Distribution Agreement

Flexible Semiparametric Regression Methods for Observational Follow-up Studies

By

Xiaoyan Sun
Doctor of Philosophy

Biostatistics

____
Limin Peng, Ph.D.
Advisor

____
Yijian Huang, Ph.D.
Committee Member

____
Amita K. Manatunga, Ph.D.
Committee Member

____
Michele Marcus, Ph.D.
Committee Member

Accepted:

____
Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

____
Date

Flexible Semiparametric Regression Methods for Observational Follow-up Studies

By

Xiaoyan Sun B.S., Zhejiang University, 2009

Advisor: Limin Peng, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements of the degree of
Doctor of Philosophy
in Biostatistics
2014

Abstract

Flexible Semiparametric Regression Methods for Observational Follow-up Studies

By Xiaoyan Sun

Observational follow-up studies often present various challenges that can complicate statistical analysis, such as complex censoring mechanism, missing observations, and highly skewed measurements. In my dissertation, we have developed flexible semiparametric regression methods for three different complex data scenarios.

The first one is recurrent events setting subject to window observation, which arises when the observation of recurrent event is not available before the follow-up starts and after the follow-up ends. We adopt the accelerated recurrent time model (Huang and Peng, 2009), and develop two estimators for window observed recurrent event data. We illustrate our method via an analysis of the time to expected frequency of pseudomonas aeruginosa (PA) infection in Cystic Fibrosis (CF) children through the use of the US CF Foundation Patient Registry (CFFPR).

The second project is about longitudinal data with skewed outcome subject to left censoring and following an informative intermittent missing pattern, which is motivated by the Michigan Long-Term Polybrominated Biphenyls (PBB's) Study. In this work, we consider quantile regression modeling for the data from such longitudinal studies. We adopt an appropriate censored quantile regression technique to handle left censoring and employ the idea of inverse probability weighting to tackle the issue associated with informative intermittent missing data. We evaluate our method by simulation studies. The proposed method is applied to the Michigan PBB study to investigate the PBB decay profile.

The third data scenario is longitudinal data with skewed outcome subject to left censoring and irregular outcome-dependent follow-up. For example, in the Michigan PBB study, serum samples were not taken at a set of common time points but at irregular time intervals. In this work, we propose an inverse intensity-ratio weighted least absolute deviation estimator in censored quantile regression. This approach yields consistent estimates of the quantile regression parameters provided that the model for the follow-up visit process has been correctly specified. The proposed method is also applied to the Michigan PBB study to investigate the PBB decay profile.

Flexible Semiparametric Regression Methods for Observational Follow-up Studies

By

Xiaoyan Sun
B.S., Zhejiang University, 2009

Advisor: Limin Peng, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements of the degree of
Doctor of Philosophy
in Biostatistics
2014

Acknowledgments

To my advisor Dr. Limin Peng.
To the committee members and all teachers.
To my family.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background

Observational follow-up studies often present various challenges that can complicate statistical analysis, such as complex censoring mechanism, missing observations, and highly skewed measurements. Many semiparametric regression models, such as quantile regression, has received increased attention for their capability of handling skewed data and allowing for varying covariate effects. In my dissertation, we have studied three different data scenarios. The first one is recurrent events setting subject to window observation, which arises when the observation of recurrent event is not available before the follow-up starts and after the follow-up ends. The second one is focused on longitudinal data with skewed outcomes subject to left censoring plus outcome-dependent intermittent missingness. The third one deals with longitudinal skewed measurements subject to left censoring and observed only at irregular outcome-dependent follow-up times. As elaborated later, considerable statistical challenges are involved in developing statistical methods for appropriately analyzing the complex data scenarios described above.

