ABSTRACT

Title of Dissertation: PROPORTIONAL HAZARDS MODEL FOR RIGHT CENSORED SURVIVAL DATA WITH LONGITUDINAL COVARIATES

Yuyin Shi
Doctor of Philosophy, 2023

Dissertation Directed by: Professor Joan Jian-Jian Ren
Department of Mathematics

The proportional hazards model is one of the most widely used tools in analyzing survival data. In medical and epidemiological studies, the interrelationship between time-to-event variable and longitudinal covariates is often the primary research interest. Thus, joint modeling of survival data and longitudinal data has received very much attention in statistical literature, but it's a considerably difficult problem due to censoring on the survival time and that the longitudinal covariate process is in fact a completely unknown and not completely observed stochastic process. Up to now, all existing works made parametric or semi-parametric assumptions on the longitudinal covariate process, and resulting inferences critically depends on validity of these not justifiable assumptions. This dissertation does not make any parametric or semi-parametric assumptions on the longitudinal covariate process. We use the empirical likelihood method to derive the maximum likelihood estimator (MLE) for the proportional hazards model based on right censored survival data with longitudinal covariates. Computation algorithm is developed here and our simulation studies show that our MLE performs very well.

PROPORTIONAL HAZARDS MODEL FOR RIGHT CENSORED SURVIVAL DATA WITH LONGITUDINAL COVARIATES

by

Yuyin Shi

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2023

Dissertation Committee:
Chair: Prof. Joan Jian-Jian Ren
Members: Prof. Sandra Cerrai
Prof. Takumi Saegusa
Prof. Grace L. Yang
Prof. Xin He, Dean's representative

Acknowledgments

I would like to express my great gratitude and appreciation to my advisor Dr. Joan Jian-Jian Ren, who provided me with countless suggestions and invaluable guidance throughout my dissertation research and each stage of my Ph.D. studies. Her insights and knowledge have always inspired me and I truly learned a lot from her.

I would also like to give my special thanks to my dissertation committee members: Professors Sandra Cerrai, Xin He, Takumi Saegusa and Grace L. Yang, for sharing their valuable knowledge and helpful review and advice of my dissertation.

Last but not least, I would like to thank my dear father Qi Shi, my husband Bing Liu, and my aunt, uncle and cousin, who have always supported and helped me whenever I needed. I would not have achieved what I have today without their continuous and unconditional support.

Table of Contents

List of Tables

Chapter 1: Introduction

1.1 Introduction

Survival analysis is a research field in statistics which plays a vital role in biomedical research, reliability studies, finance data analysis, etc. The focus of the studies in this field is concerned with the *lifetime variable* T and related statistical inference problems, such as estimation, testing, modeling, etc.. For references, see Cox and Oakes (1984).

In practice, often the interest of study is to investigate the effect of covariates $\boldsymbol{Z}$ on survival time T, i.e., how covariates, such as gender, age, treatments, etc., impact lifetime variable T. In some data analysis, researchers make parametric assumptions of the distribution of T. However, it is usually not possible to make assessment of the distribution of T, say, in medical clinical trials due to the lack of sufficient prior data. Another big challenge in survival analysis is that lifetime variable T is often subject to censoring; that is T may not be observed for all individuals during the clinical trials. In this context, the semi-parametric and non-parametric methods are desirable and widely considered acceptable to be employed in survival data analysis.

Since 1972, the Cox model (Cox, 1972) has become one of the most widely used tools in the analysis of survival data, and Cox's partial likelihood estimator $\hat{\beta}_c$ has been considered as the standard estimate for regression parameter β_0 in statistical literature.

However, the Cox partial likelihood estimator $\hat{\beta}_c$ is not based on the likelihood in the ordinary sense, and is only formulated for right censored data, which cannot be directly extended to other types of censored data, such as doubly censored data, interval censored data, etc. Moreover, the efficiency results of Cox's partial likelihood estimator are only asymptotic, and for samples with small or moderate sample size the loss in precision from using the partial likelihood is rather large; see Cox and Oakes (1984). To address these issues, Ren and Zhou (2011) used empirical likelihood (Owen, 1988), the nonparametric likelihood approach, to derive the full likelihood function for the Cox model with right censored data, and obtained the maximum likelihood estimator (MLE) for (β_0, F_0) and the relevant asymptotic properties of the MLE, where β_0 is the regression parameter and F_0 is the baseline distribution in Cox's model. They explicitly established the relation between full likelihood and partial likelihood, and showed that in simulation studies, for small or moderate sample size the MLE performs favorably over Cox's partial likelihood estimator. For other types of censored data, Ren (2001, 2008b) developed the concept of weighted empirical likelihood, which is a unified approach that is applicable to various types of censored data; see Ren (2008b), Ren and Wang (2022).

During clinical trials and medical researches, the event time of interest and longitudinal covariates are often recorded together. In these studies, the relationship between survival time T and longitudinal covariates is often the primary interest of the research. In fact, the longitudinal data is an unknown and not completely observed stochastic process, which poses great challenge in the survival data analysis. Due to the challenges encountered in some clinical trials on AIDS and cancer studies, statisticians started considering to use the method of joint modeling of survival data and longitudinal data; see

Tsiatis and Davidian (2004), Zeng and Cai (2005), Ding and Wang (2008), Su and Wang (2012), among others; and see Li and Ren (2011) for more recent review of the literature on this subject. It is well known that all existing works rely on certain parametric or semi-parametric assumptions on the longitudinal covariate processes. Thus, the conclusions of statistical inferences critically depend on the validity of these assumptions.

In this dissertation, we do not make any parametric or semi-parametric assumptions on the survival time and the longitudinal covariate process. We use the empirical likelihood method (Owen, 1988) to derive the MLE for the proportional hazards (PH) model for right censored survival data with longitudinal covariates. The computation algorithm is developed in our research and some simulation results are presented. We also conduct real data analysis for smoking cessation data and liver disease data.

This dissertation is organized as follows: the rest of Chapter 1 briefly describes the concepts of right censored data, Cox model and longitudinal data, introduces right censored survival data with longitudinal covariates based on some real data examples, and provides a review of existing works on joint modeling right censored survival data with longitudinal covariates, then presents our proposed proportional hazards model in the same context; Chapter 2 gives a brief review of empirical likelihood (Owen, 1988, 2001) and the empirical likelihood for the Cox model (Ren and Zhou, 2011); Chapter 3 derives the MLE for our proposed PH model for right censored survival data with longitudinal covariates; Chapter 4 develops computing algorithm and presents some results of simulation studies on our MLE; Chapter 5 gives the algorithm for data with ties among observed survival time, and conducts real data analysis for smoking cessation data and liver disease data; and Chapter 6 provides some concluding remarks and discussions.

1.2 Right Censored Data

One major challenge in survival analysis is the presence of censoring. The most frequently encountered type of censoring in practice is right censoring, thus analysis of right censored data has been a major area of research in statistical literature. In practical situations, right censoring occurs due to various reasons, such as drop-outs, the end time of the study or any other possible factors, which is described below.

Consider a random sample from lifetime variable T:

$$T_1, T_2, \ldots, T_n, \tag{1.1}$$

and F_T is the distribution function (d.f.) of T. If T_i's in (1.1) are subject to right censoring, then the observed data for sample (1.1) are:

$$(V_i, \delta_i), \quad i = 1, \ldots, n, \tag{1.2}$$

where

$$V_i = \begin{cases} T_i, & \text{if } T_i \leq C_i,\ \delta_i = 1 \\ C_i, & \text{if } T_i > C_i,\ \delta_i = 0 \end{cases} \tag{1.3}$$

and C_i is the right censoring variable and is assumed to be independent of T_i.

The likelihood function or the empirical likelihood function, for F_T with right censored data (1.2) is given by:

$$L_R(F) = \prod_{i=1}^{n} [F(V_i) - F(V_i-)]^{\delta_i} [1 - F(V_i)]^{1-\delta_i}, \tag{1.4}$$

where F is an arbitrary d.f., and the nonparametric maximum likelihood estimator (NPMLE) for F_T is the solution that maximizes this likelihood function $L_R(F)$. It was shown that estimator $\hat{F}_{KM}$ by Kaplan and Meier (1958) is the NPMLE for F_T, which is given by

$$\hat{F}_{KM}(t) = 1 - \prod_{V_{(i)} \leq t} (1 - \frac{\delta_{(i)}}{n - i + 1}), \quad t \leq V_{(n)}, \tag{1.5}$$

where $V_{(1)} \leq \ldots \leq V_{(n)}$ is the order statistic of $V_1, \ldots, V_n$, and $\delta_{(i)}$ is the corresponding δ for $V_{(i)}$. Note that if there are ties among the V_i's, then the uncensored V_i's with $\delta_i = 1$ are ranked ahead of the censored V_i's with $\delta_i = 0$.

Stute and Wang (1993) showed that $||\hat{F}_{KM} - F_T|| \overset{\text{a.s.}}{\to} 0$, as $n \to \infty$, and Gill (1983) showed that $\sqrt{n}(\hat{F}_{KM} - F_T)$ weakly converges to a centered Gaussian process as $n \to \infty$.

1.3 Cox Model

Cox model is one of the most widely used tools in survival analysis to explore the relationship between the lifetime variable T and the explanatory or covariate variables. The Cox model is a specific form of the proportional hazards model, and is popular because it makes no parametric assumptions of the underlying distribution function. In this section, we begin with introducing some basic concepts in survival analysis and the model set-up of proportional hazards model, then in Sections 1.3.1-1.3.2, we briefly review the Cox model with time-independent and time-dependent covariates, respectively.

Let T be the non-negative continuous lifetime variable with d.f.:

$$F_T(t) = P(T \leq t), \quad t > 0, \tag{1.6}$$

and its p.d.f. is given by:

$$f_T(t) = F_T^{'}(t), \quad t > 0. \tag{1.7}$$

The *survival function* of T is given by:

$$\bar{F}_T(t) = 1 - F_T(t) = P(T > t), \quad t > 0, \tag{1.8}$$

which is the survival probability of T lasting beyond time point t. Another important characterization of the distribution of T is the *hazard function*, or instantaneous rate of occurrence of the event, which is defined as:

$$h_T(t) = \lim_{\Delta \to 0^+} \frac{P(t < T \leq t + \Delta \mid T > t)}{\Delta}, \quad t > 0. \tag{1.9}$$

The interpretation of the hazard function is that given an individual survives to time t, the instantaneous rate of "death" or "event" at time t is given by $h_T(t)$. Accordingly, the *cumulative hazard function* of T is given by:

$$H_T(t) = \int_0^t h_T(u)du, \quad t > 0. \tag{1.10}$$

.

If we know any one of the functions $\bar{F}_T(t)$, $f_T(t)$, $h_T(t)$, or $H_T(t)$, we can derive the other three functions, because

$$\begin{aligned} h_T(t) &= \lim_{\Delta \to 0^+} \frac{P(t < T \leq t + \Delta \mid T > t)}{\Delta} = \lim_{\Delta \to 0^+} \frac{P(t < T \leq t + \Delta)}{\Delta P(T > t)} \\ &= \frac{1}{\bar{F}_T(t)} \lim_{\Delta \to 0^+} \frac{F_T(t + \Delta) - F_T(t)}{\Delta} = \frac{f_T(t)}{\bar{F}_T(t)}, \end{aligned}$$

which gives

$$h_T(t) = \frac{f_T(t)}{\bar{F}_T(t)} = -\frac{d}{dt}(\ln \bar{F}_T(t)), \tag{1.11}$$

thus (1.10) becomes

$$H_T(t) = \int_0^t h_T(u)du = -\int_0^t \frac{d}{du}(\ln \bar{F}_T(u)) = -\ln \bar{F}_T(u)\Big|_0^t = -\ln \bar{F}_T(t),$$

$$\Longrightarrow \qquad H_T(t) = -\ln \bar{F}_T(t) \quad \text{or} \quad \bar{F}_T(t) = e^{-H_T(t)}. \tag{1.12}$$

Notice that (1.11)-(1.12) imply

$$f_T(t) = h_T(t)e^{-H_T(t)}. \tag{1.13}$$

In practice, survival data usually involves an explanatory covariate vector

$$\boldsymbol{Z} = (Z_1, \dots, Z_k), \tag{1.14}$$

which may include continuous factors such as age and blood pressure at initial time, and discrete factors such as gender and treatment group. This covariate vector $\boldsymbol{Z}$ does not change over time for any individual. To investigate the association between survival time T and explanatory variable $\boldsymbol{Z}$, the proportional hazards model is a frequently used approach, where we do not make specific parametric assumptions on the distribution of T.

Denote the random sample from $(T, \boldsymbol{Z})$ as:

$$(T_1, \boldsymbol{Z}_1), \dots (T_n, \boldsymbol{Z}_n). \tag{1.15}$$

The proportional hazards model is expressed by

$$h(t \mid \boldsymbol{z}) = h_0(t)\psi(\boldsymbol{z}), \quad t > 0, \tag{1.16}$$

where $\boldsymbol{Z}$ is the time-independent covariate vector, $h_0(t)$ is an arbitrary baseline hazard function, $h(t \mid \boldsymbol{z})$ is the conditional hazard function of T given $\boldsymbol{Z} = \boldsymbol{z}$, and $\psi(\boldsymbol{z})$ is a function of $\boldsymbol{z}$ satisfying

$$\psi(\boldsymbol{z}) = \begin{cases} 1, & \text{if } \boldsymbol{z} = \mathbf{0} \\ \\ > 0, & \forall \boldsymbol{z} \neq \mathbf{0}. \end{cases} \tag{1.17}$$

The meaning of proportional hazards model (1.16) is that we assume $h(t \mid \boldsymbol{z})$ and $h_0(t)$ are proportional up to the function $\psi(\boldsymbol{z})$.

In the case of $\boldsymbol{Z}$ having more than a few distinct values, a parametric form of $\psi(\boldsymbol{z})$ in proportional hazards model (1.16) can be helpful, and it is usually parameterized as $\psi(\boldsymbol{z}) = \psi(\boldsymbol{z}; \boldsymbol{\beta})$, where $\boldsymbol{\beta}$ is a parameter which can help us to interpret the effect of $\boldsymbol{z}$ to survival time T in the statistical analysis. From (1.17), we require such parametric function $\psi(\boldsymbol{z}; \boldsymbol{\beta})$ to satisfy:

$$\begin{cases} \psi(\boldsymbol{z}; \boldsymbol{\beta}) > 0, \quad \forall \boldsymbol{z} \neq \mathbf{0} \text{ and } \forall \boldsymbol{\beta} \\ \\ \psi(\mathbf{0}; \boldsymbol{\beta}) = 1, \quad \forall \boldsymbol{\beta} \end{cases} . \tag{1.18}$$

1.3.1 Cox Model with Time-independent Covariate

A natural choice of function $\psi(\boldsymbol{z};\boldsymbol{\beta})$ in (1.18) is given by Cox (1972):

$$\psi(\boldsymbol{z};\boldsymbol{\beta}) = e^{\boldsymbol{\beta}^\top \boldsymbol{z}}. \tag{1.19}$$

Thus, the proportional hazards model (1.16) with above $\psi(\boldsymbol{z}) = \psi(\boldsymbol{z};\boldsymbol{\beta})$ becomes:

$$h(t \mid \boldsymbol{z}) = h_0(t)e^{\boldsymbol{\beta}_0^\top \boldsymbol{z}}, \quad t > 0, \tag{1.20}$$

which is called the *Cox model*. In (1.20), $\boldsymbol{\beta}_0$ is the regression parameter, $\boldsymbol{Z}$ is the time-independent covariate vector, $h_0(t)$ is the arbitrary baseline hazard function with a corresponding d.f. $F_0(t)$, and $h(t \mid \boldsymbol{z})$ is the conditional hazard function of T given $\boldsymbol{Z} = \boldsymbol{z}$.

If the covariate vector $\boldsymbol{Z}$ is a scalar, Cox model (1.20) becomes

$$h(t \mid z) = h_0(t)e^{\beta_0 z}, \quad t > 0. \tag{1.21}$$

As noted in Kalbfleisch and Prentice (2002), Ren and Zhou (2011), among others, Cox model (1.21) is equivalent to the following Lehmann family property:

$$\bar{F}(t \,|\, z) = [\bar{F}_0(t)]^{e^{\beta_0 z}}, \tag{1.22}$$

where $\bar{F}(t \,|\, z)$ is the conditional survival function of T given $Z = z$ with $f(t \,|\, z)$ and $f_0(t)$ as the density functions of $F(t \,|\, z)$ and $F_0(t)$, respectively.

To see (1.22), recall $\bar{F}_T(t) = e^{-H_T(t)}$ in (1.12) for any lifetime variable T, and we start with Cox model (1.21): for $t > 0$,

$$
\begin{aligned}
& h(t \,|\, z) = h_0(t) e^{\beta_0 z} \quad \Leftrightarrow \quad \int_0^t h(u \,|\, z) du = \int_0^t h_0(u) e^{\beta_0 z} du \\
\Leftrightarrow \quad & H(t \,|\, z) = e^{\beta_0 z} H_0(t) \quad \Leftrightarrow \quad e^{-H(t \,|\, z)} = e^{-e^{\beta_0 z} H_0(t)} \\
\Leftrightarrow \quad & \bar{F}(t \,|\, z) = [\bar{F}_0(t)]^{e^{\beta_0 z}},
\end{aligned}
$$

where $H(t \,|\, z)$ is the conditional cumulative hazard function of T given $Z = z$, and $H_0(t)$ is the cumulative hazard function corresponding to F_0.

From Lehmann family property (1.22), we know that the Cox model (1.21) is equivalent to the following: for given $Z = z$,

$$
f(t \,|\, z) = e^{\beta_0 z} f_0(t) [\bar{F}_0(t)]^{e^{\beta_0 z} - 1}. \tag{1.23}
$$

For the case when T_i in (1.15) is subject to right censoring, the actually observed survival data is given by

$$
(V_1, \delta_1, \boldsymbol{Z}_1), \ldots, (V_n, \delta_n, \boldsymbol{Z}_n), \tag{1.24}
$$

where (V_i, δ_i) is the observed right censored data (1.2) on T_i and $\boldsymbol{Z}_i$'s are as in (1.14)-(1.15). Ren and Zhou (2011) used above equation (1.23) to derive the full (empirical) likelihood function for the Cox model (1.20) with right censored data (1.24).

The Cox's partial likelihood (Cox, 1972) has been widely used for estimating the regression coefficient β_0 in Cox model (1.20) with right censored data (1.24). Efron (1977) and Oakes (1977) discussed the asymptotic efficiency and derived the asymptotic variance of Cox's partial likelihood estimator $\hat{\beta}_c$. The strong consistency and asymptotic normality for the partial likelihood estimator were established by Tsiatis (1981). But the efficiency results on partial likelihood estimator is only asymptotic (Cox and Oakes, 1984). Ren and Zhou (2011) derived the full likelihood function for the Cox model with right censored data (1.24) using the empirical likelihood, and made explicit connections between the full likelihood and the Cox's partial likelihood. Ren and Zhou (2011) derived the full likelihood based MLE $(\hat{\beta}_n, \hat{F}_n)$ for (β_0, F_0) in Cox model (1.20), and established the asymptotic normality and efficiency of the MLE $\hat{\beta}_n$.

1.3.2 Cox Model with Time-dependent Covariate

In some situation, the explanatory covariate $\boldsymbol{Z}$ in (1.14) is time-dependent, denoted as $\boldsymbol{Z}(t)$; that is the covariate that may change over time t, and $\boldsymbol{Z}(t)$ is a stochastic process. In practice, the time-dependent covariate is a repeated measurement on a subject or perhaps a change in the subject's treatment.

The extension of the Cox model (1.20) to allow time-dependent covariates $\boldsymbol{Z}(t)$ is given by

$$h(t \mid \boldsymbol{Z}(t)) = h_0(t)e^{\boldsymbol{\beta}_0^\top \boldsymbol{Z}(t)}, \quad t > 0, \tag{1.25}$$

where $\boldsymbol{Z}(t)$ is assumed to be completely known for all t in the statistical literature. In other words, $\boldsymbol{Z}(t)$ is under total parametric assumptions; see Cox (1972) and Cox and Oakes (1984). Since the values of variable $\boldsymbol{Z}(t)$ depend on time t, the hazard ratio is also time-dependent, which is different from the conventional Cox model (1.20), where the hazard ratio $\frac{h(t \mid \boldsymbol{Z})}{h_0(t)} = e^{\boldsymbol{\beta}_0^\top \boldsymbol{z}}$ only depends on the covariates $\boldsymbol{z}$, but not on time t.

1.4 Longitudinal Data

In clinical research, longitudinal studies play an important role in understanding the development and persistence of disease. The defining feature of a longitudinal study is that individuals are measured repeatedly through time, thereby permitting the direct assessment of changes in the response variable over time. For instance, in AIDS research (Pawitan and Self, 1993), a measurement for immune system, called CD4 counts, was recorded for patients once every 3 months, which gives longitudinal data; see data example at the end of this section.

Longitudinal studies are in contrast to cross-sectional studies, where measurements are obtained at only a single point in time. By obtaining measurements of the same individuals repeatedly through time, longitudinal studies can address fundamental questions concerning the assessment of changes within-individual in the response variable, and also it can address whether these changes within-individual in the response are related to certain selected covariates.

Longitudinal data analysis requires special statistical methods because the set of observations on one individual tends to be intercorrelated. Such correlation must be taken into account to draw valid scientific inferences, see Diggle, Heagerty, Liang, and Zeger (2002) for examples and models.

A primary research interest in longitudinal studies is to explore the interrelationships between time-to-event variable and longitudinal covariates. The patterns of the longitudinal covariates are of additional interest. The Cox model has been used widely in the analysis of time to diagnosis or death from disease, and the relationship between a time-dependent covariate and survival times can also be evaluated via the Cox model with time-dependent covariate, but the parametric assumption on the time-dependent covariate $\boldsymbol{Z}(t)$ in (1.25) usually is uncheckable for its validity.

Below is a data example on longitudinal data $\boldsymbol{X}(t)$ related to lifetime variable T.

AIDS Data Example

In an AIDS cohort study Pawitan and Self (1993), a total of 249 healthy homosexual males were recruited between July 1984 and July 1985, and T is time to progression to AIDS after infection. A battery of immunologic tests were performed at baseline and repeated about once every 3 months to record the longitudinal covariates $\boldsymbol{X}(t) = (X_1(t), \ldots, X_k(t))$, where t is time and $X_1 =$ [CD4 counts], $X_2 =$ [CD4/CD8 ratio], etc., because CD4 counts and CD4/CD8 ratio are important measures of immune system. A decrease in CD4 counts and CD4/CD8 ratio indicates a weakened immune system, making the individual more susceptible to infections and diseases. The focus of the research was to find the relation between longitudinal covariates and the event time T to AIDS of HIV infected patients.

1.5 Right Censored Survival Data with Longitudinal Covariate

Let T be a lifetime variable, and let $X(t)$ be a covariate process that is observable only at certain time points for each individual. We consider the following survival data with a longitudinal covariate:

$$(T_i, X_i(t)), \quad i = 1, 2, \ldots, n, \tag{1.26}$$

where T_i is the survival time of the ith individual, and $X_i(t)$ is the longitudinal covariate process of the ith individual.

For the ith individual, covariate process $X_i(t)$ is only observed at time points $t_{i1} < t_{i2} < \ldots < t_{iJ_i}$, thus the actually observed data on $X_i(t)$ is given by

$$\mathcal{X}_i = (X_i(t_{i1}), \ldots, X_i(t_{iJ_i})), \tag{1.27}$$

where $X_i(t_{ij})$ is the observed longitudinal covariate at time t_{ij}, $j = 1, \ldots, J_i$.

If the survival time T_i is subject to right censoring, then the actually observed data on T_i is (V_i, δ_i) given by (1.2), C_i is the right censoring variable with d.f. F_C, and C_i is independent of T_i and the covariate process $X_i(t)$. Thus, in practice the actually observed right censored survival data on (1.26) are given by:

$$(V_i, \delta_i, \mathcal{X}_i), \quad i = 1, \ldots, n, \tag{1.28}$$

where (V_i, δ_i) and $\mathcal{X}_i$ are given by (1.2) and (1.27), respectively.

Note that in practice, we do not observe anything on $X_i(t)$ for $t > V_i$ in (1.28), thus the time points observed on $X_i(t)$, i.e., $X_i(t_{ij})$, satisfy $t_{iJ_i} \le V_i$. In turn, the actually observed data in (1.28) on data $(T_i, X_i(t))$'s in (1.26) is given by

$$(V_i, \delta_i, \mathcal{X}_i^{V_i}), \quad i = 1, \ldots, n, \tag{1.29}$$

where $\mathcal{X}_i^t = \{X_i(t_{ij}) \,|\, t_{ij} \le t\}$ represents the vector of all observed $X_i(s)$ with $s \le t$.

Real Data Examples

Smoking Cessation Data

In a smoking cessation study (Shiffman et al., 1996), 151 smokers who had quit for at least 24 hours were participated, and T_i is time to the first lapse, which is subject to right censoring due to drop-outs and the end of the study. During the study, each participant was randomly sampled 5 times per day to record the measurements on covariate process such as $X(t) =$ [Urge to Smoke], etc. For the ith participant, the actually observed data on $X_i(t)$ is $\mathcal{X}_i = (X_i(t_{i1}), \ldots, X_i(t_{iJ_i}))$, where $X_i(t_{ij})$ is the observed level of urge to smoke of the ith participant at time t_{ij}, $j = 1, \ldots, J_i$. Because the timing of measurement of urge to smoke is random, each participant was measured randomly at different time points t_{ij}'s, and the total number J_i of the measurements taken on the covariate process $X_i(t)$ for each participant were also different due to lapse, the drop-outs, the length of period in which the individual has participated, and other factors. In an empirical analysis of Smoking Cessation Data (Li, Root, and Shiffman, 2006), one of the research interests was the relationship between time to lapse and the changes in the level of urge to smoke. To characterize the historic change patterns of a participant's level of urge to smoke, a summary measurement of X^t was chosen to be the average of $X_i(t_{ij})$'s in a particular time interval prior to V_i. This is an example of right censored survival data with intensive longitudinal covariates (1.29).

Child Development Data

In the study conducted by Cole et al. (2011), children of 18 months of age were invited for four lab visits at age 18, 24, 36 and 48 months, and T_i is the age at which a specified development goal is achieved for a child. Here, T_i is subject to double censoring, because some children already achieved the development goal at the first visit which is left censored, and some children didn't achieve it at the end of the study or drop-outs which is right censored. For each child, an 8-minute test was performed at each lab visit, and the emotion variables such as anger, duration of anger, calm, etc., were recorded every 15 seconds. For instance, the actually observed data on anger, $X_i(t)$, is $\mathcal{X}_i = (X_i(t_{i1}), \ldots, X_i(t_{iJ_i}))$, where $X_i(t_{ij})$ is the observed amount of anger of the ith child at time t_{ij}, $j = 1, \ldots, J_i$. To characterize the historic change patterns of a child's anger expression during the 8-minutes test at a particular visit, one of the modeling functions used by Cole et al. (2011) is the total amount of anger of ith child in a particular time interval (t_0, t) and t_0 is the starting time of the test. This is an example of doubly censored survival data with intensive longitudinal covariates.

Liver Disease Data

In a Primary Biliary Cirrhosis (PBC) clinical trial (Murtaugh et al., 1994) conducted by Mayo Clinic between 1974 and 1984, a total of 312 PBC patients, i.e., liver disease patients, were referred to Mayo Clinic during that ten-year interval and participated in the randomized trial of the drug D-penicillamine. Patients often had abnormalities in their blood tests, such as elevated and gradually increased serum bilirubin. Thus, investigating the association between bilirubin and death due to PBC is of great medical interest. In this case, T_i is the survival time of a PBC patient, which is subject to right censoring due to drop-outs, the end of the study, liver transplantation and other factors. For each patient, longitudinal covariates including serum bilirubin and albumin were measured at follow-up lab visits at 6 months, 1 year and annually thereafter. For instance, the observed data on serum bilirubin, $X_i(t)$, is $\mathcal{X}_i = (X_i(t_{i1}), \ldots, X_i(t_{iJ_i}))$, where $X_i(t_{ij})$ is the observed amount of serum bilirubin of the ith patient at time t_{ij}, $j = 1, \ldots, J_i$. To characterize the historic change patterns of a patient's level of serum bilirubin, a summary measurement can be chosen to be the average of $X_i(t_{ij})$'s in a particular time interval prior to V_i. This is also an example of right censored survival data with longitudinal covariates (1.29).

1.6 Existing Methods and Challenges

The existing methods for joint modeling right censored survival data and longitudinal covariates are mainly on the line of Tsiatis and Davidian (2004), Zeng and Cai (2005), Ding and Wang (2008) and Su and Wang (2012), among others. Tsiatis and Davidian (2004), Li and Ren (2011) provided review of the literature. This section first describes the joint modeling framework, then summarizes some overview key points in the works mentioned above.

For right censored survival data with longitudinal covariate (1.29), denoting

$$X^t = \{X(s); 0 \leq s < t\} \tag{1.30}$$

as the history of the longitudinal process up to time t, the joint modeling survival data and longitudinal data assumes the following proportional hazards model:

$$h(t \mid Z_i, X_i^t) = h_0(t) \exp(\beta_Z Z_i + \beta_X X_i(t)), \tag{1.31}$$

where β_Z and β_X are regression parameters, $h(t \mid Z_i, X_i^t)$ is the conditional hazard function of T given $Z = Z_i$, $X^t = X_i^t$, and $h_0(t)$ is an arbitrary baseline hazard function with corresponding d.f. F_0.

Note that model (1.31) is different from the usual Cox model with time-dependent covariates (1.25). In (1.31), $X_i(t)$ is only observed at time t_{ij}, while in the Cox model with time-dependent covariates (1.25), $X_i(t)$ is assumed to be observed fully at any time

t; see Kalbfleisch and Prentice (2002), Tsiatis and Davidian (2004), Li and Ren (2011).

Since the covariate process $X_i(t)$ is only observable at time points $t_{i1} < \ldots < t_{iJ_i} \leq V_i$, given as (1.29), various parametric or semi-parametric assumptions on $X_i(t)$ were considered by various authors in order to estimate the covariate process $X_i(t)$.

In the review paper of Tsiatis and Davidian (2004), a simple linear model was considered to characterize $X_i(t)$ in terms of subject-specific random effects:

$$X_i(t) = \alpha_{0i} + \alpha_{1i}t, \quad \alpha_i = (\alpha_{0i}, \alpha_{1i}). \tag{1.32}$$

More flexible models is also considered:

$$X_i(t) = \alpha_i \omega(t), \tag{1.33}$$

where $\omega(t)$ is a known functions of time. In (1.32) and (1.33), α_i is usually assumed to be normally distributed to represent within-subject variation.

In Zeng and Cai (2005)'s work, their model assumptions are not exactly on the line of the setting in this section. The main difference is that in their work, the covariate process $X(t)$ is assumed to be fully observed while $X(t)$ in (1.31) is only observed at time points t_{ij}. And they assumed the covariate process $X(t)$ is conditional on the normally distributed random effects. Under the model assumptions, Zeng and Cai (2005) established the consistency and asymptotic normality of the resulting MLE.

Ding and Wang (2008) use the multiplicative random effect model with population mean function $\mu(t) = \mathrm{E}[X_i(t)]$:

$$X_i(t) = b_i \mu(t), \quad \mathrm{E}(b_i) = 1, \tag{1.34}$$

where b_i is the random effect variable and its distribution is assumed to be known. In contrast to (1.33) used in other works, the mean function $\mu(t)$ here is assumed to be completely unknown. Specifically, mean function $\mu(t)$ is approximated by B-spline basis functions and the number of knots and degrees are selected based on a version of the Akaike information criterion (AIC). The specification of the distribution of the random effect b_i is required in their work.

For Su and Wang (2012), they considered survival data that are subject to both left truncation and right censoring on T_i's in (1.26), and they considered

$$X_i(t) = g(t) A_i^*, \tag{1.35}$$

where $g(\cdot)$ is a known function and A_i^* is a random effect with known distribution; that is they assume a common known parametric density function $f_A^*(\cdot \mid \alpha)$ with an unknown parameter α for A_i^*.

It should be noticed that in models (1.32)-(1.35), $X_i(t)$ depends critically on validity of the parametric or semi-parametric assumptions which are very difficult to justify in practice. And theoretical properties of the existing procedures are only available under these unjustifiable parametric or semi-parametric assumptions.

One should also note that in joint modeling (1.31), the regression covariates used on observed $\mathcal{X}_i$ in (1.29) is $X_i(t)$ which does not reflect the history of the observed longitudinal covariates, i.e., does not reflect the historic change patterns in observed $\boldsymbol{\mathcal{X}}_i$. In the literature, there have not been any modeling procedures that directly study the relationship between survival time and within-subject historic patterns of change, nor have there been any works on joint modeling doubly censored or interval censored survival data together with longitudinal covariates. Moreover, there are some more complicated problems involving the intensive longitudinal covariates and the multi-phase intensive longitudinal covariates where considerably more repeated measurements are taken within a relative short period of time; see Smoking Cessation Data example and Child Development Data example described in Section 1.5. These problems have not been well studied in statistical literature.

Challenges. In practice, the longitudinal covariates $X_i(t)$ in (1.31) is in fact a completely unknown and not completely observed stochastic process. This, plus censored observation (V_i, δ_i) on survival time T_i, poses huge challenge to developing statistical methodology for joint modeling (1.31) based on data (1.29) using nonparametric approach, which has not been done in statistical literature, although everyone knows this would be the most desirable way from practical view point.

1.7 Our Proposed Model and Methods

In this dissertation, we consider the right censored survival data with longitudinal covariates (1.29), and propose the following proportional hazards (PH) model with covariate process $X(t)$:

$$h(t \,|\, Z, X^t) = h_0(t) \exp\big\{\alpha_0 Z + \beta_0 Z(t)\big\}, \tag{1.36}$$

where as (1.30), $X^t = \{X(s) \,|\, s \le t\}$ represents the history of the covariate process up to time t, Z is the time-independent covariate, $Z(t) = \varphi(X^t)$ is a modeling function or summary measurement which characterizes the historic change patterns in observed longitudinal covariates prior to time t, $h_0(t)$ is the baseline hazard function with corresponding d.f. F_0, (α_0, β_0) is the regression parameter, and $h(t \,|\, Z, X^t)$ is the conditional hazard function of lifetime variable T given Z and X^t in the following sense:

$$h(t \,|\, Z, X^t) = \lim_{\Delta \to 0} \Delta^{-1} P\{t \le T < t + \Delta \,|\, T \ge t, Z, X^t\}, \tag{1.37}$$

which is the conditional probability of failure at time t given that the failure does not occur prior to time t and that the covariate process $X(\cdot)$ has the history of the process up to time t, denoted by X^t as in (1.30) and above for (1.36).

Note that our consideration of $Z(t) = \varphi(X^t)$ in (1.36) is motived by the Smoking Cessation Data and Child Development Data described in Section 1.5, and our proposed PH model (1.36) includes (1.31) as a special case with $Z(t) = X(t)$. Due to the complexity of model (1.36) and that our main interest is $Z(t)$ in this study, for the rest of this dissertation we focus on the following model for simplicity of presentation:

$$h(t \,|\, X^t) = h_0(t) \exp\left\{\beta_0 Z(t)\right\}. \tag{1.38}$$

Also, note that $X_i^{V_i} \neq \mathcal{X}_i^{V_i}$, because as explained in (1.27), $X_i(t)$ is not known except at points t_{ij}'s, and the covariate process $X_i^{V_i}$ is not entirely observed. The longitudinal covariate $\mathcal{X}_i^{V_i}$ is what's actually observed on covariate process $X_i^{V_i}$. The PH model (1.38) for right censored survival data with longitudinal covariates (1.29) is a very different problem from the Cox model with time-dependent covariates $X(t)$ as stated in (1.25), which assumes that $X(t)$ is entirely known for all t; see Chapter 4 and Chapter 6 in Kalbfleisch and Prentice (2002).

