_2 are covariate values as before, ξ follows exponential distribution with unit mean, and ϵ follows standard normal distribution. Note that the conditional variance of $\tilde{T}$ is given by $\text{Var}(\tilde{T}|z_1, z_2) = I(z_2 = 0) + (1+b_2)^2 I(z_2 = 1)$ and this is not homoscedastic when $b_2 \neq 0$. We follow the same procedure for data generation for model (6.10) as described for the AFT model setting and we choose the combination $c_1 = 0.15$, $c_2 = 5$, $c_3 = 0.25$ and $c_4 = 3$ for mild censoring and same combination except that $c_4 = 5$ for heavy censoring. We fix $b_1 = 0$ and $b_2 = -1.5$. We see that the model assumption given in (6.1) holds good for model (6.10) with $g(x) = e^x$ and $\boldsymbol{\beta_0}(\tau) = (\beta_0(\tau), \beta_1(\tau), \beta_2(\tau))^\top$, where $\beta_0(\tau) = Q_\epsilon(\tau)$, $\beta_1(\tau) = b_1 = 0$ and $\beta_2(\tau) = Q_{b_2\xi+\epsilon}(\tau) - Q_\epsilon(\tau)$. Tables 6.3 - 6.4 show the results obtained for regression quantile estimation. It is evident from the tables that biases are small, the empirical standard deviations and resampling based standard deviations are close with each other and coverage probabilities are close to 1. Here also we can observe that the bias, AERSD and ESD decrease when the sample size increases. Further, when the censoring rate increases they increase, while the CP decreases, as in the case of AFT model. Tables 6.7 - 6.8 present the results of testing of hypotheses concerning significance of the regression quantile process and constancy of the covariate effects. It is seen that the probability of type-I error is close to nominal value 0.05, the estimated average covariate effects of z_1 and z_2 are close to the true values and the average resampling-based standard deviation and empirical standard deviations are

in agreement. Moreover, we can observe that the ERR shows better results with increase in sample size, but it worsens with increase in censoring percentage, as we observed in the case of AFT model. Further, the presence of heteroscedasticity is well captured by our estimation procedure which is evident from plots of estimates of coefficient b_2 in Figures 6.3 - 6.4. The estimates of intercept term and b_1 are also displayed. The mean estimated effects obtained from the proposed method (in dashed lines) and the true coefficients (in continuous lines) are close to each other, suggesting a satisfactory performance under the finite sample setting.

6.4 Data Analysis

The proposed quantile regression method is applied to a real data set studied by Copelan et al. (1991). The study assesses the impact of bone marrow transplantation, which is a standard treatment for acute leukemia patients in prolonging their life lengths. A brief discussion about the study is given in Klein and Moeschberger (2005) and a data set consisting of 137 leukemia patients who are treated with bone marrow transplantation undergoing various treatment conditions and disease conditions is provided therein. The original study is concerned over the recovery following bone marrow transplantation, which is depending on several risk factors known at the time of transplantation such as age at transplantation, patient's gender, waiting time from the diagnosis to transplantation etc. as well as the post-transplantation history of the patients, like the possible development of acute graft-versus-host disease that typically occurs within the first 100 days following transplantation etc. The study was carried out with allogeneic marrow transplants for 99 patients with acute myeloctic leukemia (AML) and 38 patients with acute lymphoblastic leukemia



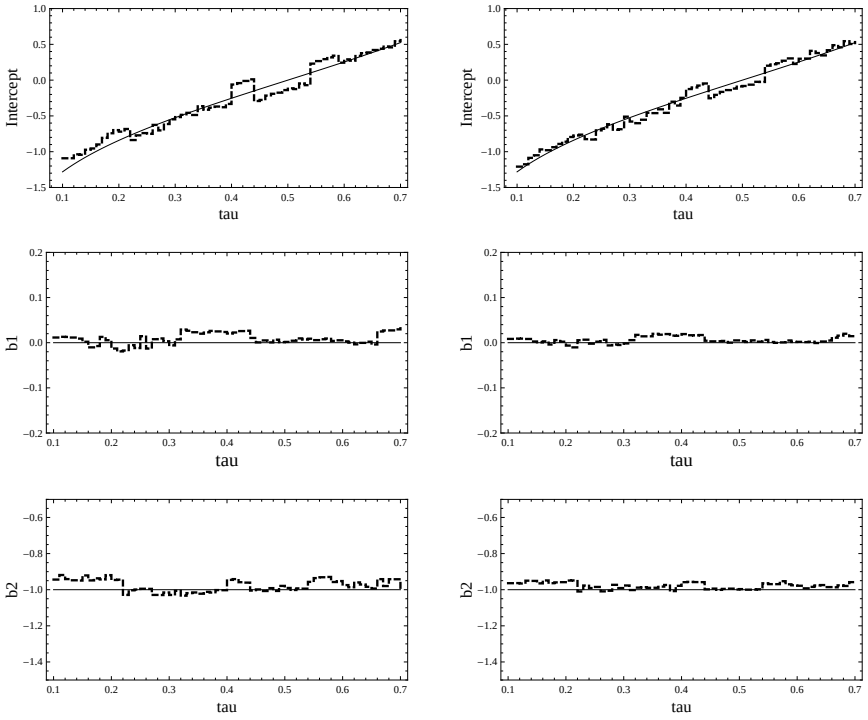
(a) $n = 50$ (b) $n = 100$

FIGURE 6.1: Plots of true coefficients (in continuous lines) and their estimated regression quantiles (in dashed lines) for AFT model under mild censoring.

(ALL). All these patients were treated in four different hospitals, two each in USA and Australia. Among them, patients in Australia were given a graft-versus-host prophylactic combining methotrexate (MTX) with cyclosporine whereas the other two hospitals didn't administer this prophylactic. Further, each patient was put in one among three disease groups- ALL group with 38 patients, AML low-risk group with 54 patients and AML high-risk group with 45 patients. For the illustration purpose we consider six risk factors which are listed below.

(i) Development status of acute graft-versus-host disease, denoted by GVHD, coded as 1 if developed and 0 if not developed,

(ii) Waiting time to transplant, denoted by WT, coded as 1 if waiting time is more than 1 year and 2 otherwise,



(a) $n = 50$ (b) $n = 100$

FIGURE 6.2: Plots of true coefficients (in continuous lines) and their estimated regression quantiles (in dashed lines) for AFT model under heavy censoring.

(iii) Age coded as 1 if age is less than 30 years and 2 otherwise,

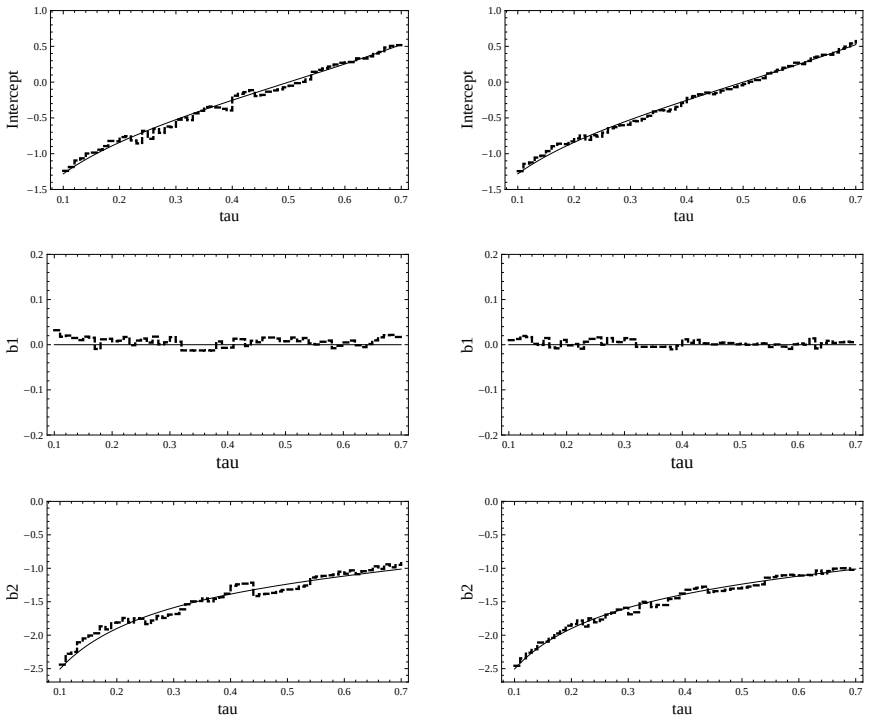
(iv) Patient gender, coded as 1 for female, 0 for male,

(v) Disease group, in three categories labeled 0 if patient belonging to ALL group, 1 if in AML low-risk group and 2 if in AML high-risk group, and

(vi) The treatment strategy with prophylactic, coded as 1 if MTX is given and 0 otherwise.

We now generate censoring intervals using the method described in Section 6.3, with $c_1 = 0.5$, $c_2 = 4.8$, $c_3 = 0.4$ and $c_4 = 3.5$. The resulting data set includes about 25% censored observations and the data is analyzed by the method discussed in Section 6.2 with $g(x) = x$, $x \in \mathbb{R}$.