In the first project, we have studied the accelerated recurrence time model for recurrent events data subject to window observation. Recurrent events are frequently encountered in biomedical research. Examples include tumor recurrences, asthmatic attacks, and hospitalizations. In some observational studies, the observation of recurrent events is constrained to an observation window between the start of follow-up and the last follow-up visit. An example is the US Cystic Fibrosis Foundation Patient Registry (CFFPR) study. Cystic Fibrosis (CF) is a life-limiting genetic disease without known cure yet, affecting about 30,000 people in the United States (Cystic Fibrosis Foundation, 2011). For CF patients, lung infections will result in damaged and lower lung function. Pseudomonas aeruginosa (PA), the most important pathogen that shortens survival of CF patients, infects more than half of people with CF (Cystic Fibrosis Foundation, 2011). The first 10 years were known to be an important potentially beneficial period for early diagnosis and new therapies (Campbell and White, 2005; Gross et al., 2006). Therefore, it is of scientific interest to investigate the association between the timing of PA infection recurrences in the first 10 years and its risk factors. Though the onset and remissions of PA infections are usually well monitored during CFFPR follow-up, no record of PA infections is available before the registry entry. The missing information on PA infection before registry entry is not ignorable because most of CF children did not entry the registry at or shortly after birth due to delayed diagnosis after birth or delayed entry to the registry. At the same time the observation of PA infection is also terminated at the last follow-up. Such a recurrent event setting is the focus of the first project in my dissertation.

The second and the third projects are concerned with regression analyses of longitudinal measurements, which have skewed distributions and are subject to left censoring. Our motivating example is the Michigan Long-Term Polybrominated Biphenyl (PBB) Study. PBBs are manufactured chemicals that accidentally mixed with animal feed during 1973-1974. Residents on Michigan farms and neighboring communities were exposed to PBBs by consuming meat, milk, and other food products from contaminated animals. This study was established following exposure to PBB's. Since the initial enrollment period (1976 - 1978), the Michigan Department of Community Health (MDCH) has periodically contacted cohort members to obtain additional serum samples to measure PBB concentration levels. PBBs are stable, persistent halogenated organic pollutants with extremely long half-lives that have been shown to have suggested effects on several diseases in animal studies. Participants in this study continued to have measurable serum PBB concentration levels more than 20 years later. Thus, it is of interest to understand the pattern by which PBB is eliminated from the body. In the longitudinal data collected in the Michigan PBB study, serum PBB concentration measurements are highly right skewed (Figure 1.1). The measurements are also left censored due to laboratory assay detection. If the serum PBB concentration is less than 1 pbb, it is recorded as 1 pbb.

Subject's follow-up pattern may depend on observed PBB concentrations. Figure 1.2 presents the distribution of the first PBB measurement in each group of subjects with the same number of visits. It is clear that the first PBB measurement is higher in groups with more visits than groups with fewer visits. We take two different perspectives to handle the statistical issues from the outcome-dependent follow-up pattern. In project 2, we divide the continuous time scale into prespecified time-intervals so that we formulate the PBB data as longitudinal data with fixed visit times. When a subject did not come for a follow-up visit in a given time-interval, the longitudinal outcome at the given time point is treated as missing. Under this view, we encounter informative missing data because the missing pattern of longitudinal outcome are not completely random, as implied by Figure 1.2. To handle the missing data, we adpoted the assumption of missing at random (MAR), which allows the data missingness to be related to the observed responses, but is assumed to be conditionally independent of the missing responses (Little and Rubin, 1987). In the PBB study, participants were given their PBB levels after they were analyzed. It is reasonable to make the assumption that the missing pattern is related to the observed measurements and is independent of the missing measurements given the observed measurements. In the second projects, we study the estimation and inference of a quantile regression model for longitudinal measurements subject to informative intermittent missing and left censoring.