In this dissertation, we do not make any parametric or semi-parametric assumptions on the covariate process $X(t)$, and we use empirical likelihood to derive the MLE for (β_0, F_0) in the proportional hazards model (1.38) based on right censored survival data with longitudinal covariates (1.29).

Chapter 2: Empirical Likelihood

For parametric likelihood methods, it is assumed that the random sample is from a known distribution except for unknown parameters to be determined. However, in many scenarios such as medical clinical trials, etc, often we do not have any prior data to know which is the appropriate parametric family to use, and often such assumptions are very difficult or impossible to be verified.

Empirical likelihood (EL) (Owen, 1988) is a nonparametric likelihood method for statistical inference problems. It allows statisticians to use the likelihood methods without requiring specification of distribution family for the data. In this chapter, we briefly review the concept of empirical likelihood and some important results for censored data, then the empirical likelihood method for the Cox model.

2.1 Empirical Likelihood

EL method was introduced by Owen (1988) as a device for constructing confidence regions based on random samples without any parametric assumption. It has been shown that EL has many advantages over the usual approximation and bootstrap methods; see Hall and La Scala (1990), DiCiccio, Hall, and Romano (1991), among others. Not only does empirical likelihood provide data-determined shapes for confidence regions, but also it can easily incorporate known constraints on parameters and make inferences without the estimation of variance because the nonparametric version of Wilk's theorem was established by Owen (1988). Below, we outline Owen's EL approach.

Let

$$X_1, \ldots, X_n \tag{2.1}$$

be a random sample from a d.f. F_0. The *empirical likelihood function* by Owen (1988) for F_0 is given by

$$L(F) = \prod_{i=1}^{n} [F(X_i) - F(X_i-)], \tag{2.2}$$

where F is any d.f. The nonparametric maximum likelihood estimate (NPMLE) for F_0 is the d.f. that maximizes likelihood function $L(F)$ in (2.2) over all d.f.'s. It has been shown that the NPMLE for F_0 is the empirical distribution function of $X_1, \ldots, X_n$ given by

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I\{X_i \leq x\}, \quad -\infty < x < \infty. \tag{2.3}$$

The *empirical likelihood ratio function* is defined as:

$$R(F) = \frac{L(F)}{L(F_n)}. \tag{2.4}$$

Owen (1988) showed that the empirical likelihood ratio function can be used to construct confidence intervals for statistical functional $T(F_0) = \theta_0$, and that under some mild conditions, the nonparametric version of Wilk's theorem holds; that is the log empirical likelihood ratio converges to a chi-square distribution, which is parameter free distribution, thus desirable in practice.

For more references on empirical likelihood, see DiCiccio, Hall, and Romano (1991), Chen and Hall (1993), Chen and Qin (1993), Qin and Lawless (1994), Mykland (1995), Murphy and van der Vaart (1997), Owen (2001), Ren (2001), Zhou (2005), Chen and Cui (2006), Ren (2008a,b), Wang and Chen (2009), Ren and Zhou (2011), among others.

However, the empirical likelihood function given in (2.2) is formulated for completely observed data (2.1), which can not be applied directly to different types of censored data. In survival analysis, the survival data is often not completely observed. For right censored data (1.2), Kaplan and Meier (1958) gave the NPMLE $\hat{F}_{KM}$ for F_0; see review given in Section 1.2. He, Liang, Shen and Yang (2016) constructed the empirical likelihood based confidence intervals for statistical functionals of the underlaying distribution F_0 for right censored data, and showed that the log likelihood ratio converges to chi-squared distribution.

For doubly censored data, the observed data (V_i, δ_i) on X_i is given by

$$V_i = \begin{cases} T_i, & \text{if } D_i < X_i \leq C_i,\ \delta_i = 1 \\ C_i, & \text{if } X_i > C_i,\ \delta_i = 2 \\ D_i, & \text{if } X_i \leq D_i,\ \delta_i = 3 \end{cases} \tag{2.5}$$

where C_i and D_i are right and left censoring variables, respectively, and they are independent of X_i with $P\{D_i < C_i\} = 1$. Doubly censored data has been encountered in many important scientific research, such as child development data (Leiderman et al., 1973), breast cancer data (Ren and Gu, 1997), etc. Based on the empirical likelihood function for F_0, Mykland and Ren (1996) derived the NPMLE $\hat{F}_{MR}$ for F_0 and showed $\hat{F}_{MR}$ can be numerically computed by the EM algorithm. Chang and Yang (1987) showed that $||\hat{F}_{MR} - F_0|| \stackrel{\text{a.s.}}{\to} 0$, as $n \to \infty$, and Gu and Zhang (1993) showed that $\sqrt{n}(\hat{F}_{MR} - F_0)$ weakly converges to a centered Gaussian process as $n \to \infty$, under regularity conditions.

Under Case 2 interval censoring, the observed data (C_i, D_i, δ_i) on X_i is given by

$$\delta_i = \begin{cases} 1, & \text{if } D_i < X_i \leq C_i, \\ 2, & \text{if } X_i > C_i, \\ 3, & \text{if } X_i \leq D_i, \end{cases} \tag{2.6}$$

where C_i and D_i are independent of X_i with $P\{D_i < C_i\} = 1$. Based on the empirical likelihood function for F_0, Groeneboom and Wellner (1992) derived the NPMLE $\hat{F}_{GW}$ for F_0 and showed $||\hat{F}_{GW} - F_0|| \stackrel{\text{a.s.}}{\to} 0$, as $n \to \infty$.

To deal with various types of censored data, such as right censored data, doubly censored data and interval censored data, Ren (2001) developed a new and unified approach called *weighted empirical likelihood* and applied it to a general setting of two-sample semi-parametric models (Ren, 2008b), which include biased sampling models and logistic regression models as special cases. For various types of censored data aforementioned, Ren (2008b) derived the weighted empirical likelihood-based MLE $(\hat{\theta}_n, \hat{F}_n)$ for the underlying parameter θ_0 and distribution F_0, and established the strong consistency of $(\hat{\theta}_n, \hat{F}_n)$ and the asymptotic normality of $(\hat{\theta}_n, \hat{F}_n)$ under some conditions.

All above nice results demonstrate the importance and desirability from practical view point of the empirical likelihood approach, especially the problem involving doubly censored data and interval censored data which are very challenging mathematically.

2.2 Empirical Likelihood for Cox Model

Consider Cox model (1.21) with right censored survival data (1.24) with scaler Z:

$$(V_i, \delta_i, Z_i), \quad i = 1, \ldots, n, \tag{2.7}$$

where (V_i, δ_i) is given in (1.2)-(1.3) and Z_i is the time-independent covariate of ith individual. As described in Section 1.1 and Section 1.3.1, Cox's partial likelihood estimator $\hat{\beta}_c$ is not based on the likelihood in ordinary sense and is only asymptotic efficient.

Ren and Zhou (2011) used empirical likelihood and Lehmann family properties (1.22)-(1.23) to derive the full likelihood function for (β_0, F_0) in Cox model (1.21) with right censored survival data (2.7). Their full likelihood function is given by

$$L(\beta, F) = \prod_{i=1}^{n} (c_i dF(V_i))^{\delta_i} (\bar{F}(V_i))^{c_i - \delta_i}, \tag{2.8}$$

where $c_i = \exp(Z_i \beta)$. They used empirical likelihood parameterization, then explicitly profiled out nuisance parameter F_0 and obtained the full-profile likelihood estimating function for β_0:

$$\psi_{RZ}(\beta) = n^{-1} \sum_{i=1}^{n} \delta_i \Big(Z_i + e_i \log \frac{d_i - 1}{d_i} \Big), \tag{2.9}$$

where $d_i = \sum_{j=i}^{n} c_j$ and $e_i = \sum_{j=i}^{n} Z_j c_j$. Ren and Zhou (2011) showed that the leading term of Taylor's expansion of the full-profile likelihood estimating function $\psi_{RZ}(\beta)$ in (2.9) is the Cox's partial likelihood estimating function $\varphi_c(\beta)$, and their relation is given by

$$\psi_{RZ}(\beta) = \varphi_c(\beta) + O_p\Big(\frac{\log n}{n}\Big), \tag{2.10}$$

where $\varphi_c(\beta)$ is the partial likelihood estimating function. Ren and Zhou (2011) showed that the log full-likelihood ratio has an asymptotic chi-squared distribution, which is the nonparametric version of Wilk's Theorem for right censored data (2.7).

For the Cox model (1.21), if T_i's in (2.7) are subject to various types of censoring, including right censoring, doubly censoring, interval censoring, etc., Ren and Wang (2022) extended the concept of weighted empirical likelihood (Ren, 2001, 2008b) to multivariate case, and derived the weighted empirical likelihood-based MLE (WMLE) $(\hat{\beta}_n, \hat{F}_n)$ for the Cox model that is applicable to various types of censored data under a unified framework. They established the asymptotic properties of WMLE: $\sqrt{n}(\hat{\beta}_n - \beta_0) \stackrel{\mathcal{D}}{\to} N(0, \sigma_w^2)$, as $n \to \infty$, and $\sqrt{n}(\hat{F}_n - F_0)$ weakly converges to a centered Gaussian process.

Chapter 3: MLE for Jointly Modeling Survival and Longitudinal Data

This chapter considers our proposed proportional hazards model (1.38) based on right censored survival data with longitudinal covariates (1.29). In Section 3.1.1, we first derive the similar results to the Lehmann family properties (1.22)-(1.23) for the PH model (1.38), then in Section 3.1.2 we use the Lehmann family properties to derive the empirical likelihood function for (β_0, F_0) in PH model (1.38) based on right censored survival data with longitudinal covariates (1.29). Section 3.2 derives the empirical likelihood estimating function for β_0 under certain restriction and derives the approximated estimating function for β_0. Section 3.3 derives the empirical likelihood estimating function for β_0 without the restriction in Section 3.2. All proofs are given in Section 3.4

3.1 Empirical Likelihood for Joint Modeling

3.1.1 Lehmann Family Property for Joint Modeling

In Section 1.3.1, Lehmann family properties (1.22)-(1.23) for the Cox model (1.21) has been derived. Here, we derive the results similar to the Lehmann family properties for our proposed PH model (1.38).

From (4.4) on page 96 of Kalbfleisch and Prentice (2002), we know that for X^t given by (1.30) and for any fixed $t > 0$, $f(t \,|\, X^t)$ is given by

$$
\begin{aligned}
f(t \,|\, X^t) = h(t \,|\, X^t)\bar{F}(t \,|\, X^t) &= \lim_{\Delta \to 0} \frac{P\{t \le T < t + \Delta \,|\, T \ge t, X^t\}}{\Delta} \bar{F}(t \,|\, X^t) \\
&= \lim_{\Delta \to 0} \frac{P\{t \le T < t + \Delta \,|\, X^t\}}{\Delta P\{T \ge t \,|\, X^t\}} \bar{F}(t \,|\, X^t) \\
&= \lim_{\Delta \to 0} \Delta^{-1} P\{t \le T < t + \Delta \,|\, X^t\},
\end{aligned} \tag{3.1}
$$

where from page 197 of Kalbfleisch and Prentice (2002),

$$
\bar{F}(t \,|\, X^t) = P\{T \ge t \,|\, X^t\}. \tag{3.2}
$$

Assume that the covariate process $X(t)$ is external, that is $X(t)$ for any $s < t$ is not affected by the occurrence of failure at time s. Then from (6.8) on page 196 of Kalbfleisch and Prentice (2002), we have that for partition points of interval $[0, t]$: $0 = s_0 < s_1 < \ldots < s_N = t$ which have equal width $\Delta = \frac{t}{N}$ for each subinterval $[s_{j-1}, s_j]$:

$$
\begin{aligned}
\int_0^t f(s \,|\, X^s) ds &= o(1) + \sum_{k=0}^{N-1} f(s_k \,|\, X^{s_k}) \Delta \\
&\overset{(3.1)}{=\!=} o(1) + \sum_{k=0}^{N-1} \Big(P\{s_k \le T < s_k + \Delta \,|\, X^{s_k}\} + r_k \Delta \Big) \\
&\overset{\text{(6.8) of K-P}}{=\!=\!=} o(1) + \sum_{k=0}^{N-1} \Big(P\{s_k \le T < s_k + \Delta \,|\, X^t\} + r_k \Delta \Big) \\
&= o(1) + P\{0 \le T < t \,|\, X^t\} + \left(\sum_{k=0}^{N-1} r_k \right) \frac{t}{N} \\
&= o(1) + P\{0 \le T < t \,|\, X^t\} + O(\max_k |r_k|)
\end{aligned} \tag{3.3}
$$

where r_k is the reminder term from (3.1) at each s_k, and we assume the uniform convergence of $\Delta^{-1}P\{t \leq T < t+\Delta \,|\, X^t\}$ to $f(t \,|\, X^t)$, i.e.,

$$\sup_t |\Delta^{-1}P\{t \leq T < t+\Delta \,|\, X^t\} - f(t \,|\, X^t)| \to 0, \quad \text{as } \Delta \to 0, \tag{AS0}$$

which implies that for any $\epsilon > 0$, we have $\max_k |r_k| \leq \epsilon$ for sufficiently small Δ in (3.3). Hence, equation (3.3) implies for any t,

$$\int_0^t f(s \,|\, X^s)ds = P\{0 \leq T < t \,|\, X^t\} = F(t \,|\, X^t) \tag{3.4}$$

which is equivalent to

$$\frac{d}{dt}\Big(F(t \,|\, X^t)\Big) = f(t \,|\, X^t). \tag{3.5}$$

From (3.1)-(3.2), (3.4)-(3.5), and $\bar{F}(t \,|\, X^t) + F(t \,|\, X^t) \equiv 1$, we know

$$\begin{aligned} H(t \,|\, X^t) &= \int_0^t h(s \,|\, X^s)ds = \int_0^t \frac{f(s \,|\, X^s)}{\bar{F}(s \,|\, X^s)}ds \\ &= -\int_0^t \frac{d\bar{F}(s \,|\, X^s)}{\bar{F}(s \,|\, X^s)} = -\ln \bar{F}(t \,|\, X^t), \end{aligned} \tag{3.6}$$

which by model assumption (1.38) gives

$$\bar{F}(t \,|\, X^t) = \exp\{-H(t \,|\, X^t)\} = \exp\Big\{-\int_0^t h_0(s)\exp(Z(s)\beta_0)ds\Big\}. \tag{3.7}$$

This equation (3.7) is the ”Lehmann family properties” for our PH model (1.38), which is similar to the Lehmann family properties (1.22)-(1.23) for Cox model (1.21).

Note that when $Z(t)$ in (3.7) does not depend on t, say $Z(t) = Z$ for all $t > 0$, then (3.7) is exactly the same as the Lehmann family property (1.22) for Cox model (1.21), and the PH model (1.38) is the same as the Cox model (1.21).

3.1.2 Empirical Likelihood for Joint Modeling

Based on the "Lehmann family properties" in (3.7), we derive the empirical likelihood function for (β_0, F_0) in our proposed PH model (1.38) for right censored survival data with longitudinal covariates (1.29) as below.

Consider the following conditional p.d.f. of (V, δ) in (1.29) given X^t:

$$g(t, \delta \,|\, X^t) = \lim_{\Delta \to 0} \Delta^{-1} P\{t \le V < t + \Delta,\ \delta = \delta \,|\, X^t\}, \tag{3.8}$$

then we have the following lemma with proof given in Section 3.4.

Lemma 3.1

$$\begin{cases} g(t, 1 \,|\, X^t) = f(t \,|\, X^t)\bar{F}_C(t) \\ g(t, 0 \,|\, X^t) = \bar{F}(t \,|\, X^t) f_C(t) \end{cases} \tag{3.9}$$

where $f(t \,|\, X^t)$ and $\bar{F}(t \,|\, X^t)$ are given by (3.1) and (3.2) respectively, and $F_C(t)$ is the d.f. of right censoring variable with p.d.f. $f_C(t) = F_C'(t)$.

Thus, the likelihood for (V_i, δ_i) in (1.29) given $X_i^{V_i}$, $1 \le i \le n$, is given by

$$\prod_{i=1}^{n} g(V_i, \delta_i \,|\, X_i^{V_i}) = \prod_{i=1}^{n} \big(f(V_i \,|\, X_i^{V_i})\bar{F}_C(V_i)\big)^{\delta_i} \big(\bar{F}(V_i \,|\, X_i^{V_i}) f_C(V_i)\big)^{1-\delta_i}, \tag{3.10}$$

which is proportional to

$$L(V, \delta, X^V) = \prod_{i=1}^{n} \left(f(V_i \,|\, X_i^{V_i})\right)^{\delta_i} \left(\bar{F}(V_i \,|\, X_i^{V_i})\right)^{1-\delta_i}. \tag{3.11}$$

Note that this likelihood in (3.11) is not the likelihood for observed survival data (1.29) because the covariate process $X_i^{V_i}$ in (3.11) is not entirely observed, only the longitudinal covariate $\mathcal{X}_i^{V_i}$ is observed, that is as pointed out in Section 1.7, we have $X_i^{V_i} \neq \mathcal{X}_i^{V_i}$. This likelihood given by (3.11) is strictly based on the model assumptions and definitions in (1.36)-(1.37) without Z assuming that $(V_i, \delta_i, X_i^{V_i})$ is observed for each individual.

Also, note that $\bar{F}(t \,|\, X^t)$ in (3.11) is given by (3.7), which cannot be calculated unless the entire covariate process $X(s)$ is known up to time t. Thus, for likelihood inference problems, we need to estimate the covariate process $X_i^{V_i}$ based on observed longitudinal covariates $\mathcal{X}_i^{V_i}$, then obtain an estimated likelihood of $L(V, \delta, X^V)$ based on observed survival data $(V_i, \delta_i, \mathcal{X}_i^{V_i})$ given in (1.29).

To estimate the last term in (3.7) for the ith individual, we notice that $X_i(t)$ is not known entirely, and it is only observed at time points: $t_{i1} < t_{i2} < \ldots < t_{iJ_i}$. A natural way to deal with this is to estimate $X_i(t)$ via a discrete function with jumps only at points t_{ij}'s; in turn, we have a step function $\hat{Z}_i(t)$ as an estimate for $Z_i(t) = \varphi(X_i^t)$, which only changes its value at time points: $0 = t_{i0} < t_{i1} < t_{i2} < \ldots < t_{iJ_i}$:

$$\hat{Z}_i(t) = Z_i(t_{iJ_i}) I\{t \geq t_{iJ_i}\} + \sum_{j=1}^{J_i - 1} Z_i(t_{ij}) I\{t_{ij} \leq t < t_{i,j+1}\}, \quad 0 \leq t \leq V_i, \tag{3.12}$$

where $\hat{Z}_i(t) = 0$ for $t \in [0, t_{i1})$ and $\hat{Z}_i(t) = Z_i(t_{ij})$ for any $t_{ij} \le t < t_{i,j+1}$. Thus, using $\hat{Z}_i(t)$ as an estimate of $Z_i(t)$ in (3.7), for any $t_{iJ_i} \le t \le V_i$:

$$
\begin{aligned}
&-\int_0^t h_0(s)\exp(\hat{Z}_i(s)\beta_0)ds \\
&\quad = -\sum_{k=1}^{J_i}\int_{t_{i,k-1}}^{t_{ik}} h_0(s)\exp(\hat{Z}_i(s)\beta_0)ds - \int_{t_{iJ_i}}^{t} h_0(s)\exp(\hat{Z}_i(s)\beta_0)ds \\
&\quad = -\sum_{k=1}^{J_i}\int_{t_{i,k-1}}^{t_{ik}} h_0(s)\exp(Z_i(t_{i,k-1})\beta_0)ds - \int_{t_{iJ_i}}^{t} h_0(s)\exp(Z_i(t_{iJ_i})\beta_0)ds \\
&\quad = -\sum_{k=1}^{J_i}\exp(Z_i(t_{i,k-1})\beta_0)\int_{t_{i,k-1}}^{t_{ik}} \frac{f_0(s)}{\bar{F}_0(s)}ds - \exp(Z_i(t_{iJ_i})\beta_0)\int_{t_{iJ_i}}^{t} \frac{f_0(s)}{\bar{F}_0(s)}ds \\
&\quad = \sum_{k=1}^{J_i}\exp(Z_i(t_{i,k-1})\beta_0)\ln \bar{F}_0(s)\Big|_{t_{i,k-1}}^{t_{ik}} + \exp(Z_i(t_{iJ_i})\beta_0)\ln \bar{F}_0(s)\Big|_{t_{iJ_i}}^{t} \\
&\quad = \sum_{k=1}^{J_i}\exp(Z_i(t_{i,k-1})\beta_0)\ln \frac{\bar{F}_0(t_{ik})}{\bar{F}_0(t_{i,k-1})} + \exp(Z_i(t_{iJ_i})\beta_0)\ln \frac{\bar{F}_0(t)}{\bar{F}_0(t_{iJ_i})}, \qquad (3.13)
\end{aligned}
$$

where we use notations: $0 = \hat{Z}_i(0) = \hat{Z}_i(t_{i0}) = Z_i(t_{i0}) = Z_i(0)$ and $f_0(t)$ is the p.d.f. of $F_0(t)$. With the proof given in Section 3.4, we have the following lemma.

Lemma 3.2 *For any $t_{iJ_i} \le t \le V_i$,*

$$
\bar{F}(t \,|\, X_i^t) \approx \bar{F}(t \,|\, \mathcal{X}_i^t) = [\bar{F}_0(t)]^{c_i} \prod_{k=1}^{J_i} \left(\bar{F}_0(t_{ik})\right)^{c_{i,k-1}-c_{ik}}, \qquad (3.14)
$$

where $c_{ik} = e^{\beta Z_i(t_{ik})}$ and $c_i = c_{iJ_i}$ with $\beta = \beta_0$.

Hence, we have for any $t_{iJ_i} \le t \le V_i$,

$$
f(t \,|\, X_i^t) \approx f(t \,|\, \mathcal{X}_i^t) = c_i f_0(t)[\bar{F}_0(t)]^{c_i - 1} \prod_{k=1}^{J_i} \left(\bar{F}_0(t_{ik})\right)^{c_{i,k-1}-c_{ik}}, \qquad (3.15)
$$

where due to (3.5), we have $f(t \,|\, \mathcal{X}_i^t)$ given in (3.15) as the derivative of $\bar{F}(t \,|\, \mathcal{X}_i^t)$ with respect to t, and $f(t \mid \mathcal{X}_i^t)$ is an estimate of $f(t \mid X_i^t)$.

Now, in likelihood function (3.11), we replace terms $f(t \mid X_i^t)$ and $\bar{F}(t \mid X_i^t)$ by $f(t \mid \mathcal{X}_i^t)$ in (3.15) and $\bar{F}(t \mid \mathcal{X}_i^t)$ in (3.14), respectively, then obtain the estimated likelihood of $L(V, \delta, X^V)$ for observed survival data with longitudinal covariates $(V_i, \delta_i, \mathcal{X}_i^{V_i})$ in (1.29) as follows:

$$\begin{aligned}
&\prod_{i=1}^{n}\big(f(V_i \mid X_i^{V_i})\big)^{\delta_i}\big(\bar{F}(V_i \mid X_i^{V_i})\big)^{1-\delta_i} \approx \prod_{i=1}^{n}\big(f(V_i \mid \mathcal{X}_i^{V_i})\big)^{\delta_i}\big(\bar{F}(V_i \mid \mathcal{X}_i^{V_i})\big)^{1-\delta_i} \\
&\quad = \prod_{i=1}^{n}\left(c_i f_0(V_i)[\bar{F}_0(V_i)]^{c_i-1}\prod_{k=1}^{J_i}\big(\bar{F}_0(t_{ik})\big)^{c_{i,k-1}-c_{ik}}\right)^{\delta_i}\left([\bar{F}_0(V_i)]^{c_i}\prod_{k=1}^{J_i}\big(\bar{F}_0(t_{ik})\big)^{c_{i,k-1}-c_{ik}}\right)^{1-\delta_i} \\
&\quad = \prod_{i=1}^{n}\Big(c_i f_0(V_i)\Big)^{\delta_i}\Big(\bar{F}_0(V_i)\Big)^{c_i-\delta_i}\left(\prod_{j=1}^{J_i}\big(\bar{F}_0(t_{ij})\big)^{d_{ij}}\right),
\end{aligned}$$

where $d_{ij} = c_{i,j-1} - c_{ij}$. Therefore, the estimated likelihood function for (β_0, F_0) in our proposed PH model (1.38) with right censored survival data with longitudinal covariates (1.29) is given by

$$L(\beta, F) = \prod_{i=1}^{n}\Big(c_i dF(V_i)\Big)^{\delta_i}\Big(\bar{F}(V_i)\Big)^{c_i-\delta_i}\left(\prod_{j=1}^{J_i}\big(\bar{F}(t_{ij})\big)^{d_{ij}}\right), \tag{3.16}$$

where F is an arbitrary d.f., $dF(t) = F(t) - F(t-)$, and from notation $Z_i(0)$ in (3.13),

$$\begin{cases}
c_{ik} = e^{\beta Z_i(t_{ik})}, \quad k = 0, 1, 2, \ldots, J_i \\
c_{i0} = e^{\beta Z_i(t_{i0})} = e^{\beta Z_i(0)} = e^0 = 1 \\
c_i = c_{iJ_i} = e^{\beta Z_i(t_{iJ_i})} \\
d_{ij} = c_{i,j-1} - c_{ij} = e^{\beta Z_i(t_{i,j-1})} - e^{\beta Z_i(t_{ij})}.
\end{cases} \tag{3.17}$$

The MLE $(\hat{\beta}_n, \hat{F}_n)$ for (β_0, F_0) in our proposed PH model (1.38) is the solution that maximizes above estimated likelihood function $L(\beta, F)$ given in (3.16).

Remark 1. From observed data (1.29), we do not know whether $Z_i(0) = 0$. But in (3.13), based on (3.12) we have $\hat{Z}_i(t) = 0$ for $\forall t \in [t_{i0}, t_{i1}) = [0, t_{i1})$ which implies $\hat{Z}_i(0) = 0$. In (3.13), we have $\hat{Z}_i(s) = 0$ for $s \in [t_{i0}, t_{i1})$ and $\hat{Z}_i(s) = Z_i(t_{ij})$ for $s \in [t_{ij}, t_{i,j+1})$, where $j = 1, \ldots, J_i$. Thus, the use of notation: $\hat{Z}_i(t_{i0}) = Z_i(t_{i0}) = Z_i(0) = 0$ in (3.13) is for convenience sake and is again used in (3.17) for c_{i0}.

3.2 Estimating Function under Restriction

Here, we use the empirical likelihood parameterization (Owen, 1988) to profile out nuisance parameter F_0, then obtain the profile likelihood estimating function for β_0 and the MLE $(\hat{\beta}_n, \hat{F}_n)$ for (β_0, F_0) in our proposed PH model (1.38) based on right censored survival data with longitudinal covariates (1.29).

Without loss of generality, we assume that there are no ties among V_i's in observed right censored survival data with longitudinal covariates (1.29), and assume that

$$V_1 < V_2 < \ldots < V_n. \tag{3.18}$$

If we restrict all possible candidates for the MLE of F_0 to those d.f.'s that assign all their probability masses to points V_i's and interval (V_n, ∞), then the d.f. F under consideration in likelihood function (3.16) is given by

$$F(t) = \sum_{i=1}^{n} p_i I\{V_i \leq t\}, \tag{3.19}$$

where $p_i = F(V_i) - F(V_i-)$ for $1 \leq i \leq n$, and $\sum_{i=1}^{n+1} p_i = 1$ with $0 \leq p_{n+1} \leq 1$ as the probability mass of F on interval (V_n, ∞). For any d.f. F given by (3.19), we let

$$\eta_{kij} = I\{V_k \leq t_{ij} < V_{k+1}\}, \quad V_{n+1} = \infty, \tag{3.20}$$

then $F(t_{ij}) = 0$ for $0 \leq t_{ij} < V_1$, and all $t_{ij} \leq V_i \leq V_n$, and we have in (3.16):

$$\begin{aligned}\prod_{i=1}^{n}\prod_{j=1}^{J_i}(\bar{F}(t_{ij}))^{d_{ij}} &= \prod_{i=1}^{n}\prod_{j=1}^{J_i}\prod_{k=1}^{n}(\bar{F}(t_{ij}))^{\eta_{kij}d_{ij}} = \prod_{i=1}^{n}\prod_{j=1}^{J_i}\prod_{k=1}^{n}(\bar{F}(V_k))^{\eta_{kij}d_{ij}} \\ &= \prod_{k=1}^{n}\prod_{i=1}^{n}\prod_{j=1}^{J_i}(\bar{F}(V_k))^{\eta_{kij}d_{ij}} = \prod_{k=1}^{n}(\bar{F}(V_k))^{d_{0k}}, \end{aligned} \tag{3.21}$$

where $d_{0k} = \sum_{i=1}^{n}\sum_{j=1}^{J_i}\eta_{kij}d_{ij}$, in turn, we have

$$L(\beta, F) = \prod_{i=1}^{n}\Big(c_i dF(V_i)\Big)^{\delta_i}\Big(\bar{F}(V_i)\Big)^{d_i-\delta_i}, \tag{3.22}$$

where

$$\begin{aligned} d_i = c_i + d_{0i} &= \exp(\beta Z_i(t_{iJ_i})) + \sum_{k=1}^{n}\sum_{j=1}^{J_k} d_{kj} I\{V_i \le t_{kj} < V_{i+1}\} \\ &= \exp(\beta Z_i(t_{iJ_i})) + \sum_{k=1}^{n}\sum_{j=1}^{J_k}\big(e^{\beta Z_k(t_{k,j-1})} - e^{\beta Z_k(t_{kj})}\big) I\{V_i \le t_{kj} < V_{i+1}\}. \end{aligned} \tag{3.23}$$

Thus, for any d.f. F given by (3.19), likelihood function (3.22) can be written as

$$L(\beta, F) = \prod_{i=1}^{n}\Big(c_i p_i\Big)^{\delta_i}\Big(\sum_{j=i+1}^{n+1} p_j\Big)^{d_i-\delta_i} \equiv L(\beta, \boldsymbol{p}), \tag{3.24}$$

where $\boldsymbol{p} = (p_1, \ldots, p_{n+1})$, c_i is given by (3.17) and d_i is given by (3.23). Hence, the MLE $(\hat{\beta}_n, \hat{F}_n)$ for (β_0, F_0) in our proposed PH model (1.38) with data (1.29) is the solution that maximizes $L(\beta, F)$ in (3.24).

Similar to the treatment in Ren and Zhou (2011), for now we assume that in (3.23) for some β:

$$d_i \ge 1 \quad \text{for all } 1 \le i \le n, \tag{AS1}$$

then we consider the following transformations for $L(\beta, \boldsymbol{p})$ in (3.24):

$$\begin{cases} a_i = \dfrac{p_i}{b_i}, \quad \text{for } i = 1, \ldots, n \\ \\ b_i = \sum_{j=i}^{n+1} p_j, \quad \text{for } i = 1, \ldots, n \end{cases} \tag{3.25}$$

and we have the following lemma with proof given in Section 3.4.

Lemma 3.3 *Under assumption (AS1), we have*

$$L(\beta, \boldsymbol{p}) = \prod_{i=1}^{n} (c_i a_i)^{\delta_i} (1 - a_i)^{e_i - \delta_i} \equiv L_1(\boldsymbol{a}; \beta), \tag{3.26}$$

which is maximized at

$$\hat{a}_i = \frac{\delta_i}{e_i}, \quad i = 1, 2, \ldots, n, \tag{3.27}$$

then likelihood function $L(\beta, F)$ in (3.24) is maximized by

$$[1 - \hat{F}_n(t; \beta)] = \prod_{V_i \le t} \frac{e_i - \delta_i}{e_i}, \tag{3.28}$$

where $e_i = \sum_{j=i}^{n} d_j$, c_i and d_i are given by (3.17) and (3.23), respectively.

Remark 2. Under assumption (AS1) that $d_i \ge 1$ for all $1 \le i \le n$, likelihood function $L(\beta, F)$ in (3.24) has a finite maximum value. Also, (AS1) implies $\hat{F}_n(t; \beta)$ given by (3.28) is a proper d.f. with all terms on the right-hand side of equation (3.28) between 0 and 1, because $e_i \ge d_n \ge 1$ for all $1 \le i \le n$.

Remark 3. Here is an example showing that if (AS1) does not hold, the maximum value of $L(\beta, F)$ given by (3.24) is not finite.Consider the case of $n = 2$ with $\delta_1 = \delta_2 = 1$,

$d_1 \geq 1$ and $d_2 < 1$ in (3.24). Then, noticing that $p_1 + p_2 + p_3 = 1$, we know that (3.24) can be written as:

$$
\begin{aligned}
L(\beta, F) &= \Big(c_1 p_1 (p_2 + p_3)^{d_1 - 1}\Big)\Big(c_2 p_2 (p_3)^{d_2 - 1}\Big) = \frac{c_1 c_2 p_1 p_2 (p_2 + p_3)^{d_1 - 1}}{p_3^{1-d_2}} \\
&\geq \frac{c_1 c_2 p_1 p_2^{d_1}}{p_3^{1-d_2}} = \frac{c_1 c_2 p_1 (1 - p_1 - p_3)^{d_1}}{p_3^{1-d_2}}.
\end{aligned}
$$

Thus, for any fixed $p_1 \in (0, 1)$, we have

$$
L(\beta, F) \geq \lim_{p_3 \to 0+} \frac{c_1 c_2 p_1 (1 - p_1 - p_3)^{d_1}}{p_3^{1-d_2}} = \frac{c_1 c_2 p_1 (1 - p_1)^{d_1}}{0+} = \infty,
$$

which means that $L(\beta, F)$ does not have a finite maximum value.

Estimating Function under Restriction

Now from (3.26)-(3.27), we obtain the profile likelihood function for β_0 under assumption (AS1):

$$
\ell(\beta) = \log L(\beta, \hat{F}_n(\cdot\,;\beta)) = \log\left(\prod_{i=1}^{n}\left(\frac{c_i}{e_i}\right)^{\delta_i}\left(\frac{e_i - \delta_i}{e_i}\right)^{e_i - \delta_i}\right), \tag{3.29}
$$

where

$$
e_i = d_i + \ldots + d_n, \quad i = 1, 2, \ldots, n. \tag{3.30}
$$

Thus, the MLE for β_0 is the solution $\hat{\beta}_n$ which maximizes the value of $\ell(\beta)$ in (3.29), and the MLE for F_0 is given by $\hat{F}_n(t) = \hat{F}_n(t; \hat{\beta}_n)$.