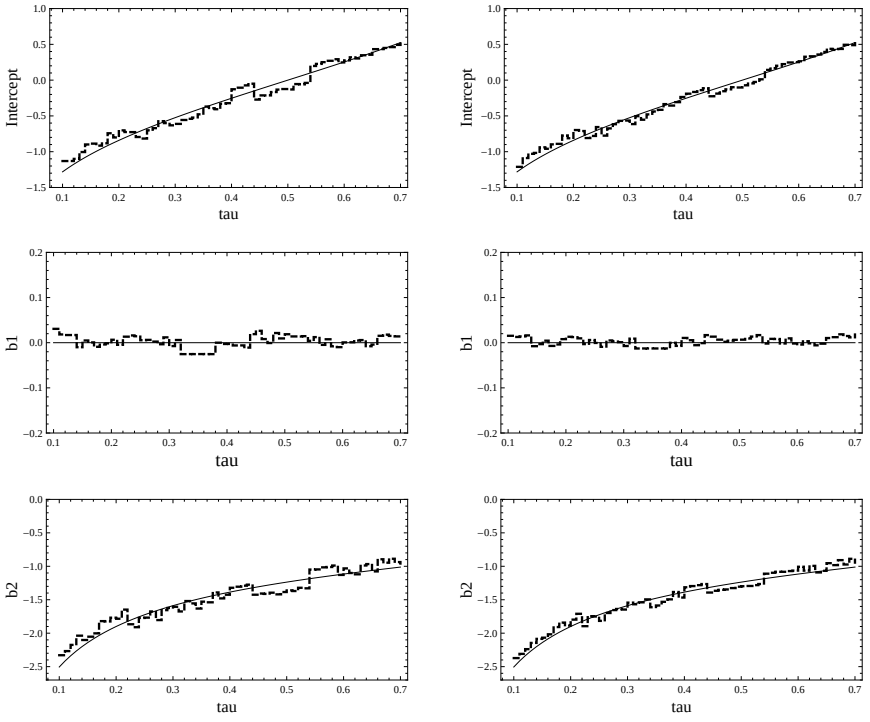
Figure 6.5 displays the estimated regression quantiles for $\boldsymbol{\beta}_0(\tau)$ for $\tau \in [l, u]$



(a) $n = 50$ (b) $n = 100$

FIGURE 6.3: Plots of true coefficients (in continuous lines) and their estimated regression quantiles (in dashed lines) for heteroscedastic model under mild censoring.

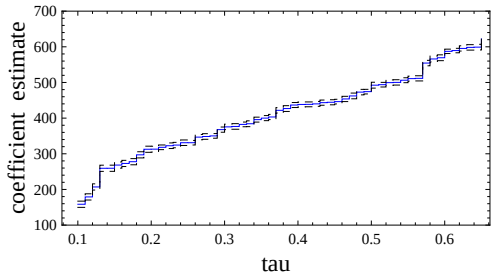
where $l = 0.1$ and $u = 0.65$ (in continuous lines) along with their 95% confidence intervals (in dashed lines) corresponding to model (6.1). The group of male patients of age less than 30 years, belonging to ALL group, having not developed GVHD, and not received prophylactic MTX is chosen as the reference comparison group. Figure 6.5(i) shows that about 10% of those patients belonging to the reference comparison group have died soon after 155 days from transplantation and 65% of them died soon after 580 days from transplantation. Figure 6.5(ii) shows negative values of coefficient corresponding to development of GVHD implying a life shortening effect of GVHD development. Similarly the longevity in waiting time from diagnosis to transplantation also affects negatively as shown in Figure 6.5(iii) and



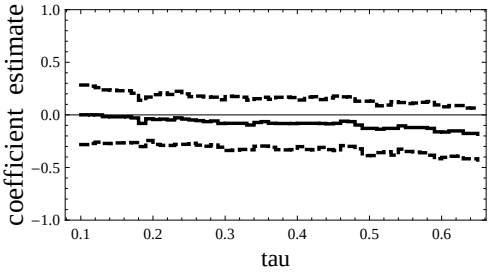
(a) $n = 50$ (b) $n = 100$

FIGURE 6.4: Plots of true coefficients (in continuous lines) and their estimated regression quantiles (in dashed lines) for heteroscedastic model under heavy censoring.

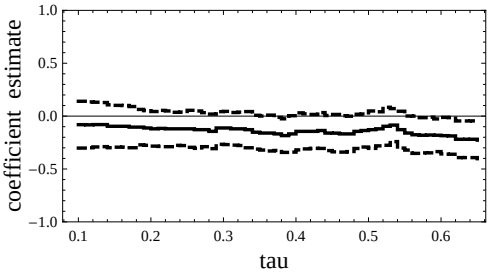
this effect worsens rapidly towards the upper quartile range. From Figure 6.5(iv) we observe that higher age is also a threat factor, but it behaves more or less similar for some ranges, soon after a decline, and then worsens faster. It is evident from Figure 6.5(v) that females tend to have shorter life as compared to males, as the corresponding coefficient lie below zero towards higher τ values. However, this gender difference is not prominent at an early stage. Figure 6.5(vi) conveys that the AML-low risk group has possibly a constant effect in many subintervals, which need to be tested for confirmation, but it is clear that this too has an adverse effect like its high-risk counterpart as shown in Figure 6.5(vii), except that the AML-high risk group tends to be more severe in regaining health after transplantation. Finally,



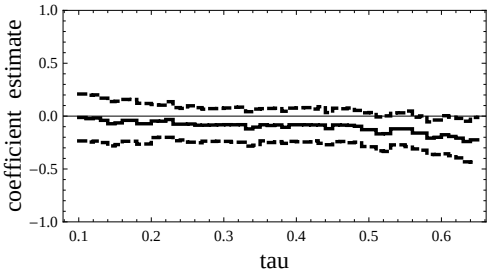
(i) Intercept



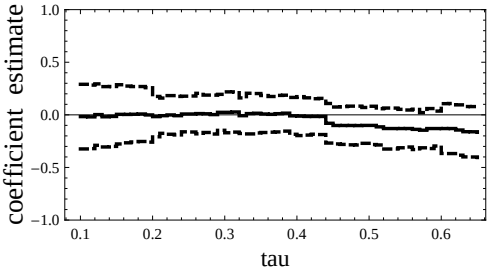
(ii) GVHD developed



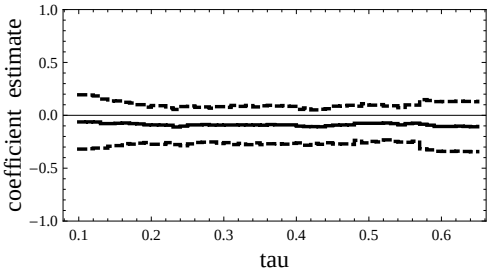
(iii) Waiting time more than one year



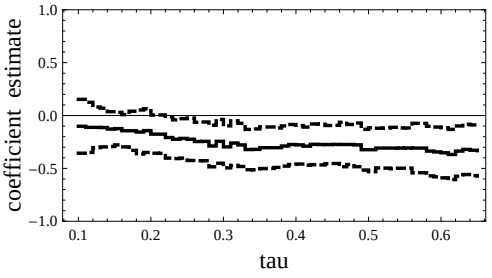
(iv) Age more than 30 years



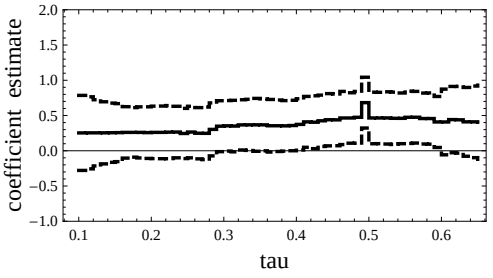
(v) Female



(vi) AML low-risk



(vii) AML high-risk



(viii) MTX given

FIGURE 6.5: Plots of estimated regression quantiles (continuous lines) and their 95% confidence intervals (in dashed lines).

it is evident from Figure 6.5(viii) that administration of MTX as a prophylactic is useful in improving the post-transplantation condition of patients and therefore its use can be advocated.

To validate these visual trends, we need to carry out formal statistical tests on the significance of covariate effects based on the average quantile effects across $\tau \in [0.1, 0.65]$ as described earlier. The adverse effects caused by GVHD and AML-low risk are marginally significant with P-values respectively 0.073 and 0.08 but all others are significant with P-values 0.032, 0.002, 0.007, < 0.001 and <0.001 respectively for the cases of WT more than 1 year, age more than 30 years, females, AML high-risk and MTX given. The constancy test is also performed with the weight function $\tilde{\Theta}(t) = I(t < (l+u)/2)$. All covariates except AML low-risk shows significant non-constancy, whereas AML low-risk shows a P-value of 0.089 and this is confirmed by altering the weight function to $\tilde{\Theta}(t) = I(t > (l+u)/2)$. We also checked the overall fit of the model (6.1) to our data set by using the martingale based diagnostic method discussed previously. Following Peng and Huang (2008), we choose the $q_0(\cdot)$ as a quadratic function of patients age, i.e., $q_0(Age) = ((Age - 28.78)^2/52)$ where 28.78 and 52 are the mean and maximum ages of those patients with exact lifetimes in the original dataset consisting of 137 patients. The supremum-norm lack of fit test described in Section-6.2 is carried out and obtained a P-value of 0.727, suggesting a reasonable fit for the data. This is confirmed with a quadratic weight function similarly defined for the covariate WT and the resulting P-value of 0.67 showing a reasonable model fit.

6.5 Conclusions

In this chapter, we developed a quantile regression model for a general censoring scheme called middle-censoring. The martingale feature of the observed data is utilized to develop inference procedures and to establish the asymptotic properties of the estimator of regression quantiles in Section 6.2. The proposed method made use of monotone estimating equations for sequentially solving the regression quantile process. Simulation studies in Section 6.3 indicated that the proposed methods work well with finite samples. In Section 6.4, the model was applied to a real data set.

The regression quantile estimator $\hat{\boldsymbol{\beta}}(\tau)$ is found to have larger variability for the values of τ close to zero. But this instability, as shown by Peng and Huang (2008), has no serious impact on the estimation at larger values of τ. In simulation studies, we have considered two models for the lifetime variable. One can also use several other models for this purpose and accordingly different weight functions may be used for hypotheses testing problems, provided such weight functions give higher power values.

Chapter 7

Bayesian Analysis of Middle-censored Lifetime Data

7.1 Introduction

Bayesian inference has become increasingly popular due to recent developments in computation, the ability to fit a wide range of models and to produce intuitive interpretations of the results. It provides a convenient way of implementing the scientific method for learning about the survival model, that the experimenter is interested about. Bayesian approach provides a natural and effective way of incorporating prior information with data, within a solid decision theoretical framework and we utilize this prior information for future analysis. A concise description of basic concepts involved in Bayesian analysis was given in Chapter 1. When we gather new information about the unknown parameter, the posterior distribution is updated by treating the existing posterior as prior and all inferences logically follow from Bayes theorem. It provides inferences that are conditional on the data and are exact, without reliance on asymptotic approximation. Moreover, small sample inference procedures are developed as in the same manner as with a large sample. Another

[1]The results in this chapter have been communicated as entitled "Bayesian analysis of middle-censored lifetime data in the presence of covariates"(see Prasad and Sankaran (2017))

useful feature of Bayesian procedure is that it obeys the likelihood principle. In Chapter 2 we mentioned the advantage of using parametric models for analyzing lifetime data and stated the features of Weibull distribution in modeling a lifetime data. Several works appeared in literature on modeling and analysis of Weibull lifetime data. Recently, Pradhan and Kundu (2014) used the Weibull distribution to model lifetime data under the interval censoring scheme. Motivated by these, in the present chapter, we propose a Bayesian analysis of middle-censored lifetime data in the presence of covariates, where we assume that the lifetime variate follows Weibull distribution.