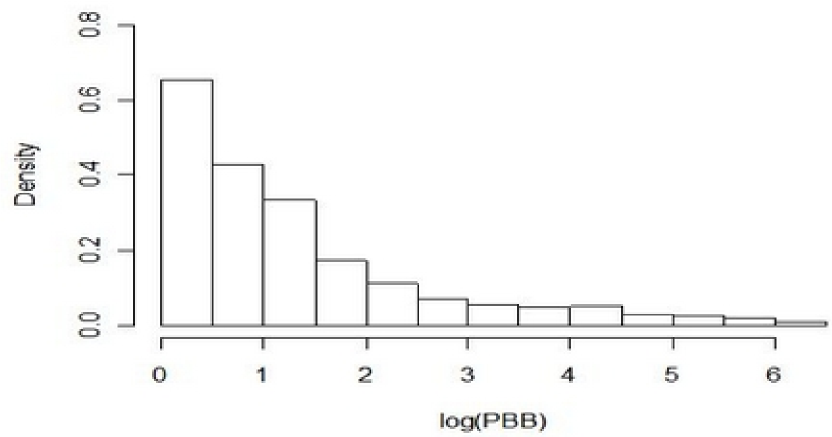

Figure 1.1: Distribution of PBB concentration measurements after logrithm transformation

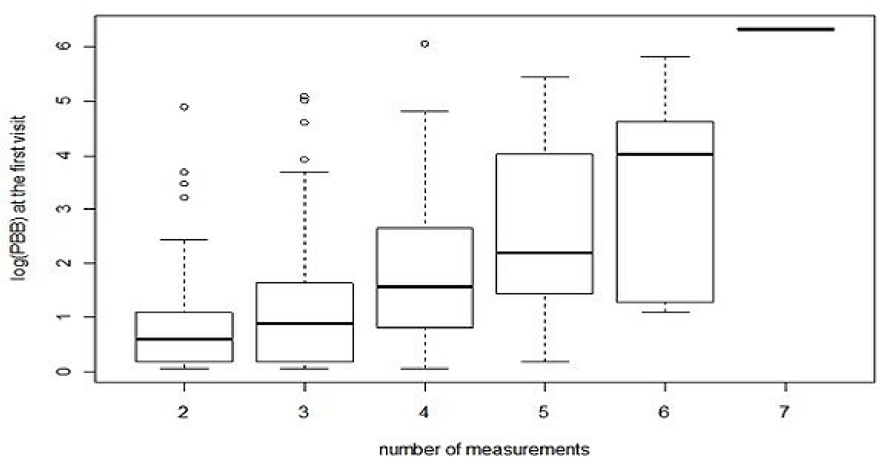

Figure 1.2: Distribution of log (PBB) at the first visit versus number of measurements

In the third project, we treat visit times as continuous and the PBB data are then formulated as longitudinal data with irregular follow-up times. By this approach, we avoid the time interval division, which may involve arbitrariness. To handle outcome-dependent follow-up, we adopt separate modeling for visit times and longitudinal outcomes. We model follow-up visits via a recurrent event process that follows a proportional intensity model. We develop a marginal quantile regression method for left censored longitudinal data that appropriately handles the outcome-dependent irregular follow-up.

In all projects, we have adopted semiparametric quantile regression models that are more flexible than traditional models. Quantile regression can deal with skewed data without imposing distributional assumptions, such as normality. Also quantile regression formulates covariate effects separately on different quantiles and do not require them to be constant over different quantile levels. This feature may help detect inhomogeneous risk patterns. For example, in the Michigan PBB study, quantile regression can capture the characteristic that the upper end of PBB concentration distribution tends to decrease faster than the lower end of PBB distribution while standard linear mixed models cannot. The accelerated recurrence time model shares the same spirit as quantile regression. The accelerated recurrence time model specifies covariate effects on time to expected frequencies. This model does not imposing any assumption about the pattern of time to expected frequencies for the reference group. It also has the flexibility to allow covariate to have different effects on time to different expected frequencies.

In my dissertation, we develop new semiparametric regression methods, pursuing the advantages described above while appropriately handling the data complexities present in real studies. In the next section, we present literature reviews separately on regression for recurrent event data and regression for longitudinal data. An outline of my dissertation is given in the end of this chapter.

1.2 Literature Review

1.2.1 Existing Work on Regression for Recurrent Event Data

Let $\mathbf{T} = \{T^{(1)}, T^{(2)}, \ldots\}$ be the recurrent event times. Let $\mathbf{Z}$ be the associated $p \times 1$ vector of covariates. The corresponding counting process to the recurrent event times is denoted by $N(t) = \sum_{j=1}^{\infty} I(T^{(j)} \leq t)$. The regression analysis of recurrent event time data, concerned by the first project, has been investigated in literature.