Differentiating $\ell(\beta)$, the profile likelihood estimating function $\psi_n(\beta)$ is given by:

$$\begin{aligned}\psi_n(\beta) &\equiv n^{-1}\ell^{'}(\beta)\\ &= n^{-1}\frac{d}{d\beta}\Bigg(\sum_{i=1}^{n}\big(\delta_i(\log c_i - \log e_i) + (e_i-\delta_i)(\log(e_i-\delta_i) - \log e_i)\big)\Bigg).\end{aligned} \tag{3.31}$$

Notice that above c_i's, d_i's and e_i's are given by (3.17), (3.23) and (3.30), respectively, thus they are functions of β. With notations $c_i^{'} = \frac{dc_i}{d\beta}$, $d_i^{'} = \frac{dd_i}{d\beta}$ and $e_i^{'} = \frac{de_i}{d\beta}$, we know $\frac{c_i^{'}}{c_i} = Z_i(t_{iJ_i})$ and we compute (3.31) as follows:

$$\begin{aligned}\psi_n(\beta) &= n^{-1}\sum_{i=1}^{n}\left(\delta_i\frac{c_i^{'}}{c_i} - \delta_i\frac{e_i^{'}}{e_i} + e_i^{'}(\log(e_i-\delta_i) - \log e_i) + (e_i-\delta_i)\left(\frac{e_i^{'}}{e_i-\delta_i} - \frac{e_i^{'}}{e_i}\right)\right)\\ &= n^{-1}\sum_{i=1}^{n}\left(\delta_i Z_i(t_{iJ_i}) + e_i^{'}\log\frac{e_i-\delta_i}{e_i}\right) = n^{-1}\sum_{i=1}^{n}\delta_i\left(Z_i(t_{iJ_i}) + e_i^{'}\log\frac{e_i-1}{e_i}\right).\end{aligned}$$

Therefore, we obtain the profile likelihood estimating function $\psi_n(\beta)$ for β_0 given by:

$$\psi_n(\beta) = n^{-1}\sum_{i=1}^{n}\delta_i\left(Z_i(t_{iJ_i}) + e_i^{'}\log\frac{e_i-1}{e_i}\right), \tag{3.32}$$

where

$$\begin{cases} e_i &= \sum\limits_{l=i}^{n}\left(e^{\beta Z_l(t_{lJ_l})} + \sum\limits_{k=1}^{n}\sum\limits_{j=1}^{J_k}\left(e^{\beta Z_k(t_{k,j-1})} - e^{\beta Z_k(t_{kj})}\right)I\{V_l \le t_{kj} < V_{l+1}\}\right)\\ e_i^{'} &= \sum\limits_{l=i}^{n}\Bigg(Z_l(t_{lJ_l})e^{\beta Z_l(t_{lJ_l})}\\ &\quad + \sum\limits_{k=1}^{n}\sum\limits_{j=1}^{J_k}\left(Z_k(t_{k,j-1})e^{\beta Z_k(t_{k,j-1})} - Z_k(t_{kj})e^{\beta Z_k(t_{kj})}\right)I\{V_l \le t_{kj} < V_{l+1}\}\Bigg)\end{cases}. \tag{3.33}$$

The solution of the estimating equation $\psi_n(\beta) = 0$ is the MLE $\hat{\beta}_n$ of β_0.

It should be noticed that above estimating function $\psi_n(\beta)$ is only defined for those β that satisfies condition (AS1). The generalized version of $\psi_n(\beta)$ without assumption (AS1) will be given in Section 3.3.

Leading Term of Taylor's Expansion of $\psi_n(\beta)$

Since Ren and Zhou (2011) showed that the estimating function based on Cox's partial likelihood is the the leading term of Taylor's expansion of that based on the full likelihood, we consider the Taylor's expansion of the estimating function $\psi_n(\beta)$ given in (3.33) as follows.

Consider Taylor's expansion of $\log(1+x)$ at $x = 0$: $\log(1+x) = x - \frac{x^2}{2(1+\xi)^2}$, where ξ is between 0 and x. We have the following for $\psi_n(\beta)$ in (3.32):

$$\begin{aligned}\psi_n(\beta) &= n^{-1}\sum_{i=1}^{n}\delta_i\bigg(Z_i(t_{iJ_i}) + e_i^{'}\bigg(-\frac{1}{e_i} - \frac{1}{2(1+\xi_i)^2e_i^2}\bigg)\bigg)\\ &= \varphi_n(\beta) - n^{-1}\sum_{i=1}^{n}\bigg(\frac{\delta_i e_i^{'}}{2(1+\xi_i)^2e_i^2}\bigg),\end{aligned} \tag{3.34}$$

where ξ_i is between 0 and $-\frac{1}{e_i}$, and $\varphi_n(\beta)$ is the leading term of Taylor's expansion of $\psi_n(\beta)$ in (3.32) and gives the *approximated estimating function* given by:

$$\varphi_n(\beta) = n^{-1}\sum_{i=1}^{n}\delta_i\bigg(Z_i(t_{iJ_i}) - \frac{e_i^{'}}{e_i}\bigg). \tag{3.35}$$

The solution of the approximated estimating equation $\varphi_n(\beta) = 0$ is the *approximated maximum likelihood estimator* (MLE) $\tilde{\beta}_n$ for β_0.

With the proof given in Section 3.4, we have the following lemma to simplify the expression of e_i's and $e_i^{'}$'s given in (3.33), and Lemma 3.5 shows the monotonicity of $\varphi_n(\beta)$ given in (3.35).

Lemma 3.4 *We can simplify the expression of e_i and $e_i^{'}$ in (3.33) as follows:*

$$\begin{cases} e_i & = \sum_{k=i}^{n} A_{ki} = \sum_{k=i}^{n} \left(e^{\beta Z_k(t_{kJ_k})} I\{\mathcal{E}_{ki} = \emptyset\} + e^{\beta Z_k(t_{km_{ki}})} I\{\mathcal{E}_{ki} \neq \emptyset\} \right) \\ e_i{}' & = \sum_{k=i}^{n} \left(Z_k(t_{kJ_k}) e^{\beta Z_k(t_{kJ_k})} I\{\mathcal{E}_{ki} = \emptyset\} + Z_k(t_{km_{ki}}) e^{\beta Z_k(t_{km_{ki}})} I\{\mathcal{E}_{ki} \neq \emptyset\} \right) \end{cases} \tag{3.36}$$

where $\mathcal{E}_{ki} = \left\{ t_{kj} \,|\, t_{kj} \geq V_i, j \in \{1, \ldots, J_k\} \right\}$ is a subset of $\{t_{k1}, t_{k2}, \ldots, t_{kJ_k}\}$, and

$$\begin{cases} A_{ki} = \begin{cases} e^{\beta Z_k(t_{kJ_k})} & \textit{if} \quad t_{kj} < V_i, \ \forall \ j \in \{1, \ldots, J_k\} \\ e^{\beta Z_k(t_{km_{ki}})} & \textit{if} \quad \exists \ t_{kj} \geq V_i, \ \textit{for some } j \in \{1, \ldots, J_k\} \end{cases} & (3.37) \\ m_{ki} = n_{ki} - 1 & (3.38) \end{cases}$$

for n_{ki} given by

$$n_{ki} = \min\{ \, j \mid t_{kj} \geq V_i, \ j = 1, \ldots, J_k\}, \tag{3.39}$$

noting that n_{ki} is not defined unless $\mathcal{E}_{ki} \neq \emptyset$.

Lemma 3.5 *Function $\varphi_n(\beta)$ given by (3.35) is non-increasing in $\beta \in \mathbb{R}$.*

3.3 Estimating Function without Restriction

As pointed out in Remark 3 of Section 3.2, assumption (AS1) is needed so that likelihood function $L(\beta, F)$ in (3.24) has a finite maximum value, and $L(\beta, F)$ in (3.24) is maximized by (3.28) via transformation (3.25). Here, we make adjustment to remove assumption (AS1) in estimating function for β_0.

First, notice that the key of achieving result in Section 3.2 is that transformation (3.25) reduces $L(\beta, F) = L(\beta, \boldsymbol{p})$ in (3.24) into the form $L_1(\boldsymbol{a}; \beta)$ given by (3.26). In such process, the derivations in the proof of Lemma 3.3 actually require all $p_i \in (0, 1)$ and $a_i \in (0, 1)$. Since for $\boldsymbol{p} = (p_1, \ldots, p_{n+1})$ and $\boldsymbol{a} = (a_1, \ldots, a_n)$, we know that in (3.24) and (3.26), assuming finite maximum value exists:

$$\max_{p_i \in [0,1]} L(\beta, \boldsymbol{p}) = \sup_{p_i \in (0,1)} L(\beta, \boldsymbol{p}) \overset{(3.26)}{=\!=} \sup_{a_i \in (0,1)} L_1(\boldsymbol{a}; \beta) = \max_{a_i \in [0,1]} L_1(\boldsymbol{a}; \beta). \tag{3.40}$$

This means that the finite maximum value of $L(\beta, \boldsymbol{p})$ exists for fixed β is equivalent to that of $L_1(\boldsymbol{a}; \beta)$. Thus, assumption (AS1) can be relaxed to the following assumption:

$$e_i \geq 1 \quad \text{for all } 1 \leq i \leq n, \tag{AS2}$$

which ensures that $L(\beta, \boldsymbol{p})$ in (3.24) and $L_1(\boldsymbol{a}; \beta)$ in (3.26) both have finite maximum value for fixed β.

Above (AS2) ensures that in (3.27) we have $\hat{a}_i \in [0, 1]$ and obviously (AS2) is a weaker condition than (AS1). Further, our Lemma 3.4 in Section 3.2 shows that all e_i's are all nonnegative for any β and we know that all $e_i \geq 1$ if $\beta \geq 0$ and all $Z_i(t_{ij}) \geq 0$. Thus, here we use ”shifting” method proposed in Ren and Zhou (2011), Ren and Wang (2022) to achieve (AS2) for β being restricted on a compact set.

For the ”shifting” method based on Remark 1 and 2 in Ren and Zhou (2011), we consider any constant $M \in \mathbb{R}$ and rewrite PH model (1.38) as follows:

$$h(t \,|\, X^t) = h_1(t)e^{\beta_0 Z(t)+M}, \tag{3.41}$$

where

$$h_1(t) \equiv h_M(t) = h_0(t)e^{-M}, \tag{3.42}$$

where $F_1(t)$ is the d.f. corresponding to $h_1(t)$ in (3.42) and has $f_1(t)$ as its p.d.f. Model (3.41) does not affect any derivations in (3.1)-(3.6) and (3.7) becomes

$$\bar{F}(t \,|\, X^t) = \exp\{-H(t \,|\, X^t)\} = \exp\Big\{-\int_0^t h_1(s)e^{\beta_0 Z(s)+M} ds\Big\}, \tag{3.43}$$

which is the ”Lehmann family property” for rewritten model formula (3.41). Thus, in the derivation of the likelihood function in Section 3.1.2, (3.8)-(3.12) are not affected by (3.41) and the formula corresponding to (3.13) becomes for $\forall\, t_{iJ_i} \leq t \leq V_i$:

$$\begin{aligned}
&-\int_0^t h_1(s)e^{\beta_0 \hat{Z}_i(s)+M} ds \\
&\qquad = -\sum_{k=1}^{J_i}\int_{t_{i,k-1}}^{t_{ik}} h_1(s)e^{\beta_0 \hat{Z}_i(s)+M} ds - \int_{t_{iJ_i}}^{t} h_1(s)e^{\beta_0 \hat{Z}_i(s)+M} ds \\
&\qquad = \sum_{k=1}^{J_i} e^{\beta_0 Z_i(t_{i,k-1})+M} \ln \frac{\bar{F}_1(t_{ik})}{\bar{F}_1(t_{i,k-1})} + e^{\beta_0 Z_i(t_{iJ_i})+M} \ln \frac{\bar{F}_1(t)}{\bar{F}_1(t_{iJ_i})},
\end{aligned} \tag{3.44}$$

where $F_1(t)$ is the d.f. corresponding to $h_1(t)$ in (3.42) and has $f_1(t)$ as its p.d.f. This gives the following formula corresponding to (3.14): for $\forall\, t_{iJ_i} \le t \le V_i$,

$$\bar{F}(t\,|\,X_i^t) \approx \bar{F}(t\,|\,\mathcal{X}_i^t) = \exp\Big\{-\int_0^t h_1(s)e^{\beta_0 \hat{Z}_i(s)+M}ds\Big\}$$
$$= [\bar{F}_1(t)]^{C_{iJ_i}} \prod_{k=1}^{J_i} \big(\bar{F}_1(t_{ik})\big)^{C_{i,k-1}-C_{ik}} = [\bar{F}_1(t)]^{C_i} \prod_{k=1}^{J_i} \big(\bar{F}_1(t_{ik})\big)^{C_{i,k-1}-C_{ik}}, \qquad (3.45)$$

where for $\beta = \beta_0$, we have

$$C_{ik} = e^{\beta Z_i(t_{ik})+M} \quad \text{and} \quad C_i = C_{iJ_i}. \qquad (3.46)$$

Thus, the formula corresponding to (3.15) is: for $\forall\, t_{iJ_i} \le t \le V_i$,

$$f(t\,|\,X_i^t) \approx f(t\,|\,\mathcal{X}_i^t) = C_i f_1(t)[\bar{F}_1(t)]^{C_i-1} \prod_{k=1}^{J_i} \big(\bar{F}_1(t_{ik})\big)^{C_{i,k-1}-C_{ik}}. \qquad (3.47)$$

In turn, (3.45)-(3.47) give the estimated likelihood of $L(V, \delta, X^V)$ in (3.11) as follows:

$$\prod_{i=1}^n \big(f(V_i\,|\,X_i^{V_i})\big)^{\delta_i}\big(\bar{F}(V_i\,|\,X_i^{V_i})\big)^{1-\delta_i} \approx \prod_{i=1}^n \big(f(V_i\,|\,\mathcal{X}_i^{V_i})\big)^{\delta_i}\big(\bar{F}(V_i\,|\,\mathcal{X}_i^{V_i})\big)^{1-\delta_i}$$
$$= \prod_{i=1}^n \Big(C_i f_1(V_i)[\bar{F}_1(V_i)]^{C_i-1} \prod_{k=1}^{J_i} \big(\bar{F}_1(t_{ik})\big)^{C_{i,k-1}-C_{ik}}\Big)^{\delta_i} \Big([\bar{F}_1(V_i)]^{C_i} \prod_{k=1}^{J_i} \big(\bar{F}_1(t_{ik})\big)^{C_{i,k-1}-C_{ik}}\Big)^{1-\delta_i}$$
$$= \prod_{i=1}^n \Big(C_i f_1(V_i)\Big)^{\delta_i} \Big(\bar{F}_1(V_i)\Big)^{C_i-\delta_i} \Big(\prod_{j=1}^{J_i} \big(\bar{F}_1(t_{ij})\big)^{D_{ij}}\Big),$$

where $D_{ij} = C_{i,j-1} - C_{ij}$. Therefore, the estimated likelihood function for (β_0, F_1) in the PH model (3.41) is given by

$$L_M(\beta, F) = \prod_{i=1}^n \Big(C_i dF(V_i)\Big)^{\delta_i} \Big(\bar{F}(V_i)\Big)^{C_i-\delta_i} \Big(\prod_{j=1}^{J_i} \big(\bar{F}(t_{ij})\big)^{D_{ij}}\Big), \qquad (3.48)$$

where F is an arbitrary d.f., $dF(t) = F(t) - F(t-)$, and from notation $Z_i(0)$ in (3.13):

$$\begin{cases} C_{ik} = e^{\beta Z_i(t_{ik})+M}, \quad k = 0, 1, 2, \ldots, J_i \\ C_{i0} = e^{\beta Z_i(t_{i0})+M} = e^{\beta Z_i(0)+M} = e^M \\ C_i = C_{iJ_i} = e^{\beta Z_i(t_{iJ_i})+M} \\ D_{ij} = C_{i,j-1} - C_{ij} = e^{\beta Z_i(t_{i,j-1})+M} - e^{\beta Z_i(t_{ij})+M}. \end{cases} \tag{3.49}$$

The relation between $L(\beta, F)$ given by (3.16) and $L_M(\beta, F)$ given by (3.48) is as follows. Notice that from (3.42) we have that $\forall\, t > 0$:

$$\begin{aligned} &\int_0^t h_1(s)ds = e^{-M} \int_0^t h_0(s)ds \quad \Leftrightarrow \quad H_1(t) = e^{-M} H_0(t) \\ \Leftrightarrow \quad & e^{-H_1(t)} = [e^{-H_0(t)}]^{e^{-M}} \qquad\qquad \Leftrightarrow \quad \bar{F}_1(t) = [\bar{F}_0(t)]^{e^{-M}}. \end{aligned} \tag{3.50}$$

Thus, we have from (3.48):

$$\begin{aligned} L_M(\beta, F_1) &= \prod_{i=1}^n \Big(C_i dF_1(V_i)\Big)^{\delta_i} \Big(\bar{F}_1(V_i)\Big)^{C_i - \delta_i} \left(\prod_{j=1}^{J_i} \big(\bar{F}_1(t_{ij})\big)^{D_{ij}}\right) \\ &= \prod_{i=1}^n \Big(C_i d\Big(1 - [\bar{F}_0(V_i)]^{e^{-M}}\Big)\Big)^{\delta_i} \Big([\bar{F}_0(V_i)]^{e^{-M}}\Big)^{C_i - \delta_i} \left(\prod_{j=1}^{J_i} \Big([\bar{F}_0(t_{ij})]^{e^{-M}}\Big)^{D_{ij}}\right) \\ &= \prod_{i=1}^n \Big(C_i e^{-M} [\bar{F}_0(V_i)]^{e^{-M}-1} dF_0(V_i)\Big)^{\delta_i} \Big(\bar{F}_0(V_i)\Big)^{(C_i - \delta_i)e^{-M}} \left(\prod_{j=1}^{J_i} \Big([\bar{F}_0(t_{ij})]\Big)^{D_{ij} e^{-M}}\right) \\ &= \prod_{i=1}^n \Big(c_i [\bar{F}_0(V_i)]^{e^{-M}-1} dF_0(V_i)\Big)^{\delta_i} \Big(\bar{F}_0(V_i)\Big)^{c_i - \delta_i e^{-M}} \left(\prod_{j=1}^{J_i} \Big([\bar{F}_0(t_{ij})]\Big)^{d_{ij}}\right) \\ &= \prod_{i=1}^n \Big(c_i dF_0(V_i)\Big)^{\delta_i} \Big(\bar{F}_0(V_i)\Big)^{c_i - \delta_i} \left(\prod_{j=1}^{J_i} \Big([\bar{F}_0(t_{ij})]\Big)^{d_{ij}}\right) = L(\beta, F_0) \end{aligned}$$

because we have

$$\begin{cases} c_i \ = e^{\beta Z_i(t_{iJ_i})} = e^{\beta Z_i(t_{iJ_i})+M} \times e^{-M} = C_i e^{-M} \\ d_{ij} = e^{\beta Z_i(t_{i,j-1})} - e^{\beta Z_i(t_{ij})} = \left(e^{\beta Z_i(t_{i,j-1})+M} - e^{\beta Z_i(t_{ij})+M}\right) e^{-M} = D_{ij} e^{-M}. \end{cases} \tag{3.51}$$

Hence, from (3.16) and (3.48) we have for any d.f.'s G and F:

$$L_M(\beta, G) = L(\beta, F), \quad \text{where} \quad \bar{G} = \left[\bar{F}\right]^{e^{-M}}. \tag{3.52}$$

This means that if we can choose a proper constant M such that (AS2) holds for $L_M(\beta, G)$ when β is restricted on certain compact set, then we can find $\hat{G}_n(t;\beta)$ for given β that maximizes $L_M(\beta, G)$, and obtain the profile likelihood function for β: $\ell_M(\beta) = \log L_M(\beta, \hat{G}_n(\cdot;\beta))$, which gives the estimating function $\psi_M(\beta) = n^{-1}\ell'_M(\beta)$. Thus, the solution $\hat{\beta}_n$ of equation $\psi_M(\beta) = 0$ is the MLE for β_0 and is well-defined; in turn, $\hat{G}_n(t) = \hat{G}_n(t;\hat{\beta}_n) \equiv \hat{F}_1(t)$ is the MLE for F_1, and by (3.52), the MLE $\hat{F}_n(t)$ for F_0 is given by

$$\bar{\hat{F}}_n(t) = \left[\bar{\hat{F}}_1(t)\right]^{e^M} = \left[\bar{\hat{G}}_n(t;\hat{\beta}_n)\right]^{e^M}. \tag{3.53}$$

In Section 3.2, (3.18)-(3.20) are not affected by (3.41) and the corresponding formula to (3.21) becomes:

$$\begin{aligned} \prod_{i=1}^{n}\prod_{j=1}^{J_i}(\bar{F}(t_{ij}))^{D_{ij}} &= \prod_{i=1}^{n}\prod_{j=1}^{J_i}\prod_{k=1}^{n}(\bar{F}(t_{ij}))^{\eta_{kij}D{ij}} = \prod_{i=1}^{n}\prod_{j=1}^{J_i}\prod_{k=1}^{n}(\bar{F}(V_k))^{\eta_{kij}D_{ij}} \\ &= \prod_{k=1}^{n}\prod_{i=1}^{n}\prod_{j=1}^{J_i}(\bar{F}(V_k))^{\eta_{kij}D_{ij}} = \prod_{k=1}^{n}(\bar{F}(V_k))^{D_{0k}}, \end{aligned} \tag{3.54}$$

where $D_{0k} = \sum_{i=1}^{n} \sum_{j=1}^{J_i} \eta_{kij} D_{ij}$, and from (3.48) the formula corresponding to (3.22) is given by:

$$L_M(\beta, F) = \prod_{i=1}^{n} \Big(C_i dF(V_i) \Big)^{\delta_i} \Big(\bar{F}(V_i) \Big)^{D_i - \delta_i}, \tag{3.55}$$

where

$$\begin{aligned} D_i &= C_i + D_{0i} \\ &= e^{\beta Z_i(t_{iJ_i})+M} + \sum_{k=1}^{n} \sum_{j=1}^{J_k} \big(e^{\beta Z_k(t_{k,j-1})+M} - e^{\beta Z_k(t_{kj})+M} \big) I\{V_i \le t_{kj} < V_{i+1}\}. \end{aligned} \tag{3.56}$$

Thus, for any d.f. G given by (3.19), likelihood function (3.55) can be written as

$$L_M(\beta, G) = \prod_{i=1}^{n} \Big(C_i p_i \Big)^{\delta_i} \Big(\sum_{j=i+1}^{n+1} p_j \Big)^{D_i - \delta_i} \equiv L_M(\beta, \boldsymbol{p}), \tag{3.57}$$

where C_i is given by (3.49) and D_i is given by (3.56).

From the derivation of (3.28), we know when constant M is properly chose such that (AS2) holds for $L_M(\beta, G)$ for fixed β being restricted on a compact set, likelihood function $L_M(\beta, G)$ is maximized by

$$\Big[1 - \hat{G}_n(t; \beta) \Big] = \prod_{V_i \le t} \frac{E_i - \delta_i}{E_i}, \tag{3.58}$$

where from (3.28) regarding e_i, E_i is given by:

$$E_i = D_i + \ldots + D_n, \tag{3.59}$$

for D_i's given by (3.56). The assumption (AS2) for $L_M(\beta, G)$ is for E_i's given by

$$E_i \geq 1 \quad \text{for all } 1 \leq i \leq n. \tag{AS3}$$

We know that condition (AS3) implies $\hat{G}_n(t;\beta)$ given by (3.58) is a proper d.f. Hence, in practice we just need to choose data based M to satisfy (AS3), and such choice of M will be given in Section 4.1.

In turn, parallel to (3.29), the profile likelihood function for β_0 based on $\hat{G}_n(\cdot;\beta)$ is given by:

$$\begin{aligned}
\ell_M(\beta) = \log L_M(\beta, \hat{G}_n(\cdot\,;\beta)) &= \log\left(\prod_{i=1}^n \left(\frac{C_i}{E_i}\right)^{\delta_i}\left(\frac{E_i-\delta_i}{E_i}\right)^{E_i-\delta_i}\right)\\
&= \sum_{i=1}^n \Big(\delta_i(\log C_i - \log E_i) + (E_i-\delta_i)\big[\log(E_i-\delta_i) - \log E_i\big]\Big).
\end{aligned}$$

Thus, the profile likelihood estimating function corresponding to (3.32) is given by:

$$\psi_M(\beta) = \frac{1}{n}\sum_{i=1}^n \delta_i\left(Z_i(t_{iJ_i}) + E_i' \log\frac{E_i-1}{E_i}\right), \tag{3.60}$$

where

$$\begin{cases}
E_i = \sum\limits_{l=i}^n \left(e^{\beta Z_l(t_{lJ_l})+M} + \sum\limits_{k=1}^n\sum\limits_{j=1}^{J_k}\left(e^{\beta Z_k(t_{k,j-1})+M} - e^{\beta Z_k(t_{kj})+M}\right)I\{V_l \leq t_{kj} < V_{l+1}\}\right)\\
E_i' = \sum\limits_{l=i}^n \Bigg(Z_l(t_{lJ_l})e^{\beta Z_l(t_{lJ_l})+M}\\
\qquad + \sum\limits_{k=1}^n\sum\limits_{j=1}^{J_k}\left(Z_k(t_{k,j-1})e^{\beta Z_k(t_{k,j-1})+M} - Z_k(t_{kj})e^{\beta Z_k(t_{kj})+M}\right)I\{V_l \leq t_{kj} < V_{l+1}\}\Bigg).
\end{cases} \tag{3.61}$$

From (3.33) and (3.61), the relation between e_i and E_i, $e_i^{'}$ and $E_i^{'}$ are given by:

$$E_i = e^M e_i \quad \text{and} \quad E_i^{'} = e^M e_i^{'}. \tag{3.62}$$

The solution of the estimating equation $\psi_M(\beta) = 0$ is the MLE $\hat{\beta}_n$ of β_0.

In Section 4.1, we show that for any fixed M, the leading term of Taylor's expansion of $\psi_M(\beta)$ in (3.60) gives the exact same approximated estimating function $\varphi_n(\beta)$ given in (3.35).

3.4 Proofs

Proof of Lemma 3.1 To see (3.9), note that for $\delta = 1$,

$$g(t, 1 \,|\, X^t) = \lim_{\Delta \to 0} \Delta^{-1} P\{t \le V < t + \Delta,\, \delta = 1 \,|\, X^t\}$$

$$= \lim_{\Delta \to 0} \Delta^{-1} P\{t \le T < t + \Delta,\, T \le C \,|\, X^t\},$$

and we have

$$P\{t \le T < t + \Delta,\, T \le C \,|\, X^t\} \le P\{t \le T < t + \Delta,\, t \le C \,|\, X^t\}$$

$$= P\{t \le T < t + \Delta \,|\, X^t\} P\{t \le C \,|\, X^t\}$$

$$= P\{t \le T < t + \Delta \,|\, X^t\} P\{t \le C\} = P\{t \le T < t + \Delta \,|\, X^t\} \bar{F}_C(t),$$

$$P\{t \le T < t + \Delta,\, T \le C \,|\, X^t\} \ge P\{t \le T < t + \Delta,\, t + \Delta \le C \,|\, X^t\}$$

$$= P\{t \le T < t + \Delta \,|\, X^t\} P\{t + \Delta \le C \,|\, X^t\}$$

$$= P\{t \le T < t + \Delta \,|\, X^t\} P\{t + \Delta \le C\} = P\{t \le T < t + \Delta \,|\, X^t\} \bar{F}_C(t + \Delta),$$

which by (3.1) imply the first equation in (3.9) because

$$\lim_{\Delta \to 0} \frac{P\{t \le T < t + \Delta \,|\, X^t\}}{\Delta} \bar{F}_C(t + \Delta) \le g(t, 1 \,|\, X^t) \le \lim_{\Delta \to 0} \frac{P\{t \le T < t + \Delta \,|\, X^t\}}{\Delta} \bar{F}_C(t).$$

Also, note that for $\delta = 0$,

$$g(t, 0 \,|\, X^t) = \lim_{\Delta \to 0} \Delta^{-1} P\{t \le V < t + \Delta,\, \delta = 0 \,|\, X^t\}$$

$$= \lim_{\Delta \to 0} \Delta^{-1} P\{t \le C < t + \Delta,\, T > C \,|\, X^t\},$$

and we have

$$P\{t \le C < t+\Delta,\, T > C \,|\, X^t\} \le P\{t \le C < t+\Delta,\, T \ge t \,|\, X^t\}$$

$$= P\{t \le C < t+\Delta \,|\, X^t\} P\{T \ge t \,|\, X^t\} \overset{(3.2)}{=} P\{t \le C < t+\Delta\} \bar{F}(t \,|\, X^t),$$

$$P\{t \le C < t+\Delta,\, T > C \,|\, X^t\} \ge P\{t \le C < t+\Delta,\, T \ge t+\Delta \,|\, X^t\}$$

$$= P\{t \le C < t+\Delta \,|\, X^t\} P\{T \ge t+\Delta \,|\, X^t\} \overset{(3.2)}{=} P\{t \le C < t+\Delta\} \bar{F}(t+\Delta \,|\, X^t),$$

which by (3.1) imply the second equation in (3.9) because

$$\lim_{\Delta \to 0} \frac{P\{t \le C < t+\Delta\}}{\Delta} \bar{F}(t+\Delta \,|\, X^t) \le g(t, 0 \,|\, X^t) \le \lim_{\Delta \to 0} \frac{P\{t \le C < t+\Delta\}}{\Delta} \bar{F}(t \,|\, X^t).$$

□

Proof of Lemma 3.2 Replacing the last exponent in (3.7) by (3.13), we obtain $\bar{F}(t \,|\, \mathcal{X}_i^t)$ as an estimate of $\bar{F}(t \,|\, X_i^t)$, for any $t_{iJ_i} \le t \le V_i$,

$$\begin{aligned}
\bar{F}(t \,|\, X_i^t) \approx \bar{F}(t \,|\, \mathcal{X}_i^t) &= \exp\Big\{ -\int_0^t h_0(s) \exp(\hat{Z}_i(s)\beta_0) ds \Big\} \\
&= \exp\left\{ \sum_{k=1}^{J_i} \exp(Z_i(t_{i,k-1})\beta_0) \ln \frac{\bar{F}_0(t_{ik})}{\bar{F}_0(t_{i,k-1})} + \exp(Z_i(t_{iJ_i})\beta_0) \ln \frac{\bar{F}_0(t)}{\bar{F}_0(t_{iJ_i})} \right\} \\
&= \left(\prod_{k=1}^{J_i} \exp\left\{ e^{Z_i(t_{i,k-1})\beta_0} \ln \frac{\bar{F}_0(t_{ik})}{\bar{F}_0(t_{i,k-1})} \right\} \right) \exp\left\{ e^{Z_i(t_{iJ_i})\beta_0} \ln \frac{\bar{F}_0(t)}{\bar{F}_0(t_{iJ_i})} \right\} \\
&= \left(\prod_{k=1}^{J_i} \left\{ \frac{\bar{F}_0(t_{ik})}{\bar{F}_0(t_{i,k-1})} \right\}^{c_{i,k-1}} \right) \left\{ \frac{\bar{F}_0(t)}{\bar{F}_0(t_{iJ_i})} \right\}^{c_{iJ_i}} \\
&= \frac{[\bar{F}_0(t_{i1})]^{c_{i0}}}{[\bar{F}_0(t_{i0})]^{c_{i0}}} \times \frac{[\bar{F}_0(t_{i2})]^{c_{i1}}}{[\bar{F}_0(t_{i1})]^{c_{i1}}} \times \ldots \times \frac{[\bar{F}_0(t_{iJ_i})]^{c_{i,J_i-1}}}{[\bar{F}_0(t_{i,J_i-1})]^{c_{i,J_i-1}}} \times \frac{[\bar{F}_0(t)]^{c_{iJ_i}}}{[\bar{F}_0(t_{iJ_i})]^{c_{iJ_i}}} \\
&= [\bar{F}_0(t)]^{c_{iJ_i}} \prod_{k=1}^{J_i} \big(\bar{F}_0(t_{ik})\big)^{c_{i,k-1}-c_{ik}} = [\bar{F}_0(t)]^{c_i} \prod_{k=1}^{J_i} \big(\bar{F}_0(t_{ik})\big)^{c_{i,k-1}-c_{ik}}, \qquad (3.63)
\end{aligned}$$

where $c_{ik} = e^{\beta Z_i(t_{ik})}$ and $c_i = c_{iJ_i}$ with $\beta = \beta_0$. □

Proof of Lemma 3.3 Consider any fixed β satisfying (AS1) for c_i's and d_i's given by (3.17) and (3.23), respectively. From (3.25), we have the following:

$$a_i = \frac{p_i}{b_i}, \quad b_i = \sum_{j=i}^{n+1} p_j, \tag{3.64}$$

then we have

$$b_1 = 1, \quad b_{n+1} = p_{n+1}, \quad b_{i+1} = b_i - p_i, \quad 1 - a_i = \frac{b_{i+1}}{b_i}. \tag{3.65}$$

Note that

$$\prod_{i=1}^{n}(1-a_i) \overset{(3.65)}{=} \prod_{i=1}^{n} \frac{b_{i+1}}{b_i} = \frac{b_2}{b_1}\,\frac{b_3}{b_2}\dots\frac{b_{n+1}}{b_n} = b_{n+1}, \tag{3.66}$$

and

$$\begin{aligned}
\prod_{i=1}^{n}(a_i)^{d_i}(1-a_i)^{n-h_i} &\overset{(3.65)}{=} \left(\prod_{i=1}^{n}(a_i)^{d_i}\right)\prod_{i=1}^{n}\left(\frac{b_{i+1}}{b_i}\right)^{n-h_i} \\
&= \left(\prod_{i=1}^{n}(a_i)^{d_i}\right)\left(\frac{b_2}{b_1}\right)^{n-h_1}\left(\frac{b_3}{b_2}\right)^{n-h_2}\dots\left(\frac{b_{n+1}}{b_n}\right)^{n-h_n} \\
&= \left(\prod_{i=1}^{n}(a_i)^{d_i}\right)(b_2)^{h_2-h_1}(b_3)^{h_3-h_2}\dots(b_n)^{h_n-h_{n-1}}(b_{n+1})^{n-h_n} \\
&= \left(\prod_{i=1}^{n}(a_i)^{d_i}\right)(b_1)^{d_1}(b_2)^{d_2}\dots(b_n)^{d_n}(b_{n+1})^{n-h_n} = \left(\prod_{i=1}^{n}(a_i)^{d_i}(b_i)^{d_i}\right)(b_{n+1})^{n-n\bar{d}} \\
&\overset{(3.64)}{=} (b_{n+1})^{n-n\bar{d}}\prod_{i=1}^{n}(p_i)^{d_i},
\end{aligned} \tag{3.67}$$

where $h_i = d_1 + d_2 + \dots + d_i$ and $\bar{d} = n^{-1}\sum_{i=1}^{n} d_i$. From (3.66) and (3.67), we can write likelihood function $L(\beta, F) = L(\beta, \boldsymbol{p})$ in (3.24) as

$$
\begin{aligned}
L(\beta, \boldsymbol{p}) &\overset{(3.64)}{=} \prod_{i=1}^{n} \left(c_i p_i\right)^{\delta_i} \left(b_{i+1}\right)^{d_i-\delta_i} = \prod_{i=1}^{n} \left((c_i)^{\delta_i}(p_i)^{d_i}\right) \left(\frac{b_{i+1}}{p_i}\right)^{d_i-\delta_i} \\
&\overset{(3.64)}{=} \prod_{i=1}^{n} \left((c_i)^{\delta_i}(p_i)^{d_i}\right) \left(\frac{b_{i+1}}{a_i b_i}\right)^{d_i-\delta_i} \overset{(3.65)}{=} \prod_{i=1}^{n} \left((c_i)^{\delta_i}(p_i)^{d_i}\right) \left(\frac{1-a_i}{a_i}\right)^{d_i-\delta_i} \\
&= \left(\prod_{i=1}^{n}(c_i)^{\delta_i}(p_i)^{d_i}\right) \frac{\prod_{i=1}^{n}(a_i)^{\delta_i}(1-a_i)^{n-\delta_i-(d_1+\ldots+d_{i-1})}}{\prod_{i=1}^{n}(a_i)^{d_i}(1-a_i)^{n-(d_1+\ldots+d_i)}} \\
&\overset{(3.67)}{=} \left(\prod_{i=1}^{n}(c_i)^{\delta_i}(p_i)^{d_i}\right) \frac{\prod_{i=1}^{n}(a_i)^{\delta_i}(1-a_i)^{n-\delta_i-(d_1+\ldots+d_{i-1})}}{(b_{n+1})^{n-n\bar{d}}\prod_{i=1}^{n}(p_i)^{d_i}} \\
&\overset{(3.66)}{=} \left(\prod_{i=1}^{n}(c_i)^{\delta_i}\right) \frac{\prod_{i=1}^{n}(a_i)^{\delta_i}(1-a_i)^{n-\delta_i-(d_1+\ldots+d_{i-1})}}{\prod_{i=1}^{n}(1-a_i)^{n-n\bar{d}}} \\
&= \prod_{i=1}^{n}(c_i)^{\delta_i}(a_i)^{\delta_i}(1-a_i)^{d_i+\ldots+d_n-\delta_i} \\
&= \prod_{i=1}^{n}(c_i a_i)^{\delta_i}(1-a_i)^{e_i-\delta_i} \equiv L_1(\boldsymbol{a};\beta),
\end{aligned}
\tag{3.68}
$$

where $\boldsymbol{a} = (a_1, \ldots, a_n)$ and $e_i = d_i + \ldots + d_n$, which gives (3.26) in Lemma 3.3.