The rest of the chapter is organized as follows. In Section 7.2, we introduce a regression model for the lifetime variate T, whose baseline distribution is assumed to be Weibull. We consider different prior distributions and the estimation of the parameters is carried out. Section 7.3 reports simulation studies to evaluate the performance of the estimators with finite sample size. An application of the proposed method to a real life data is given in Section 7.4. The chapter ends with a brief conclusion in Section 7.5.

7.2 The Model and Inference Procedure

Let the lifetime variate T be middle-censored by the random censoring interval (U, V) as before. We assume that given the covariate $\boldsymbol{z}$, T is independent of (U, V). Thus we observe $(X, \delta, \boldsymbol{z})$, where $X = T$ if $\delta = 1$, and $X = (U, V)$ if $\delta = 0$. We assume that the baseline distribution of T is Weibull with shape parameter α

and scale parameter γ with probability density function given by

$$f_0(t) = \alpha\gamma t^{\alpha-1}\exp(-\gamma t^{\alpha}), \quad t > 0;\ \alpha, \gamma > 0. \tag{7.1}$$

We introduce the covariate effect through the scale parameter by putting $\gamma = \exp(\boldsymbol{z}^{\top}\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is the unknown $p \times 1$ vector of regression parameters. Thus the probability density function of T given $\boldsymbol{z}$ is given by

$$f(t|\boldsymbol{z}) = \alpha \exp(\boldsymbol{z}^{\top}\boldsymbol{\theta})t^{\alpha-1}\exp(-\exp(\boldsymbol{z}^{\top}\boldsymbol{\theta})t^{\alpha}), \quad t > 0;\ \alpha > 0, \boldsymbol{\theta} \in \mathbb{R}_p. \tag{7.2}$$

The observed data consists of n independent and identically distributed replicates $(X_i, \delta_i, \boldsymbol{z}_i)$ for $i = 1, 2, ..., n$, of $(X, \delta, \boldsymbol{z})$. The likelihood function corresponding to the observed data is given by

$$L(\boldsymbol{\psi}|\text{data}) = \prod_{i=1}^{n} f(t_i|\boldsymbol{z}_i)^{\delta_i}[S(u_i|\boldsymbol{z}_i) - S(v_i|\boldsymbol{z}_i)]^{1-\delta_i}, \tag{7.3}$$

where $\boldsymbol{\psi} = (\alpha, \boldsymbol{\theta}^{\top})^{\top}$. Using (7.2), the likelihood function in (7.3) can be rewritten as

$$\begin{aligned} L(\boldsymbol{\psi}|\text{data}) = \prod_{i=1}^{n_1} \alpha \exp(\boldsymbol{z}_i^{\top}\boldsymbol{\theta})t_i^{\alpha-1}\exp(-\exp(\boldsymbol{z}_i^{\top}\boldsymbol{\theta})t_i^{\alpha}) \cdot \\ \prod_{i=n_1+1}^{n} \left[\exp(-\exp(\boldsymbol{z}_i^{\top}\boldsymbol{\theta})u_i^{\alpha}) - \exp(-\exp(\boldsymbol{z}_i^{\top}\boldsymbol{\theta})v_i^{\alpha})\right], \end{aligned} \tag{7.4}$$

where we assumed that first n_1 observations are exact lifetimes and remaining n_2 are censored intervals, with $n_1 + n_2 = n$.

We now make appropriate prior specifications for the parameters α and $\boldsymbol{\theta}$. The parameter α shall be assigned with the distribution of any nonnegative continuous

valued random variable. Common choices include informative priors like gamma or beta distributions or any non-informative prior like an improper density. There is no range restriction for $\boldsymbol{\theta}$ and hence we may choose, for example, p-variate normal distribution as its prior distribution. Denote these distributions respectively by $\pi_1(\cdot)$ and $\pi_2(\cdot)$. Then the joint posterior distribution may be written as

$$\pi(\boldsymbol{\psi}|\text{data}) \propto L(\boldsymbol{\psi}|\text{data})\pi_1(\alpha)\pi_2(\boldsymbol{\theta}). \tag{7.5}$$

We usually choose prior distributions in such a way that the resultant posterior distribution is easy to analyze. For the estimation of a parametric function of interest, say $\varphi(\boldsymbol{\psi})$, the decision-theoretic approach to statistical inference requires the specification of a loss function, which represents the loss incurred by estimating $\varphi(\boldsymbol{\psi})$ with a specified course of action, say ϱ. The Bayesian version of this approach leads to the minimization of the Bayes' risk. The most commonly used loss function is the quadratic loss function $(\varphi(\boldsymbol{\psi})-\varrho)^2$. For this loss function, the Bayes' estimator of $\varphi(\boldsymbol{\psi})$ is the posterior mean given by

$$E(\varphi(\boldsymbol{\psi})|\text{data}) = \int_{\mathscr{X}} \varphi(\boldsymbol{\psi})\pi(\boldsymbol{\psi}|\text{data})\text{d}\boldsymbol{\psi}, \tag{7.6}$$

where $\mathscr{X}$ is the parameter space given by $\mathbb{R}^+ \times \mathbb{R}^p$. In general, the closed form expression for the density given in (7.5) will not be available. This prompt us to resort to Monte Carlo simulation methods to infer about characteristics of the posterior distribution like the one given in (7.6).

In the present study, we exploit the importance sampling method (Robert and Casella (2013)) to evaluate the integral given in (7.6). The method calls for sampling from any arbitrary density $g_*(\cdot)$ whose support contains that of $\pi(\cdot)$. Then

the integral in (7.6) can be rewritten as

$$E(\varphi(\boldsymbol{\psi})|\text{data}) = \int_{\mathscr{X}} \varphi(\boldsymbol{\psi}) \frac{\pi(\boldsymbol{\psi}|\text{data})}{g_*(\boldsymbol{\psi})} g_*(\boldsymbol{\psi}) \text{d}\boldsymbol{\psi}. \tag{7.7}$$

This method, therefore, attracts considerable interest since it permits much freedom on the choice of the instrumental distribution $g_*(\cdot)$, which is usually chosen from among a class of distributions which are easy to simulate with the help of modern softwares. By applying the strong law of large numbers, the integral in (7.7) can be approximated by using a weighted sum as follows

$$E(\varphi(\boldsymbol{\psi})|\text{data}) \approx \frac{\sum_{j=1}^{m} \varphi(\boldsymbol{\psi}_j)\pi(\boldsymbol{\psi}_j|\text{data})/g_*(\boldsymbol{\psi}_j)}{\sum_{j=1}^{m} \pi(\boldsymbol{\psi}_j|\text{data})/g_*(\boldsymbol{\psi}_j)}, \tag{7.8}$$

where $\boldsymbol{\psi}_1, \boldsymbol{\psi}_2, ..., \boldsymbol{\psi}_m$ is a random sample drawn from $g_*(\boldsymbol{\psi})$. Note that the direct approximation $\frac{1}{m}\sum_{j=1}^{m} \varphi(\boldsymbol{\psi}_j)\pi(\boldsymbol{\psi}_j|\text{data})/g_*(\boldsymbol{\psi}_j)$ too converges almost surely to (7.7). However, the probability density function $g_*(\cdot)$ need to have thicker tail as compared to $\pi(\cdot)$ so as to have a finite posterior variance. The major advantage of approximation given in (7.8) is that it does not require any such assumption.

7.3 Simulation Studies

Simulation studies are carried out to assess the finite sample behavior of the estimator. In the present study, we consider a single covariate, say z, which is generated from uniform distribution over $[0, 10]$. Let θ be the corresponding regression parameter. We generate lifetimes T given z from Weibull distribution with shape parameter α and scale parameter $\exp(z\theta)$, having probability density function as

TABLE 7.1: Bias, SD, MSE and Len(CR) of the estimates with prior parameters $\mu_0 = 0.5, \sigma_0^2 = 100$.