One well known approach is through intensity function which represents the instantaneous probability of an event conditional on the process history. The intensity function is mathematically defined as

$$\lambda(t|H(t)) = \lim_{\Delta \to 0} \frac{Pr\{\Delta N(t) = 1|H(t)\}}{\Delta t}, \tag{1.1}$$

where $H(t) = \{\mathbf{Z}, N(s) : 0 \leq s < t\}$ denote the history up to time t. One popular way to model the association between the intensity function and covariate is the proportional intensity model proposed by Andersen and Gill (1982) which takes the form

$$\lambda(t|\mathbf{Z}) = \lambda_0(t) e^{\boldsymbol{\beta}^T \mathbf{Z}(t)}, \tag{1.2}$$

where $\lambda_0(\cdot)$ is an unspecified baseline intensity function and $\boldsymbol{\beta}$ is a vector of unknown regression parameters. Andersen and Gill (1982) assumes zero intra-individual correlation among recurrent events when estimating the coefficients which may not be appropriate in many applications. A useful approach to accommodating the intra-individual correlation is to incorporate a random effect also called frailty γ into model

(1.2):

$$\lambda(t|\mathbf{Z}, \gamma) = \gamma\lambda_0(t)e^{\boldsymbol{\beta}^T\mathbf{Z}(t)}. \tag{1.3}$$

This model has been investigated by Nielsen et al. (1992) and Oakes (1992), among others. Frailty is usually specified to follow a distribution, such as Gamma distribution.

Another popular intensity models is through modeling waiting time or gap time between two adjacent recurrent events. This type of methods is often adopted when events are relatively infrequent. The simplest example is assuming that the gap times are independent within one subject,

$$\lambda(t|H(t)) = h\left(t - T_{N(t^-)}\right), \tag{1.4}$$

where $h(\cdot)$ is the hazard function for the gap times between events which are independent and identically distributed. Models for gap time have also been investigated in literature. In dealing with the Gap times between adjacent recurrent events, the main challenge is dependent censoring. When the overall follow-up time is subject to independent censoring, the gap time except the first one are subject to dependent censoring. Prentice et al. (1981) developed a regression for gap times based on proportional hazard model but only applicable when recurrent event times are conditionally independent, given the covariate. Regression procedures without the requirement for conditional independence among recurrent event times also have been investigated. For example, Huang (2002) proposed a regression procedure for gap times based on accelerated failure time model. Schaubel and Cai (2004) developed an estimating equation for fitting proportional hazard model for gap times.

Conditional models based on intensity function usually requires assumption about dependence structure among recurrent events within a subject. This potential drawback makes robust marginal models based on expected event frequency be popular alternative in practice. For example, proportional mean and rate models formulate covariate effects on the mean frequency function $E[N(t)|\mathbf{Z}]$ or the rate function $E[dN(t)|\mathbf{Z}]$. These models have been studied by many authors, such as Pepe and Cai (1993), Lawless and Nadeau (1995), and Lin et al. (2000), and been extended to models with a more general class of transformation by Lin et al. (2001).

The model we adopted for project 1 falls into the model category that uses mean/rate rather than intensity. A directly relevant model is Lin et al. (1998)'s accelerated failure time model (AFT) for recurrent events data. Specifically, Lin et al. (1998) proposed to specify covariate effects on the frequency of recurrences as expanding or contracting the time scale, $E(N(t)|\mathbf{Z}) = \mu_0(e^{\boldsymbol{\beta}_0'\mathbf{Z}}t)$, where $\mu_0(\cdot)$ is an unspecified continuous function. As pointed out by D. R. Cox (Reid (1994), p. 450), "accelerated life models are in many ways more appealing [than the proportional hazards model] because of their quite direct physical interpretation".