Then by taking logarithm we obtain:

$$
\log L_1(\boldsymbol{a};\beta) = \sum_{i=1}^{n}[\delta_i(\log c_i + \log a_i) + (e_i - \delta_i)\log(1-a_i)].
$$

Setting $\frac{\partial \log L_1}{\partial a_i} = 0$, we obtain

$$
\frac{\delta_i}{a_i} - \frac{e_i - \delta_i}{1 - a_i} = 0 \quad \Longrightarrow \quad \hat{a}_i = \frac{\delta_i}{e_i}, \quad i = 1, 2, \ldots, n. \tag{3.69}
$$

Also, we have

$$
\frac{\partial^2(\log L_1)}{\partial a_j \partial a_i} = \begin{cases} 0 & \text{if} \quad i \neq j \\ -\frac{\delta_i}{a_i^2} - \frac{e_i-\delta_i}{(1-a_i)^2} < 0 & \text{if} \quad i = j \end{cases}
$$

because $e_i = d_i + \ldots + d_n \geq d_n \geq 1$ under the condition (AS1). Thus, for any fixed β satisfying (AS1), $\hat{a}_i$ in (3.27) maximizes $L_1(\boldsymbol{a};\beta)$ in (3.26) and (AS1) implies all

$0 \le \hat{a}_i \le 1$, which gives (3.27) in Lemma 3.3.

From (3.64)-(3.65), we can find $\hat{p}_i$'s corresponding to $\hat{a}_i$'s as follows:

$$\hat{p}_1 = \hat{a}_1\hat{b}_1 = \hat{a}_1, \qquad \hat{p}_2 = \hat{a}_2\hat{b}_2 = \hat{a}_2(\hat{b}_1 - \hat{p}_1) = \hat{a}_2(1-\hat{a}_1)$$

$$\hat{p}_3 = \hat{a}_3\hat{b}_3 = \hat{a}_3(\hat{b}_2 - \hat{p}_2) = \hat{a}_3(\hat{b}_1 - \hat{p}_1 - \hat{p}_2) = \hat{a}_3(1-\hat{a}_1-\hat{a}_2(1-\hat{a}_1)) = \hat{a}_3(1-\hat{a}_2)(1-\hat{a}_1)$$

$$\vdots$$

$$\hat{p}_i = \hat{a}_i(1-\hat{a}_{i-1})\dots(1-\hat{a}_1) = \hat{a}_i\prod_{j=1}^{i-1}(1-\hat{a}_j) = \frac{\delta_i}{e_i}\prod_{j=1}^{i-1}\left(1-\frac{\delta_j}{e_j}\right). \tag{3.70}$$

Thus, for any fixed β satisfying (AS1), $L(\beta, F) = L(\beta, \boldsymbol{p})$ is maximized by $\hat{F}_n(t;\beta) = \sum_{i=1}^n \hat{p}_i I\{V_i \le t\}$, where $\hat{p}_i$ is given by (3.70). Notice that for any $1 \le k \le n$, we have that for any $V_k \le t < V_{k+1}$, (3.70) gives

$$\begin{aligned}
\hat{F}_n(t;\beta) &= \sum_{i=1}^n \hat{p}_i I\{V_i \le t\} = \sum_{i=1}^n \hat{p}_i I\{V_i \le V_k\} = \sum_{i=1}^k \hat{p}_i = \sum_{i=1}^k \left[\frac{\delta_i}{e_i}\prod_{j=1}^{i-1}\left(1-\frac{\delta_j}{e_j}\right)\right] \\
&= \sum_{i=1}^k \left[\left(1+\frac{\delta_i}{e_i}-1\right)\prod_{j=1}^{i-1}\left(1-\frac{\delta_j}{e_j}\right)\right] = \sum_{i=1}^k \left[\prod_{j=1}^{i-1}\left(1-\frac{\delta_j}{e_j}\right) - \prod_{j=1}^{i}\left(1-\frac{\delta_j}{e_j}\right)\right] \\
&= \prod_{j=1}^{0}\left(1-\frac{\delta_j}{e_j}\right) - \prod_{j=1}^{k}\left(1-\frac{\delta_j}{e_j}\right) = 1 - \prod_{j=1}^{k}\left(1-\frac{\delta_j}{e_j}\right)
\end{aligned}$$

which implies

$$1 - \hat{F}_n(t;\beta) = \prod_{j=1}^{k}\left(1-\frac{\delta_j}{e_j}\right), \quad \text{for } V_k \le t < V_{k+1},\ 1 \le k \le n,$$

which finishes the proof of Lemma 3.3. $\square$

Proof of Lemma 3.4 From (3.17) and (3.20) we know

$$d_{kj} = e^{\beta Z_k(t_{k,j-1})} - e^{\beta Z_k(t_{kj})}, \quad \eta_{lkj} = I\{V_l \le t_{kj} < V_{l+1}\}, \quad V_{n+1} = \infty. \tag{3.71}$$

Let

$$N = \max\{J_1, \dots, J_n\} \quad \text{and} \quad S_i = \sum_{l=i}^{n} e^{\beta Z_l(t_{lJ_l})}, \tag{3.72}$$

then e_i in (3.33) can be written as follows:

$$\begin{aligned}
e_i &= \sum_{l=i}^{n} e^{\beta Z_l(t_{lJ_l})} + \sum_{l=i}^{n}\sum_{k=1}^{n}\sum_{j=1}^{J_k} \left(e^{\beta Z_k(t_{k,j-1})} - e^{\beta Z_k(t_{kj})}\right) I\{V_l \le t_{kj} < V_{l+1}\} \\
&\overset{(3.71)-(3.72)}{=\!=\!=} S_i + \sum_{l=1}^{n}\sum_{k=1}^{n}\sum_{j=1}^{N} d_{kj}\eta_{lkj} I\{l \ge i\} I\{j \le J_k\} \\
&= S_i + \sum_{k=1}^{n}\sum_{j=1}^{N}\sum_{l=1}^{n} d_{kj}\eta_{lkj} I\{l \ge i\} I\{j \le J_k\} \\
&= S_i + \sum_{k=1}^{n}\sum_{j=1}^{N} d_{kj} I\{j \le J_k\} \left(\sum_{l=1}^{n} \eta_{lkj} I\{l \ge i\} \right) \\
&\overset{(3.71)}{=\!=} S_i + \sum_{k=1}^{n}\sum_{j=1}^{J_k} d_{kj} \left(\sum_{l=i}^{n} I\{V_l \le t_{kj} < V_{l+1}\} \right) \\
&\overset{(3.18)}{=\!=} S_i + \sum_{k=1}^{n}\sum_{j=1}^{J_k} d_{kj} I\{V_i \le t_{kj} < V_{n+1}\} \\
&\overset{(3.72)}{=\!=} S_i + \sum_{k=1}^{n}\sum_{j=1}^{J_k} d_{kj} I\{V_i \le t_{kj}\}.
\end{aligned} \tag{3.73}$$

For $i = 1$, (3.73) can be written as:

$$\begin{aligned}
e_1 &= S_1 + \sum_{k=1}^{n}\sum_{j=1}^{J_k} d_{kj} I\{V_1 \le t_{kj}\} \\
&\overset{(3.72)}{=\!=} \sum_{l=1}^{n} e^{\beta Z_l(t_{lJ_l})} + \sum_{k=1}^{n}\sum_{j=1}^{J_k} d_{kj} I\{V_1 \le t_{kj}\} \\
&= \sum_{k=1}^{n} e^{\beta Z_k(t_{kJ_k})} + \sum_{k=1}^{n}\sum_{j=1}^{J_k} d_{kj} I\{V_1 \le t_{kj}\} \\
&= \sum_{k=1}^{n} \left(e^{\beta Z_k(t_{kJ_k})} + \sum_{j=1}^{J_k} d_{kj} I\{V_1 \le t_{kj}\} \right).
\end{aligned} \tag{3.74}$$

For $2 \leq i \leq n$, from (3.73) we have

$$e_i = S_i + \sum_{k=1}^{i-1} \sum_{j=1}^{J_k} d_{kj}(1 - I\{V_i > t_{kj}\}) + \sum_{k=i}^{n} \sum_{j=1}^{J_k} d_{kj} I\{V_i \leq t_{kj}\}. \tag{3.75}$$

From (1.27) and (1.29), we know $0 = t_{i0} < t_{i1} < t_{i2} < \ldots < t_{iJ_i} \leq V_i$. Thus, for case of $i \geq 2$ and from (3.18), we have for $k = 1, \ldots, i-1$ and $j = 1, \ldots, J_k$:

$$t_{k1} < t_{k2} < \ldots < t_{kJ_k} \leq V_k \leq V_{i-1} < V_i \quad \Rightarrow \quad I\{V_i > t_{kj}\} = 1. \tag{3.76}$$

In turn, we have in (3.75):

$$\sum_{k=1}^{i-1} \sum_{j=1}^{J_k} d_{kj}(1 - I\{V_i > t_{kj}\}) = 0. \tag{3.77}$$

This implies that for $i \geq 2$, e_i in (3.75) can be expressed as:

$$\begin{aligned} e_i &= S_i + \sum_{k=i}^{n} \sum_{j=1}^{J_k} d_{kj} I\{V_i \leq t_{kj}\} \\ &\overset{(3.72)}{=\!=} \sum_{l=i}^{n} e^{\beta Z_l(t_{lJ_l})} + \sum_{k=i}^{n} \sum_{j=1}^{J_k} d_{kj} I\{V_i \leq t_{kj}\} \\ &= \sum_{k=i}^{n} e^{\beta Z_k(t_{kJ_k})} + \sum_{k=i}^{n} \sum_{j=1}^{J_k} d_{kj} I\{V_i \leq t_{kj}\} \\ &= \sum_{k=i}^{n} \left(e^{\beta Z_k(t_{kJ_k})} + \sum_{j=1}^{J_k} d_{kj} I\{V_i \leq t_{kj}\} \right). \end{aligned} \tag{3.78}$$

Hence, from (3.74) and (3.78), we obtain (3.36) for e_i, $i = 1, 2, \ldots, n$:

$$e_i = \sum_{k=i}^{n} \left(e^{\beta Z_k(t_{kJ_k})} + \sum_{j=1}^{J_k} d_{kj} I\{V_i \leq t_{kj}\} \right) = \sum_{k=i}^{n} A_{ki}, \tag{3.79}$$

where

$$A_{ki} = e^{\beta Z_k(t_{kJ_k})} + \sum_{j=1}^{J_k} d_{kj} I\{V_i \le t_{kj}\}$$

$$\overset{(3.71)}{=\!=\!=} e^{\beta Z_k(t_{kJ_k})} + \sum_{j=1}^{J_k} \left(e^{\beta Z_k(t_{k,j-1})} - e^{\beta Z_k(t_{kj})}\right) I\{V_i \le t_{kj}\}. \tag{3.80}$$

Next, we simplify the expression of A_{ki} given in (3.80). If $\mathcal{E}_{ki} = \emptyset$ in (3.36), then $t_{kj} < V_i$ for $\forall\, j \in \{1, \ldots, J_k\}$, in turn, we have $I\{V_i \le t_{kj}\} = 0$ for all $j \in \{1, \ldots, J_k\}$, which implies (3.80) becomes

$$A_{ki} = e^{\beta Z_k(t_{kJ_k})}. \tag{3.81}$$

If $\mathcal{E}_{ki} \ne \emptyset$ in (3.36), then $\exists\, j \in \{1, \ldots, J_k\}$ such that $V_i \le t_{kj}$, in turn, for n_{ki} given by (3.39), we have $V_i \le t_{kn_{ki}} < t_{k,n_{ki}+1} < \ldots < t_{kJ_k}$, in turn, (3.80) becomes

$$\begin{aligned}
A_{ki} &= e^{\beta Z_k(t_{kJ_k})} + \sum_{j=n_{ki}}^{J_k} \left(e^{\beta Z_k(t_{k,j-1})} - e^{\beta Z_k(t_{kj})}\right) \\
&= e^{\beta Z_k(t_{kJ_k})} + e^{\beta Z_k(t_{k,n_{ki}-1})} - e^{\beta Z_k(t_{kn_{ki}})} + e^{\beta Z_k(t_{kn_{ki}})} - e^{\beta Z_k(t_{k,n_{ki}+1})} \\
&\quad + \ldots + e^{\beta Z_k(t_{k,J_k-2})} - e^{\beta Z_k(t_{k,J_k-1})} + e^{\beta Z_k(t_{k,J_k-1})} - e^{\beta Z_k(t_{k,J_k})} \\
&= e^{\beta Z_k(t_{k,n_{ki}-1})} \overset{(3.38)}{=\!=\!=} e^{\beta Z_k(t_{km_{ki}})}.
\end{aligned} \tag{3.82}$$

Thus, (3.36) for e_i follows from above (3.81)-(3.82), and the part of (3.36) for $e_i^{'}$ is also obvious. $\square$

Proof of Lemma 3.5 Denoting $e_i^{''} = \frac{d}{d\beta}(e_i^{'})$, we compute the derivative of $\varphi_n(\beta)$ in (3.35) as follows:

$$\varphi_n^{'}(\beta) = n^{-1}\sum_{i=1}^{n}\delta_i\left(\frac{-e_i^{''}e_i + (e_i^{'})^2}{(e_i)^2}\right), \tag{3.83}$$

where e_i's and $e_i^{'}$'s are given by (3.36)-(3.37), and we have

$$e_i{''} = \frac{d}{d\beta}(e_i^{'}) = \frac{d}{d\beta}\left(\sum_{k=i}^{n} B_{ki}A_{ki}\right) = \sum_{k=i}^{n}(B_{ki})^2 A_{ki}, \tag{3.84}$$

where

$$B_{ki} = Z_k(t_{kJ_k})I\{\mathcal{E}_{ki} = \emptyset\} + Z_k(t_{km_{ki}})I\{\mathcal{E}_{ki} \neq \emptyset\}. \tag{3.85}$$

From (3.36) and (3.84), we know that Cauchy-Schwarz inequality implies:

$$\begin{aligned}(e_i{'})^2 &= \left(\sum_{k=i}^{n} B_{ki}A_{ki}\right)^2 = \left[\sum_{k=i}^{n}\sqrt{A_{ki}}\left(B_{ki}\sqrt{A_{ki}}\right)\right]^2 \\ &\leq \left(\sum_{k=i}^{n} A_{ki}\right)\left(\sum_{k=i}^{n}(B_{ki})^2 A_{ki}\right) = e_i e_i^{''}.\end{aligned} \tag{3.86}$$

This means that in (3.83) we have $\varphi_n^{'}(\beta) \leq 0$, in turn, $\varphi_n(\beta)$ is non-increasing in β. □

Remark 4. Since we have strict inequality in (3.86) unless for all k and i: $\sqrt{A_{ki}} = B_{ki}\sqrt{A_{ki}} \Leftrightarrow B_{ki} = 1$, we know that except very unusual cases with the data given by (1.29), we usually have $\varphi_n^{'}(\beta) < 0$, thus $\varphi_n(\beta)$ given by (3.35) is decreasing in β.

Chapter 4: Computing Algorithm and Simulation Studies

This chapter considers the MLE $\hat{\beta}_n$ based on estimating function $\psi_n(\beta)$ in (3.32) under assumption (AS2), then the MLE $\hat{\beta}_n$ based on $\psi_M(\beta)$ in (3.60) with proper choice of M. The approximated MLE $\tilde{\beta}_n$ based on the approximated estimating function $\varphi_n(\beta)$ in (3.35) is also considered.

We conduct simulation studies to evaluate the performances of $\hat{\beta}_n$ and $\tilde{\beta}_n$. Section 4.1 outlines our simulation algorithm and the choice of M; Section 4.2.1 describes the simulation setting for Case 1 and Section 4.2.2 presents the simulation results; Section 4.3.1 describes the simulation setting for Case 2 and Section 4.3.2 presents the simulation results. Section 4.4 provides concluding remarks and some discussions.

4.1 Algorithm and Choice of M

Let $\hat{\beta}_n$ be the solution of equation $\psi_n(\beta) = 0$ or $\psi_M(\beta) = 0$ and let $\tilde{\beta}_n$ be the solution of equation $\varphi_n(\beta) = 0$, where $\psi_n(\beta)$, $\psi_M(\beta)$ and $\varphi_n(\beta)$ are given by (3.32), (3.60) and (3.35), respectively. Since $\tilde{\beta}_n$ always exists and since preliminary simulation results show that $\hat{\beta}_n$ and $\tilde{\beta}_n$ are usually very close, in our computing algorithm we compute $\tilde{\beta}_n$ first, then when needed, we choose data-based M to satisfy (AS2) or (AS3) based on the following mathematical idea.

Note that from (3.34) we have

$$\psi_n(\beta) = \varphi_n(\beta) - n^{-1}\sum_{i=1}^{n}\left(\frac{\delta_i e_i^{'}}{2(1+\xi_i)^2 e_i^2}\right) \tag{4.1}$$

and from (3.60) we also have

$$\begin{aligned}
\psi_M(\beta) &= n^{-1}\sum_{i=1}^{n}\delta_i\left(Z_i(t_{iJ_i}) + E_i^{'}\left(-\frac{1}{E_i} - \frac{1}{2(1+\xi_i)^2 E_i^2}\right)\right)\\
&= n^{-1}\sum_{i=1}^{n}\delta_i\left(Z_i(t_{iJ_i}) - \frac{E_i^{'}}{E_i} - \frac{E_i^{'}}{2(1+\xi_i)^2 E_i^2}\right)\\
&\overset{(3.62)}{=\!=} n^{-1}\sum_{i=1}^{n}\delta_i\left(Z_i(t_{iJ_i}) - \frac{e^M e_i^{'}}{e^M e_i} - \frac{E_i^{'}}{2(1+\xi_i)^2 E_i^2}\right)\\
&= \varphi_n(\beta) - n^{-1}\sum_{i=1}^{n}\left(\frac{\delta_i E_i^{'}}{2(1+\xi_i)^2 E_i^2}\right).
\end{aligned} \tag{4.2}$$

From, (3.36) and (3.62), we know e_i's and E_i's are nonnegative if $Z_i(t_{ij})$'s are nonnegative. Thus, we have

$$\psi_n(\beta) < \varphi_n(\beta) \quad \text{and} \quad \psi_M(\beta) < \varphi_n(\beta) \quad \text{for } \forall\beta. \tag{4.3}$$

Steps to Choose M:

Step 1. Compute $e_1, \ldots, e_n$ at $\beta = \tilde{\beta}_n - \delta$ and order them from smallest to largest as $e_{(1)} \le e_{(2)} \ldots \le e_{(n)}$.

Step 2. Choose $e^* = \min_{1\le i\le n}\{e_{(i)} \mid e_{(i)} \ge 10^{-7}\}$ and compute $M_0 = -\log e^*$. If $e_{(1)} > 1$ (i.e., (AS2) holds), then no need to choose M_0.

(i) For $\beta_0 > 0$, we search for $\hat{\beta}_n$ in interval $(\tilde{\beta}_n - \delta, \tilde{\beta}_n)$; when no M_0 is used, we use $\psi_n(\beta)$ given by (3.32) to find $\hat{\beta}_n$; for $\psi_n(\beta)$ or $\psi_{M_0}(\beta)$, increase δ until $\psi_n(\tilde{\beta}_n - \delta) > 0$ or $\psi_{M_0}(\tilde{\beta}_n - \delta) > 0$. Try $\delta = \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1$, etc., and start with $\delta = \frac{1}{4}$. We use bisection methods with stopping precision 0.001.

(ii) For $\boldsymbol{\beta_0} = \mathbf{0}$, use above choice of M_0 and $\psi_{M_0}(\beta)$ given by (3.60) to search $\hat{\beta}_n^0$ in $(\tilde{\beta}_n - \delta, \tilde{\beta}_n)$. If $\psi_{M_0}(\tilde{\beta}_n - \delta) < 0$, then increase δ as above until $\psi_{M_0}(\tilde{\beta}_n - \delta) > 0$.

(iii) For $\boldsymbol{\beta_0} < \mathbf{0}$, use above choice of M_0 and $\psi_{M_0}(\beta)$ given by (3.60) to search $\hat{\beta}_n^0$ in $(\tilde{\beta}_n - \delta, \tilde{\beta}_n)$. If $\psi_{M_0}(\tilde{\beta}_n - \delta) < 0$, then increase as above δ until $\psi_{M_0}(\tilde{\beta}_n - \delta) > 0$.

Step 3. If Step 2 uses M_0 to find $\hat{\beta}_n^0$ such that $\psi_{M_0}(\hat{\beta}_n^0) = 0$, adjust to $M = M_0 + 1$, then use $\psi_M(\beta)$ to search solution $\hat{\beta}_n$ of equation $\psi_M(\beta) = 0$ in interval $\beta \in (\hat{\beta}_n^0, \tilde{\beta}_n)$.

Step 4. Check if there are more than two $E_i \leq 1$ when computing $\psi_M(\hat{\beta}_n)$.

$$\begin{cases} \text{If no,} & \text{then keep } \hat{\beta}_n. \\ \text{If yes,} & \text{adjust } M = M_0 + 1 + a \text{ and use } \psi_M(\beta) \text{ to search solution } \hat{\beta}_n \text{ of} \\ & \text{equation } \psi_M(\beta) = 0 \text{ in interval } \beta \in (\hat{\beta}_n^0, \tilde{\beta}_n). \text{ Try } a = 1, 2, 3, \ldots, \text{ etc.} \\ & \text{until there are no more than two } E_i \leq 1 \text{ when computing } \psi_M(\hat{\beta}_n) \end{cases}$$

NOTE:

(A) Above choice of M to avoid more than two $e_i \leq 1$ or $E_i \leq 1$, it is to avoid computing floating error, which is difficult to control in intensive Monte Carlo simulation studies. In real data analysis we conduct in Chapter 5, there are no such issues.

(B) The rational of our method for choosing M_0 in Step 2 is as follows. From Lemma 3.4 and (3.62), we have $E_i = e^M e_i$ and E_i is nondecreasing in β with nonnegative $Z_i(t_{ij})$'s. For above choice of M_0 about E_i in Lemma 3.4, for any $\beta \geq \tilde{\beta}_n - \delta$:

$$
\begin{aligned}
E_i &= e^{M_0} \sum_{k=i}^{n} \Big(e^{\beta Z_k(t_{kJ_k})} I\{\mathcal{E}_{ki} = \emptyset\} + e^{\beta Z_k(t_{km_{ki}})} I\{\mathcal{E}_{ki} \neq \emptyset\} \Big) \\
&\geq e^{M_0} \sum_{k=i}^{n} \Big(e^{(\tilde{\beta}_n - \delta) Z_k(t_{kJ_k})} I\{\mathcal{E}_{ki} = \emptyset\} + e^{(\tilde{\beta}_n - \delta) Z_k(t_{km_{ki}})} I\{\mathcal{E}_{ki} \neq \emptyset\} \Big) \qquad (4.4) \\
&= e^{M_0} e_i^{**}
\end{aligned}
$$

where e_i^{**} is the e_i with $\beta = \tilde{\beta}_n - \delta$. If $\exists\ e_i^{**} \leq 10^{-7}$, our simulation results show that there are only very few such e_i^{**}'s that are less than $e^* = \min_{1 \leq i \leq n}\{e_{(i)}^{**} \mid e_{(i)}^{**} \geq 10^{-7}\}$, thus by using $M_0 = -\log e^*$, almost all of the E_i's satisfy:

$$E_i \geq e^{M_0} e_i^{**} \geq e^{M_0} \cdot e^* = e^{-\log e^*} \cdot e^* = \tfrac{1}{e^*} \cdot e^* = 1.$$

(C) The reason that we search $\hat{\beta}_n^0$ in interval $(\tilde{\beta}_n - \delta, \tilde{\beta}_n)$ when $\psi_n(\tilde{\beta}_n - \delta) > 0$ or $\psi_{M_0}(\tilde{\beta}_n - \delta) > 0$ is because (4.3) implies $\psi_n(\tilde{\beta}_n) < 0$ and $\psi_{M_0}(\tilde{\beta}_n) < 0$ for $\forall\ M_0 > 0$.

(D) The reason for using $M = M_0 + 1$ and searching $\hat{\beta}_n$ in interval $(\hat{\beta}_n^0, \tilde{\beta}_n)$ is the following. From (3.60), (4.4) and $E_i^{'} = e^M e_i^{'}$, we have for $E_i > 1$:

$$
\begin{aligned}
\frac{\partial \psi_M(\beta)}{\partial M} &= \frac{1}{n} \sum_{i=1}^{n} \delta_i \Big(E_i^{'} \log\Big(\frac{E_i - 1}{E_i}\Big) + E_i^{'} \cdot \frac{E_i}{E_i - 1} \cdot \frac{E_i^2 - (E_i - 1)E_i}{E_i^2} \Big) \\
&= \frac{1}{n} \sum_{i=1}^{n} \delta_i E_i^{'} \Big(\log\Big(1 - \frac{1}{E_i}\Big) + \frac{1}{E_i - 1} \Big) \\
&= \frac{1}{n} \sum_{i=1}^{n} \delta_i E_i^{'} \Big(\log\Big(1 - \frac{1}{E_i}\Big) + \frac{1/E_i}{1 - 1/E_i} \Big) \\
&= \frac{1}{n} \sum_{i=1}^{n} \delta_i E_i^{'} g\Big(\frac{1}{E_i}\Big), \qquad (4.5)
\end{aligned}
$$

where $g(x) = \log(1-x) + \frac{x}{1-x}$, $0 < x < 1$. Since

$$
\begin{aligned}
g^{'}(x) &= -\frac{1}{1-x} + \frac{1-x-x\cdot(-1)}{(1-x)^2} = \frac{-1}{1-x} + \frac{1}{(1-x)^2} \\
&= \frac{-(1-x)+1}{(1-x)^2} = \frac{x}{(1-x)^2} > 0,
\end{aligned}
\tag{4.6}
$$

we know ψ_M is increasing in M for any fixed β. Thus, from $M > M_0$ and (4.3), we have

$$
\psi_M(\hat{\beta}_n^0) > \psi_{M_0}(\hat{\beta}_n^0) = 0 \quad \text{and} \quad \psi_M(\tilde{\beta}_n) < \varphi_n(\tilde{\beta}_n) = 0, \tag{4.7}
$$

which means $\psi_M(\beta) = 0$ has a solution in interval $(\hat{\beta}_n^0, \tilde{\beta}_n)$. The reason for using $M = M_0 + 1$ is due to (4.5)-(4.7) and that the choice of M_0 is conservative which has cases $E_i \leq 1$ occasionally, so M is not too big and can eliminative cases $E_i \leq 1$.

(E) From (4.2), we know that for $\forall\beta$,

$$
\psi_M(\beta) \to \varphi_n(\beta), \quad \text{as } M \to \infty. \tag{4.8}
$$

In fact, under some conditions we can show above convergence in (4.8) is uniformly for $|\beta| \leq M_\beta$. Thus, the use of $M = M_0 + 1$ moves away from M_0 enough to avoid the occurrence of $E_i \leq 1$, but not too far from M_0 and not too big which makes $\psi_M(\beta)$ not too close to $\varphi_n(\beta)$.

Computing $\psi_n(\beta)$ or $\psi_M(\beta)$ and $\varphi_n(\beta)$ based on Data (1.29):

Step 1. Compute $\boldsymbol{Z}_i = (Z(t_{i1}), \ldots, Z(t_{iJ_i}))$, $i = 1, \ldots, n$, with

$$Z(t_{ij}) = \varphi(\mathcal{X}_i^{t_{ij}}),\ j = 1, \ldots, J_i. \tag{4.9}$$

Step 2. Order V_i's into $V_1 < V_2 < \ldots < V_n$, and $(\delta_i, J_i, \boldsymbol{t}_i, \boldsymbol{Z}_i)$ follow V_i's position, where $\boldsymbol{t}_i = (t_{i1}, \ldots, t_{iJ_i})$, $i = 1, \ldots, n$.

Step 3. Obtain $J_{\max} = \max\{J_1, \ldots, J_n\}$ and $N = J_{\max} \times n$, then create vectors $\boldsymbol{Z}_{\text{full}}$ and $\boldsymbol{t}_{\text{full}}$, which both have dimension N. Initialize the value of all elements in $\boldsymbol{Z}_{\text{full}}$ and $\boldsymbol{t}_{\text{full}}$ to be 0, then for each $i = 1, \ldots, n$ and for each $j = 1, \ldots, J_i$, let

$$q = (i-1)J_{\max} + j, \quad \text{then } \boldsymbol{Z}_{\text{full}}[q] = Z(t_{ij}), \quad \boldsymbol{t}_{\text{full}}[q] = t_{ij}.$$

Also obtain vector $\boldsymbol{ZZ} = (Z_1, , \ldots, Z_n)$, where

$$Z_i = Z(t_{iJ_i}), \tag{4.10}$$

for each $i = 1, \ldots, n$, we have

$$\text{for } q = (i-1)J_{\max} + J_i, \quad \boldsymbol{Z}_{\text{full}}[q] = Z_i. \tag{4.11}$$

Step 4. Compute $\psi_n(\beta)$ in (3.32) or $\psi_M(\beta)$ in (3.60) and $\varphi_n(\beta)$ in (3.35):

for any given β and M, set $\psi_n = 0$, $\psi_M = 0$, $\varphi_n = 0$, then

for each $i = 1, \dots, n$

$e_i = 0, \; e_i^{'} = 0$

for each $k = i, \dots, n$

$a = 0, \; b = 0$

(a) for each $j = 1, \dots, J_k$

$q = (k-1)J_{\max} + j$

if $t_{\text{full}}[q] \geq V_i$, then $qq = q - 1$

if $qq = (k-1)J_{\max}$, then $a = 1, \; b = 0$,

else $a = e^{\beta \boldsymbol{Z}_{\text{full}}[qq]}, \; b = \boldsymbol{Z}_{\text{full}}[qq]$, end if. Go to **(b)**

else continue **(a)**

(b) if $a = 0$, then $q = (k-1)J_{\max} + J_k$, $a = e^{\beta \boldsymbol{Z}_{\text{full}}[q]}, \; b = \boldsymbol{Z}_{\text{full}}[q] = Z_k$,

end if.

then $e_i = e_i + a$, $e_i^{'} = e_i + a * b$

then

$$\begin{cases} \psi_n = \psi_n + n^{-1}\delta_i\Big(Z_i + e_i^{'} \log \frac{e_i - 1}{e_i}\Big) \\ \psi_M = \psi_M + n^{-1}\delta_i\Big(Z_i + e^M e_i^{'} \log \frac{e^M e_i - 1}{e^M e_i}\Big) \\ \varphi_n = \varphi_n + \delta_i\Big(Z_i - \frac{e_i^{'}}{e_i}\Big) \end{cases}.$$

Thus, the final sums are $\psi_n(\beta)$, $\psi_M(\beta)$ and $\varphi_n(\beta)$ as given by (3.32), (3.60) and (3.35), respectively.

4.2 Case 1: Use Average Summary Measurement

Here, we consider using $Z(t) = \varphi(X^t)$ in our proposed PH model (1.38) as average summary measurement, which is motivated by the Smoking Cessation Data described in Section 1.5. We call this Case 1 in our simulation studies.

Below, Section 4.2.1 describes the specific simulation setting for our Case 1 studies, and outlines the steps to generate data (1.29) for such case, while Section 4.2.2 presents the simulation results under our Case 1 setting.

4.2.1 Case 1 Simulation Setting

Here, we consider the summary measurement as the average of $X(t;\omega)$ in time interval $[0, t]$. Recall the average value of function f over interval $[a, b]$ is given by

$$\Big\{ \text{average value of } f \text{ on } [a, b] \Big\} = \frac{1}{b-a} \int_a^b f(x)dx. \tag{4.12}$$

Thus, in our proposed PH model (1.38), we consider the following Case 1 data setting for our simulation studies:

- $F_0 = \text{Exp}(1)$, which is the baseline d.f., thus $h_0(t) = 1$ in equation (3.7)
- $F_C = \text{Exp}(3)$, which is the d.f. of right censoring variable in (1.2)
- $X(t) = X(t;\omega) = \omega t$, which is the longitudinal covariate process with $\omega \sim U(0, 1)$
- $Z(t) = \varphi(X^t) = t^{-1} \int_0^t X(s)ds = t^{-1} \int_0^t \omega s ds = t^{-1} \Big(\frac{1}{2}\omega s^2 \Big|_0^t \Big) = \frac{\omega t}{2}$, which is the average of $X(t;\omega)$ in time interval $[0, t]$

- $\bar{F}(t \mid X^t)$ given by (3.7) is derived based on the PH model assumption (1.38), where $h(t \mid X^t)$ is the conditional hazard function of T given X^t, and for this particular case considered here we have $X^t = \{X(s) \mid s \leq t\} = \{\omega s \mid s \leq t\}$. Thus, $F(t \mid X^t)$ given by (3.7) is something we can use to generate a conditional r.v. T given X^t, denoted by $[\, T \mid X^t \,]$.

Based on above choices, we have the following in (3.7): for $\beta_0 \neq 0$,

$$\begin{aligned} \bar{F}(t \mid X^t) &= \exp\left\{ -\int_0^t h_0(s) e^{Z(s)\beta_0} ds \right\} = \exp\left\{ -\int_0^t e^{\beta_0 \omega s/2} ds \right\} \\ &= \exp\left\{ -\frac{2e^{\beta_0 \omega s/2}}{\beta_0 \omega}\Big|_0^t \right\} = \exp\left\{ \frac{2(1 - e^{\beta_0 \omega t/2})}{\beta_0 \omega} \right\}, \end{aligned} \tag{4.13}$$

which implies

$$F(t \mid X^t) = 1 - \exp\left\{ \frac{2(1 - e^{\beta_0 \omega t/2})}{\beta_0 \omega} \right\} \equiv G(t). \tag{4.14}$$

We have the following in above $G(t)$:

$$\begin{aligned} &G(0) = 1 - e^0 = 0 \\ &G(\infty) = 1 - \lim_{t \to \infty} \exp\left\{ \frac{2(1 - e^{\beta_0 \omega t/2})}{\beta_0 \omega} \right\} = 1 - e^{\frac{2}{\beta_0 \omega}} \lim_{t \to \infty} \exp\left\{ \frac{-2e^{\beta_0 \omega t/2}}{\beta_0 \omega} \right\}. \end{aligned} \tag{4.15}$$

Notice that for any given $\omega > 0$, above $G(t)$ is an increasing function because

$$\begin{aligned} G^{'}(t) &= -\left(\frac{-2e^{\beta_0 \omega t/2}}{\beta_0 \omega} \right) \left(\frac{\beta_0 \omega}{2} \right) \exp\left\{ \frac{2(1 - e^{\beta_0 \omega t/2})}{\beta_0 \omega} \right\} \\ &= e^{\beta_0 \omega t/2} \exp\left\{ \frac{2(1 - e^{\beta_0 \omega t/2})}{\beta_0 \omega} \right\} > 0. \end{aligned}$$

For $\beta_0 > 0$, from (4.15) we have $G(t)$ in (4.14) satisfies

$$\begin{aligned}
&G(0) = 1 - e^0 = 0 \\
&G(\infty) = 1 - e^{\frac{2}{\beta_0\omega}} \lim_{t\to\infty} \exp\left\{\frac{-2e^{\beta_0\omega t/2}}{\beta_0\omega}\right\} = 1 - e^{\frac{2}{\beta_0\omega}} \exp\left\{\frac{-2e^{+\infty}}{\beta_0\omega}\right\} \\
&\qquad = 1 - e^{\frac{2}{\beta_0\omega}} e^{-\infty} = 1 - e^{\frac{2}{\beta_0\omega}} \cdot 0 = 1.
\end{aligned}$$

Thus, $G(t)$ is a proper d.f. when $\beta_0 > 0$.