		$n = 50$				$n = 100$			
α	θ	Bias	SD	MSE	Len(CR)	Bias	SD	MSE	Len(CR)
Mild censoring									
1.10	-1.00	0.017	0.77	0.0033	0.084 (0.925)	0.015	0.69	0.0019	0.073 (0.926)
1.10	1.00	0.030	0.67	0.0053	0.096 (0.938)	0.025	0.64	0.0019	0.088 (0.941)
2.25	-0.50	0.036	0.59	0.0059	0.085 (0.957)	0.033	0.50	0.0016	0.081 (0.958)
2.25	0.05	0.034	0.37	0.0065	0.086 (0.894)	0.031	0.35	0.0017	0.081 (0.895)
3.00	1.50	0.031	0.93	0.0036	0.079 (0.947)	0.027	0.87	0.0009	0.072 (0.951)
3.00	0.30	0.003	0.40	0.0046	0.073 (0.936)	0.001	0.29	0.0015	0.069 (0.938)
0.85	-0.05	0.030	0.55	0.0024	0.093 (0.895)	0.027	0.53	0.0011	0.087 (0.899)
0.85	0.10	0.031	0.79	0.0062	0.079 (0.934)	0.027	0.63	0.0017	0.070 (0.938)
4.00	0.04	0.032	0.17	0.0042	0.115 (0.957)	0.030	0.11	0.0015	0.112 (0.958)
4.00	-0.20	0.043	0.27	0.0039	0.051 (0.932)	0.040	0.19	0.0009	0.041 (0.935)
Moderate censoring									
1.10	-1.00	0.021	0.90	0.0050	0.085 (0.921)	0.016	0.88	0.0036	0.078 (0.925)
1.10	1.00	0.034	0.84	0.0084	0.098 (0.937)	0.026	0.80	0.0031	0.091 (0.940)
2.25	-0.50	0.039	0.71	0.0075	0.086 (0.954)	0.035	0.65	0.0029	0.083 (0.956)
2.25	0.05	0.037	0.64	0.0089	0.088 (0.891)	0.033	0.63	0.0030	0.082 (0.894)
3.00	1.50	0.036	1.05	0.0057	0.080 (0.944)	0.029	1.00	0.0020	0.075 (0.949)
3.00	0.30	0.008	0.66	0.0087	0.074 (0.933)	0.003	0.60	0.0032	0.073 (0.937)
0.85	-0.05	0.033	0.81	0.0069	0.094 (0.890)	0.028	0.72	0.0030	0.091 (0.898)
0.85	0.10	0.032	0.83	0.0077	0.080 (0.931)	0.029	0.74	0.0035	0.074 (0.936)
4.00	0.04	0.036	0.44	0.0056	0.117 (0.954)	0.032	0.37	0.0032	0.113 (0.956)
4.00	-0.20	0.046	0.61	0.0070	0.052 (0.927)	0.042	0.51	0.0025	0.045 (0.933)
Heavy censoring									
1.10	-1.00	0.026	0.94	0.0088	0.086 (0.916)	0.018	0.90	0.0047	0.081 (0.923)
1.10	1.00	0.036	0.98	0.0118	0.099 (0.935)	0.028	0.94	0.0050	0.095 (0.939)
2.25	-0.50	0.042	1.05	0.0125	0.088 (0.953)	0.037	0.88	0.0041	0.087 (0.955)
2.25	0.05	0.041	0.79	0.0106	0.090 (0.887)	0.035	0.71	0.0043	0.087 (0.892)
3.00	1.50	0.038	1.21	0.0094	0.082 (0.942)	0.031	1.05	0.0031	0.078 (0.948)
3.00	0.30	0.011	0.78	0.0130	0.076 (0.929)	0.005	0.74	0.0046	0.075 (0.935)
0.85	-0.05	0.037	0.99	0.0081	0.096 (0.888)	0.029	0.85	0.0048	0.092 (0.896)
0.85	0.10	0.035	0.83	0.0089	0.082 (0.928)	0.030	0.80	0.0054	0.077 (0.935)
4.00	0.04	0.039	0.91	0.0088	0.118 (0.952)	0.033	0.78	0.0042	0.116 (0.954)
4.00	-0.20	0.050	0.92	0.0088	0.053 (0.923)	0.043	0.88	0.0040	0.049 (0.931)

TABLE 7.2: Bias, SD, MSE and Len(CR) of the estimates with prior parameters $\mu_0 = -0.5, \sigma_0^2 = 100$.

		$n = 50$				$n = 100$			
α	θ	Bias	SD	MSE	Len(CR)	Bias	SD	MSE	Len(CR)
Mild censoring									
1.00	-0.500	0.013	0.44	0.0051	0.109 (0.907)	0.010	0.28	0.0014	0.099 (0.910)
1.00	1.000	0.041	0.51	0.0058	0.091 (0.889)	0.036	0.26	0.0009	0.085 (0.891)
2.00	-0.080	0.009	0.44	0.0030	0.077 (0.925)	0.004	0.25	0.0008	0.072 (0.929)
2.00	0.050	0.030	0.56	0.0039	0.108 (0.903)	0.028	0.30	0.0012	0.100 (0.907)
2.75	-0.010	0.026	0.49	0.0031	0.055 (0.928)	0.023	0.28	0.0012	0.047 (0.930)
2.75	0.300	0.029	0.51	0.0058	0.053 (0.949)	0.026	0.31	0.0019	0.048 (0.951)
3.50	0.025	0.029	0.47	0.0026	0.117 (0.922)	0.025	0.34	0.0010	0.107 (0.925)
3.50	-0.750	0.033	0.52	0.0025	0.053 (0.928)	0.031	0.32	0.0011	0.047 (0.930)
5.00	0.010	0.015	0.50	0.0052	0.079 (0.941)	0.011	0.42	0.0017	0.069 (0.943)
5.00	-0.250	0.018	0.47	0.0058	0.126 (0.932)	0.014	0.37	0.0016	0.120 (0.934)
Moderate censoring									
1.00	-0.500	0.014	0.57	0.0075	0.110 (0.904)	0.012	0.39	0.0034	0.102 (0.909)
1.00	1.000	0.045	0.73	0.0094	0.093 (0.887)	0.037	0.36	0.0025	0.087 (0.890)
2.00	-0.080	0.013	0.56	0.0042	0.079 (0.923)	0.006	0.45	0.0023	0.075 (0.927)
2.00	0.050	0.033	0.65	0.0088	0.110 (0.901)	0.029	0.52	0.0023	0.105 (0.906)
2.75	-0.010	0.028	0.60	0.0057	0.056 (0.926)	0.025	0.42	0.0028	0.050 (0.928)
2.75	0.300	0.032	0.64	0.0088	0.055 (0.945)	0.028	0.47	0.0038	0.051 (0.949)
3.50	0.025	0.032	0.57	0.0046	0.119 (0.920)	0.026	0.44	0.0025	0.111 (0.924)
3.50	-0.750	0.034	0.63	0.0064	0.054 (0.927)	0.033	0.52	0.0024	0.048 (0.928)
5.00	0.010	0.016	0.57	0.0095	0.080 (0.939)	0.012	0.50	0.0033	0.073 (0.941)
5.00	-0.250	0.022	0.74	0.0079	0.127 (0.930)	0.015	0.61	0.0026	0.123 (0.932)
Heavy censoring									
1.00	-0.500	0.018	0.67	0.0116	0.111 (0.900)	0.013	0.52	0.0049	0.107 (0.908)
1.00	1.000	0.049	0.83	0.0139	0.095 (0.883)	0.038	0.55	0.0042	0.091 (0.889)
2.00	-0.080	0.015	0.71	0.0067	0.081 (0.921)	0.007	0.49	0.0038	0.079 (0.925)
2.00	0.050	0.036	0.76	0.0111	0.112 (0.896)	0.031	0.49	0.0035	0.108 (0.904)
2.75	-0.010	0.029	0.74	0.0097	0.057 (0.922)	0.027	0.56	0.0047	0.052 (0.927)
2.75	0.300	0.033	0.81	0.0115	0.056 (0.943)	0.030	0.57	0.0057	0.054 (0.947)
3.50	0.025	0.036	0.76	0.0091	0.120 (0.915)	0.028	0.63	0.0041	0.115 (0.923)
3.50	-0.750	0.036	0.80	0.0086	0.056 (0.924)	0.034	0.78	0.0042	0.052 (0.927)
5.00	0.010	0.019	0.64	0.0120	0.081 (0.936)	0.014	0.61	0.0046	0.078 (0.939)
5.00	-0.250	0.026	0.86	0.0113	0.128 (0.928)	0.017	0.77	0.0044	0.127 (0.931)

TABLE 7.3: Bias, SD, MSE and Len(CR) of the estimates with prior parameters $m_0 = 0.75, p_0 = 2, \mu_0 = -0.5, \sigma_0^2 = 100$.

		$n = 50$				$n = 100$			
α	θ	Bias	SD	MSE	Len(CR)	Bias	SD	MSE	Len(CR)
Mild censoring									
0.85	-1.00	0.049	0.46	0.0028	0.117 (0.917)	0.046	0.31	0.0015	0.108 (0.919)
0.85	1.00	0.048	0.56	0.0055	0.105 (0.908)	0.044	0.28	0.0020	0.099 (0.910)
1.10	-0.50	0.012	0.32	0.0038	0.121 (0.917)	0.009	0.26	0.0019	0.118 (0.918)
1.10	0.05	0.013	0.50	0.0031	0.086 (0.928)	0.009	0.42	0.0015	0.084 (0.929)
2.25	1.50	0.005	0.37	0.0055	0.076 (0.907)	0.002	0.24	0.0020	0.074 (0.911)
2.25	0.30	0.017	0.43	0.0039	0.114 (0.940)	0.013	0.38	0.0015	0.111 (0.942)
3.00	-0.05	0.036	0.51	0.0065	0.109 (0.929)	0.032	0.33	0.0019	0.109 (0.931)
3.00	0.10	0.045	0.46	0.0031	0.103 (0.914)	0.043	0.35	0.0019	0.097 (0.916)
4.00	0.04	0.036	0.52	0.0058	0.079 (0.925)	0.035	0.45	0.0010	0.071 (0.929)
4.00	-0.20	0.011	0.57	0.0063	0.077 (0.948)	0.007	0.50	0.0020	0.072 (0.951)
Moderate censoring									
0.85	-1.00	0.050	0.69	0.0045	0.118 (0.913)	0.048	0.48	0.0026	0.111 (0.917)
0.85	1.00	0.050	0.64	0.0102	0.106 (0.904)	0.045	0.51	0.0032	0.102 (0.909)
1.10	-0.50	0.015	0.51	0.0088	0.123 (0.914)	0.011	0.73	0.0034	0.119 (0.917)
1.10	0.05	0.015	0.80	0.0080	0.088 (0.923)	0.010	0.69	0.0032	0.085 (0.928)
2.25	1.50	0.009	0.42	0.0101	0.078 (0.906)	0.003	0.61	0.0038	0.075 (0.909)
2.25	0.30	0.019	0.47	0.0068	0.115 (0.935)	0.014	0.42	0.0025	0.115 (0.941)
3.00	-0.05	0.037	0.67	0.0096	0.111 (0.926)	0.033	0.42	0.0030	0.110 (0.929)
3.00	0.10	0.050	0.57	0.0051	0.080 (0.911)	0.045	0.47	0.0039	0.074 (0.915)
4.00	0.04	0.042	0.74	0.0082	0.104 (0.922)	0.037	0.60	0.0023	0.100 (0.927)
4.00	-0.20	0.015	0.81	0.0076	0.078 (0.946)	0.009	0.63	0.0031	0.076 (0.949)
Heavy censoring									
0.85	-1.00	0.054	0.81	0.0061	0.119 (0.909)	0.050	0.60	0.0039	0.115 (0.915)
0.85	1.00	0.053	0.93	0.0123	0.108 (0.899)	0.046	0.90	0.0050	0.105 (0.907)
1.10	-0.50	0.017	0.98	0.0137	0.125 (0.911)	0.012	0.72	0.0045	0.121 (0.916)
1.10	0.05	0.019	0.87	0.0114	0.090 (0.922)	0.012	0.84	0.0046	0.087 (0.926)
2.25	1.50	0.014	0.81	0.0140	0.080 (0.905)	0.005	0.73	0.0054	0.076 (0.908)
2.25	0.30	0.022	0.77	0.0113	0.117 (0.933)	0.016	0.62	0.0044	0.116 (0.940)
3.00	-0.05	0.041	0.87	0.0122	0.113 (0.921)	0.035	0.67	0.0043	0.112 (0.928)
3.00	0.10	0.053	0.82	0.0091	0.081 (0.909)	0.046	0.95	0.0059	0.078 (0.913)
4.00	0.04	0.046	0.98	0.0107	0.106 (0.921)	0.038	0.91	0.0035	0.103 (0.926)
4.00	-0.20	0.017	0.90	0.0097	0.080 (0.943)	0.010	0.77	0.0042	0.079 (0.947)