It would be desirable to have the flexibility to accommodate varying effects of covariates. However, Lin et al. (1998)'s AFT model does not have such a capability. In fact, there are relatively limited work on marginal recurrent event models with varying covariate effects. Fine et al. (2004) proposed temporal process regression allowing time-varying covariate effects on the mean frequency function. Chiang and Wang (2009) proposed a proportional rate model with time-varying coefficients. Estimators are obtained through maximizing the kernel weighted partial likelihood function which requires smoothing parameters.

More recently, Huang and Peng (2009) proposed an accelerated recurrence time model allowing for varying covariate effects. Define the inverse function of the mean frequency function as $\tau_{\mathbf{Z}}(u) = \inf\{t : E(N(t)|\mathbf{Z}) > u\}$, which can be easily interpreted as time to expected frequency u. The accelerated recurrence time model assumes varying covariate effects dependent on the expected frequency:

$$\log \tau_{\mathbf{Z}}(u) = \boldsymbol{\beta}_0(u)^T \mathbf{X}, \qquad \forall u \in (0, \infty), \tag{1.5}$$

where $\mathbf{X} = (1, \mathbf{Z}^T)^T$. This model could be viewed as a generalization of the accelerated failure time model for counting process (Lin et al., 1998). When all components of $\boldsymbol{\beta}_0(u)$ except the intercept are constant in u, the accelerated recurrence time model reduces to the accelerated failure time model for counting processes. Compared with other varying-coefficient models, the accelerated recurrence time model retains the appealing advantage of direct physical interpretation that covariate effects are specified on each recurrence instead of on the mean frequency or rate function. The modeling strategy for varying covariates effects in the accelerated recurrence time model shares the same spirit as quantile regression. Quantile regression model for survival data is a special case of the accelerated recurrence time model when each individual may experience at most one event. The accelerated recurrence time model has the flexibility allowing covariates to have different effects on time to different expected frequencies. Compared to Lin et al. (1998)'s accelerated failure time model, the new model could provide a more comprehensive view of covariate effects.

Note that, Huang and Peng (2009) only considered observation windows starting from zero for recurrent events data analysis. It is not straightforward to extend Huang and Peng (2009)'s work to handle the more realistic recurrent events data setting where the observation of recurrent event may not start right from the time origin. This hurdles the exploration of the CFFPR data based on the flexible accelerated recurrence time model. This constitutes the motivation for the first project

1.2.2 Existing Work on Regression Analysis of Longitudinal Data

The second and third projects are focused on longitudinal data with repeated measurements where correlation among measurements on the same subject is often not ignorable. There are two popular approaches to analyzing such data. One is to use mixed effects model to handle intra-subject correlation through assuming random effects. Let y_{ij} be the jth response of the ith subject and $\mathbf{x}_i$ be its covariate vector. For example, one may assume that

$$y_{ij} = \alpha_i + \boldsymbol{\beta}\mathbf{x}_i + \epsilon_{ij}, \tag{1.6}$$

where α_i is random intercept and ϵ_{ij} is measurement error, both with distribution fully specified. Estimates of $\boldsymbol{\beta}$ can be obtained through maximizing the likelihood. For example, model (1.6) with normal random effects has been studied by Laird and Ware (1982) and Ware (1985). The other commonly used method for analyzing longitudinal data is called generalized estimating equation (GEE) developed by Liang and Zeger (1986). The advantage of this method is that it does not require the correct specification of the intra-individual correlation. When the correlation matrix is misspecified, the resulting estimator is still consistent though less efficient compared to the estimator resulting from an estimating equation with correct correlation matrix.

Missing data is not unusual in longitudinal data. A subject may dropout in the middle of the entire follow-up or a subject may be missing at one follow-up visit but return at the next visit. Missing data in longitudinal studies has been investigated by many literatures. Rubin (1976) defined three missing mechanisms. Missing completely at random (MCAR) when missingness is unrelated with the data. Missing at random (MAR) if missingness depends on the observed data only (given the observed data, missingness is unrelated with the unobserved data). Missing not at random (MNAR) that missingness depends on the unobserved data, given the observed data. It is well known that MCAR and MAR are ignorable in the likelihood and Bayesian approaches, while MAR is not ignorable in marginal regression methods, such as generalized estimating equations (GEE's) (Ibrahim and Molenberghs, 2009). Robins et al. (1995) proposed an inverse probability weighted GEE under the MAR assumption that yields consistent and asymptotically normal estimators when the dependence of missingness on the observed past is correctly specified.