For $\beta_0 = 0$, we have the following in (3.7)

$$\begin{aligned}
\bar{F}(t \,|\, X^t) &= \exp\left\{ - \int_0^t h_0(s) e^{Z(s)\beta_0} ds \right\} = \exp\left\{ - \int_0^t 1 ds \right\} \\
&= \exp\left\{ - s\Big|_0^t \right\} = e^{-t},
\end{aligned} \tag{4.16}$$

which implies

$$F(t \,|\, X^t) = 1 - e^{-t} \equiv G(t). \tag{4.17}$$

Notice that above $G(t)$ is the d.f. of Exp(1), thus $G(t)$ is a proper d.f. when $\beta_0 = 0$.

For $\beta_0 < 0$, from (4.15) we have $G(t)$ in (4.14) satisfies

$$\begin{aligned}
&G(0) = 1 - e^0 = 0 \\
&G(\infty) = 1 - e^{\frac{2}{\beta_0\omega}} \lim_{t\to\infty} \exp\left\{\frac{-2e^{\beta_0\omega t/2}}{\beta_0\omega}\right\} = 1 - e^{\frac{2}{\beta_0\omega}} \exp\left\{\frac{-2e^{-\infty}}{\beta_0\omega}\right\} \\
&\qquad = 1 - e^{\frac{2}{\beta_0\omega}} e^0 = 1 - e^{\frac{2}{\beta_0\omega}} \equiv g_1(\omega, \beta_0) < 1,
\end{aligned} \tag{4.18}$$

which implies $G(t)$ is not a proper d.f. when $\beta_0 < 0$. From (4.18), we know $g_1(\omega, \beta_0)$ depends on the values of β_0 and ω, where $\omega \sim U(0, 1)$.

From (3.7) and (1.38), we obtain that the p.d.f. $g(t)$ of $G(t)$, which is given by: $g(t) = h(t \,|\, X^t)\bar{G}(t)$. This means that if Y is a r.v. with d.f. $G(t)$, the hazard function of Y is equal to $h(t \,|\, X^t)$, because $h_Y(t) = \frac{g(t)}{\bar{G}(t)} = h(t \,|\, X^t)$. Since $h(t \,|\, X^t)$ is the conditional hazard function of T given X^t, we know $Y \stackrel{D}{=} [\, T \,|\, X^t \,]$.

Recall: If $X \sim F_X$, then let $Y = F_X^{-1}(U)$, where $U \sim U(0,1)$, we have

$$P(Y \leq y) = P(F_X^{-1}(U) \leq y) = P(U \leq F_X(y)) = F_X(y),$$

which means that r.v. Y has d.f. F_X, thus Y is a r.v. with d.f. F_X. From $G(t) = F(t \,|\, X^t)$ given by (4.14), we know that $Y = G^{-1}(U)$ is a r.v. generated from d.f. $G(t) = F(t \,|\, X^t)$.

Derivation for G^{-1} in Case 1:

To find G^{-1}, we compute the following:

$$\begin{aligned} G(t) = u \;\Rightarrow\; u = 1 - \exp\left\{ \frac{2(1 - e^{\beta_0 \omega t/2})}{\beta_0 \omega} \right\} \;\Rightarrow\; & \frac{2(1 - e^{\beta_0 \omega t/2})}{\beta_0 \omega} = \log(1-u) \\ \Rightarrow\; & e^{\beta_0 \omega t/2} = 1 - \frac{\beta_0 \omega}{2} \log(1-u) \\ \Rightarrow\; & t = \frac{2}{\beta_0 \omega} \log\left(1 - \frac{\beta_0 \omega}{2} \log(1-u)\right) = G^{-1}(u). \end{aligned} \tag{4.19}$$

Hence, for $U \sim U(0,1)$, random variable $Y = \frac{2}{\beta_0 \omega} \log\left(1 - \frac{\beta_0 \omega}{2} \log(1-U)\right)$ is generated from above d.f. $G(t) = F(t \,|\, X^t)$ in (4.14), and Y is a conditional r.v. T given X^t.

For $\beta_0 > 0$, $G^{-1}(U)$ in (4.19) is given by

$$G^{-1}(U) = \frac{2}{\beta_0 \omega} \log\left(1 - \frac{\beta_0 \omega}{2} \log(1-U)\right). \tag{4.20}$$

For $\beta_0 = 0$, from (4.17) we compute the following:

$$G(t) = u \;\Rightarrow\; u = 1 - e^{-t} \Rightarrow\; t = -\log(1-u) = G^{-1}(u).$$

Thus, when $\beta_0 = 0$, $G^{-1}(U)$ is given by

$$G^{-1}(U) = -\log(1-U). \tag{4.21}$$

From (4.18), for $\beta_0 < 0$ and $U \leq 1 - e^{\frac{2}{\beta_0 \omega}}$, $G^{-1}(U)$ in (4.19) is given by

$$G^{-1}(U) = \frac{2}{\beta_0 \omega} \log\Big(1 - \frac{\beta_0 \omega}{2} \log(1-U)\Big). \tag{4.22}$$

For $\beta_0 < 0$ and $U > 1 - e^{\frac{2}{\beta_0 \omega}}$, $G^{-1}(U) = \infty$ in (4.19).

Steps to Generate Data (1.29) for Case 1:

Step 1. Generate $\omega_i \sim U(0,1)$

Step 2. Generate $U \sim U(0,1)$ and compute

$$T_i = \begin{cases} \frac{2}{\beta_0 \omega_i} \log\big(1 - \frac{\beta_0 \omega_i}{2} \log(1-U)\big) & \text{if } \beta_0 > 0 \\ \\ -\log(1-U) & \text{if } \beta_0 = 0 \end{cases}$$

$$T_i \overset{(4.18)}{=\!=} \begin{cases} \frac{2}{\beta_0 \omega_i} \log\big(1 - \frac{\beta_0 \omega_i}{2} \log(1-U)\big) & \text{if } \beta_0 < 0,\; U \leq 1 - e^{\frac{2}{\beta_0 \omega_i}} \\ \\ \infty & \text{if } \beta_0 < 0,\; U > 1 - e^{\frac{2}{\beta_0 \omega_i}} \end{cases}$$

Step 3. Generate $U \sim U(0,1)$ and compute $C_i = -3\log(1-U)$

Step 4. Obtain (V_i, δ_i) based on Step 2-3, where $V_i = \min\{T_i, C_i\}$ and $\delta_i = I\{T_i \le C_i\}$, which is the first part of data (1.29).

Step 5. Let $N_i = [V_i] + 1$, where $[V_i]$ is the integer part of V_i, then for a chosen integer m, compute $J_i = mN_i$. Generate t_{ij}'s as follows.

$$\left.\begin{array}{ll} \text{Method } 1: & \text{Generate } J_i \text{ r.v.'s from } U(0,V_i); \text{ i.e., for } j=1,\ldots,J_i, \\ & \text{generate } U_j \sim U(0,1) \text{ and let } t_{ij} = U_j V_i. \\ \text{Method } 2: & \text{Generate } J_i \text{ time points } t_{ij} = \frac{j}{J_i} V_i,\ j=1,\ldots,J_i;\ \text{i.e.,} \\ & t_{ij} \text{ are evenly distributed time points in interval } (0,V_i). \end{array}\right\} \tag{4.23}$$

Then order them into $t_{i1} < \ldots < t_{iJ_i}$, which gives

$$\mathcal{X}_i^{V_i} = \{X_i(t_{ij}) \,|\, j=1,\ldots,J_i\} = \{\omega_i t_{ij} \,|\, j=1,\ldots,J_i\}, \tag{4.24}$$

which is the last part of data (1.29).

Step 6. Compute $Z(t_{ij})$ in equation (4.9):

$$Z(t_{ij}) = \varphi(\mathcal{X}_i^{t_{ij}}) = \frac{\sum_{k=1}^{j}(t_{ik} - t_{i,k-1})X_i(t_{i,k-1})}{t_{ij}}, \tag{4.25}$$

where $j = 1,2,\ldots,J_i+1$ with $t_{i,J_i+1} = V_i$ for Method 1 in (4.23) and $j = 1,2,\ldots,J_i$ for Method 2 in (4.23).

NOTE:

(A) Above T_i has hazard function $h(t \mid X^t)$, which satisfies PH model assumption (1.38), thus T_i is a conditional r.v. T given X^t. Generating T_i this way ensures that PH model assumption (1.38) is satisfied for data $(T_i, X_i(t;\omega))$, which is mentioned in (1.29).

(B) Above integer N_i is the number of days before V_i for the ith individual in Smoking Cessation Data example, and integer m is the total number of time points at which the longitudinal covariate process is observed on a given day. Thus, the larger m is, the closer $\mathcal{X}_i^{V_i}$ in (1.29) is to $X_i^{V_i}$.

(C) Above $G(t)$ is not always a proper d.f.

(D) In Step 5, Method 1 generate t_{ij}'s randomly and Method 2 generate fixed t_{ij}'s for any given V_i. Thus, we can control the time interval length for Method 2.

4.2.2 Simulation Results

For Case 1 simulation studies, we consider $\beta_0 = 1, 0, -1$, and we follow Steps 1-6 described in Section 4.2.1 to generate 1000 samples with sample size $n = 100, 200, 500,$ 1000 and $m = 5, 10, 20, 30, 40, 50$, respectively. Note that in Step 5, we generate t_{ij}'s by using two different methods given in (4.23). For each case, we compute the simulation mean and s.d. for $\hat{\beta}_n$ and $\tilde{\beta}_n$, and we also compute (Ave. $\hat{\beta}_n$) $\pm$ s.d. and (Ave. $\tilde{\beta}_n$) $\pm$ s.d. with Δ for interval (a, b), where

$$\Delta = \max\{|a - \beta_0|, |b - \beta_0|\}, \tag{4.26}$$

represents the largest distance between end points and β_0. When the simulation s.d. varies a lot, relying solely on the simulation mean to evaluate the estimator's performance is not sufficient, thus Δ provides better assessment on how close the interval is to β_0.

Table 4.2.1 presents the simulation mean and s.d. of $\hat{\beta}_n$ and $\tilde{\beta}_n$ for Case 1 data setting with $\beta_0 = 1$ and $t_{ij} \sim U(0, V_i)$, while Table 4.2.2 gives (Ave. $\hat{\beta}_n$) $\pm$ s.d. and (Ave. $\tilde{\beta}_n$) $\pm$ s.d. with Δ for simulation results in Table 4.2.1. Similarly, Tables 4.2.3-4.2.4 display results for Case 1 data setting with $\beta_0 = 1$ and t_{ij}'s evenly chosen; Tables 4.2.5-4.2.12 display results accordingly for $\beta_0 = 0, -1$, respectively.

Table 4.2.1 Case 1: with $\beta_0 = 1$ and Random Observation Time Points

Ave. of $\hat{\beta}_n$ & $\tilde{\beta}_n$ (s.d.) for Case 1: $\beta_0 = 1, t_{ij} \sim U(0, V_i)$

Number of Simulation Loop: 1000

	$n = 100$		$n = 200$	
	Ave. $\hat{\beta}_n$ (s.d.)	Ave. $\tilde{\beta}_n$ (s.d.)	Ave. $\hat{\beta}_n$ (s.d.)	Ave. $\tilde{\beta}_n$ (s.d.)
$m = 5$	2.223 (1.18)	2.484 (1.13)	2.113 (0.76)	2.263 (0.74)
$m = 10$	1.525 (1.13)	1.819 (1.14)	1.496 (0.76)	1.683 (0.74)
$m = 20$	1.188 (1.03)	1.475 (1.08)	1.059 (0.73)	1.266 (0.73)
$m = 30$	0.961 (1.05)	1.246 (1.11)	1.039 (0.75)	1.254 (0.75)
$m = 40$	0.979 (1.07)	1.260 (1.13)	0.950 (0.75)	1.160 (0.75)
$m = 50$	0.927 (1.09)	1.209 (1.16)	0.959 (0.71)	1.171 (0.71)
Censoring %	22.4%		22.3%	

	$n = 500$		$n = 1000$	
	Ave. $\hat{\beta}_n$ (s.d.)	Ave. $\tilde{\beta}_n$ (s.d.)	Ave. $\hat{\beta}_n$ (s.d.)	Ave. $\tilde{\beta}_n$ (s.d.)
$m = 5$	2.023 (0.46)	2.123 (0.46)	2.019 (0.30)	2.084 (0.29)
$m = 10$	1.466 (0.42)	1.582 (0.41)	1.472 (0.31)	1.557 (0.30)
$m = 20$	1.134 (0.43)	1.260 (0.42)	1.169 (0.30)	1.266 (0.29)
$m = 30$	1.048 (0.43)	1.176 (0.42)	1.072 (0.29)	1.172 (0.29)
$m = 40$	1.003 (0.44)	1.132 (0.43)	1.030 (0.30)	1.133 (0.30)
$m = 50$	0.974 (0.44)	1.103 (0.44)	0.991 (0.30)	1.096 (0.30)
Censoring %	22.3%		22.3%	

Table 4.2.2 Case 1: Measurement of Coverage for Table 4.2.1

(Ave. $\hat{\beta}_n$) $\pm$ s.d. and (Ave. $\tilde{\beta}_n$) $\pm$ s.d. with Δ

	$m = 5$		$m = 10$	
	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ
$n = 100$	(1.043, 3.403) 2.403	(1.354, 3.614) 2.614	(0.395, 2.655) 1.655	(0.679, 2.959) 1.959
$n = 200$	(1.353, 2.873) 1.873	(1.523, 3.003) 2.003	(0.736, 2.256) 1.256	(0.943, 2.423) 1.423
$n = 500$	(1.563, 2.483) 1.483	(1.663, 2.583) 1.583	(1.046, 1.886) 0.886	(1.172, 1.992) 0.992
$n = 1000$	(1.719, 2.319) 1.319	(1.794, 2.374) 1.374	(1.162, 1.782) 0.782	(1.257, 1.857) 0.857

	$m = 20$		$m = 30$	
	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ
$n = 100$	(0.158, 2.218) 1.218	(0.395, 2.555) 1.555	(-0.089, 2.011) 1.089	(0.136, 2.356) 1.356
$n = 200$	(0.329, 1.789) 0.789	(0.536, 1.996) 0.996	(0.289, 1.789) 0.789	(0.504, 2.004) 1.004
$n = 500$	(0.704, 1.564) 0.564	(0.840, 1.680) 0.680	(0.618, 1.478) 0.478	(0.756, 1.596) 0.596
$n = 1000$	(0.869, 1.469) 0.469	(0.976, 1.556) 0.556	(0.782, 1.362) 0.362	(0.882, 1.462) 0.462

	$m = 40$		$m = 50$	
	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ
$n = 100$	(-0.091, 2.049) 1.091	(0.130, 2.390) 1.390	(-0.163, 2.017) 1.163	(0.049, 2.369) 1.369
$n = 200$	(0.200, 1.700) 0.800	(0.410, 1.910) 0.910	(0.249, 1.669) 0.751	(0.461, 1.881) 0.881
$n = 500$	(0.563, 1.443) 0.443	(0.702, 1.562) 0.562	(0.534, 1.414) 0.466	(0.663, 1.543) 0.543
$n = 1000$	(0.730, 1.330) 0.330	(0.833, 1.433) 0.433	(0.691, 1.291) 0.309	(0.796, 1.396) 0.396

Table 4.2.3 Case 1: with $\beta_0 = 1$ and Evenly Chosen Observation Time Points

Ave. of $\hat{\beta}_n$ & $\tilde{\beta}_n$ (s.d.) for Case 1: $\beta_0 = 1$, t_{ij} evenly chosen

Number of Simulation Loop: 1000

	$n = 100$		$n = 200$	
	Ave. $\hat{\beta}_n$ (s.d.)	Ave. $\tilde{\beta}_n$ (s.d.)	Ave. $\hat{\beta}_n$ (s.d.)	Ave. $\tilde{\beta}_n$ (s.d.)
$m = 5$	3.264 (1.74)	3.533 (1.69)	2.941 (1.08)	3.102 (1.06)
$m = 10$	1.792 (1.29)	2.098 (1.30)	1.712 (0.86)	1.911 (0.84)
$m = 20$	1.244 (1.16)	1.553 (1.20)	1.233 (0.78)	1.455 (0.77)
$m = 30$	1.091 (1.12)	1.396 (1.17)	1.102 (0.76)	1.313 (0.76)
$m = 40$	1.017 (1.10)	1.321 (1.16)	1.039 (0.74)	1.253 (0.75)
$m = 50$	0.975 (1.09)	1.271 (1.16)	1.000 (0.74)	1.212 (0.75)
Censoring %	22.4%		22.3%	
	$n = 500$		$n = 1000$	
	Ave. $\hat{\beta}_n$ (s.d.)	Ave. $\tilde{\beta}_n$ (s.d.)	Ave. $\hat{\beta}_n$ (s.d.)	Ave. $\tilde{\beta}_n$ (s.d.)
$m = 5$	2.672 (0.61)	2.786 (0.61)	2.586 (0.42)	2.680 (0.43)
$m = 10$	1.643 (0.51)	1.767 (0.50)	1.622 (0.35)	1.722 (0.35)
$m = 20$	1.238 (0.47)	1.365 (0.46)	1.241 (0.33)	1.343 (0.33)
$m = 30$	1.117 (0.46)	1.246 (0.45)	1.127 (0.32)	1.231 (0.32)
$m = 40$	1.059 (0.45)	1.190 (0.45)	1.070 (0.32)	1.173 (0.32)
$m = 50$	1.024 (0.45)	1.154 (0.45)	1.037 (0.32)	1.143 (0.32)
Censoring %	22.3%		22.3%	

Table 4.2.4 Case 1: Measurement of Coverage for Table 4.2.3

(Ave. $\hat{\beta}_n$) $\pm$ s.d. and (Ave. $\tilde{\beta}_n$) $\pm$ s.d. with Δ

	$m = 5$		$m = 10$	
	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ
$n = 100$	(1.524, 5.004) 4.004	(1.843, 5.223) 4.223	(0.502, 3.082) 2.082	(0.798, 3.398) 2.398
$n = 200$	(1.861, 4.021) 3.021	(2.042, 4.162) 3.162	(0.852, 2.572) 1.572	(1.071, 2.751) 1.751
$n = 500$	(2.062, 3.282) 2.282	(2.176, 3.396) 2.396	(1.133, 2.153) 1.153	(1.267, 2.267) 1.267
$n = 1000$	(2.166, 3.006) 2.006	(2.250, 3.110) 2.110	(1.272, 1.972) 0.972	(1.372, 2.072) 1.072
	$m = 20$		$m = 30$	
	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ
$n = 100$	(0.084, 2.404) 1.404	(0.353, 2.753) 1.753	(-0.029, 2.211) 1.211	(0.226, 2.566) 1.566
$n = 200$	(0.453, 2.013) 1.013	(0.685, 2.225) 1.225	(0.342, 1.862) 0.862	(0.553, 2.073) 1.073
$n = 500$	(0.768, 1.708) 0.708	(0.905, 1.825) 0.825	(0.657, 1.577) 0.577	(0.796, 1.696) 0.696
$n = 1000$	(0.911, 1.571) 0.571	(1.013, 1.673) 0.673	(0.807, 1.447) 0.447	(0.911, 1.551) 0.551
	$m = 40$		$m = 50$	
	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ
$n = 100$	(-0.083 2.117) 1.117	(0.161, 2.481) 1.481	(-0.115, 2.065) 1.115	(0.111, 2.431) 1.431
$n = 200$	(0.299, 1.779) 0.779	(0.503, 2.003) 1.003	(0.260, 1.740) 0.740	(0.462, 1.962) 0.962
$n = 500$	(0.609, 1.509) 0.509	(0.740, 1.640) 0.640	(0.574, 1.474) 0.474	(0.704, 1.604) 0.604
$n = 1000$	(0.750, 1.390) 0.390	(0.853, 1.493) 0.493	(0.717, 1.357) 0.357	(0.823, 1.463) 0.463

Table 4.2.5 Case 1: with $\beta_0 = 0$ and Random Observation Time Points

Ave. of $\hat{\beta}_n$ & $\tilde{\beta}_n$ (s.d.) for Case 1: $\beta_0 = 0, t_{ij} \sim U(0, V_i)$

Number of Simulation Loop: 1000

	$n = 100$		$n = 200$	
	Ave. $\hat{\beta}_n$ (s.d.)	Ave. $\tilde{\beta}_n$ (s.d.)	Ave. $\hat{\beta}_n$ (s.d.)	Ave. $\tilde{\beta}_n$ (s.d.)
$m = 5$	1.129 (0.98)	1.333 (1.01)	0.968 (0.63)	1.131 (0.63)
$m = 10$	0.471 (0.87)	0.662 (0.96)	0.404 (0.59)	0.554 (0.63)
$m = 20$	0.100 (0.90)	0.269 (0.98)	0.169 (0.60)	0.303 (0.64)
$m = 30$	0.050 (0.92)	0.216 (0.99)	0.066 (0.60)	0.199 (0.64)
$m = 40$	-0.009 (0.94)	0.156 (1.01)	-0.003 (0.60)	0.125 (0.63)
$m = 50$	-0.062 (0.97)	0.091 (1.04)	-0.057 (0.60)	0.072 (0.64)
Censoring %	25.2%		25.0%	
	$n = 500$		$n = 1000$	
	Ave. $\hat{\beta}_n$ (s.d.)	Ave. $\tilde{\beta}_n$ (s.d.)	Ave. $\hat{\beta}_n$ (s.d.)	Ave. $\tilde{\beta}_n$ (s.d.)
$m = 5$	0.895 (0.37)	1.006 (0.36)	0.872 (0.26)	0.953 (0.25)
$m = 10$	0.395 (0.34)	0.504 (0.36)	0.379 (0.24)	0.475 (0.25)
$m = 20$	0.122 (0.35)	0.215 (0.37)	0.134 (0.22)	0.207 (0.25)
$m = 30$	0.060 (0.34)	0.148 (0.36)	0.081 (0.23)	0.152 (0.25)
$m = 40$	0.032 (0.35)	0.120 (0.37)	0.016 (0.23)	0.081 (0.25)
$m = 50$	-0.011 (0.37)	0.074 (0.39)	0.009 (0.23)	0.074 (0.25)
Censoring %	25.1%		25.0%	

Table 4.2.6 Case 1: Measurement of Coverage for Table 4.2.5

(Ave. $\hat{\beta}_n$) $\pm$ s.d. and (Ave. $\tilde{\beta}_n$) $\pm$ s.d. with Δ

	$m = 5$		$m = 10$	
	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ
$n = 100$	(0.149, 2.109) 2.109	(0.323, 2.343) 2.343	(-0.399, 1.341) 1.341	(-0.298, 1.622) 1.622
$n = 200$	(0.338, 1.598) 1.598	(0.501, 1.761) 1.761	(-0.186, 0.994) 0.994	(-0.076, 1.184) 1.184
$n = 500$	(0.525, 1.265) 1.265	(0.646, 1.366) 1.366	(0.055, 0.735) 0.735	(0.144, 0.864) 0.864
$n = 1000$	(0.612, 1.132) 1.132	(0.703, 1.203) 1.203	(0.139, 0.619) 0.619	(0.225, 0.725) 0.725
	$m = 20$		$m = 30$	
	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ
$n = 100$	(-0.800, 1.000) 1.000	(-0.711, 1.249) 1.249	(-0.870, 0.970) 0.970	(-0.774, 1.206) 1.206
$n = 200$	(-0.431, 0.769) 0.769	(-0.337, 0.943) 0.943	(-0.534, 0.666) 0.666	(-0.441, 0.839) 0.839
$n = 500$	(-0.228, 0.472) 0.472	(-0.155, 0.585) 0.585	(-0.280, 0.400) 0.400	(-0.212, 0.508) 0.508
$n = 1000$	(-0.086, 0.354) 0.354	(-0.043, 0.457) 0.457	(-0.149, 0.311) 0.311	(-0.098, 0.402) 0.402
	$m = 40$		$m = 50$	
	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ
$n = 100$	(-0.949, 0.931) 0.949	(-0.854, 1.166) 1.166	(-1.032, 0.908) 1.032	(-0.949, 1.131) 1.131
$n = 200$	(-0.603, 0.597) 0.603	(-0.505, 0.755) 0.755	(-0.657, 0.543) 0.657	(-0.568, 0.712) 0.712
$n = 500$	(-0.318, 0.382) 0.382	(-0.250, 0.490) 0.490	(-0.381, 0.359) 0.381	(-0.316, 0.464) 0.464
$n = 1000$	(-0.214, 0.246) 0.246	(-0.169, 0.331) 0.331	(-0.221, 0.239) 0.239	(-0.176, 0.324) 0.324

Table 4.2.7 Case 1: with $\beta_0 = 0$ and Evenly Chosen Observation Time Points

Ave. of $\hat{\beta}_n$ & $\tilde{\beta}_n$ (s.d.) for Case 1: $\beta_0 = 0$, t_{ij} evenly chosen

Number of Simulation Loop: 1000

	$n = 100$		$n = 200$	
	Ave. $\hat{\beta}_n$ (s.d.)	Ave. $\tilde{\beta}_n$ (s.d.)	Ave. $\hat{\beta}_n$ (s.d.)	Ave. $\tilde{\beta}_n$ (s.d.)
$m = 5$	1.576 (1.32)	1.807 (1.34)	1.307 (0.84)	1.480 (0.84)
$m = 10$	0.566 (1.02)	0.770 (1.10)	0.470 (0.67)	0.631 (0.71)
$m = 20$	0.190 (0.97)	0.366 (1.05)	0.157 (0.63)	0.302 (0.67)
$m = 30$	0.079 (0.96)	0.251 (1.04)	0.068 (0.63)	0.204 (0.67)
$m = 40$	0.023 (0.96)	0.191 (1.03)	0.023 (0.62)	0.156 (0.66)
$m = 50$	-0.005 (0.95)	0.159 (1.03)	-0.001 (0.62)	0.130 (0.66)
Censoring %	25.2%		25.0%	
	$n = 500$		$n = 1000$	
	Ave. $\hat{\beta}_n$ (s.d.)	Ave. $\tilde{\beta}_n$ (s.d.)	Ave. $\hat{\beta}_n$ (s.d.)	Ave. $\tilde{\beta}_n$ (s.d.)
$m = 5$	1.126 (0.48)	1.248 (0.47)	1.077 (0.33)	1.176 (0.33)
$m = 10$	0.411 (0.38)	0.527 (0.40)	0.401 (0.27)	0.501 (0.28)
$m = 20$	0.145 (0.36)	0.244 (0.38)	0.156 (0.25)	0.234 (0.27)
$m = 30$	0.065 (0.36)	0.159 (0.38)	0.084 (0.25)	0.155 (0.27)
$m = 40$	0.028 (0.36)	0.119 (0.38)	0.049 (0.25)	0.117 (0.27)
$m = 50$	0.007 (0.36)	0.097 (0.38)	0.029 (0.25)	0.096 (0.27)
Censoring %	25.1%		25.0%	

Table 4.2.8 Case 1: Measurement of Coverage for Table 4.2.7

(Ave. $\hat{\beta}_n$) $\pm$ s.d. and (Ave. $\tilde{\beta}_n$) $\pm$ s.d. with Δ

	$m = 5$		$m = 10$	
	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ
$n = 100$	(0.256, 2.896) 2.896	(0.467, 3.147) 3.147	(-0.454, 1.586) 1.586	(-0.330, 1.870) 1.870
$n = 200$	(0.467, 2.147) 2.147	(0.640, 2.320) 2.320	(-0.200, 1.140) 1.140	(-0.079, 1.341) 1.341
$n = 500$	(0.646, 1.606) 1.606	(0.778, 1.718) 1.718	(0.031, 0.791) 0.791	(0.127, 0.927) 0.927
$n = 1000$	(0.747, 1.407) 1.407	(0.846, 1.506) 1.506	(0.131, 0.671) 0.671	(0.221, 0.781) 0.781
	$m = 20$		$m = 30$	
	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ
$n = 100$	(-0.780, 1.160) 1.160	(-0.684, 1.416) 1.416	(-0.881, 1.039) 1.039	(-0.789, 1.291) 1.291
$n = 200$	(-0.473, 0.787) 0.787	(-0.368, 0.972) 0.972	(-0.562, 0.698) 0.698	(-0.466, 0.874) 0.874
$n = 500$	(-0.215, 0.505) 0.505	(-0.136, 0.624) 0.624	(-0.295, 0.425) 0.425	(-0.221, 0.539) 0.539
$n = 1000$	(-0.094, 0.406) 0.406	(-0.036, 0.504) 0.504	(-0.166, 0.334) 0.334	(-0.115, 0.425) 0.425
	$m = 40$		$m = 50$	
	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ
$n = 100$	(-0.937, 0.983) 0.983	(-0.839, 1.221) 1.221	(-0.955, 0.945) 0.955	(-0.871, 1.189) 1.189
$n = 200$	(-0.597, 0.643) 0.643	(-0.504, 0.816) 0.816	(-0.621, 0.619) 0.621	(-0.530, 0.790) 0.790
$n = 500$	(-0.332, 0.388) 0.388	(-0.261, 0.499) 0.499	(-0.353, 0.367) 0.367	(-0.283, 0.477) 0.477
$n = 1000$	(-0.201, 0.299) 0.299	(-0.115, 0.425) 0.425	(-0.221, 0.279) 0.279	(-0.174, 0.366) 0.366

Table 4.2.9 Case 1: with $\beta_0 = -1$ and Random Observation Time Points

Ave. of $\hat{\beta}_n$ & $\tilde{\beta}_n$ (s.d.) for Case 1: $\beta_0 = -1, t_{ij} \sim U(0, V_i)$

Number of Simulation Loop: 1000

	$n = 100$		$n = 200$	
	Ave. $\hat{\beta}_n$ (s.d.)	Ave. $\tilde{\beta}_n$ (s.d.)	Ave. $\hat{\beta}_n$ (s.d.)	Ave. $\tilde{\beta}_n$ (s.d.)
$m = 5$	0.092 (0.84)	0.212 (0.89)	-0.012 (0.50)	0.071 (0.53)
$m = 10$	-0.576 (0.86)	-0.482 (0.92)	-0.591 (0.52)	-0.536 (0.54)
$m = 20$	-0.854 (0.88)	-0.771 (0.92)	-0.873 (0.56)	-0.824 (0.58)
$m = 30$	-0.957 (0.92)	-0.876 (0.95)	-0.980 (0.57)	-0.930 (0.59)
$m = 40$	-1.038 (0.96)	-0.955 (0.99)	-0.993 (0.60)	-0.944 (0.61)
$m = 50$	-1.032 (0.94)	-0.953 (0.97)	-1.028 (0.59)	-0.980 (0.60)
Censoring %	30.1%		30.0%	
	$n = 500$		$n = 1000$	
	Ave. $\hat{\beta}_n$ (s.d.)	Ave. $\tilde{\beta}_n$ (s.d.)	Ave. $\hat{\beta}_n$ (s.d.)	Ave. $\tilde{\beta}_n$ (s.d.)
$m = 5$	-0.075 (0.30)	-0.027 (0.32)	-0.091 (0.20)	-0.062 (0.21)
$m = 10$	-0.587 (0.33)	-0.562 (0.34)	-0.578 (0.22)	-0.562 (0.23)
$m = 20$	-0.815 (0.34)	-0.791 (0.34)	-0.821 (0.23)	-0.807 (0.23)
$m = 30$	-0.911 (0.33)	-0.889 (0.34)	-0.897 (0.23)	-0.879 (0.23)
$m = 40$	-0.967 (0.34)	-0.946 (0.35)	-0.937 (0.23)	-0.922 (0.23)
$m = 50$	-0.946 (0.35)	-0.923 (0.35)	-0.936 (0.24)	-0.919 (0.24)
Censoring %	30.1%		30.0%	

Table 4.2.10 Case 1: Measurement of Coverage for Table 4.2.9

(Ave. $\hat{\beta}_n$) $\pm$ s.d. and (Ave. $\tilde{\beta}_n$) $\pm$ s.d. with Δ

	$m = 5$		$m = 10$	
	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ
$n = 100$	(-0.748, 0.932) 1.932	(-0.678, 1.102) 2.102	(-1.436, 0.284) 1.284	(-1.402, 0.438) 1.438
$n = 200$	(-0.512, 0.488) 1.488	(-0.459, 0.601) 1.601	(-1.111, -0.071) 0.929	(-1.076, 0.004) 1.004
$n = 500$	(-0.375, 0.225) 1.225	(-0.347, 0.293) 1.293	(-0.917, -0.257) 0.743	(-0.902, -0.222) 0.778
$n = 1000$	(-0.291, 0.109) 1.109	(-0.272, 0.148) 1.148	(-0.798, -0.358) 0.642	(-0.792, -0.332) 0.668
	$m = 20$		$m = 30$	
	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ
$n = 100$	(-1.734, 0.026) 1.026	(-1.691, 0.149) 1.149	(-1.877, -0.037) 0.963	(-1.826, 0.074) 1.074
$n = 200$	(-1.433, -0.313) 0.687	(-1.404, -0.244) 0.756	(-1.550, -0.410) 0.590	(-1.520, -0.340) 0.660
$n = 500$	(-1.155, -0.475) 0.525	(-1.131, -0.451) 0.549	(-1.241, -0.581) 0.419	(-1.229, -0.549) 0.451
$n = 1000$	(-1.051, -0.591) 0.409	(-1.037, -0.577) 0.423	(-1.127, -0.667) 0.333	(-1.109, -0.649) 0.351
	$m = 40$		$m = 50$	
	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ
$n = 100$	(-1.998, -0.078) 0.998	(-1.945, 0.035) 1.035	(-1.972, -0.092) 0.972	(-1.923, 0.017) 1.017
$n = 200$	(-1.593, -0.393) 0.607	(-1.554, -0.334) 0.666	(-1.618, -0.438) 0.618	(-1.580, -0.380) 0.620
$n = 500$	(-1.307, -0.627) 0.373	(-1.296, -0.596) 0.404	(-1.296, -0.596) 0.404	(-1.273, -0.573) 0.427
$n = 1000$	(-1.167, -0.707) 0.293	(-1.152, -0.692) 0.308	(-1.176, -0.696) 0.304	(-1.159, -0.679) 0.321

Table 4.2.11 Case 1: with $\beta_0 = -1$ and Evenly Chosen Observation Time Points

Ave. of $\hat{\beta}_n$ & $\tilde{\beta}_n$ (s.d.) for Case 1: $\beta_0 = -1$, t_{ij} evenly chosen

Number of Simulation Loop: 1000

	$n = 100$		$n = 200$	
	Ave. $\hat{\beta}_n$ (s.d.)	Ave. $\tilde{\beta}_n$ (s.d.)	Ave. $\hat{\beta}_n$ (s.d.)	Ave. $\tilde{\beta}_n$ (s.d.)
$m = 5$	0.179 (1.04)	0.322 (1.10)	0.015 (0.63)	0.118 (0.68)
$m = 10$	-0.580 (0.92)	-0.479 (0.97)	-0.607 (0.59)	-0.542 (0.62)
$m = 20$	-0.882 (0.92)	-0.795 (0.96)	-0.864 (0.61)	-0.808 (0.63)
$m = 30$	-0.974 (0.92)	-0.890 (0.96)	-0.937 (0.61)	-0.883 (0.63)
$m = 40$	-1.017 (0.93)	-0.934 (0.96)	-0.974 (0.61)	-0.921 (0.63)
$m = 50$	-1.043 (0.93)	-0.963 (0.96)	-0.996 (0.61)	-0.943 (0.63)
Censoring %	30.1%		30.0%	
	$n = 500$		$n = 1000$	
	Ave. $\hat{\beta}_n$ (s.d.)	Ave. $\tilde{\beta}_n$ (s.d.)	Ave. $\hat{\beta}_n$ (s.d.)	Ave. $\tilde{\beta}_n$ (s.d.)
$m = 5$	-0.087 (0.33)	-0.025 (0.36)	-0.098 (0.23)	-0.059 (0.25)
$m = 10$	-0.629 (0.33)	-0.599 (0.34)	-0.627 (0.24)	-0.607 (0.24)
$m = 20$	-0.850 (0.34)	-0.825 (0.34)	-0.841 (0.24)	-0.824 (0.24)
$m = 30$	-0.918 (0.34)	-0.893 (0.35)	-0.906 (0.24)	-0.889 (0.24)
$m = 40$	-0.949 (0.34)	-0.925 (0.35)	-0.935 (0.24)	-0.919 (0.24)
$m = 50$	-0.968 (0.34)	-0.944 (0.35)	-0.952 (0.24)	-0.936 (0.25)
Censoring %	30.1%		30.0%	

Table 4.2.12 Case 1: Measurement of Coverage for Table 4.2.11

(Ave. $\hat{\beta}_n$) $\pm$ s.d. and (Ave. $\tilde{\beta}_n$) $\pm$ s.d. with Δ

	$m = 5$		$m = 10$	
	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ
$n = 100$	(-0.861, 1.219) 2.219	(-0.778, 1.422) 2.422	(-1.500, 0.340) 1.340	(-1.449, 0.491) 1.491
$n = 200$	(-0.615, 0.645) 1.645	(-0.562, 0.798) 1.798	(-1.197, -0.017) 0.983	(-1.162, 0.078) 1.078
$n = 500$	(-0.417, 0.243) 1.243	(-0.385, 0.335) 1.335	(-0.959, -0.299) 0.701	(-0.939, -0.259) 0.741
$n = 1000$	(-0.328, 0.132) 1.132	(-0.309, 0.191) 1.191	(-0.867, -0.387) 0.613	(-0.847, -0.367) 0.633
	$m = 20$		$m = 30$	
	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ
$n = 100$	(-1.802, 0.038) 1.038	(-1.755, 0.165) 1.165	(-1.894, -0.054) 0.946	(-1.850, 0.070) 1.070
$n = 200$	(-1.474, -0.254) 0.746	(-1.438, -0.178) 0.822	(-1.547, -0.327) 0.673	(-1.513, -0.253) 0.747
$n = 500$	(-1.190, -0.510) 0.490	(-1.165, -0.485) 0.515	(-1.258, -0.578) 0.422	(-1.243, -0.543) 0.457
$n = 1000$	(-1.081, -0.601) 0.399	(-1.064, -0.584) 0.416	(-1.146, -0.666) 0.334	(-1.129, -0.649) 0.351
	$m = 40$		$m = 50$	
	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ
$n = 100$	(-1.947, -0.087) 0.947	(-1.894, 0.026) 1.026	(-1.973, -0.113) 0.973	(-1.923, -0.003) 0.997
$n = 200$	(-1.584, -0.364) 0.636	(-1.551, -0.291) 0.709	(-1.606, -0.386) 0.614	(-1.573, -0.313) 0.687
$n = 500$	(-1.289, -0.609) 0.391	(-1.275, -0.575) 0.425	(-1.308, -0.628) 0.372	(-1.294, -0.594) 0.406
$n = 1000$	(-1.175, -0.695) 0.305	(-1.159, -0.679) 0.321	(-1.192, -0.712) 0.288	(-1.186, -0.686) 0.314

4.3 Case 2: Use Total Summary Measurement

Here, we consider using $Z(t) = \varphi(X^t)$ in our proposed PH model (1.38) as total summary measurement, which is motivated by the Child Development Data described in Section 1.5. We call this Case 2 in our simulation studies.