TABLE 7.4: Bias, SD, MSE and Len(CR) of the estimates with prior parameters $m_0 = 0.25, p_0 = 1.5, \mu_0 = -0.5, \sigma_0^2 = 100$.

		$n = 50$				$n = 100$			
α	θ	Bias	SD	MSE	Len(CR)	Bias	SD	MSE	Len(CR)
Mild censoring									
1.00	-0.500	0.044	0.57	0.0032	0.078 (0.932)	0.040	0.42	0.0013	0.073 (0.933)
1.00	1.000	0.030	0.60	0.0059	0.057 (0.930)	0.026	0.33	0.0015	0.049 (0.931)
2.00	-0.080	0.017	0.46	0.0060	0.056 (0.955)	0.014	0.30	0.0018	0.049 (0.958)
2.00	0.050	0.026	0.67	0.0062	0.079 (0.890)	0.022	0.41	0.0019	0.074 (0.895)
2.75	-0.010	0.021	0.75	0.0058	0.047 (0.920)	0.017	0.52	0.0015	0.042 (0.921)
2.75	0.300	0.020	0.52	0.0037	0.092 (0.908)	0.018	0.27	0.0013	0.090 (0.912)
3.50	0.025	0.036	0.36	0.0030	0.077 (0.949)	0.032	0.29	0.0014	0.076 (0.950)
3.50	-0.750	0.013	0.49	0.0054	0.062 (0.947)	0.009	0.40	0.0017	0.058 (0.948)
5.00	0.010	0.012	0.78	0.0050	0.095 (0.958)	0.009	0.69	0.0010	0.090 (0.959)
5.00	-0.250	0.045	0.65	0.0052	0.121 (0.948)	0.041	0.51	0.0015	0.119 (0.952)
Moderate censoring									
1.00	-0.500	0.049	0.68	0.0073	0.080 (0.929)	0.041	0.56	0.0024	0.074 (0.931)
1.00	1.000	0.033	0.85	0.0084	0.058 (0.927)	0.027	0.48	0.0029	0.054 (0.929)
2.00	-0.080	0.022	0.60	0.0095	0.057 (0.950)	0.015	0.49	0.0029	0.050 (0.957)
2.00	0.050	0.028	0.83	0.0096	0.081 (0.886)	0.024	0.61	0.0033	0.077 (0.893)
2.75	-0.010	0.025	0.87	0.0072	0.049 (0.917)	0.019	0.58	0.0029	0.046 (0.920)
2.75	0.300	0.024	0.64	0.0053	0.094 (0.903)	0.019	0.45	0.0029	0.091 (0.911)
3.50	0.025	0.040	0.58	0.0079	0.079 (0.946)	0.034	0.46	0.0029	0.077 (0.948)
3.50	-0.750	0.018	0.54	0.0085	0.063 (0.944)	0.011	0.50	0.0036	0.062 (0.947)
5.00	0.010	0.016	0.86	0.0095	0.097 (0.955)	0.010	0.72	0.0023	0.093 (0.958)
5.00	-0.250	0.049	0.73	0.0093	0.123 (0.947)	0.043	0.55	0.0029	0.120 (0.950)
Heavy censoring									
1.00	-0.500	0.052	0.85	0.0115	0.081 (0.925)	0.042	0.69	0.0035	0.079 (0.930)
1.00	1.000	0.035	0.97	0.0124	0.060 (0.924)	0.029	0.66	0.0045	0.056 (0.928)
2.00	-0.080	0.023	0.89	0.0136	0.059 (0.947)	0.017	0.80	0.0044	0.055 (0.955)
2.00	0.050	0.032	0.88	0.0143	0.083 (0.885)	0.026	0.80	0.0050	0.080 (0.892)
2.75	-0.010	0.027	0.93	0.0119	0.051 (0.914)	0.020	0.77	0.0041	0.047 (0.918)
2.75	0.300	0.028	0.77	0.0070	0.096 (0.899)	0.021	0.65	0.0046	0.093 (0.910)
3.50	0.025	0.043	0.85	0.0125	0.081 (0.945)	0.035	0.61	0.0044	0.078 (0.946)
3.50	-0.750	0.021	0.91	0.0103	0.065 (0.940)	0.012	0.79	0.0050	0.064 (0.945)
5.00	0.010	0.019	0.91	0.0116	0.099 (0.952)	0.011	0.83	0.0034	0.098 (0.956)
5.00	-0.250	0.052	0.87	0.0124	0.124 (0.944)	0.045	0.75	0.0044	0.123 (0.949)

defined in (7.1), with γ replaced by $\exp(z\theta)$. Corresponding to each observation on T given z, a random censoring interval (U, V) is generated where U and $Y = V - U$ are assumed to be independent exponential variates with means $\lambda_1^{-1} = 15$ and $\lambda_2^{-1} = 10$ respectively. If we find $T \notin (U, V)$ then T is selected as the sample observation, otherwise we choose the interval (U, V) as the observation. We consider three different censoring rates viz., 10% (mild censoring), 20% (moderate censoring) and 30% (heavy censoring) for comparison purpose. Our aim is to obtain the Bayes' estimate of the regression parameter θ under the squared error loss function. The prior distribution for the parameter α is first assumed to be noninformative having density function given by $\pi_1(\alpha) \propto \alpha^{-1}, 0 < \alpha < \infty$. The prior density $\pi_2(\cdot)$ for θ is selected as normal with mean μ_0 and variance σ_0^2. Using these priors we evaluate the posterior mean and posterior standard deviation of the parameters by using the importance sampling method described in previous section with sample sizes $n = 50$ and $n = 100$. The pattern of results on simulation study seem to be similar for different combinations of prior parameters and we show the results for two such combinations in Tables 7.1 - 7.2. It is evident that the performance of the estimator is quite satisfactory in terms of average absolute bias (Bias), posterior standard deviation (SD) and estimated mean squared error (MSE) for different combinations of prior parameters. It can be observed that as the sample size increases, the bias, SD and MSE decrease. Moreover, when the censoring percentage increases the estimator becomes less efficient in terms of bias, SD and MSE, as expected. We also report average length of 95% credible intervals (Len) along with their coverage rates (CR). It is evident that as the sample size increases, the average length of the credible intervals decrease for all the three censoring percentages, while as the censoring rate increases, the average length of the credible intervals increase. In all these cases the

coverage percentages seem to be fairly large. Next we consider gamma prior with shape parameter m_0 and scale parameter p_0, having probability density function given by $\pi_1'(\alpha|m_0, p_0) \propto \alpha^{m_0-1}\exp(-p_0\alpha)$ for the parameter α. The parameter θ assumes the same normal prior specification as earlier. The results obtained while analyzing the same data sets using these priors with two specific combinations of the prior parameters are summarized in Tables 7.3 - 7.4. The analysis shows similar results as that in the case with noninformative prior. Other combinations of prior parameters also show similar inferences.

Selection of prior distributions are usually made in such a way that the resultant posterior density is easy to analyze, as mentioned earlier. However, in practical situations if they happen to depart much from target distribution, the convergence to the latter becomes slow and computations involved will be tedious. Similarly the selection of the candidate distribution requires careful attention. It is to be ensured that its support contains that of the posterior density. Moreover it may require enormous amount of computation to achieve an accurate approximations of the quantities of interest, if some theoretical conditions are not met (see Robert and Casella (2013), Section 3.3).

7.4 Illustrative Data Analysis

In this section, we apply our model to a real life data. We consider the data on survival times in months for 65 multiple myeloma patients studied by Krall et al. (1975). The complete data set is given in Lawless (2003). Following Jammalamadaka and Mangalam (2003), and Iyer et al. (2008), we first selected all exact lifetimes for our

TABLE 7.5: Estimate, SD and CI of parameters.

	Estimate	SD	CI
$\boldsymbol{\mu}_0 = (-0.5, -0.2)^\top, \boldsymbol{\Sigma}_0 = 100\mathcal{I}_2$			
α	1.102	0.302	(0.6174, 1.5986)
θ_1	0.092	0.159	(0.0642, 0.1232)
θ_2	0.102	0.288	(0.0281, 0.2165)
$\boldsymbol{\mu}_0 = (0.5, -0.1)^\top, \boldsymbol{\Sigma}_0 = 100\mathcal{I}_2$			
α	1.085	0.269	(0.6346, 1.6487)
θ_1	0.106	0.200	(0.0727, 0.1618)
θ_2	0.091	0.251	(0.0447, 0.1743)
$\boldsymbol{\mu}_0 = (1.0, 1.0)^\top, \boldsymbol{\Sigma}_0 = 100\mathcal{I}_2$			
α	0.991	0.238	(0.5813, 1.4908)
θ_1	0.097	0.161	(0.0476, 0.1064)
θ_2	0.094	0.270	(0.0207, 0.2281)

study. A Weibull model is fitted and the P-value is observed to be 0.8142 under the Kolmogorov-Smirnov test. Therefore this model assumption can not be ruled out. Motivated by the method adopted by Jammalamadaka and Mangalam (2003), and Iyer et al. (2008), we then artificially middle-censored 25% of the lifetimes as described in previous section with $\lambda_1^{-1} = 12$ and $\lambda_2^{-1} = 7$. This new data set consisting of exact lifetimes as well as censored intervals is further checked for its suitability under the Weibull model assumption as described below. We first consider a likelihood function at baseline level, given by $L_0(\alpha, \gamma) = \prod_{i=1}^{n_1} f_0(t_i) \prod_{i=n_1+1}^{n} (S_0(u_i) - S_0(v_i))$. The distribution parameters α and γ are estimated using the EM algorithm. The estimators thus obtained are denoted by $\hat{\alpha}_0$ and $\hat{\gamma}_0$. We then obtain an estimator of distribution function denoted by $\breve{F}_0(t)$ from the Nelson-Aalen-type estimator of baseline cumulative hazard function corresponding to the middle-censoring scheme as we did in Chapter 2. A PP-plot is drawn with the distribution function of Weibull distribution with parameters as $\hat{\alpha}_0$ and $\hat{\gamma}_0$, say $F_0(t; \hat{\alpha}_0, \hat{\gamma}_0)$ along the Y-axis and the distribution function $\breve{F}_0(t)$ along the X-axis. If the model assumption holds, then a straight line passing through origin making an angle 45 degrees with the X-axis is

expected. Figure 7.1 shows the graph obtained for the censored data set considered. The plot seems to be close to the line, which validate the distribution assumption.

Although there are five covariates in the original data, we take only two covariates viz., logarithm of blood urea nitrogen measurement, denoted by z_1, and serum calcium measurement at diagnosis, denoted by z_2. By our model assumption, we introduce the covariate effects through the scale parameter $\gamma_i = \exp(\boldsymbol{z}_i^\top \boldsymbol{\theta})$, where $\boldsymbol{z}_i = (z_{1i}, z_{2i})^\top$ denotes the observed covariate for the i'th individual and $\boldsymbol{\theta} = (\theta_1, \theta_2)^\top$ is the corresponding vector of regression parameters. We set $\varphi(\boldsymbol{\psi}) = (\theta_1, \theta_2)^\top$ and estimate this under the squared error loss function. The prior distribution of α is assumed to be noninformative as described in Section 7.3 and the prior distribution of $\boldsymbol{\theta}$ is taken to be bivariate normal with mean vector $\boldsymbol{\mu}_0$ and dispersion matrix $\boldsymbol{\Sigma}_0 = \sigma^2 \mathcal{I}_2$, where σ^2 is the prior variance of each regression parameter and $\mathcal{I}_2$ is the identity matrix of order 2. For finding the Bayes' estimate, we make use of the importance sampling method as before. We find that sampling from the candidate density $g_*(\cdot)$ becomes easy to implement if we let $g_*(\cdot)$ to be the product of two independent densities, say $g_1(\cdot)$ and $g_2(\cdot)$, where $g_1(\cdot)$ is a gamma density with shape parameter m and scale parameter p; and $g_2(\cdot)$ is a bivariate normal density with mean $\boldsymbol{\mu}$ and dispersion matrix $\boldsymbol{\Sigma}$. Various combinations of prior parameters are considered. We compute the estimate of each parameter, its posterior standard deviation (SD) and 95% credible intervals (CI). Table 7.5 shows the results obtained for few such combinations. The results show similarity among various choices of prior parameters. The estimate of α is close to one and the estimates of the regression parameters have positive values which suggest that the covariates have adverse effects on lifespan of individuals under study. Several other combinations of prior distributions are also used and they all show similar impact of covariates.

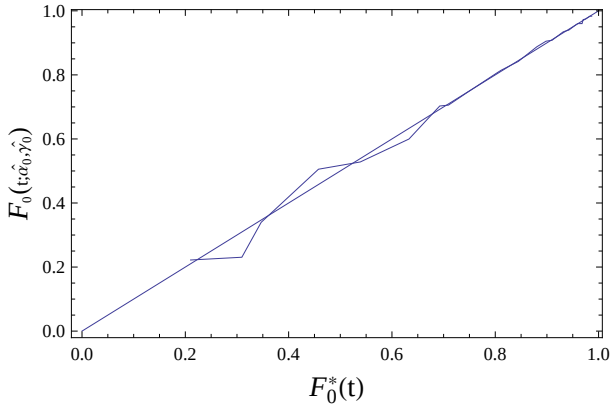


FIGURE 7.1: Plot of $F_0(t;\hat{\alpha}_0,\hat{\gamma}_0)$ against $\breve{F}_0(t)$.

7.5 Conclusion

In this chapter, we considered Bayesian analysis of lifetime data in the presence of covariates under middle-censored setup. We assumed a Weibull baseline distribution for the lifetime variable in Section 7.2. Non-informative as well as informative prior densities are considered for the shape parameter. Multivariate normal prior is selected for the regression parameter. Since the resulting posterior density seems to be difficult to sample from, we employed importance sampling method for obtaining the posterior summary. Extensive simulation studies were carried out in Section 7.3 to assess the finite sample properties of the estimator under different sample sizes with varying censoring rates. The simulation studies indicated satisfactory performance of the proposed method. In Section 7.4, we applied the model to a real life data concerned with lifetimes of multiple myeloma patients.

Chapter 8

Summary and Future Work

8.1 Summary

Middle-censored data arise naturally in many situations, where a random middle part on the support of a lifetime variate is missed to observe. The existing literature in middle-censored data analysis mainly focuses on the estimation of the distribution function of the lifetime variate without the presence of covariates. There are many situations where the lifetime data are observed along with covariates. The existing statistical methods for various censoring schemes such as left, right, interval, and double censoring are not appropriate in such contexts and therefore new regression models along with advanced methodology are required for the analysis of such a data. In view of this, in this study, we have proposed several parametric and semiparametric regression models for middle-censored lifetime data.

In Chapter 2, we have introduced proportional hazards regression model in parametric context and the regression parameters are estimated via the method of maximum likelihood. Asymptotic normality of the estimator was established. Extensive simulation studies were carried out to assess the performance of the estimator under finite sample setting. The procedure was applied to a real life data set.

In Chapter 3, we considered semiparametric approach for proportional hazards regression problem, where we only assumed that the baseline distribution function is absolutely continuous. We developed an iterative algorithm for estimating unknown baseline survival function and the regression parameters. Large sample properties of the estimators were established. Simulation studies were carried out, which indicated that the proposed estimators are performing well. The estimation procedure was illustrated through a real life dataset.

In Chapter 4, we presented a parametric estimation procedure for additive hazards regression model, where the baseline distribution was assumed to be exponentiated exponential. We proposed maximum likelihood estimator for the regression parameter and established its asymptotic normality. We assessed the performance of the estimator by conducting simulation studies. The proposed method was applied to a real life data set.

In Chapter 5, we introduced a semiparametric version of additive hazards regression problem. Two different methods of inference were proposed. Large sample properties of the estimators were established. Simulation studies were carried out to assess the finite sample performance of the estimators. The utility of the proposed methods was demonstrated through a real data set.

In Chapter 6, we proposed a quantile regression model for analyzing middle-censored lifetime data. A grid-based sequential estimation method for estimating the regression quantiles was developed by using the martingale structure of the observed data. Second stage inference procedures were developed using resampling technique. Asymptotic properties of the estimator were established. The estimator

possesses desirable properties such as consistency and asymptotic normality. A simulation study was conducted to examine the finite sample behavior of the estimator. The proposed method was well demonstrated using a real data set.

Finally in Chapter 7, we have introduced and studied Bayesian approach for analyzing middle-censored lifetime data, where we related the covariates with the lifetime variate through the scale parameter. The estimation of parameters were carried out via Monte Carlo simulation method. Simulation studies were carried out and the proposed model was applied to a real life data set.

8.2 Future Work

In many occasions, there may arise multiple causes of failure, which can be observed along with covariates or explanatory variables. Competing risks models can be employed is such contexts. The proposed regression models can be extended to this setup, which merits a future research.

The well known class of parametric models in survival analysis is log-location scale family of distributions. The parametric regression models, which we have considered in Chapter 2 and Chapter 4 can be generalized by choosing log-location scale type distribution as the baseline lifetime model.

Frailty or random effect models are often employed in survival analysis to model unobserved covariate, to incorporate the intra-class association, serial correlation and other forms of dependence. Frailty models are derived under conditional independence assumption, by specifying a latent variable that act multiplicatively

on the baseline hazard (see Hougaard (2000)). Various parametric families of frailty models may be generated by taking different distributional form for frailty random variable. Inference procedure for such frailty models are complicated due to censoring. The semiparametric regression models described here can be extended to frailty setup under middle-censoring, which is another area yet to be studied.

Throughout this study, we assumed that the lifetime variate and censoring interval are independently distributed, given covariates. In practice, this assumption may seem to be restrictive and in such instances one can relax this requirement, which leads to a dependent setup. The regression models for middle-censored data under such a dependent setup is not yet discussed, and this can be explored in future research.

In medical research, often our interest may be in inferring the remaining life years of a patient, given the patient history. The remaining life years of a patient could be prolonged by treating or preventing a disease by a medical intervention. Existing statistical methods based on hazard function are not appropriate in such contexts as these methods are often cumbersome and not straightforward, especially when the remaining lifetimes need to be evaluated in the middle of an observation period. Further, the remaining lifetimes at a specific time point estimated from such semiparametric models may heavily depend on model assumptions that affect the entire observation period. Motivated by this, researchers have developed regression models based on mean remaining life in literature. For some fundamental works on this topic, one could refer to Zahedi (1991), Oakes and Dasu (2003), Maguluri and Zhang (1994), and Chen and Cheng (2005). These works can be extended to

the case of lifetime data subject to middle-censoring, which will be reported in a separate work.

For a right censored survival data, Jung et al. (2009) developed a quantile remaining life regression model and studied its properties. In Chapter 6, we have developed a quantile regression model for analyzing middle-censored survival data. The method developed there can further be extended to accommodate quantile remaining life, which is another area of interest for future research.