Below, Section 4.3.1 describes the specific simulation setting for our Case 2 studies and outlines the steps to generate data (1.29) for such case, while Section 4.3.2 presents the simulation results under our Case 2 setting.

4.3.1 Case 2 Simulation Setting

Here, we consider the summary measurement as the total of $X(t;\omega)$ in time interval $(0,t)$. Let points $s_0 = 0 < s_1 < \ldots < s_N = t$ evenly divide interval $(0,t)$ in to N subintervals with equal length $\Delta = \frac{t}{N}$. Then, the approximated amount of $X(t;\omega)$ in time interval $(s, s+\Delta)$ is $X(s;\omega)\Delta$, thus the approximated total amount of $X(t;\omega)$ in time interval $(0,t)$ is $S_N = \sum_{i=1}^{N} X(s_i;\omega)\Delta = \frac{t\sum_{i=1}^{N} X(s_i;\omega)}{N}$, which is a Riemann sum. Let $N \to \infty$, from the definition of definite integral we have:

$$Z(t) = \varphi(X^t) = \lim_{N\to\infty} S_N = \int_0^t X(s;\omega)ds.$$

Thus, in our proposed PH model (1.38), we consider the following Case 2 data setting for our simulation studies:

- $F_0 = 1 - e^{-t^2/2}$, which is the baseline d.f., thus $f_0(t) = te^{-t^2/2}$, $h_0(t) = t$ in equation (3.7)

- $F_C = \text{Exp}(3)$, which is the d.f. of right censoring variable in (1.2)

- $X(t) = X(t;\omega) = \frac{\omega t}{2}$, which is the longitudinal covariate process with $\omega \sim U(0,1)$

- $Z(t) = \varphi(X^t) = \int_0^t \frac{\omega s}{2} ds = \frac{\omega t^2}{4}$, which is the total of $X(t;\omega)$ in time interval $(0,t)$

- From Case 1, we know $F(t\,|\,X^t)$ given by (3.7) is something we can use to generate a conditional r.v. T given X^t, denoted by $[\,T\,|\,X^t\,]$.

Based on above choices, we have the following in (3.7): for $\beta_0 \neq 0$,

$$\begin{aligned}\bar{F}(t\,|\,X^t) &= \exp\left\{-\int_0^t h_0(s)e^{Z(s)\beta_0}ds\right\} = \exp\left\{-\int_0^t se^{\beta_0\omega s^2/4}ds\right\}\\ &= \exp\left\{-\frac{e^{\beta_0\omega s^2/4}}{\beta_0\omega/2}\bigg|_0^t\right\} = \exp\left\{\frac{1-e^{\beta_0\omega t^2/4}}{\beta_0\omega/2}\right\},\end{aligned} \tag{4.27}$$

which implies

$$F(t\,|\,X^t) = 1-\exp\left\{\frac{1-e^{\beta_0\omega t^2/4}}{\beta_0\omega/2}\right\} \equiv G(t). \tag{4.28}$$

We have the following in above $G(t)$:

$$\begin{aligned}&G(0) = 1-e^0 = 0\\ &G(\infty) = 1-\lim_{t\to\infty}\exp\left\{\frac{1-e^{\beta_0\omega t^2/4}}{\beta_0\omega/2}\right\} = 1-e^{\frac{2}{\beta_0\omega}}\lim_{t\to\infty}\exp\left\{\frac{-e^{\beta_0\omega t^2/4}}{\beta_0\omega/2}\right\}.\end{aligned} \tag{4.29}$$

Notice that for any given $\omega > 0$, above $G(t)$ is an increasing function because

$$\begin{aligned}G^{'}(t) &= -\left(\frac{-e^{\beta_0\omega t^2/4}}{\beta_0\omega/2}\right)\left(\beta_0\frac{\omega t}{2}\right)\exp\left\{\frac{1-e^{\beta_0\omega t^2/4}}{\beta_0\omega/2}\right\}\\ &= \exp\left\{\frac{1-e^{\beta_0\omega t^2/4}}{\beta_0\omega/2}\right\}e^{\beta_0\omega t^2/4}t > 0.\end{aligned}$$

For $\beta_0 > 0$, from (4.29) we have $G(t)$ in (4.28) satisfies

$$
\begin{aligned}
G(0) &= 1 - e^0 = 0 \\
G(\infty) &= 1 - e^{\frac{2}{\beta_0\omega}} \lim_{t\to\infty} \exp\left\{\frac{-e^{\beta_0\omega t^2/4}}{\beta_0\omega/2}\right\} = 1 - e^{\frac{2}{\beta_0\omega}} \exp\left\{\frac{-e^{+\infty}}{\beta_0\omega/2}\right\} \\
&= 1 - e^{\frac{2}{\beta_0\omega}} e^{-\infty} = 1 - e^{\frac{2}{\beta_0\omega}} \cdot 0 = 1.
\end{aligned}
$$

Thus, $G(t)$ is a proper d.f. when $\beta_0 > 0$.

For $\beta_0 = 0$, we have the following in (3.7)

$$
\begin{aligned}
\bar{F}(t \,|\, X^t) &= \exp\left\{ -\int_0^t h_0(s) e^{Z(s)\beta_0} ds \right\} = \exp\left\{ -\int_0^t s ds \right\} \\
&= \exp\left\{ -\frac{s^2}{2}\Big|_0^t \right\} = e^{-t^2/2}, \qquad (4.30)
\end{aligned}
$$

which implies

$$
F(t \,|\, X^t) = 1 - e^{-t^2/2} \equiv G(t). \qquad (4.31)
$$

Notice that above $G(t)$ is an increasing function satisfying

$$
G(0) = 1 - e^0 = 0 \quad \text{and} \quad G(\infty) = 1 - e^{-\infty} = 1.
$$

Thus, $G(t)$ is a proper d.f. when $\beta_0 = 0$.

For $\beta_0 < 0$, from (4.29) we have $G(t)$ in (4.28) satisfies

$$
\begin{aligned}
G(0) &= 1 - e^0 = 0 \\
G(\infty) &= 1 - e^{\frac{2}{\beta_0\omega}} \lim_{t\to\infty} \exp\left\{\frac{-e^{\beta_0\omega t^2/4}}{\beta_0\omega/2}\right\} = 1 - e^{\frac{2}{\beta_0\omega}} \exp\left\{\frac{-e^{-\infty}}{\beta_0\omega/2}\right\} \\
&= 1 - e^{\frac{2}{\beta_0\omega}} \cdot e^0 = 1 - e^{\frac{2}{\beta_0\omega}} \equiv g_2(\omega, \beta_0) < 1, \qquad (4.32)
\end{aligned}
$$

which implies $G(t)$ is a not proper d.f. when $\beta_0 < 0$. From (4.32), we know $g_2(\omega, \beta_0)$ depends on the value of β_0 and ω, where $\omega \sim U(0,1)$.

Derivation for G^{-1} in Case 2:

From $G(t) = F(t \mid X^t)$ given by (4.28), we know that $Y = G^{-1}(U)$ is a r.v. generated from d.f. $G(t)$, where $U \sim U(0,1)$. To find G^{-1}, we compute the following:

$$
\begin{aligned}
G(t) = u \;&\Rightarrow\; u = 1 - \exp\left\{\frac{1 - e^{\beta_0 \omega t^2/4}}{\beta_0 \omega/2}\right\} \;\Rightarrow\; \frac{1 - e^{\beta_0 \omega t^2/4}}{\beta_0 \omega/2} = \log(1-u) \\
&\Rightarrow\; e^{\beta_0 \omega t^2/4} = 1 - \frac{\beta_0 \omega}{2} \log(1-u) \\
&\Rightarrow\; t = \left[\tfrac{4}{\beta_0 \omega} \log\left(1 - \frac{\beta_0 \omega}{2} \log(1-u)\right)\right]^{1/2} = G^{-1}(u).
\end{aligned}
\tag{4.33}
$$

Hence, for $U \sim U(0,1)$, random variable $Y = \left[\tfrac{4}{\beta_0 \omega} \log\left(1 - \tfrac{\beta_0 \omega}{2} \log(1-U)\right)\right]^{1/2}$ is generated from above d.f. $G(t) = F(t \mid X^t)$ in (4.33), and Y is a conditional r.v. T given X^t.

For $\beta_0 > 0$, $G^{-1}(U)$ in (4.33) is given by

$$
G^{-1}(U) = \left[\tfrac{4}{\beta_0 \omega} \log\left(1 - \tfrac{\beta_0 \omega}{2} \log(1-U)\right)\right]^{1/2}. \tag{4.34}
$$

For $\beta_0 = 0$, from (4.31) we compute the following:

$$
G(t) = u \;\Rightarrow\; u = 1 - e^{-t^2/2} \Rightarrow\; -\frac{t^2}{2} = \log(1-u) \Rightarrow\; t = \left[-2\log(1-u)\right]^{1/2}.
$$

Thus, when $\beta_0 = 0$, $G^{-1}(U)$ is given by

$$
G^{-1}(U) = \left[-2\log(1-U)\right]^{1/2}. \tag{4.35}
$$

From (4.32), for $\beta_0 < 0$ and $U \leq 1 - e^{\frac{2}{\beta_0 \omega}}$, $G^{-1}(U)$ in (4.33) is given by

$$G^{-1}(U) = \left[\frac{4}{\beta_0 \omega} \log\left(1 - \frac{\beta_0 \omega}{2} \log(1-U)\right)\right]^{1/2}. \tag{4.36}$$

For $\beta_0 < 0$ and $U > 1 - e^{\frac{2}{\beta_0 \omega}}$, $G^{-1}(U) = \infty$ in (4.33).

Steps to Generate Data (1.29) for Case 2:

Step 1. Generate $\omega_i \sim U(0, 1)$

Step 2. Generate $U \sim U(0, 1)$ and compute

$$T_i = \begin{cases} \left[\frac{4}{\beta_0 \omega_i} \log\left(1 - \frac{\beta_0 \omega_i}{2} \log(1-U)\right)\right]^{1/2} & \text{if } \beta_0 > 0 \\ \left[-2\log(1-U)\right]^{1/2} & \text{if } \beta_0 = 0 \end{cases}$$

$$T_i \overset{(4.32)}{=\!=} \begin{cases} \left[\frac{4}{\beta_0 \omega_i} \log\left(1 - \frac{\beta_0 \omega_i}{2} \log(1-U)\right)\right]^{1/2} & \text{if } \beta_0 < 0,\ U \leq 1 - e^{\frac{2}{\beta_0 \omega_i}} \\ \infty & \text{if } \beta_0 < 0,\ U > 1 - e^{\frac{2}{\beta_0 \omega_i}} \end{cases}$$

Step 3. Generate $U \sim U(0, 1)$ and compute $C_i = -3\log(1-U)$.

Step 4. Obtain (V_i, δ_i) based on Step 2-3, where $V_i = \min\{T_i, C_i\}$ and $\delta_i = I\{T_i \leq C_i\}$, which is the first part of data (1.29).

Step 5. Let $N_i = [V_i]+1$, where $[V_i]$ is the integer part of V_i, then for a chosen integer m, compute $J_i = mN_i$. Generate t_{ij}'s as in (4.23), then order them into $t_{i1} < \ldots < t_{iJ_i}$, which gives $\mathcal{X}_i^{V_i} = \{X_i(t_{ij}) \,|\, j = 1, \ldots, J_i\} = \{\frac{\omega_i}{2} t_{ij} \,|\, j = 1, \ldots, J_i\}$, which is the last part of data (1.29).

Step 6. Compute $Z(t_{ij})$ in equation (4.9):

$$Z(t_{ij}) = \varphi(\mathcal{X}_i^{t_{ij}}) = \sum_{k=1}^{j}(t_{ik} - t_{i,k-1})X_i(t_{i,k-1}). \tag{4.37}$$

NOTE: $G(t)$ is not a proper d.f. when $\beta_0 < 0$ because from (4.32) we have $g_2(\omega, \beta_0) < 1$. If $u \in (g_2(\omega, \beta_0), 1)$, then $G^{-1}(u)$ does not exists and $T_i = G^{-1}(u) = \infty$, which gives $V_i = \min\{T_i, C_i\} = C_i$. For Case 1 with $\beta_0 < 0$, we have the same situation.

4.3.2 Simulation Results

Similar to Case 1, for Case 2 simulation studies, we consider $\beta_0 = 1, 0, -1$, and we follow Steps 1-6 described in Section 4.3.1 to generate 1000 samples with sample size $n = 100, 200, 500, 1000$ and $m = 5, 10, 20, 30, 40, 50$, respectively. For each case, we compute the simulation mean and s.d. for $\hat{\beta}_n$ and $\tilde{\beta}_n$, and we also compute (Ave. $\hat{\beta}_n$) $\pm$ s.d. and (Ave. $\tilde{\beta}_n$) $\pm$ s.d. with Δ for interval (a, b), where Δ is given in (4.26).

Table 4.3.1 presents the simulation mean and s.d. of $\hat{\beta}_n$ and $\tilde{\beta}_n$ for Case 2 data setting with $\beta_0 = 1$ and $t_{ij} \sim U(0, V_i)$, while Table 4.3.2 gives gives (Ave. $\hat{\beta}_n$) $\pm$ s.d. and (Ave. $\tilde{\beta}_n$) $\pm$ s.d. with Δ for simulation results in Table 4.3.1. Similarly, Tables 4.3.3-4.3.4 display results for Case 2 with $\beta_0 = 1$ and t_{ij}'s evenly chosen; Tables 4.3.5-4.3.12 display results accordingly for $\beta_0 = 0, -1$, respectively.

Table 4.3.1 Case 2: with $\beta_0 = 1$ and Random Observation Time Points

Ave. of $\hat{\beta}_n$ & $\tilde{\beta}_n$ (s.d.) for Case 2: $\beta_0 = 1, t_{ij} \sim U(0, V_i)$

Number of Simulation Loop: 1000

	$n = 100$		$n = 200$	
	Ave. $\hat{\beta}_n$ (s.d.)	Ave. $\tilde{\beta}_n$ (s.d.)	Ave. $\hat{\beta}_n$ (s.d.)	Ave. $\tilde{\beta}_n$ (s.d.)
$m = 5$	0.848 (1.05)	1.135 (1.13)	0.759 (0.70)	0.974 (0.72)
$m = 10$	0.762 (1.13)	1.064 (1.21)	0.806 (0.75)	1.028 (0.78)
$m = 20$	0.778 (1.17)	1.078 (1.25)	0.807 (0.74)	1.036 (0.77)
$m = 30$	0.775 (1.13)	1.067 (1.21)	0.816 (0.74)	1.044 (0.76)
$m = 40$	0.750 (1.05)	1.049 (1.14)	0.820 (0.72)	1.042 (0.74)
$m = 50$	0.796 (1.14)	1.083 (1.22)	0.843 (0.77)	1.073 (0.79)
Censoring %	31.2%		31.1%	
	$n = 500$		$n = 1000$	
	Ave. $\hat{\beta}_n$ (s.d.)	Ave. $\tilde{\beta}_n$ (s.d.)	Ave. $\hat{\beta}_n$ (s.d.)	Ave. $\tilde{\beta}_n$ (s.d.)
$m = 5$	0.873 (0.43)	1.007 (0.42)	0.840 (0.29)	0.947 (0.29)
$m = 10$	0.889 (0.44)	1.029 (0.43)	0.887 (0.29)	0.999 (0.29)
$m = 20$	0.871 (0.44)	1.012 (0.43)	0.909 (0.30)	1.022 (0.30)
$m = 30$	0.890 (0.45)	1.031 (0.44)	0.885 (0.31)	0.999 (0.30)
$m = 40$	0.871 (0.47)	1.011 (0.46)	0.885 (0.30)	0.999 (0.30)
$m = 50$	0.882 (0.44)	1.020 (0.43)	0.904 (0.30)	1.015 (0.30)
Censoring %	31.1%		31.0%	

Table 4.3.2 Case 2: Measurement of Coverage for Table 4.3.1

(Ave. $\hat{\beta}_n$) $\pm$ s.d. and (Ave. $\tilde{\beta}_n$) $\pm$ s.d. with Δ

	$m = 5$		$m = 10$	
	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ
$n = 100$	(-0.202, 1.898) 1.202	(0.005, 2.265) 1.265	(-0.368, 1.892) 1.368	(-0.146, 2.274) 1.274
$n = 200$	(0.059, 1.459) 0.941	(0.254, 1.694) 0.746	(0.056, 1.556) 0.944	(0.248, 1.808) 0.808
$n = 500$	(0.443, 1.303) 0.557	(0.587, 1.427) 0.427	(0.449, 1.329) 0.551	(0.599, 1.459) 0.459
$n = 1000$	(0.550, 1.130) 0.450	(0.657, 1.237) 0.343	(0.597, 1.177) 0.403	(0.709, 1.289) 0.291
	$m = 20$		$m = 30$	
	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ
$n = 100$	(-0.392, 1.948) 1.392	(-0.172, 2.328) 1.328	(-0.355, 1.905) 1.355	(-0.143, 2.277) 1.277
$n = 200$	(0.067, 1.547) 0.933	(0.266, 1.806) 0.806	(0.076, 1.556) 0.924	(0.284, 1.804) 0.804
$n = 500$	(0.431, 1.311) 0.569	(0.582, 1.442) 0.442	(0.440, 1.340) 0.560	(0.591, 1.471) 0.471
$n = 1000$	(0.609, 1.209) 0.391	(0.722, 1.322) 0.322	(0.575, 1.195) 0.425	(0.699, 1.299) 0.301
	$m = 40$		$m = 50$	
	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ
$n = 100$	(-0.300, 1.800) 1.300	(-0.091, 2.189) 1.189	(-0.344, 1.936) 1.344	(-0.137, 2.303) 1.303
$n = 200$	(0.100, 1.540) 0.900	(0.302, 1.782) 0.782	(0.073, 1.613) 0.927	(0.283, 1.863) 0.863
$n = 500$	(0.401, 1.341) 0.599	(0.551, 1.471) 0.471	(0.442, 1.322) 0.558	(0.590, 1.450) 0.450
$n = 1000$	(0.585, 1.185) 0.415	(0.699, 1.299) 0.301	(0.604, 1.204) 0.396	(0.715, 1.315) 0.315

Table 4.3.3 Case 2: with $\beta_0 = 1$ and Evenly Chosen Observation Time Points

Ave. of $\hat{\beta}_n$ & $\tilde{\beta}_n$ (s.d.) for Case 2: $\beta_0 = 1$, t_{ij} evenly chosen

Number of Simulation Loop: 1000

	$n = 100$		$n = 200$	
	Ave. $\hat{\beta}_n$ (s.d.)	Ave. $\tilde{\beta}_n$ (s.d.)	Ave. $\hat{\beta}_n$ (s.d.)	Ave. $\tilde{\beta}_n$ (s.d.)
$m = 5$	2.823 (1.62)	3.127 (1.57)	2.531 (1.01)	2.710 (0.98)
$m = 10$	1.696 (1.30)	2.034 (1.32)	1.607 (0.86)	1.833 (0.84)
$m = 20$	1.222 (1.19)	1.544 (1.25)	1.198 (0.79)	1.435 (0.79)
$m = 30$	1.079 (1.17)	1.396 (1.23)	1.073 (0.77)	1.311 (0.78)
$m = 40$	1.006 (1.15)	1.311 (1.22)	1.014 (0.76)	1.244 (0.78)
$m = 50$	0.973 (1.14)	1.271 (1.22)	0.980 (0.76)	1.208 (0.77)
Censoring %	31.2%		31.1%	
	$n = 500$		$n = 1000$	
	Ave. $\hat{\beta}_n$ (s.d.)	Ave. $\tilde{\beta}_n$ (s.d.)	Ave. $\hat{\beta}_n$ (s.d.)	Ave. $\tilde{\beta}_n$ (s.d.)
$m = 5$	2.300 (0.56)	2.418 (0.56)	2.227 (0.39)	2.320 (0.39)
$m = 10$	1.553 (0.50)	1.681 (0.49)	1.533 (0.35)	1.635 (0.35)
$m = 20$	1.207 (0.47)	1.342 (0.46)	1.211 (0.33)	1.319 (0.33)
$m = 30$	1.100 (0.46)	1.235 (0.45)	1.110 (0.33)	1.220 (0.32)
$m = 40$	1.044 (0.46)	1.180 (0.46)	1.059 (0.32)	1.170 (0.32)
$m = 50$	1.009 (0.46)	1.147 (0.45)	1.028 (0.32)	1.139 (0.32)
Censoring %	31.1%		31.0%	

Table 4.3.4 Case 2: Measurement of Coverage for Table 4.3.3

(Ave. $\hat{\beta}_n$) $\pm$ s.d. and (Ave. $\tilde{\beta}_n$) $\pm$ s.d. with Δ

	$m = 5$		$m = 10$	
	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ
$n = 100$	(1.203, 4.443) 3.443	(1.557, 4.697) 3.697	(0.396, 2.996) 1.996	(0.714, 3.354) 2.354
$n = 200$	(1.521, 3.541) 2.541	(1.730, 3.690) 2.690	(0.747, 2.467) 1.467	(0.993, 2.673) 1.673
$n = 500$	(1.740, 2.860) 1.860	(1.858, 2.978) 1.978	(1.053, 2.053) 1.053	(1.191, 2.171) 1.171
$n = 1000$	(1.837, 2.617) 1.617	(1.930, 2.710) 1.710	(1.183, 1.883) 0.883	(1.285, 1.985) 0.985
	$m = 20$		$m = 30$	
	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ
$n = 100$	(0.032, 2.412) 1.412	(0.294, 2.794) 1.794	(-0.091, 2.249) 1.249	(0.166, 2.626) 1.626
$n = 200$	(0.408, 1.988) 0.988	(0.645, 2.225) 1.225	(0.303, 1.843) 0.843	(0.531, 2.091) 1.091
$n = 500$	(0.737, 1.677) 0.677	(0.882, 1.802) 0.802	(0.640, 1.560) 0.560	(0.785, 1.685) 0.685
$n = 1000$	(0.881, 1.541) 0.541	(0.989, 1.649) 0.649	(0.780, 1.440) 0.440	(0.900, 1.540) 0.540
	$m = 40$		$m = 50$	
	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ
$n = 100$	(-0.144, 2.156) 1.156	(0.091, 2.531) 1.531	(-0.167, 2.113) 1.167	(0.051, 2.491) 1.491
$n = 200$	(0.254, 1.774) 0.774	(0.464, 2.024) 1.024	(0.220, 1.740) 0.780	(0.438, 1.978) 0.978
$n = 500$	(0.584, 1.504) 0.504	(0.720, 1.640) 0.640	(0.549, 1.469) 0.469	(0.697, 1.597) 0.597
$n = 1000$	(0.739, 1.379) 0.379	(0.850, 1.490) 0.490	(0.708, 1.348) 0.348	(0.819, 1.459) 0.459

Table 4.3.5 Case 2: with $\beta_0 = 0$ and Random Observation Time Points

Ave. of $\hat{\beta}_n$ & $\tilde{\beta}_n$ (s.d.) for Case 2: $\beta_0 = 0, t_{ij} \sim U(0, V_i)$

Number of Simulation Loop: 1000

	$n = 100$		$n = 200$	
	Ave. $\hat{\beta}_n$ (s.d.)	Ave. $\tilde{\beta}_n$ (s.d.)	Ave. $\hat{\beta}_n$ (s.d.)	Ave. $\tilde{\beta}_n$ (s.d.)
$m = 5$	-0.111 (0.96)	0.037 (1.03)	-0.041 (0.55)	0.083 (0.59)
$m = 10$	-0.112 (0.99)	0.046 (1.07)	-0.106 (0.61)	0.019 (0.65)
$m = 20$	-0.128 (1.03)	0.032 (1.12)	-0.123 (0.65)	0.003 (0.69)
$m = 30$	-0.105 (1.01)	0.053 (1.09)	-0.127 (0.61)	-0.004 (0.66)
$m = 40$	-0.171 (1.00)	-0.021 (1.07)	-0.166 (0.63)	-0.044 (0.67)
$m = 50$	-0.192 (1.03)	-0.036 (1.12)	-0.109 (0.61)	0.018 (0.65)
Censoring %	32.8%		32.7%	
	$n = 500$		$n = 1000$	
	Ave. $\hat{\beta}_n$ (s.d.)	Ave. $\tilde{\beta}_n$ (s.d.)	Ave. $\hat{\beta}_n$ (s.d.)	Ave. $\tilde{\beta}_n$ (s.d.)
$m = 5$	-0.030 (0.33)	0.052 (0.35)	-0.026 (0.23)	0.039 (0.25)
$m = 10$	-0.058 (0.34)	0.024 (0.36)	-0.056 (0.23)	0.007 (0.25)
$m = 20$	-0.079 (0.33)	0.005 (0.36)	-0.060 (0.23)	0.003 (0.25)
$m = 30$	-0.095 (0.36)	-0.012 (0.38)	-0.059 (0.23)	0.002 (0.24)
$m = 40$	-0.097 (0.35)	-0.013 (0.37)	-0.068 (0.24)	-0.003 (0.25)
$m = 50$	-0.072 (0.34)	0.012 (0.36)	-0.063 (0.24)	0.000 (0.25)
Censoring %	32.7%		32.6%	

Table 4.3.6 Case 2: Measurement of Coverage for Table 4.3.5

(Ave. $\hat{\beta}_n$) $\pm$ s.d. and (Ave. $\tilde{\beta}_n$) $\pm$ s.d. with Δ

	$m = 5$		$m = 10$	
	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ
$n = 100$	(-1.071, 0.849) 1.071	(-0.993, 1.067) 1.067	(-1.102, 0.878) 1.102	(-1.024, 1.116) 1.116
$n = 200$	(-0.591, 0.509) 0.591	(-0.507, 0.673) 0.673	(-0.716, 0.504) 0.716	(-0.631, 0.669) 0.669
$n = 500$	(-0.360, 0.300) 0.360	(-0.298, 0.402) 0.402	(-0.398, 0.282) 0.398	(-0.336, 0.384) 0.384
$n = 1000$	(-0.256, 0.204) 0.256	(-0.211, 0.289) 0.289	(-0.286, 0.174) 0.286	(-0.243, 0.257) 0.257
	$m = 20$		$m = 30$	
	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ
$n = 100$	(-1.158, 0.902) 1.158	(-1.088, 1.152) 1.152	(-1.115, 0.905) 1.115	(-1.037, 1.143) 1.143
$n = 200$	(-0.773, 0.527) 0.773	(-0.687, 0.693) 0.693	(-0.737, 0.483) 0.737	(-0.664, 0.656) 0.664
$n = 500$	(-0.409, 0.251) 0.409	(-0.355, 0.365) 0.365	(-0.455, 0.265) 0.455	(-0.392, 0.368) 0.392
$n = 1000$	(-0.290, 0.170) 0.290	(-0.247, 0.253) 0.253	(-0.289, 0.171) 0.289	(-0.238, 0.242) 0.242
	$m = 40$		$m = 50$	
	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ
$n = 100$	(-1.171, 0.829) 1.171	(-1.091, 1.049) 1.091	(-1.222, 0.838) 1.222	(-1.156, 1.084) 1.156
$n = 200$	(-0.796, 0.464) 0.796	(-0.714, 0.626) 0.714	(-0.719, 0.501) 0.719	(-0.632, 0.668) 0.668
$n = 500$	(-0.447, 0.253) 0.447	(-0.383, 0.357) 0.383	(-0.412, 0.268) 0.412	(-0.348, 0.372) 0.372
$n = 1000$	(-0.308, 0.172) 0.308	(-0.253, 0.247) 0.253	(-0.303, 0.177) 0.303	(-0.250, 0.250) 0.250

Table 4.3.7 Case 2: with $\beta_0 = 0$ and Evenly Chosen Observation Time Points

Ave. of $\hat{\beta}_n$ & $\tilde{\beta}_n$ (s.d.) for Case 2: $\beta_0 = 0$, t_{ij} evenly chosen

Number of Simulation Loop: 1000

	$n = 100$		$n = 200$	
	Ave. $\hat{\beta}_n$ (s.d.)	Ave. $\tilde{\beta}_n$ (s.d.)	Ave. $\hat{\beta}_n$ (s.d.)	Ave. $\tilde{\beta}_n$ (s.d.)
$m = 5$	1.330 (1.19)	1.561 (1.24)	1.031 (0.74)	1.216 (0.76)
$m = 10$	0.545 (0.99)	0.747 (1.08)	0.429 (0.63)	0.584 (0.68)
$m = 20$	0.204 (0.96)	0.374 (1.05)	0.151 (0.61)	0.294 (0.66)
$m = 30$	0.093 (0.96)	0.259 (1.05)	0.061 (0.60)	0.202 (0.65)
$m = 40$	0.034 (0.96)	0.196 (1.05)	0.019 (0.60)	0.156 (0.65)
$m = 50$	0.002 (0.97)	0.161 (1.05)	-0.007 (0.60)	0.127 (0.65)
Censoring %	32.8%		32.7%	
	$n = 500$		$n = 1000$	
	Ave. $\hat{\beta}_n$ (s.d.)	Ave. $\tilde{\beta}_n$ (s.d.)	Ave. $\hat{\beta}_n$ (s.d.)	Ave. $\tilde{\beta}_n$ (s.d.)
$m = 5$	0.849 (0.42)	0.977 (0.42)	0.801 (0.29)	0.904 (0.28)
$m = 10$	0.352 (0.36)	0.466 (0.38)	0.337 (0.24)	0.433 (0.26)
$m = 20$	0.130 (0.35)	0.226 (0.37)	0.135 (0.23)	0.213 (0.26)
$m = 30$	0.061 (0.34)	0.153 (0.37)	0.071 (0.23)	0.146 (0.25)
$m = 40$	0.022 (0.35)	0.111 (0.37)	0.039 (0.24)	0.109 (0.25)
$m = 50$	0.002 (0.35)	0.089 (0.37)	0.019 (0.24)	0.088 (0.25)
Censoring %	32.7%		32.6%	