List of Published/Accepted Papers

1. Jammalamadaka, S. R., Prasad, S., and Sankaran, P. G. (2016). A semiparametric regression model for analysis of middle-censored lifetime data. *Statistica*, LXXVI(1), 27–40.

2. Sankaran, P. G., and Prasad, S. (2014). Weibull regression model for analysis of middle-censored lifetime data. *Journal of Statistics and Management Systems*, 17 (5-6), 433–443.

3. Sankaran, P. G., and Prasad, S. (2017a). Additive risks regression model for middle-censored exponentiated-exponential lifetime data. *Communications in Statistics - Simulation and Computation*, to appear.

4. Sankaran, P. G., and Prasad, S. (2017b). An additive risks regression model for the analysis of middle-censored lifetime data. *Statistics in Transition-new series*, to appear.

References

Aalen, O. O. (1989). A linear regression model for the analysis of lifetimes. *Statistics in Medicine*, *8*(8), 907–925.

Aalen, O. O. (1993). Further results on the nonparametric linear regression model in survival analysis. *Statistics in Medicine*, *12*(17), 1569–1588.

Abuzaid, A., Abu El-Qumsan, M., & El-Habil, A. (2017). On the robustness of right and middle censoring schemes in parametric survival models. *Communications in Statistics-Simulation and Computation*, *46*(3), 1771–1780.

Ahuja, J., & Nash, S. W. (1967). The generalized Gompertz-Verhulst family of distributions. *Sankhyā: The Indian Journal of Statistics, Series A*, *29*, 141–156.

Akaike, H. (1974). A new look at the statistical model identification. *Automatic Control, IEEE Transactions on Automatic Control*, *19*(6), 716–723.

Andersen, P. K., Borgan, O., Gill, R. D., & Keiding, N. (2012). *Statistical Models Based on Counting Processes*. Springer Science & Business Media, New York.

Andersen, P. K., & Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *The Annals of Statistics*, *10*(4), 1100–1120.

Aranda-Ordaz, F. J. (1983). An extension of the proportional hazards model for grouped data. *Biometrics*, *39*(1), 109–117.

Bain, L. (1976). *Statistical Analysis of Reliability and Life-testing Models*. Marcel and Dekker Inc., New York.

Beadle, G. F., Come, S., Henderson, I. C., Silver, B., Hellman, S., & Harris, J. R.

(1984). The effect of adjuvant chemotherapy on the cosmetic results after primary radiation treatment for early stage breast cancer. *International Journal of Radiation Oncology Biology Physics*, *10*(11), 2131–2137.

Bennett, N., Iyer, S. K., & Jammalamadaka, S. R. (2017). Analysis of gamma and Weibull lifetime data under a general censoring scheme and in the presence of covariates. *Communications in Statistics-Theory and Methods*, *46*(5), 2277–2289.

Borgan, Ø., & Langholz, B. (1997). Estimation of excess risk from case-control data using Aalen's linear regression model. *Biometrics*, *53*(2), 690–697.

Breslow, N. E. (1972). Discussion of paper of D. R. Cox. *Journal of Royal Statistical Society Series B*, *34*, 216–217.

Breslow, N. E., & Day, N. E. (1980). Conditional logistic regression for matched sets. *Statistical Methods in Cancer Research*, *1*, 248–279.

Buckley, J., & James, I. (1979). Linear regression with censored data. *Biometrika*, *66*, 429–436.

Cai, T., & Cheng, S. (2004). Semiparametric regression analysis for doubly censored data. *Biometrika*, *91*, 277–290.

Chang, M. N., & Yang, G. L. (1987). Strong consistency of a nonparametric estimator of the survival function with doubly censored data. *The Annals of Statistics*, *15*(4), 1536–1547.

Chen, K., & Zhou, M. (2003). Non-parametric hypothesis testing and confidence intervals with doubly censored data. *Lifetime Data Analysis*, *9*(1), 71–91.

Chen, Y. Q., & Cheng, S. (2005). Semiparametric regression analysis of mean residual life with censored survival data. *Biometrika*, *92*(1), 19–29.

Cheng, M.-H., Ibrahim, J. G., & Sinha, D. (1999). A new Bayesian model for

survival data with a surviving fraction. *Journal of the American Statistical Association*, *94*(447), 909–919.

Copelan, E. A., Biggs, J. C., Thompson, J. M., Crilley, P., Szer, J., Klein, J. P., ... Atkinson, K. (1991). Treatment for acute myelocytic leukemia with allogeneic bone marrow transplantation following preparation with BuCy2. *Blood*, *78*(3), 838–843.

Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B*, *34*, 187–220.

Cox, D. R. (1975). Partial likelihood. *Biometrika*, *62(2)*, 269–276.

Cox, D. R., & Oakes, D. (1984). *Analysis of Survival Data*. CRC Press, London.

Cox, D. R., & Snell, E. J. (1968). A general definition of residuals(with discussion). *Journal of the Royal Statistical Society. Series B*, *30*, 248–275.

Danish, M. Y., & Arshad, I. A. (2017). Survival analysis with parametric proportional hazards model of random censorship. *American Journal of Mathematical and Management Sciences*, *36*(1), 66–83.

Davarzani, N., & Parsian, A. (2011). Statistical inference for discrete middle-censored data. *Journal of Statistical Planning and Inference*, *141*(4), 1455–1462.

Davarzani, N., Parsian, A., & Peeters, R. (2015). Statistical inference on middle-censored data in a dependent setup. *Journal of Statistical Theory and Practice*, *9*(3), 646–657.

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application* (Vol. 1). Cambridge university press.

De Gruttola, V., & Lagakos, S. W. (1989). Analysis of doubly-censored survival data, with application to AIDS. *Biometrics*, *45*, 1–11.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B*, *39*(1), 1–38.

Dörre, A., & Weißbach, R. (2017). Bayesian estimation of a proportional hazards model for double-censored durations. *Journal of Statistical Computation and Simulation*, *87*(3), 493–504.

Efron, B. (1967). The two sample problem with censored data. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 4, pp. 831–853).

Efron, B. (1982). *The jackknife, the bootstrap, and other resampling plans* (Vol. 38). Siam.

Efron, B., & Hinkley, D. V. (1978). Assessing the accuracy of the maximum likelihood estimator: Observed versus expected fisher information. *Biometrika*, *65*(3), 457–483.

Farrington, C. (2000). Residuals for proportional hazards models with interval-censored survival data. *Biometrics*, *56*(2), 473–482.

Feigl, P., & Zelen, M. (1965). Estimation of exponential survival probabilities with concomitant information. *Biometrics*, *21*(4), 826–838.

FitzSimmons, S. C. (1993). The changing epidemiology of cystic fibrosis. *The Journal of Pediatrics*, *122*(1), 1–9.

Fygenson, M., & Ritov, Y. (1994). Monotone estimating equations for censored data. *The Annals of Statistics*, *22*, 732–746.

Geskus, R. B., & Groeneboom, P. (1996). Asymptotically optimal estimation of smooth functionals for interval censoring. *Statistica Neerlandica*, *50*(1), 69–88.

Gilchrist, W. (2000). *Statistical Modelling with Quantile Functions*. CRC Press, London.

Glasser, M. (1967). Exponential survival with covariance. *Journal of the American Statistical Association*, *62*(318), 561–568.

Groeneboom, P., & Wellner, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation* (Vol. 19). Birkhäuser-Verlag, Basel.

Gupta, R. C., Gupta, P. L., & Gupta, R. D. (1998). Modeling failure time data by Lehman alternatives. *Communications in Statistics - Theory and Methods*, *27*(4), 887-904.

Gupta, R. D., & Kundu, D. (1999). Theory and methods: Generalized exponential distributions. *Australian and New Zealand Journal of Statistics*, *41*(2), 173–188.

Gupta, R. D., & Kundu, D. (2001). Generalized exponential distribution: different method of estimations. *Journal of Statistical Computation and Simulation*, *69*(4), 315–337.

Hougaard, P. (2000). *Analysis of Multivariate Survival Data*. Springer Verlag, New York.

Huang, J., & Wellner, J. A. (1995). Efficient estimation for the proportional hazards model with" case 2" interval censoring. *Technical Report No. 290, Department of Statistics, University of Washington, Seattle, USA*.

Ichida, J., Wassell, J., Keller, M., & Ayers, L. (1993). Evaluation of protocol change in burn-care management using the Cox proportional hazards model with time-dependent covariates. *Statistics in Medicine*, *12*, 301–310.

Iyer, S. K., Jammalamadaka, S. R., & Kundu, D. (2008). Analysis of middle-censored data with exponential lifetime distributions. *Journal of Statistical*

Planning and Inference, *138*(11), 3550–3560.

Jammalamadaka, S. R., & Iyer, S. K. (2004). Approximate self consistency for middle-censored data. *Journal of Statistical Planning and Inference*, *124*(1), 75–86.

Jammalamadaka, S. R., & Leong, E. (2015). Analysis of discrete lifetime data under middle-censoring and in the presence of covariates. *Journal of Applied Statistics*, *42*(4), 905–913.

Jammalamadaka, S. R., & Mangalam, V. (2003). Nonparametric estimation for middle-censored data. *Journal of Nonparametric Statistics*, *15*(2), 253–265.

Jammalamadaka, S. R., & Mangalam, V. (2009). A general censoring scheme for circular data. *Statistical Methodology*, *6*(3), 280–289.

Jammalamadaka, S. R., Prasad, S., & Sankaran, P. G. (2016). A semiparametric regression model for analysis of middle censored lifetime data. *Statistica*, *LXXVI*(1), 27–40.

Ji, S., Peng, L., Cheng, Y., & Lai, H. (2012). Quantile regression for doubly censored data. *Biometrics*, *68*(1), 101–112.

Jin, Z., Ying, Z., & Wei, L. (2001). A simple resampling method by perturbing the minimand. *Biometrika*, *88*(2), 381–390.

Jung, S.-H., Jeong, J.-H., & Bandos, H. (2009). Regression on quantile residual life. *Biometrics*, *65*(4), 1203–1212.