Table 4.3.8 Case 2: Measurement of Coverage for Table 4.3.7

(Ave. $\hat{\beta}_n$) $\pm$ s.d. and (Ave. $\tilde{\beta}_n$) $\pm$ s.d. with Δ

	$m = 5$		$m = 10$	
	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ
$n = 100$	(0.140, 2.520) 2.520	(0.321, 2.801) 2.801	(-0.445, 1.535) 1.535	(-0.333, 1.827) 1.827
$n = 200$	(0.291, 1.771) 1.771	(0.456, 1.976) 1.976	(-0.201, 1.059) 1.059	(-0.096, 1.264) 1.264
$n = 500$	(0.429, 1.269) 1.269	(0.557, 1.397) 1.397	(-0.008, 0.712) 0.712	(0.086, 0.846) 0.846
$n = 1000$	(0.511, 1.091) 1.091	(0.624, 1.184) 1.184	(0.097, 0.577) 0.577	(0.173, 0.693) 0.693
	$m = 20$		$m = 30$	
	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ
$n = 100$	(-0.756, 1.164) 1.164	(-0.676, 1.424) 1.424	(-0.867, 1.053) 1.053	(-0.791, 1.309) 1.309
$n = 200$	(-0.459, 0.761) 0.761	(-0.366, 0.954) 0.954	(-0.539, 0.661) 0.661	(-0.448, 0.852) 0.852
$n = 500$	(-0.220, 0.480) 0.480	(-0.144, 0.596) 0.596	(-0.279, 0.401) 0.401	(-0.217, 0.523) 0.523
$n = 1000$	(-0.095, 0.365) 0.365	(-0.047, 0.473) 0.473	(-0.159, 0.301) 0.301	(-0.104, 0.396) 0.396
	$m = 40$		$m = 50$	
	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ
$n = 100$	(-0.926, 0.994) 0.994	(-0.854, 1.246) 1.246	(-0.968, 0.972) 0.972	(-0.889, 1.211) 1.211
$n = 200$	(-0.581, 0.619) 0.619	(-0.494, 0.806) 0.806	(-0.607, 0.593) 0.607	(-0.523, 0.777) 0.777
$n = 500$	(-0.328, 0.372) 0.372	(-0.259, 0.481) 0.481	(-0.348, 0.352) 0.352	(-0.281, 0.459) 0.459
$n = 1000$	(-0.201, 0.279) 0.279	(-0.141, 0.359) 0.359	(-0.221, 0.259) 0.259	(-0.162, 0.338) 0.338

Table 4.3.9 Case 2: with $\beta_0 = -1$ and Random Observation Time Points

Ave. of $\hat{\beta}_n$ & $\tilde{\beta}_n$ (s.d.) for Case 2: $\beta_0 = -1, t_{ij} \sim U(0, V_i)$

Number of Simulation Loop: 1000

	$n = 100$		$n = 200$	
	Ave. $\hat{\beta}_n$ (s.d.)	Ave. $\tilde{\beta}_n$ (s.d.)	Ave. $\hat{\beta}_n$ (s.d.)	Ave. $\tilde{\beta}_n$ (s.d.)
$m = 5$	-0.982 (0.82)	-0.931 (0.83)	-0.939 (0.52)	-0.909 (0.53)
$m = 10$	-1.095 (0.83)	-1.040 (0.84)	-1.022 (0.54)	-0.997 (0.55)
$m = 20$	-1.122 (0.89)	-1.070 (0.91)	-1.071 (0.52)	-1.044 (0.53)
$m = 30$	-1.144 (0.91)	-1.089 (0.94)	-1.031 (0.55)	-1.003 (0.56)
$m = 40$	-1.105 (0.86)	-1.052 (0.89)	-1.068 (0.57)	-1.040 (0.58)
$m = 50$	-1.222 (0.95)	-1.168 (0.97)	-1.091 (0.56)	-1.065 (0.57)
Censoring %	36.5%		36.4%	
	$n = 500$		$n = 1000$	
	Ave. $\hat{\beta}_n$ (s.d.)	Ave. $\tilde{\beta}_n$ (s.d.)	Ave. $\hat{\beta}_n$ (s.d.)	Ave. $\tilde{\beta}_n$ (s.d.)
$m = 5$	-0.912 (0.28)	-0.904 (0.28)	-0.895 (0.20)	-0.886 (0.20)
$m = 10$	-1.007 (0.32)	-0.997 (0.32)	-0.963 (0.21)	-0.953 (0.21)
$m = 20$	-0.987 (0.29)	-0.978 (0.29)	-0.996 (0.21)	-0.987 (0.21)
$m = 30$	-1.007 (0.30)	-0.997 (0.30)	-0.977 (0.21)	-0.968 (0.21)
$m = 40$	-1.005 (0.31)	-0.994 (0.31)	-0.993 (0.22)	-0.985 (0.21)
$m = 50$	-1.013 (0.30)	-1.003 (0.30)	-0.989 (0.22)	-0.979 (0.22)
Censoring %	36.4%		36.3%	

Table 4.3.10 Case 2: Measurement of Coverage for Table 4.3.9

(Ave. $\hat{\beta}_n$) $\pm$ s.d. and (Ave. $\tilde{\beta}_n$) $\pm$ s.d. with Δ

	$m = 5$		$m = 10$	
	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ
$n = 100$	(-1.802, -0.162) 0.838	(-1.761, -0.101) 0.899	(-1.925, -0.265) 0.925	(-1.880, -0.200) 0.880
$n = 200$	(-1.459, -0.419) 0.581	(-1.439, -0.379) 0.621	(-1.562, -0.482) 0.562	(-1.547, -0.447) 0.553
$n = 500$	(-1.192, -0.632) 0.368	(-1.184, -0.624) 0.376	(-1.327, -0.687) 0.327	(-1.317, -0.677) 0.323
$n = 1000$	(-1.095, -0.695) 0.305	(-1.086, -0.686) 0.314	(-1.173, -0.753) 0.247	(-1.163, -0.743) 0.257
	$m = 20$		$m = 30$	
	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ
$n = 100$	(-2.012, -0.232) 1.012	(-1.980, -0.160) 0.980	(-2.054, -0.234) 1.054	(-2.029, -0.149) 1.029
$n = 200$	(-1.591, -0.551) 0.591	(-1.574, -0.514) 0.574	(-1.581, -0.481) 0.581	(-1.563, -0.443) 0.563
$n = 500$	(-1.277, -0.697) 0.303	(-1.268, -0.688) 0.312	(-1.307, -0.707) 0.307	(-1.297, -0.697) 0.303
$n = 1000$	(-1.206, -0.786) 0.214	(-1.197, -0.777) 0.223	(-1.187, -0.767) 0.233	(-1.178, -0.758) 0.242
	$m = 40$		$m = 50$	
	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ
$n = 100$	(-1.965, -0.245) 0.965	(-1.942, -0.162) 0.942	(-2.172, -0.272) 1.172	(-2.138, -0.198) 1.138
$n = 200$	(-1.638, -0.498) 0.638	(-1.620, -0.460) 0.620	(-1.651, -0.531) 0.651	(-1.635, -0.495) 0.635
$n = 500$	(-1.315, -0.695) 0.315	(-1.304, -0.684) 0.316	(-1.313, -0.713) 0.313	(-1.303, -0.703) 0.303
$n = 1000$	(-1.213, -0.773) 0.227	(-1.195, -0.775) 0.225	(-1.209, -0.769) 0.231	(-1.199, -0.759) 0.241

Table 4.3.11 Case 2: with $\beta_0 = -1$ and Evenly Chosen Observation Time Points

Ave. of $\hat{\beta}_n$ & $\tilde{\beta}_n$ (s.d.) for Case 2: $\beta_0 = -1$, t_{ij} evenly chosen

Number of Simulation Loop: 1000

	$n = 100$		$n = 200$	
	Ave. $\hat{\beta}_n$ (s.d.)	Ave. $\tilde{\beta}_n$ (s.d.)	Ave. $\hat{\beta}_n$ (s.d.)	Ave. $\tilde{\beta}_n$ (s.d.)
$m = 5$	-0.099 (0.88)	-0.011 (0.94)	-0.249 (0.50)	-0.201 (0.54)
$m = 10$	-0.599 (1.02)	-0.559 (0.88)	-0.654 (0.51)	-0.624 (0.53)
$m = 20$	-0.860 (1.00)	-0.821 (0.89)	-0.843 (0.51)	-0.817 (0.53)
$m = 30$	-0.940 (0.98)	-0.900 (0.88)	-0.913 (0.52)	-0.886 (0.53)
$m = 40$	-0.981 (0.99)	-0.947 (0.89)	-0.945 (0.52)	-0.919 (0.53)
$m = 50$	-1.013 (0.99)	-0.974 (0.90)	-0.966 (0.53)	-0.939 (0.54)
Censoring %	36.5%		36.4%	
	$n = 500$		$n = 1000$	
	Ave. $\hat{\beta}_n$ (s.d.)	Ave. $\tilde{\beta}_n$ (s.d.)	Ave. $\hat{\beta}_n$ (s.d.)	Ave. $\tilde{\beta}_n$ (s.d.)
$m = 5$	-0.331 (0.28)	-0.318 (0.29)	-0.353 (0.20)	-0.346 (0.20)
$m = 10$	-0.676 (0.29)	-0.666 (0.29)	-0.681 (0.19)	-0.674 (0.19)
$m = 20$	-0.847 (0.30)	-0.837 (0.30)	-0.835 (0.20)	-0.826 (0.20)
$m = 30$	-0.897 (0.30)	-0.887 (0.30)	-0.886 (0.20)	-0.878 (0.20)
$m = 40$	-0.927 (0.30)	-0.919 (0.30)	-0.913 (0.21)	-0.904 (0.20)
$m = 50$	-0.945 (0.31)	-0.935 (0.31)	-0.931 (0.21)	-0.924 (0.21)
Censoring %	36.4%		36.3%	

Table 4.3.12 Case 2: Measurement of Coverage for Table 4.3.11

(Ave. $\hat{\beta}_n$) $\pm$ s.d. and (Ave. $\tilde{\beta}_n$) $\pm$ s.d. with Δ

	$m = 5$		$m = 10$	
	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ
$n = 100$	(-0.979, 0.781) 1.781	(-0.951, 0.929) 1.929	(-1.619, 0.421) 1.421	(-1.439, 0.321) 1.321
$n = 200$	(-0.749, 0.251) 1.251	(-0.741, 0.339) 1.339	(-1.164, -0.144) 0.856	(-1.154, -0.094) 0.906
$n = 500$	(-0.611, -0.051) 0.949	(-0.608, -0.028) 0.972	(-0.966, -0.386) 0.614	(-0.956, -0.376) 0.624
$n = 1000$	(-0.553, -0.153) 0.847	(-0.546, -0.146) 0.854	(-0.871, -0.491) 0.509	(-0.864, -0.484) 0.516
	$m = 20$		$m = 30$	
	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ
$n = 100$	(-1.860, 0.140) 1.140	(-1.711, 0.069) 1.069	(-1.920, 0.040) 1.040	(-1.780, -0.020) 0.980
$n = 200$	(-1.353, -0.333) 0.667	(-1.347, -0.287) 0.713	(-1.433, -0.393) 0.607	(-1.416, -0.356) 0.644
$n = 500$	(-1.147, -0.547) 0.453	(-1.137, -0.537) 0.463	(-1.197, -0.597) 0.403	(-1.187, -0.587) 0.413
$n = 1000$	(-1.035, -0.635) 0.365	(-1.026, -0.626) 0.374	(-1.086, -0.686) 0.314	(-1.078, -0.678) 0.322
	$m = 40$		$m = 50$	
	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ	(Ave. $\hat{\beta}_n \pm$ s.d.) Δ	(Ave. $\tilde{\beta}_n \pm$ s.d.) Δ
$n = 100$	(-1.971, 0.009) 1.009	(-1.837, -0.057) 0.943	(-2.003, -0.023) 1.003	(-1.874, -0.074) 0.926
$n = 200$	(-1.465, -0.425) 0.575	(-1.449, -0.389) 0.611	(-1.496, -0.436) 0.564	(-1.479, -0.399) 0.601
$n = 500$	(-1.227, -0.627) 0.373	(-1.219, -0.619) 0.381	(-1.255, -0.635) 0.365	(-1.245, -0.625) 0.375
$n = 1000$	(-1.123, -0.703) 0.297	(-1.104, -0.704) 0.296	(-1.141, -0.721) 0.279	(-1.134, -0.714) 0.286

4.4 Remarks and Discussions

From Tables 4.2.1-4.2.12 and Tables 4.3.1-4.3.12, we see that right censoring percentage is high, but the performance of our algorithm is very stable. Both estimators $\hat{\beta}_n$ and $\tilde{\beta}_n$ perform well and $\hat{\beta}_n$ performs better than $\tilde{\beta}_n$ in most cases. The performances improve as the sample size n increases which is expected.

In Tables 4.2.2, 4.2.6 and 4.2.10, for Case 1 with $t_{ij} \sim U(0, V_i)$, from the values of Δ we see that the performances of $\hat{\beta}_n$ and $\tilde{\beta}_n$ are not always improving as m increases, but they are fine when m is around $m = 30$ or $m = 40$. This indicates that the performances of $\hat{\beta}_n$ and $\tilde{\beta}_n$ improve as m increases in general.

For the case that time points t_{ij} are evenly chosen, we see that $\hat{\beta}_n$ and $\tilde{\beta}_n$ clearly improve as m increases. This makes sense because the bigger the m is, the more time points we observed on the longitudinal covariates that evenly capture the covariate process $X(t)$ over interval $[0, t]$. This way the observed longitudinal covariate is more clear and more informative compared to $t_{ij} \sim U(0, V_i)$, because for $t_{ij} \sim U(0, V_i)$, time interval length is random.

For Case 2, simulation results are similar to Case 1. We observe that our simulation studies show that without parametric or semi-parametric assumptions on the longitudinal covariate $X(t)$, our nonparametric MLE still performs well for large n and m.

Chapter 5: Real Data Analysis

This chapter considers analyzing real data sets from smoking cessation and liver disease. Since there are ties among V_i's in (1.29) for real data samples, Section 5.1 derives the profile estimating function $\psi_n(\beta)$, approximated estimating function $\varphi_n(\beta)$ and adjusted estimating function $\psi_M(\beta)$ for the case that there are ties among V_i's. Section 5.2 performs data analysis for a set of Smoking Cessation Data, and Section 5.3 performs data analysis for a set of Liver Disease Data.

5.1 Algorithm for Observed Survival Time with Ties

In real data analysis, often there are ties among V_i's in (1.29) due to measurement error or round off error. Thus, the algorithm given in Chapter 4 cannot be directly used in real data analysis. To deal with this issue, let

$$U_1 < U_2 < \ldots < U_N \tag{5.1}$$

be all the distinct observations of $V_1, \ldots, V_n$, where $N \leq n$. From (3.16), we have

$$L(\beta, F) = \prod_{k=1}^{N} \prod_{V_i = U_k} \left(c_i dF(V_i)\right)^{\delta_i} \left(\bar{F}(V_i)\right)^{c_i - \delta_i} \left(\prod_{j=1}^{J_i} \left(\bar{F}(t_{ij})\right)^{d_{ij}} \right). \tag{5.2}$$

From (5.2) and (3.17), let

$$\tilde{\delta}_k = \sum_{i=1}^{n} \delta_i I\{V_i = U_k\}, \tag{5.3}$$

then

$$\begin{aligned}
\prod_{k=1}^{N} \prod_{V_i=U_k} \big(c_i dF(V_i)\big)^{\delta_i} &= \prod_{k=1}^{N} \Big(\prod_{V_i=U_k} c_i^{\delta_i} \Big) \Big(\prod_{V_i=U_k} \big(dF(V_i)\big)^{\delta_i} \Big) \\
&= \prod_{k=1}^{N} \Big(\prod_{V_i=U_k} e^{\beta Z_i(t_{iJ_i})\delta_i} \Big) \big(dF(U_k)\big)^{\tilde{\delta}_k} \\
&= \prod_{k=1}^{N} \big(e^{\beta \sum_{i=1}^{n} Z_i(t_{iJ_i})\delta_i I\{V_i=U_k\}} \big) \big(dF(U_k)\big)^{\tilde{\delta}_k} = \prod_{k=1}^{N} \big(\tilde{c}_k dF(U_k)\big)^{\tilde{\delta}_k},
\end{aligned} \tag{5.4}$$

where

$$\tilde{c}_k = \begin{cases} 1 & \text{if } \tilde{\delta}_k = 0 \\ e^{\beta \tilde{\delta}_k^{-1} \sum_{i=1}^{n} Z_i(t_{iJ_i})\delta_i I\{V_i=U_k\}} & \text{if } \tilde{\delta}_k > 0 \end{cases}. \tag{5.5}$$

Also, we have

$$\prod_{k=1}^{N} \prod_{V_i=U_k} \big(\bar{F}(V_i)\big)^{c_i-\delta_i} = \prod_{k=1}^{N} \bar{F}(U_k)^{\sum_{i=1}^{n}(c_i-\delta_i)I\{V_i=U_k\}}. \tag{5.6}$$

If we restrict all possible candidates for the MLE of F_0 to those d.f.'s that assign all their probability masses to points U_k's and interval (U_N, ∞), then the d.f. F under consideration in likelihood function (5.2) is given by

$$F(t) = \sum_{k=1}^{N} p_k I\{U_k \le t\}, \tag{5.7}$$

where $p_k = F(U_k) - F(U_k-)$ for $1 \le k \le N$, and $\sum_{l=1}^{N+1} p_k = 1$ with $0 \le p_{N+1} \le 1$ as the probability mass of F on interval (U_N, ∞). For any d.f. F given by (5.7), we let

$$\alpha_{kij} = I\{U_k \le t_{ij} < U_{k+1}\}, \quad U_{N+1} = \infty, \tag{5.8}$$

then $F(t_{ij}) = 0$ for $0 < t_{ij} < U_1$, and we have all $t_{ij} \le U_k \le U_n$ for $V_i = U_k$, thus we have in (5.2):

$$\begin{aligned}
&\prod_{k=1}^{N}\prod_{V_i=U_k}\prod_{j=1}^{J_i}(\bar{F}(t_{ij}))^{d_{ij}} = \prod_{k=1}^{N}\prod_{i=1}^{n}\left\{\prod_{j=1}^{J_i}[\bar{F}(t_{ij})]^{d_{ij}}\right\}^{I\{V_i=U_k\}} \\
&=\prod_{k=1}^{N}\prod_{i=1}^{n}\prod_{j=1}^{J_i}[\bar{F}(t_{ij})]^{d_{ij}I\{V_i=U_k\}} = \prod_{k=1}^{N}\prod_{i=1}^{n}\prod_{j=1}^{J_i}\prod_{l=1}^{N}(\bar{F}(t_{ij}))^{\alpha_{lij}d_{ij}I\{V_i=U_k\}} \\
&=\prod_{k=1}^{N}\prod_{i=1}^{n}\prod_{j=1}^{J_i}\prod_{l=1}^{N}(\bar{F}(U_l))^{\alpha_{lij}d_{ij}I\{V_i=U_k\}} = \prod_{l=1}^{N}\prod_{k=1}^{N}\prod_{i=1}^{n}\prod_{j=1}^{J_i}(\bar{F}(U_l))^{\alpha_{lij}d_{ij}I\{V_i=U_k\}} \\
&=\prod_{l=1}^{N}\prod_{k=1}^{N}(\bar{F}(U_l))^{\sum_{i=1}^{n}\sum_{j=1}^{J_i}\alpha_{lij}d_{ij}I\{V_i=U_k\}} = \prod_{l=1}^{N}(\bar{F}(U_l))^{\tilde{d}_{0l}} = \prod_{k=1}^{N}(\bar{F}(U_k))^{\tilde{d}_{0k}},
\end{aligned} \tag{5.9}$$

where for d_{ij} given in (3.17), we have that from $\sum_{l=1}^{N} I\{V_i = U_l\} = 1$,

$$\begin{aligned}
\tilde{d}_{0k} &= \sum_{l=1}^{N}\sum_{i=1}^{n}\sum_{j=1}^{J_i}\alpha_{kij}d_{ij}I\{V_i = U_l\} = \sum_{i=1}^{n}\sum_{j=1}^{J_i}\Big(\sum_{l=1}^{N} I\{V_i = U_l\}\Big)\alpha_{kij}d_{ij} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{J_i}(e^{\beta Z_i(t_{i,j-1})} - e^{\beta Z_i(t_{ij})})I\{U_k \le t_{ij} < U_{k+1}\}.
\end{aligned} \tag{5.10}$$

From (5.4), (5.6) and (5.9), we write (5.2) as below:

$$\begin{aligned}
L(\beta, F) &= \prod_{k=1}^{N}\big(\tilde{c}_k dF(U_k)\big)^{\tilde{\delta}_k}\bar{F}(U_k)^{\sum_{i=1}^{n}(c_i-\delta_i)I\{V_i=U_k\}}\bar{F}(U_k)^{\tilde{d}_{0k}} \\
&= \prod_{k=1}^{N}\big(\tilde{c}_k dF(U_k)\big)^{\tilde{\delta}_k}\big(\bar{F}(U_k)\big)^{\sum_{i=1}^{n} c_i I\{V_i=U_k\}+\tilde{d}_{0k}-\sum_{i=1}^{n}\delta_i I\{V_i=U_k\}} \\
&\overset{(5.3)}{=\!=} \prod_{k=1}^{N}\big(\tilde{c}_k dF(U_k)\big)^{\tilde{\delta}_k}\big(\bar{F}(U_k)\big)^{\tilde{d}_k-\tilde{\delta}_k},
\end{aligned} \tag{5.11}$$

where $\tilde{\delta}_k$ and $\tilde{c}_k$ are given in (5.3) and (5.5), respectively, and by (5.10)

$$\begin{aligned}
\tilde{d}_k &= \sum_{i=1}^{n} c_i I\{V_i = U_k\} + \tilde{d}_{0k} \\
&\overset{(3.17)}{=\!=} \sum_{i=1}^{n} e^{\beta Z_i(t_{iJ_i})} I\{V_i = U_k\} + \sum_{i=1}^{n}\sum_{j=1}^{J_i} \alpha_{kij} d_{ij} \\
&= \sum_{i=1}^{n} e^{\beta Z_i(t_{iJ_i})} I\{V_i = U_k\} + \sum_{i=1}^{n}\sum_{j=1}^{J_i} \left(e^{\beta Z_i(t_{i,j-1})} - e^{\beta Z_i(t_{ij})}\right) I\{U_k \le t_{ij} < U_{k+1}\}.
\end{aligned} \tag{5.12}$$

Note that (5.11) is the formula corresponding to (3.22). Thus, for any d.f. F given by (5.7), likelihood function (5.11) can be written as

$$L(\beta, F) = \prod_{k=1}^{N} \left(\tilde{c}_k p_k\right)^{\tilde{\delta}_k} \left(\sum_{j=k+1}^{N+1} p_j\right)^{\tilde{d}_k - \tilde{\delta}_k} \equiv L(\beta, \boldsymbol{p}), \tag{5.13}$$

where $\boldsymbol{p} = (p_1, \ldots, p_{N+1})$, $\tilde{\delta}_k$, $\tilde{c}_k$ and $\tilde{d}_k$ are given in (5.3), (5.5) and (5.12), respectively. The MLE $(\hat{\beta}_n, \hat{F}_n)$ for (β_0, F_0) in our proposed PH model (1.38) is the solution that maximizes $L(\beta, F)$ in (5.11) or (5.13).

From the derivation of (3.28), we assume that in (5.11) for some β:

$$\tilde{d}_k \ge \tilde{\delta}_k \quad \text{for all } 1 \le k \le N, \tag{AS4}$$

then likelihood function $L(\beta, F)$ in (5.13) is maximized by

$$[1 - \hat{F}_n(t; \beta)] = \prod_{U_k \le t} \frac{\tilde{e}_k - \tilde{\delta}_k}{\tilde{e}_k}, \tag{5.14}$$

where from (3.28) regarding e_k, $\tilde{e}_k$ is given by:

$$\tilde{e}_k = \tilde{d}_k + \ldots + \tilde{d}_N \geq \tilde{\delta}_k, \tag{5.15}$$

for $\tilde{d}_k$'s given by (5.12).

Consider any fixed β satisfying (AS4), for likelihood function $L(\beta, \boldsymbol{p})$ in (5.13), from (3.25), we let

$$a_k = \frac{p_k}{b_k}, \quad b_k = \sum_{j=k}^{N+1} p_j, \tag{5.16}$$

then we have

$$b_1 = 1, \quad b_{N+1} = p_{N+1}, \quad b_{k+1} = b_k - p_k, \quad 1 - a_k = \frac{b_{k+1}}{b_k}. \tag{5.17}$$

Note that

$$\prod_{k=1}^{N}(1-a_k) \xlongequal{(5.17)} \prod_{k=1}^{N} \frac{b_{k+1}}{b_k} = \frac{b_2}{b_1}\frac{b_3}{b_2}\ldots\frac{b_{N+1}}{b_N} = b_{N+1}, \tag{5.18}$$

and

$$\begin{aligned}
&\prod_{k=1}^{N}(a_k)^{\tilde{d}_k}(1-a_k)^{N-\tilde{h}_k} \xlongequal{(5.17)} \left(\prod_{k=1}^{N}(a_k)^{\tilde{d}_k}\right)\prod_{k=1}^{N}\left(\frac{b_{k+1}}{b_k}\right)^{N-\tilde{h}_k} \\
&= \left(\prod_{k=1}^{N}(a_k)^{\tilde{d}_k}\right)\left(\frac{b_2}{b_1}\right)^{N-\tilde{h}_1}\left(\frac{b_3}{b_2}\right)^{N-\tilde{h}_2}\ldots\left(\frac{b_{N+1}}{b_N}\right)^{N-\tilde{h}_N} \\
&= \left(\prod_{k=1}^{N}(a_k)^{\tilde{d}_k}\right)(b_2)^{\tilde{h}_2-\tilde{h}_1}(b_3)^{\tilde{h}_3-\tilde{h}_2}\ldots(b_N)^{\tilde{h}_N-\tilde{h}_{N-1}}(b_{N+1})^{N-\tilde{h}_N} \\
&= \left(\prod_{k=1}^{N}(a_k)^{\tilde{d}_k}\right)(b_1)^{\tilde{d}_1}(b_2)^{\tilde{d}_2}\ldots(b_N)^{\tilde{d}_N}(b_{N+1})^{N-\tilde{h}_N} = \left(\prod_{k=1}^{N}(a_k)^{\tilde{d}_k}(b_k)^{\tilde{d}_k}\right)(b_{N+1})^{N-N\bar{\tilde{d}}} \\
&\xlongequal{(5.16)} (b_{N+1})^{N-N\bar{\tilde{d}}}\prod_{k=1}^{N}(p_k)^{\tilde{d}_k},
\end{aligned} \tag{5.19}$$

where $\tilde{h}_k = \tilde{d}_1 + \tilde{d}_2 + \ldots + \tilde{d}_k$ and $\bar{\tilde{d}} = N^{-1}\sum_{k=1}^{N}\tilde{d}_k$. From (5.18) and (5.19), we can write likelihood function $L(\beta, F) = L(\beta, \boldsymbol{p})$ in (5.13) as

$$
\begin{aligned}
L(\beta,\boldsymbol{p}) &\xlongequal{(5.16)} \prod_{k=1}^{N} \left(\tilde{c}_k p_k\right)^{\tilde{\delta}_k} \left(b_{k+1}\right)^{\tilde{d}_k-\tilde{\delta}_k} = \prod_{k=1}^{N} \left((\tilde{c}_k)^{\tilde{\delta}_k}(p_k)^{\tilde{d}_k}\right)\left(\frac{b_{k+1}}{p_k}\right)^{\tilde{d}_k-\tilde{\delta}_k} \\
&\xlongequal{(5.16)} \prod_{k=1}^{N} \left((\tilde{c}_k)^{\tilde{\delta}_k}(p_k)^{\tilde{d}_k}\right)\left(\frac{b_{k+1}}{a_k b_k}\right)^{\tilde{d}_k-\tilde{\delta}_k} \xlongequal{(5.17)} \prod_{k=1}^{N} \left((\tilde{c}_k)^{\tilde{\delta}_k}(p_k)^{\tilde{d}_k}\right)\left(\frac{1-a_k}{a_k}\right)^{\tilde{d}_k-\tilde{\delta}_k} \\
&= \left(\prod_{k=1}^{N}(\tilde{c}_k)^{\tilde{\delta}_k}(p_k)^{\tilde{d}_k}\right)\frac{\prod_{k=1}^{N}(a_k)^{\tilde{\delta}_k}(1-a_k)^{N-\tilde{\delta}_k-(\tilde{d}_1+\ldots+\tilde{d}_{k-1})}}{\prod_{k=1}^{N}(a_k)^{\tilde{d}_k}(1-a_k)^{N-(\tilde{d}_1+\ldots+\tilde{d}_k)}} \\
&\xlongequal{(5.19)} \left(\prod_{k=1}^{N}(\tilde{c}_k)^{\tilde{\delta}_k}(p_k)^{\tilde{d}_k}\right)\frac{\prod_{k=1}^{N}(a_k)^{\tilde{\delta}_k}(1-a_k)^{N-\tilde{\delta}_k-(\tilde{d}_1+\ldots+\tilde{d}_{k-1})}}{(b_{N+1})^{N-N\bar{\tilde{d}}}\prod_{k=1}^{N}(p_k)^{\tilde{d}_k}} \\
&\xlongequal{(5.18)} \left(\prod_{k=1}^{N}(\tilde{c}_k)^{\tilde{\delta}_k}\right)\frac{\prod_{k=1}^{N}(a_k)^{\tilde{\delta}_k}(1-a_k)^{N-\tilde{\delta}_k-(\tilde{d}_1+\ldots+\tilde{d}_{k-1})}}{\prod_{k=1}^{N}(1-a_k)^{N-N\bar{\tilde{d}}}} \\
&= \prod_{k=1}^{N}(\tilde{c}_k)^{\tilde{\delta}_k}(a_k)^{\tilde{\delta}_k}(1-a_k)^{\tilde{d}_k+\ldots+\tilde{d}_N-\tilde{\delta}_k} \\
&= \prod_{k=1}^{N}(\tilde{c}_k a_k)^{\tilde{\delta}_k}(1-a_k)^{\tilde{e}_k-\tilde{\delta}_k} \equiv L_1(\boldsymbol{a};\beta),
\end{aligned} \tag{5.20}
$$

where $\boldsymbol{a} = (a_1,\ldots,a_N)$ and $\tilde{e}_k = \tilde{d}_k + \ldots + \tilde{d}_N$, then taking logarithm we obtain:

$$
\log L_1(\boldsymbol{a};\beta) = \sum_{k=1}^{N}[\tilde{\delta}_k(\log\tilde{c}_k + \log a_k) + (\tilde{e}_k - \tilde{\delta}_k)\log(1-a_k)].
$$

Setting $\frac{\partial \log L_1}{\partial a_k} = 0$, we obtain

$$
\frac{\tilde{\delta}_k}{a_k} - \frac{\tilde{e}_k - \tilde{\delta}_k}{1-a_k} = 0 \quad \Longrightarrow \quad \hat{a}_k = \frac{\tilde{\delta}_k}{\tilde{e}_k}, \quad k = 1,2,\ldots,N. \tag{5.21}
$$

Also, from (5.15), we have

$$
\frac{\partial^2(\log L_1)}{\partial a_j \partial a_k} = \begin{cases} 0 & \text{if} \quad k \neq j \\ -\frac{\tilde{\delta}_k}{a_k^2} - \frac{\tilde{e}_k-\tilde{\delta}_k}{(1-a_k)^2} < 0 & \text{if} \quad k = j \end{cases},
$$

thus $\hat{a}_k$ in (5.21) maximizes $L_1(\boldsymbol{a};\beta)$ in (5.20) and all $0 \le \hat{a}_i \le 1$.

From (5.20)-(5.21), $L(\beta, \boldsymbol{p})$ in (5.20) can be written as

$$L(\beta, \hat{F}_n(\cdot\,;\beta)) = L(\beta, \hat{\boldsymbol{p}}) = L_1(\hat{\boldsymbol{a}};\beta) = \prod_{k=1}^{N} \left(\tilde{c}_k \frac{\tilde{\delta}_k}{\tilde{e}_k}\right)^{\tilde{\delta}_k} \left(1 - \frac{\tilde{\delta}_k}{\tilde{e}_k}\right)^{\tilde{e}_k - \tilde{\delta}_k}, \tag{5.22}$$

which gives the profile likelihood function for β_0 :

$$\begin{aligned} \ell(\beta) &= \log L(\beta, \hat{F}_n(\cdot\,;\beta)) = \log \left(\prod_{k=1}^{N} \left(\tilde{c}_k \frac{\tilde{\delta}_k}{\tilde{e}_k}\right)^{\tilde{\delta}_k} \left(\frac{\tilde{e}_k - \tilde{\delta}_k}{\tilde{e}_k}\right)^{\tilde{e}_k - \tilde{\delta}_k} \right) \\ &= \sum_{k=1}^{N} \left(\tilde{\delta}_k [\log \tilde{c}_k + \log \tilde{\delta}_k - \log \tilde{e}_k] + (\tilde{e}_k - \tilde{\delta}_k) \big[\log(\tilde{e}_k - \tilde{\delta}_k) - \log \tilde{e}_k \big] \right), \end{aligned} \tag{5.23}$$

where $\tilde{\delta}_k \log \tilde{\delta}_k = 0$ if $\tilde{\delta}_k = 0$. Notice that above $\tilde{c}_k$'s, $\tilde{d}_k$'s and $\tilde{e}_k$'s are given by (5.5), (5.12) and (5.15), respectively, thus they are functions of β. With notations $\tilde{c}'_k = \frac{d\tilde{c}_k}{d\beta}$, $\tilde{d}'_k = \frac{d\tilde{d}_k}{d\beta}$ and $\tilde{e}'_k = \frac{d\tilde{e}_k}{d\beta}$, we know

$$\frac{\tilde{c}'_k}{\tilde{c}_k} \overset{(5.5)}{=\!=} \begin{cases} 0 & \text{if } \tilde{\delta}_k = 0 \\ \tilde{\delta}_k^{-1} \sum_{i=1}^{n} Z_i(t_{iJ_i}) \delta_i I\{V_i = U_k\} & \text{if } \tilde{\delta}_k > 0 \end{cases} = \begin{cases} 0 & \text{if } \tilde{\delta}_k = 0 \\ \tilde{\delta}_k^{-1} \tilde{Z}_k & \text{if } \tilde{\delta}_k > 0 \end{cases}, \tag{5.24}$$

where

$$\tilde{Z}_k = \sum_{i=1}^{n} Z_i(t_{iJ_i}) \delta_i I\{V_i = U_k\}. \tag{5.25}$$

Then we compute profile likelihood estimating function $\psi_n(\beta)$ as follows:

$$
\begin{aligned}
\psi_n(\beta) &\equiv n^{-1}\ell'(\beta) \\
&= n^{-1}\sum_{k=1}^{N}\left(\tilde{\delta}_k\left[\frac{\tilde{c}_k'}{\tilde{c}_k}-\frac{\tilde{e}_k'}{\tilde{e}_k}\right]+\tilde{e}_k'\log\frac{\tilde{e}_k-\tilde{\delta}_k}{\tilde{e}_k}+(\tilde{e}_k-\tilde{\delta}_k)\left[\frac{\tilde{e}_k'}{\tilde{e}_k-\tilde{\delta}_k}-\frac{\tilde{e}_k'}{\tilde{e}_k}\right]\right) \\
&= n^{-1}\sum_{k=1}^{N}\left(\tilde{\delta}_k\frac{\tilde{c}_k'}{\tilde{c}_k}+\tilde{e}_k'\log\frac{\tilde{e}_k-\tilde{\delta}_k}{\tilde{e}_k}\right)\overset{(5.24)}{=\!=}n^{-1}\sum_{k=1}^{N}\left(I\{\tilde{\delta}_k>0\}\tilde{Z}_k+\tilde{e}_k'\log\frac{\tilde{e}_k-\tilde{\delta}_k}{\tilde{e}_k}\right).
\end{aligned}
$$

Thus, the profile likelihood estimating function for V_i's with ties is given by:

$$
\psi_n(\beta)=n^{-1}\sum_{k=1}^{N}\left(I\{\tilde{\delta}_k>0\}\tilde{Z}_k+\tilde{e}_k'\log\frac{\tilde{e}_k-\tilde{\delta}_k}{\tilde{e}_k}\right), \tag{5.26}
$$

which is corresponding to (3.32) without ties among V_i's. From (5.3), (5.12), (5.15) and (5.25), we have

$$
\begin{cases}
\tilde{\delta}_k = \sum_{i=1}^{n}\delta_i I\{V_i=U_k\} \\
\tilde{Z}_k = \sum_{i=1}^{n} Z_i(t_{iJ_i})\delta_i I\{V_i=U_k\} \\
\tilde{e}_k = \sum_{q=k}^{N}\Big(\sum_{i=1}^{n} e^{\beta Z_i(t_{iJ_i})} I\{V_i=U_q\} \\
\qquad\qquad + \sum_{i=1}^{n}\sum_{j=1}^{J_i}\left[e^{\beta Z_i(t_{i,j-1})}-e^{\beta Z_i(t_{ij})}\right] I\{U_q\le t_{ij}<U_{q+1}\}\Big) \\
\tilde{e}_k' = \sum_{q=k}^{N}\Big(\sum_{i=1}^{n} Z_i(t_{iJ_i})e^{\beta Z_i(t_{iJ_i})} I\{V_i=U_q\} \\
\qquad\qquad + \sum_{i=1}^{n}\sum_{j=1}^{J_i}\left[Z_i(t_{i,j-1})e^{\beta Z_i(t_{i,j-1})}-Z_i(t_{ij})e^{\beta Z_i(t_{ij})}\right] I\{U_q\le t_{ij}<U_{q+1}\}\Big)
\end{cases}. \tag{5.27}
$$

The solution of above estimating equation $\psi_n(\beta)=0$ is the MLE $\hat{\beta}_n$ of β_0 in the situation when there are ties among V_i's in data (1.29).

Similar to (3.34), consider Taylor's expansion of $\log(1+x)$ at $x=0$:

$$\log(1+x) = x - \frac{x^2}{2(1+\xi)^2},$$

where ξ is between 0 and x. Thus, from Taylor's expansion we have in (5.26)

$$\begin{aligned} \psi_n(\beta) &= n^{-1} \sum_{k=1}^{N} \left(I\{\tilde{\delta}_k > 0\} \tilde{Z}_k + \tilde{e}_k' \left(-\frac{\tilde{\delta}_k}{\tilde{e}_k} - \frac{\tilde{\delta}_k^2}{2(1+\xi_k)^2 \tilde{e}_k^2} \right) \right) \\ &= \varphi_n(\beta) - n^{-1} \sum_{k=1}^{N} \frac{\tilde{\delta}_k^2 \tilde{e}_k'}{2(1+\xi_k)^2 \tilde{e}_k^2} \end{aligned} \tag{5.28}$$

where ξ_k is between 0 and $-\frac{\tilde{\delta}_k}{\tilde{e}_k}$, and $\varphi_n(\beta)$ is the leading term of Taylor's expansion of $\psi_n(\beta)$ in (5.26) and is given by:

$$\varphi_n(\beta) = n^{-1} \sum_{k=1}^{N} \left(I\{\tilde{\delta}_k > 0\} \tilde{Z}_k - \frac{\tilde{e}_k' \tilde{\delta}_k}{\tilde{e}_k} \right), \tag{5.29}$$

where $\tilde{Z}_k$, $\tilde{\delta}_k$, $\tilde{e}_k$ and $\tilde{e}_k'$ are given in (5.27).

The adjusted estimating equation corresponding to (3.60) is given by:

$$\psi_M(\beta) = n^{-1} \sum_{k=1}^{N} \left(I\{\tilde{\delta}_k > 0\} \tilde{Z}_k + \tilde{E}_k' \log \frac{\tilde{E}_k - \tilde{\delta}_k}{\tilde{E}_k} \right), \tag{5.30}$$

where similar to (3.61) we have for $\tilde{Z}_k$'s, $\tilde{\delta}_k$'s, $\tilde{e}_k$'s and $\tilde{e}_k'$'s given in (5.27),

$$\tilde{E}_k = e^M \tilde{e}_k \quad \text{and} \quad \tilde{E}_k' = e^M \tilde{e}_k', \tag{5.31}$$

which is to ensure $\tilde{E}_k = e^M \tilde{e}_k \geq \tilde{\delta}_k$ for all $k = 1, \ldots, N$ as in (5.15). The solution of the estimating equation $\psi_M(\beta) = 0$ is the MLE $\hat{\beta}_n$ of β_0 when there are ties among V_i's in data (1.29).

Simplification of $\tilde{e}_k$

Note: In (5.27), let

$$n_k = \min\{i \mid V_i \geq U_k\}, \quad k = 1, 2, \ldots, N; \tag{5.32}$$

then from

$$U_k = V_{n_k} \tag{5.33}$$

and

$$\begin{aligned}
\sum_{q=k}^{N} I\{V_i = U_q\} &= I\{V_i = U_k\} + I\{V_i = U_{k+1}\} + \ldots + I\{V_i = U_N\} \\
&= I\{V_i = U_k \text{ or } V_i = U_{k+1} \text{ or } \ldots V_i = U_N\} \\
&\overset{(5.1)}{=\!=} I\{V_i \geq U_k\} \overset{(5.33)}{=\!=} I\{V_i \geq V_{n_k}\},
\end{aligned} \tag{5.34}$$

we can write the first part of $\tilde{e}_k$ in (5.27) as:

$$\begin{aligned}
\sum_{q=k}^{N} \sum_{i=1}^{n} e^{\beta Z_i(t_{iJ_i})} I\{V_i = U_q\} &= \sum_{i=1}^{n} \Big(\sum_{q=k}^{N} I\{V_i = U_q\} \Big) e^{\beta Z_i(t_{iJ_i})} \\
&= \sum_{i=1}^{n} I\{V_i \geq V_{n_k}\} e^{\beta Z_i(t_{iJ_i})} = \sum_{i=n_k}^{n} e^{\beta Z_i(t_{iJ_i})}.
\end{aligned} \tag{5.35}$$

Before simplifying the second part of $\tilde{e}_k$ in (5.27), we first prove:

$$\sum_{l=n_k}^{n} I\{V_l \leq t_{ij} < V_{l+1}\} = \sum_{q=k}^{N} I\{U_q \leq t_{ij} < U_{q+1}\}. \tag{5.36}$$

Proof of (5.36) Note that for $V_{n_{N+1}} = \infty$ and $V_{n_{N+1}-1} = V_n$, we have

$$\sum_{l=n_k}^{n} I\{V_l \leq t_{ij} < V_{l+1}\} = \sum_{l=n_k}^{n_{k+1}-1} I\{V_l \leq t_{ij} < V_{l+1}\} + \sum_{l=n_{k+1}}^{n_{k+2}-1} I\{V_l \leq t_{ij} < V_{l+1}\}$$

$$+ \ldots + \sum_{l=n_{N-1}}^{n_N-1} I\{V_l \leq t_{ij} < V_{l+1}\} + \sum_{l=n_N}^{n_{N+1}-1} I\{V_l \leq t_{ij} < V_{l+1}\}. \quad (5.37)$$

Since for $k \leq N-1$, we have $I\{V_l \leq t_{ij} < V_{l+1}\} = 0$ for $V_l = V_{l+1}$ with $l = n_k$, $n_k+1, \ldots, n_{k+1}-2$; and for $V_l = V_{n_N}$ with $l = n_N, \ldots, n$; we know

$$\sum_{l=n_k}^{n_{k+1}-1} I\{V_l \leq t_{ij} < V_{l+1}\} = I\{V_{n_{k+1}-1} \leq t_{ij} < V_{n_{k+1}}\} = I\{V_{n_k} \leq t_{ij} < V_{n_{k+1}}\}$$

$$\overset{(5.33)}{=\!=} I\{U_k \leq t_{ij} < U_{k+1}\}, \quad (5.38)$$

thus, (5.37) can be written as

$$\sum_{l=n_k}^{N} I\{V_l \leq t_{ij} < V_{l+1}\} = I\{U_k \leq t_{ij} < U_{k+1}\} + I\{U_{k+1} \leq t_{ij} < U_{k+2}\}$$

$$+ \ldots + I\{U_{N-1} \leq t_{ij} < U_N\} + 0 \quad (5.39)$$

$$= \sum_{q=k}^{N} I\{U_q \leq t_{ij} < U_{q+1}\}. \qquad \square$$

Hence, we can write the second part in $\tilde{e}_k$ given in (5.27) as:

$$\sum_{q=k}^{N}\sum_{i=1}^{n}\sum_{j=1}^{J_i}\left(e^{\beta Z_i(t_{i,j-1})} - e^{\beta Z_i(t_{ij})}\right) I\{U_q \leq t_{ij} < U_{q+1}\}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{J_i}\left(e^{\beta Z_i(t_{i,j-1})} - e^{\beta Z_i(t_{ij})}\right)\left(\sum_{q=k}^{N} I\{U_q \leq t_{ij} < U_{q+1}\}\right)$$

$$\overset{(5.36)}{=\!=} \sum_{i=1}^{n}\sum_{j=1}^{J_i}\left(e^{\beta Z_i(t_{i,j-1})} - e^{\beta Z_i(t_{ij})}\right)\left(\sum_{l=n_k}^{n} I\{V_l \leq t_{ij} < V_{l+1}\}\right)$$

$$= \sum_{l=n_k}^{n}\sum_{i=1}^{n}\sum_{j=1}^{J_i}\left(e^{\beta Z_i(t_{i,j-1})} - e^{\beta Z_i(t_{ij})}\right) I\{V_l \leq t_{ij} < V_{l+1}\}. \quad (5.40)$$

From (5.35) and (5.40), we write (5.27) as below:

$$\begin{cases} \tilde{e}_k &= \sum\limits_{l=n_k}^{n} \Big(e^{\beta Z_l(t_{lJ_l})} + \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{J_i} \Big(e^{\beta Z_i(t_{i,j-1})} - e^{\beta Z_i(t_{ij})} \Big) I\{V_l \le t_{ij} < V_{l+1}\} \Big) \\ \tilde{e}_k' &= \sum\limits_{l=n_k}^{n} \Big(Z_l(t_{lJ_l}) e^{\beta Z_l(t_{lJ_l})} \\ & \quad + \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{J_i} \Big(Z_i(t_{i,j-1}) e^{\beta Z_i(t_{i,j-1})} - Z_i(t_{ij}) e^{\beta Z_i(t_{ij})} \Big) I\{V_l \le t_{ij} < V_{l+1}\} \Big) \end{cases} . \tag{5.41}$$

which is in the same form as e_i's and $e_i^{'}$'s given in (3.33).

From Lemma 3.4, we have the following simplified expressions for above $\tilde{e}_k$ and $\tilde{e}_k^{'}$ in (5.41) as follows:

$$\begin{cases} \tilde{e}_k &= \sum\limits_{i=n_k}^{n} \tilde{A}_{ik} = \sum\limits_{i=n_k}^{n} \Big(e^{\beta Z_i(t_{iJ_i})} I\{\tilde{\mathcal{E}}_{ik} = \emptyset\} + e^{\beta Z_i(t_{i\tilde{m}_{ik}})} I\{\tilde{\mathcal{E}}_{ik} \neq \emptyset\} \Big) \\ \tilde{e}_k{}' &= \sum\limits_{i=n_k}^{n} \Big(Z_i(t_{iJ_i}) e^{\beta Z_i(t_{iJ_i})} I\{\tilde{\mathcal{E}}_{ik} = \emptyset\} + Z_i(t_{i\tilde{m}_{ik}}) e^{\beta Z_i(t_{i\tilde{m}_{ik}})} I\{\tilde{\mathcal{E}}_{ik} \neq \emptyset\} \Big) \end{cases} \tag{5.42}$$

where $\tilde{\mathcal{E}}_{ik} = \big\{ t_{ij} \,|\, t_{ij} \ge V_{n_k}, j \in \{1, \ldots, J_i\} \big\}$ is a subset of $\{t_{i1}, t_{i2}, \ldots, t_{iJ_i}\}$, and

$$\begin{cases} \tilde{A}_{ik} = \begin{cases} e^{\beta Z_i(t_{iJ_i})} & \text{if} \quad t_{ij} < V_{n_k}, \ \forall\, j \in \{1, \ldots, J_i\} \\ e^{\beta Z_i(t_{i\tilde{m}_{ik}})} & \text{if} \quad \exists\, t_{ij} \ge V_{n_k}, \ \text{for some } j \in \{1, \ldots, J_i\} \end{cases} . & (5.43) \\ \tilde{m}_{ik} = \max\{\, j \mid t_{ij} < V_{n_k}, \ j = 0, 1, \ldots, J_i\}. & (5.44) \end{cases}$$

From Lemma 3.5, we know function $\varphi_n(\beta)$ given by (5.29) is non-increasing in $\beta \in \mathbb{R}$.

Below, we discuss the case when the use of M is needed for real data analysis, although often there is no need for use of M.

Choice of M with All Nonnegative $Z_i(t_{ij})$'s:

Step 1. Compute $M_k = \log \frac{\tilde{\delta}_k}{\tilde{e}_k}$, $k = 1, \ldots, N$ at $\beta = \tilde{\beta}_n - \delta$ and order them from smallest to largest as $M_{(1)} \leq M_{(2)} \ldots \leq M_{(N)}$.

Step 2. If $M_{(N)} < 0$, then no need to use M_0; else use $M_0 = M_{(N)} + 10^{-7}$.

Step 3. Search for $\hat{\beta}_n$ in interval $(\tilde{\beta}_n - \delta, \tilde{\beta}_n)$; when no M_0 is used, we use $\psi_n(\beta)$ given by (5.26) to find $\hat{\beta}_n$, when M_0 is used, we use $\psi_{M_0}(\beta)$ given by (5.30) to find $\hat{\beta}_n$; for $\psi_n(\beta)$ or $\psi_{M_0}(\beta)$, increase δ until $\psi_n(\tilde{\beta}_n - \delta) > 0$ or $\psi_{M_0}(\tilde{\beta}_n - \delta) > 0$. Try $\delta = \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1$, etc., and start with $\delta = \frac{1}{4}$.

NOTE: The rationale of our method for choosing M_0 in Step 2 is as follow. From (5.31) and (5.42), we have $\tilde{E}_k = e^M \tilde{e}_k$ and $\tilde{E}_k$ is nondecreasing in β with nonnegative $Z_i(t_{ij})$'s. For above choice of M_0 about $\tilde{E}_k$ in (5.42), for any $\beta \geq \tilde{\beta}_n - \delta$:

$$\tilde{E}_k = e^{M_0} \tilde{e}_k \geq e^{M_0} \tilde{e}_k^{**} > e^{M_{(N)}} \tilde{e}_k^{**} \geq \frac{\tilde{\delta}_k}{\tilde{e}_k^{**}} \tilde{e}_k^{**} = \tilde{\delta}_k,$$

where $\tilde{e}_k^{**}$ is the $\tilde{e}_k$ with $\beta = \tilde{\beta}_n - \delta$.

When some $Z_i(t_{ij})$'s are negative: from (5.42) we have for constant

$$c = \min_{i,j} \{Z_i(t_{ij})\} < 0,$$

then

$$\begin{aligned} \tilde{e}_k &= \sum_{i=n_k}^{n} \left(e^{\beta Z_i(t_{iJ_i})} I\{\tilde{\mathcal{E}}_{ik} = \emptyset\} + e^{\beta Z_i(t_{i\tilde{m}_{ik}})} I\{\tilde{\mathcal{E}}_{ik} \neq \emptyset\} \right) \\ &= e^{\beta c} \sum_{i=n_k}^{n} \left(e^{\beta (Z_i(t_{iJ_i}) - c)} I\{\tilde{\mathcal{E}}_{ik} = \emptyset\} + e^{\beta (Z_i(t_{i\tilde{m}_{ik}}) - c)} I\{\tilde{\mathcal{E}}_{ik} \neq \emptyset\} \right) = e^{\beta c} \tilde{e}_k^{*}, \end{aligned} \tag{5.45}$$

where $Z_i(t_{iJ_i}) - c \geq 0$ for $i = 1, \dots, n$, $j = 1, \dots, J_i$ and

$$\tilde{e}_k^* = \sum_{i=n_k}^{n} \left(e^{\beta(Z_i(t_{iJ_i})-c)} I\{\tilde{\mathcal{E}}_{ik} = \emptyset\} + e^{\beta(Z_i(t_{i\tilde{m}_{ik}})-c)} I\{\tilde{\mathcal{E}}_{ik} \neq \emptyset\} \right). \tag{5.46}$$

Choice of M with negative $Z_i(t_{ij})$'s:

Step 1. Compute $c = \min_{i,j}\{Z_i(t_{ij})\} < 0$.

Step 2. Compute $M_k = \log \frac{\tilde{\delta}_k}{\tilde{e}_k^*}$, $k = 1, \dots, N$ at $\beta = \tilde{\beta}_n - \delta$ and order them from smallest to largest as $M_{(1)} \leq M_{(2)} \dots \leq M_{(N)}$.

Step 3. Choose $M_0 = M_{(N)} + 10^{-7}$, then compute $M = M_0 - (\tilde{\beta}_n + \delta)c$.

Step 4. Use above choice of M and $\psi_M(\beta)$ to search for $\hat{\beta}_n$ in interval $(\tilde{\beta}_n - \delta, \tilde{\beta}_n + \delta)$; increase δ until $\psi_M(\tilde{\beta}_n - \delta) > 0$. Try $\delta = \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1$, etc., and start with $\delta = \frac{1}{4}$.

NOTE: The rationale of our method for choosing M in Step 3 is as follow. Because all $Z_i(t_{ij}) - c$ are nonnegative, $\tilde{e}_k^*$ in (5.46) is nondecreasing in β. For above choice of M about $\tilde{E}_k$ in (5.42), for any $\tilde{\beta}_n - \delta \leq \beta \leq \tilde{\beta}_n + \delta$:

$$\begin{aligned}
\tilde{E}_k &= e^M \tilde{e}_k \overset{(5.45)}{=\!=} e^M e^{\beta c} \tilde{e}_k^* = e^{M_0 - (\tilde{\beta}_n + \delta)c} e^{\beta c} \tilde{e}_k^* = e^{(\beta - (\tilde{\beta}_n + \delta))c} e^{M_0} \tilde{e}_k^* \\
&\geq e^{(\beta - (\tilde{\beta}_n + \delta))c} e^{M_0} \tilde{e}_k^{**} > e^{(\beta - (\tilde{\beta}_n + \delta))c} e^{M_{(N)}} \tilde{e}_k^{**} \geq e^{(\beta - (\tilde{\beta}_n + \delta))c} \frac{\tilde{\delta}_k}{\tilde{e}_k^{**}} \tilde{e}_k^{**} \geq \tilde{\delta}_k
\end{aligned}$$

where $\tilde{e}_k^{**}$ is the $\tilde{e}_k^*$ with $\beta = \tilde{\beta}_n - \delta$, $e^{(\beta - (\tilde{\beta}_n + \delta))c} \geq 1$ as $c < 0$ and $\beta - (\tilde{\beta}_n + \delta) \leq 0$.

5.2 Smoking Cessation Data Analysis

In this section, we consider the Smoking Cessation Data from Piper et al. (2009), which was collected from a randomized, placebo-controlled clinical trial of five smoking cessation pharmacotherapies. There were 1504 participants who were smokers and reported being motivated to quit smoking. The dataset we consider here was processed by Liu (2014), and the dataset contained records of participants for a period of two weeks after they quit smoking.

In our analysis, we are interested in lapse time T, which is the time from a participant stopped smoking to the time the participant smoked again. Such lifetime variable T is subject to right censoring due to drop-out or end of the study. There were 794 participants included in this study, and the right censoring or right truncation percentage is 59.69%. During the study, the ecological momentary assessment data were collected 4 times a day (i.e., morning, night and 2 random times) to record longitudinal covariates such as urge for smoking, negative affect and smoking cessation fatigue. For instance, the actually observed data on urge for smoking, $X_i(t)$, is $\mathcal{X}_i = (X_i(t_{i1}), \ldots, X_i(t_{iJ_i}))$, where $X_i(t_{ij})$ is the observed amount of urge of the ith participant at time t_{ij}, $j = 1, \ldots, J_i$. The treatment groups under consideration here are: placebo with $n_1 = 83$ participants, monotherapy with $n_2 = 412$ participants, and combined pharmcotherapy with $n_3 = 299$ participants.

Under our proposed PH model in (1.38), we conduct data analysis to study the effect of $X_i(t_{ij})$'s on T_i, where $X_i(t)$ is urge for smoking, negative affect, smoking cessation fatigue, respectively. First, we consider the case without using summary measurement for longitudinal covariates:

$$Z(t_{ij}) = X_i(t_{ij}), \tag{5.47}$$

and the results are displayed in Table 5.1.

Table 5.1 Estimation for Smoking Cessation Data without Using Average

		Urge		Negative Affect		Fatigue	
	#	$\hat{\beta}_n$	$\tilde{\beta}_n$	$\hat{\beta}_n$	$\tilde{\beta}_n$	$\hat{\beta}_n$	$\tilde{\beta}_n$
Whole Data Set	794	0.1012	0.1080	0.1067	0.1184	0.0287	0.0336
Placebo	83	0.1307	0.1443	0.0476	0.0720	-0.0471	-0.0354
Monotherapy	412	0.0912	0.0986	0.1116	0.1233	0.0328	0.0381
Combined Therapy	299	0.0968	0.1031	0.0969	0.1086	0.0417	0.0461

Second, under our proposed PH model (1.38), we consider the case using average summary measurement for the longitudinal covariates. This is because these longitudinal covariates are participants' momentary feelings, thus fluctuate greatly over time, and using average summary measurement gives more reliable information. In this case, we have $Z(t_{ij})$ as given in (4.25):

$$Z(t_{ij}) = \varphi(\mathcal{X}_i^{t_{ij}}) = \frac{\sum_{k=1}^{j}(t_{ik} - t_{i,k-1})X_i(t_{i,k-1})}{t_{ij}}, \tag{5.48}$$

and the results are displayed in Table 5.2.

Table 5.2 Estimation for Smoking Cessation Data Using Average Summary

		Urge		Negative Affect		Fatigue	
	#	$\hat{\beta}_n$	$\tilde{\beta}_n$	$\hat{\beta}_n$	$\tilde{\beta}_n$	$\hat{\beta}_n$	$\tilde{\beta}_n$
Whole Data Set	794	0.1171	0.1263	0.1313	0.1489	0.0332	0.0391
Placebo	83	0.1650	0.1807	0.1074	0.1369	-0.0636	-0.0470
Monotherapy	412	0.1117	0.1215	0.1074	0.1270	0.0453	0.0519
Combined Therapy	299	0.0927	0.1019	0.1406	0.1563	0.0349	0.0403

Discussion: In Liu (2014), a form of model (1.31) was used to conduct this data analysis using $X_i(t_{ij})$, which was assumed to be Gaussian process. Their estimated coefficient for urge, negative affect and fatigue in whole data set are 0.13, 0.15, 0.03, respectively. Our according estimates displayed in Table 5.1 are generally smaller, and we do not make any parametric assumptions on $X_i(t)$. Interestingly, our estimates using average summary measurement displayed in Table 5.2 for whole data set are quite similar to those estimates by Liu (2014) above mentioned, but the analysis by Liu (2014) did not consider the use of average summary measurement.

Making comparison, it is clear that for urge for smoking and negative affect, $\hat{\beta}_n$ or $\tilde{\beta}_n$ are generally bigger in Table 5.2 than those in Table 5.1. This could indicate that the use of average summary measurement provides more informative results. However, the impact of fatigue from smoking cessation is difficult to explain with or without using the average summary measurement, because $\hat{\beta}_n \approx \tilde{\beta}_n \approx 0$, which could mean no effect to T.

5.3 Liver Disease Data Analysis

In this section, we consider the Liver Disease Data, which was collected from a Primary Biliary Cirrhosis (PBC) clinical trial (Murtaugh et al., 1994) conducted by Mayo Clinic between 1974 and 1984. PBC data can be found in Therneau and Grambsch (2000) and can be accessed at Mayo Clinic website.

During the clinical trial period, 312 PBC patients were referred to Mayo Clinic during ten-year interval and participated in the randomized trial of the drug D-penicillamine. Patients often present abnormalities in their blood tests, such as elevated and gradually increased serum bilirubin. Shapiro et al. (1979) showed that after a relatively stable phase, serum bilirubin increased sharply in the months preceding death. Thus, investigating the association between bilirubin and death due to PBC is of great medical interest.

In our analysis, we consider survival time T of a PBC patient, which is the time from entering the clinical trial to death, and is subject to right censoring due to drop-outs, the end of the study, liver transplantation and other factors. Right censoring percentage of the data set is 55.13%, and for observed right censored data (V_i, δ_i) in (1.2), V_i ranged from 0.11 years to 14.32 years. For each patient, longitudinal covariate serum bilirubin was measured at follow-up lab visits at 6 months, 1 year and annually thereafter. Thus, the observed data on serum bilirubin, $X_i(t)$, is $\mathcal{X}_i = (X_i(t_{i1}), \ldots, X_i(t_{iJ_i}))$, where $X_i(t_{ij})$ is the observed amount of serum bilirubin of the ith patient at time t_{ij}, $j = 1, \ldots, J_i$. The treatment groups under consideration here are: placebo with $n_1 = 154$ patients and D-penicillamine with $n_2 = 158$ patients.

Under our proposed PH model in (1.38), we consider data analysis to study the effect of $X_i(t_{ij})$'s on T_i, where $Z(t_{ij})$ given in (5.47) is based on $\log X_i(t)$ and $X_i(t)$, respectively. The reason for considering $\log X_i(t)$ is due to clinical tradition. The results for both cases are displayed in Table 5.3.

Table 5.3 Estimation for Liver Disease Data

		log(bilirubin)		bilirubin	
	#	$\hat{\beta}_n$	$\tilde{\beta}_n$	$\hat{\beta}_n$	$\tilde{\beta}_n$
Whole Data Set	312	1.2871	1.2891	0.1506	0.1507
Treatment Group data	158	1.1582	1.1621	0.1299	0.1303
Placebo Group Gata	154	1.4072	1.4111	0.1830	0.1833

Discussion: In Ding and Wang (2008), the PH model in (1.31) was considered and $\log X_i(t)$ was used in their analysis, where as in (1.34), r.v. b_i was assumed to be normal. Their resulting estimate for β_X is $\hat{\beta}_X = 1.07$. Compared with our estimates $\hat{\beta}_n = 1.2871$ or $\tilde{\beta}_n = 1.2891$ in Table 5.3 for whole data set, $\hat{\beta}_X = 1.07$ is clearly smaller. Thus, our results without any parametric or semi-parametric assumptions on $X_i(t)$ are in stronger agreement with the conclusion by Shapiro et al. (1979). Interestingly, our results in Table 5.3 show that $\hat{\beta}_n$ and $\tilde{\beta}_n$ for treatment group are smaller than those for placebo group. This indicates that treatment is clearly efficient.

Ding and Wang (2008) didn't consider treatment group and placebo group separately, but our results give the following

$$\frac{158}{312}\Big(1.1582\Big) + \frac{154}{312}\Big(1.4072\Big) = 1.2811,$$

which is very close to $\hat{\beta}_n$ for the whole data set. This means that the difference of using PH model (1.36) and PH model (1.38) is not big for estimating β_0, which is because $Z = 0$ or 1 in (1.36) for the liver disease data considered here.

Chapter 6: Conclusion and Discussions

This dissertation does not make any parametric or semi-parametric assumptions on the longitudinal covariate process. We use the empirical likelihood method, i.e., the totally nonparametric method, to derive the MLE for our proposed proportional hazards (PH) model (1.38) based on right censored survival data with longitudinal covariates (1.29). Our simulation results show that MLE $\hat{\beta}_n$ and approximated MLE $\tilde{\beta}_n$ perform well for large sample size n and large m. These are consistent with the finding in upcoming paper by Ren and Shi (2023) that $\hat{\beta}_n$ and $\tilde{\beta}_n$ are asymptotically normal.

The computation issue has known to be challenging for the MLE in the Cox model. Here, we developed the computation algorithms for computing MLE $\hat{\beta}_n$ and approximated MLE $\tilde{\beta}_n$, and they perform well in our simulation studies. The main results in this dissertation will appear in upcoming paper by Ren, Li and Shi (2023).

In the future, we will develop algorithms for more general model setting (1.36), and will develop goodness of fit test for (1.36) and (1.38) because the MLE $\hat{F}_n$ for the baseline d.f. F_0 was given here already.

Bibliography

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

Chen, J. and Qin, J. (1993). Empirical likelihood estimation for finite populations and the effective usage of auxiliary information. *Biometrika*, Vol. 80:107–116.

Chen, S. X. and Cui, H. (2006). On bartlett correction of empirical likelihood in the presence of nuisance parameters. *Biometrika*, Vol. 93:215–220.

Chen, S. X. and Hall, P. (1993). Smoothed empirical likelihood confidence intervals for quantiles. *The Annals of Statistics*, Vol. 21:1166–1181.

Cole, P. M., Tan, P. Z., Hall, S. E., Zhang, Y., Crnic, K. A., Blair, C. B., and Li, R. (2011). Developmental changes in anger expression and attention focus: learning to wait. *Developmental psychology*, Vol. 47:1078–1089.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society*, Vol. 34:187–220.

Cox, D. R. and Oakes, D. (1984). *Analysis of Survival Data*. Chapman & Hall/CRC, New York .

DiCiccio, T., Hall, P., and Romano, J. (1991). Empirical likelihood is bartlett-correctable. *The Annals of Statistics*, Vol. 19:1053–1061.

Diggle, P. J., Heagerty, P., Liang, K.-Y., and Zeger, S. L. (2002). *Analysis of Lognitudinal Data*. Oxford University Press, Oxford.

Ding, J. and Wang, J.-L. (2008). Modeling longitudinal data with nonparametric multiplicative random effects jointly with survival data. *Biometrics*, Vol. 64:546–556.

Efron, B. (1977). The efficiency of cox's likelihood function for censored data. *Journal of the American Statistical Association*, Vol. 72:557–565.

Gill, R. (1983). Large sample behaviour of the product-limit estimator on the whole line. *The Annals of Statistics*, Vol. 11:49–58.

Groeneboom, P. and Wellner, J. A. (1992). *Information Bounds and Nonparametric Maximum likelihood Estimation*. Birkhäuser Verlag.

Gu, M. and Zhang, C. H. (1993). Asymptotic properties of self-consistent estimators based on doubly censored data. *The Annals of Statistics*, Vol. 21:611–624.

Hall, P. and La Scala, B. (1990). Methodology and algorithms of empirical likelihood. *International Statistical Review*, Vol. 58:109–127.

He, S., Liang, W., Shen, J., and Yang, G. (2016). Empirical likelihood for right censored lifetime data. *Journal of the American Statistical Association*, Vol. 111:646–655.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. John Wiley & Sons, INC.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, Vol. 53:457–481.

Leiderman, P., Babu, B., Kagia, J., Kraemer, H., and Leiderman, G. (1973). African infant precocity and some social influences during the first year. *Nature*, Vol. 242:247–249.

Li, R. and Ren, J. (2011). An overview on joint modeling of censored survival time and longitudinal data. *Analysis of High Dimensional Data*, Edited by T.Cai and X.T. Shen:195–222.

Li, R., Root, T., and Shiffman, S. (2006). A local linear estimation procedure for functional multilevel modeling. In Walls, T. A. and Schafer, J. L., editors, *Models for Intensively Longitudinal Data*, pages 63–83. Oxford University Press.

Liu, X. (2014). *Joint Modeling of Longitudinal and Survival Data: New Models, Computing Techniques and Applications*. Doctoral dissertation, Pennsylvania State University.

Murphy, S. A. and van der Vaart, A. W. (1997). Semiparametric likelihood ratio inference. *The Annals of Statistics*, Vol. 25:1471–1509.

Murtaugh, P. A., Dickson, E. R., Dam, G. M. V., Malinchoc, M., Grambsch, P. M., Langworthy, A. L., and Gips, C. H. (1994). Primary biliary cirrhosis: prediction of short-term survival based on repeated patient visits. *Hepatology*, Vol. 20:126–134.

Mykland, P. A. (1995). Dual likelihood. *The Annals of Statistics*, Vol. 23:396–421.

Mykland, P. A. and Ren, J. (1996). Algorithms for computing self-consistent and maximum likelihood estimators with doubly censored data. *The Annals of Statistics*, Vol. 24:1740–1764.

Oakes, D. (1977). The asymptotic information in censored survival data. *Biometrita*, Vol. 64:441–448.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, Vol. 75:237–249.

Owen, A. B. (2001). *Empirical Likelihood*. Chapman & Hall/CRC .

Pawitan, Y. and Self, S. (1993). Modeling disease marker processes in AIDS. *Journal of the American Statistical Association*, Vol. 88:719–726.

Piper, M. E., Smith, S. S., Schlam, T. R., Fiore, M. C., Jorenby, D. E., Fraser, D., and Baker, T. B. (2009). A randomized placebo-controlled clinical trial of five smoking cessation pharmacotherapies. *Archives of General Psychiatry*, Vol. 66:1253–1262.

Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics*, Vol. 22:300–325.

Ren, J. (2001). Weight empirical likelihood ratio confidence intervals for the mean with censored data. *Annals of the Institute of Statistical Mathematics*, Vol. 53:498–516.

Ren, J. (2008a). Smoothed weighted empirical likelihood ratio confidence intervals for quantiles. *Bernoulli*, Vol. 14:725–748.

Ren, J. (2008b). Weighted empirical likelihood in some two-sample semiparametric models with various types of censored data. *The Annals of Statistics*, Vol. 36:147–166.

Ren, J. and Gu, M. (1997). Regression M-estimators with doubly censored data. *The Annals of Statistics*, Vol. 25:2638–2664.

Ren, J., Li, R., and Shi, Y. (2023). Computation and simulation studies on the MLE for joint modeling right censored survival data with longitudinal covariates. (In preparation).

Ren, J. and Shi, Y. (2023). Empirical likelihood-based MLE in the proportional hazards model for right censored survival data with longitudinal covariates. (In preparation).

Ren, J. and Wang, Y. (2022). Multivariate weighted empirical likelihood inferences for the cox model with various types of censored data. *The Annals of Statistics*, (Revised).

Ren, J. and Zhou, M. (2011). Full likelihood inferences in the cox model: an empirical likelihood approach. *Annals of the Institute of Statistical Mathematics*, Vol. 63:1005–1018.

Shapiro, J. M., Smith, H., and Schaffner, F. (1979). Serum bilirubin: a prognostic factor in primary biliary cirrhosis. *Gut*, Vol. 20:137–140.

Shiffman, S., Gnys, M., Richards, T. J., Paty, J. A., and Hickcox, M. (1996). Temptations to smoke after quitting: a comparison of lapsers and maintainers. *Health Psychology*, Vol. 15:455–461.

Stute, W. and Wang, J.-L. (1993). The strong law under random censorship. *The Annals of Statistics*, Vol. 21:1591–1607.

Su, Y.-R. and Wang, J.-L. (2012). Modeling left-truncated and right-censored survival data with longitudinal covariates. *The Annals of Statistics*, Vol. 40:1465–1488.

Therneau, T. M. and Grambsch, P. M. (2000). *Modeling Survival Data: Extending the Cox Model*. Springer, New York .

Tsiatis, A. A. (1981). A large sample study of cox's regression model. *The Annals of Statistics*, Vol. 9:93–108.

Tsiatis, A. A. and Davidian, M. (2004). Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica*, Vol. 14:809–834.

Wang, D. and Chen, S. X. (2009). Empirical likelihood for estimating equation with missing values. *The Annals of Statistics*, Vol. 37:490–517.

Zeng, D. and Cai, J. (2005). Asymptotic results for maximum likelihood estimators in joint analysis of repeated measurements and survival time. *The Annals of Statistics*, Vol. 33:2132–2163.

Zhou, M. (2005). Empirical likelihood analysis of the rank estimator for the censored accelerated failure time model. *Biometrika*, Vol. 92:492–498.