Kalbfleisch, J. D. (1978). Nonparametric Bayesian analysis of survival time data. *Journal of the Royal Statistical Society. Series B*, *40*(2), 214–221.

Kalbfleisch, J. D., & Prentice, R. L. (2011). *The Statistical Analysis of Failure Time Data*. John Wiley & Sons, New York.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete

observations. *Journal of the American Statistical Association*, *53*(282), 457–481.

Karduan, O. (1983). Statistical analysis of male larynx-cancer patients - a case study. *Statistical Nederlandica*, *37*, 103–126.

Kim, M. Y., De Gruttola, V. G., & Lagakos, S. W. (1993). Analyzing doubly censored data with covariates, with application to AIDS. *Biometrics*, *49*, 13–22.

Klein, J. P., & Moeschberger, M. L. (2005). *Survival Analysis: Techniques for Censored and Truncated Data*. Springer Science & Business Media, New York.

Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica*, *46*, 33–50.

Krall, J. M., Uthoff, V. A., & Harley, J. B. (1975). A step-up procedure for selecting variables associated with survival. *Biometrics*, *31*, 49–57.

Krishna, H., & Goel, N. (2017). Maximum likelihood and Bayes estimation in randomly censored geometric distribution. *Journal of Probability and Statistics, Hindawi*, *2017*.

Kundu, D., & Basu, A. (2004). *Statistical Computing: Existing Methods and Recent Developments*. Narosa, New Delhi.

Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data*. John Wiley & Sons, New York.

Le, C. T. (1997). *Applied Survival Analysis*. John Wiley and Sons, New York.

Lee, E. T., & Wang, J. (2003). *Statistical Methods for Survival Data Analysis*. John Wiley & Sons, New York.

Lee, K., & Cho, Y. (2017). Bayesian and maximum likelihood estimations of the inverted exponentiated half logistic distribution under progressive type II censoring. *Journal of Applied Statistics*, *44*(5), 811–832.

Li, P., & Ling, X.-l. (2012). The additive hazard mixing models. *Acta Mathematicae Applicatae Sinica, English Series*, *28*(1), 139–148.

Lin, D. Y., Wei, L.-J., & Ying, Z. (1993). Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika*, *80*(3), 557–572.

Lin, D. Y., & Ying, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika*, *81*(1), 61–71.

Lin, D. Y., & Ying, Z. (1997). Additive hazards regression models for survival data. *Proceedings of the First Seattle Symposium in Biostatistics*, 185–198.

Maguluri, G., & Zhang, C.-H. (1994). Estimation in the mean residual life regression model. *Journal of the Royal Statistical Society. Series B*, *56*, 477–489.

Mangalam, V., Nair, G. M., & Zhao, Y. (2008). On computation of NPMLE for middle-censored data. *Statistics & Probability Letters*, *78*(12), 1452–1458.

Martz, H. F., & Waller, R. (1982). *Bayesian Reliability Analysis.* John Wiley and Sons, New York.

McKeague, I. W., & Utikal, K. J. (1991). Goodness-of-fit tests for additive hazards and proportional hazards models. *Scandinavian Journal of Statistics*, *18*(3), 177–195.

McLachlan, G., & Krishnan, T. (2007). *The EM Algorithm and Extensions.* John Wiley & Sons, New York.

Nair, N. U., Sankaran, P. G., & Balakrishnan, N. (2013). *Quantile-based Reliability Analysis.* Springer-Birkhauser, New York.

Nash, T., Williams, J., & Machin, D. (1990). TENS: does the type of stimulus really matter. *Pain Clinic*, *3*(3), 161–168.

Nelson, W. (1969). Hazard plotting for incomplete failure data. *Journal of Quality Technology*, *1*(1), 27–52.

Oakes, D., & Dasu, T. (2003). Inference for the proportional mean residual life model. *The Institute of Mathematical Statistics Lecture Notes-Monograph Series*, *43*, 105–116.

Odell, P. M., Anderson, K. M., & D'Agostino, R. B. (1992). Maximum likelihood estimation for interval-censored data using a Weibull-based accelerated failure time model. *Biometrics*, *48*(3), 951–959.

Peng, L., & Huang, Y. (2008). Survival analysis with quantile regression models. *Journal of the American Statistical Association*, *103*, 637–649.

Portnoy, S. (2003). Censored regression quantiles. *Journal of the American Statistical Association*, *98*(464), 1001–1012.

Portnoy, S., & Lin, G. (2010). Asymptotics for censored regression quantiles. *Journal of Nonparametric Statistics*, *22*(1), 115–130.

Powell, J. L. (1984). Least absolute deviations estimation for the censored regression model. *Journal of Econometrics*, *25*(3), 303–325.

Powell, J. L. (1986). Censored regression quantiles. *Journal of Econometrics*, *32*(1), 143–155.

Pradhan, B., & Kundu, D. (2014). Analysis of interval-censored data with Weibull lifetime distribution. *Sankhya B*, *76*(1), 120–139.

Prasad, S., & Sankaran, P. G. (2017). Bayesian analysis of middle-censored lifetime data in the presence of covariates. *Communicated*.

Prentice, R. L. (1973). Exponential survivals with censoring and explanatory variables. *Biometrika*, *60*(2), 279–288.

Qin, C. (2010). Estimations of parameter of exponential distribution for middle-censored data. *Journal of Hubei Normal University (Natural Science)*, *1*, 012.

Ritov, Y. (1990). Estimation in a linear regression model with censored data. *The Annals of Statistics*, *18*, 303–328.

Robert, C., & Casella, G. (2013). *Monte Carlo Statistical Methods*. Springer Science & Business Media, New York.

Sankaran, P. G., & Prasad, S. (2014). Weibull regression model for analysis of middle-censored lifetime data. *Journal of Statistics and Management Systems*, *17*(5-6), 433–443.

Sankaran, P. G., & Prasad, S. (2017a). Additive risks regression model for middle censored exponentiated-exponential lifetime data. *Communications in Statistics-Simulation and Computation*, 1–12.

Sankaran, P. G., & Prasad, S. (2017b). An additive risks regression model for the analysis of middle-censored lifetime data. *Statistics in Transition-new series*, to appear.

Sankaran, P. G., & Prasad, S. (2017c). Quantile regression model for the analysis of middle-censored lifetime data. *Communicated*.

Shang, K., & Reilly, C. (2017). Nonparametric Bayesian analysis of the 2 sample problem with censoring. *Communications in Statistics-Theory and Methods*(to appear).

Shen, P. (2010). An inverse-probability-weighted approach to the estimation of distribution function with middle-censored data. *Journal of Statistical Planning and Inference*, *140*(7), 1844–1851.

Shen, P. (2011). The nonparametric maximum likelihood estimator for middle-censored data. *Journal of Statistical Planning and Inference*, *141*(7), 2494–2499.

Sinha, D., Chen, M.-H., & Ghosh, S. K. (1999). Bayesian analysis and model

selection for interval-censored survival data. *Biometrics*, *55*(2), 585–590.

Sinha, D., Ibrahim, J. G., & Chen, M.-H. (2003). A Bayesian justification of Cox's partial likelihood. *Biometrika*, *90*(3), 629–641.

Sinha, S. K. (1986). *Reliability and Life Testing*. Wiley Eastern Ltd, New Delhi.

Sun, J. (1995). Empirical estimation of a distribution function with truncated and doubly interval-censored data and its application to AIDS studies. *Biometrics*, *51*, 1096–1104.

Sun, J. (2004). Statistical analysis of doubly interval-censored failure time data. *Advances in Survival Analysis, Handbook of Statistics*, *23*, 105–122.

Sun, J. (2006). *The Statistical Analysis of Interval-censored Failure Time Data*. Springer Science & Business Media, New York.

Sun, L., Park, D., & Sun, J. (2006). The additive hazards model for recurrent gap times. *Statistica Sinica*, *16*(3), 919–932.

Tarpey, T., & Flury, B. (1996). Self-consistency: a fundamental concept in statistics. *Statistical Science*, *11*(3), 229–243.

Thomas, D. C. (1986). Use of auxiliary information in fitting nonproportional hazards models. *Modern Statistical Methods in Chronic Disease Epidemiology,Ed. S. H. Moolgavkar and R. L. Prentice*, 197–210.

Tsai, W.-Y., & Crowley, J. (1985). A large sample study of generalized maximum likelihood estimators from incomplete data via self-consistency. *The Annals of Statistics*, *13*, 1317–1334.

Turnbull, B. W. (1974). Nonparametric estimation of a survivorship function with doubly censored data. *Journal of the American Statistical Association*, *69*(345), 169–173.

Turnbull, B. W. (1976). The empirical distribution function with arbitrarily

grouped, censored and truncated data. *Journal of the Royal Statistical Society. Series B*, *38*, 290–295.

Turnbull, B. W., Brown Jr, B. W., & Hu, M. (1974). Survivorship analysis of heart transplant data. *Journal of the American Statistical Association*, *69*(345), 74–80.

Walker, S., & Mallick, B. K. (1999). A Bayesian semiparametric accelerated failure time model. *Biometrics*, *55*(2), 477–483.

Wang, L. (2016). Estimation for exponential distribution based on competing risk middle-censored data. *Communications in Statistics-Theory and Methods*, *45*(8), 2378–2391.

Wei, L., & Gail, M. (1983). Nonparametric estimation for a scale-change with censored observations. *Journal of the American Statistical Association*, *78*(382), 382–388.

Yu, Q., Wong, G. Y., & Li, L. (2001). Asymptotic properties of self-consistent estimators with mixed interval-censored data. *Annals of the Institute of Statistical Mathematics*, *53*(3), 469–486.

Zahedi, H. (1991). Proportional mean remaining life model. *Journal of Statistical Planning and Inference*, *29*(1-2), 221–228.

Zahl, P.-H., & Tretli, S. (1997). Long-term survival of breast cancer in Norway by age and clinical stage. *Statistics in Medicine*, *16*(13), 1435–1449.

Zippin, C., & Armitage, P. (1966). Use of concomitant variables and incomplete survival information in the estimation of an exponential survival parameter. *Biometrics*, *22*(4), 665–672.