Similar results have been shown for analysis of longitudinal data with irregular outcome-dependent follow-up. Lipsitz et al. (2002) developed a likelihood-based approach for analyzing longitudinal outcomes following a multivariate Gaussian distribution. Under some mild assumptions, they showed that inferences can proceed by analyzing the observed outcome only, without modeling of the follow-up visit process. Fitzmaurice et al. (2006) extened their method to longitudinal binary data. Ryu et al. (2007) presented a Bayesian regression method of jointly modeling the follow-up visit process and the longitudinal outcome process through introducing a subject-specific latent variable for studies when both precesses are of interest. They also proposed a novel generalization of a cross-validated Bayesian procedure for model diagnostics to check whether the ignorability assumption in Lipsitz et al. (2002) is appropriate for a study. For marginal regression analysis, Lin et al. (2004)

pointed out that ignoring the dependency between follow-up times and outcomes would lead to biased estimation. In particular, for a marginal regression of the conditional mean of the longitudinal outcome, they modeled the follow-up visit process by a proportional intensity model (Andersen and Gill, 1982) and proposed an inverse intensity weighted regression approach. When the follow-up visit process is correctly modeled, their marginal regression estimator has been shown to be consistent. However, to obtain a consistent estimator of the baseline intensity function of the follow-up visit process requires kernel smoothing. To avoid estimating the baseline intensity function, Buzkova and Lumley (2007) further proposed a class of inverse intensity-ratio weighted estimators which is simple in computation and moreover, can be applied under mixture of continuous and discrete follow-up visit times.

Quantile regression for longitudinal data have been investigated by several literatures (Jung, 1996; Lipsitz et al., 1997; Koenker, 2004; Geraci and Bottai, 2007; Wang and Fygenson, 2009; Yi and He, 2009; Yuan and Yin, 2010; Lee and Kong, 2013). For example, Lipsitz et al. (1997) studied regression models for marginal quantiles and adopted estimating equations treating repeated outcomes as "independent". They further proposed inverse probability weighted estimators that account for missing at random dropouts. Koenker (2004) considered quantile regression models with subject-specific fixed effects which are intended to capture unobserved individual heterogeneity and proposed an ℓ_1 regularization estimating method to modify the inflation effect caused by the introduction of individual fixed effects. Wang and Fygenson (2009) investigated quantile regression for longitudinal outcomes that are left censored by fixed constants and developed inference procedures accouting for both censoring and intra-subject dependency. Lee and Kong (2013) presents a marginal quantile regression procedure for longitudinal data subject to left censoring and dropouts.

However, all quantile regression methods are restricted to situations when subjects have a common set of visit time points with or without monotone dropouts. These are not applicable to longitudinal data with intermittent dropouts or irregular follow-up. The goal of my second and third projects are to develop appropriate quantile regression procedures for longitudinal data with outcomes subject to left censoring plus intermittent informative dropouts or outcome-dependent follow-up.

1.3 Outline

In Chapter 2, we present two proposed estimation methods for window observed recurrent event data under the accelerated recurrence time model. First, we proposed a two-stage estimation procedure which yields a consistent initial estimator first and then derives a more efficient second stage estimator from an augmented estimating equation. Asymptotical properties, uniform consistency and uniform weak convergence, are established for the resulting estimators. Second, we proposed an estimation procedure by utilizing a mean zero stochastic process associated with recurrent event counting process. This new method enables more efficient and stable computation as compared to existing methods. We derive the asymptotic properties of the proposed estimator, and develop inference procedures. Results from simulations demonstrate good finite-sample performance of the proposed methods. We illustrate the second approach via an application to the CFFPR data.

In Chapter 3, we investigate quantile regression for longitudinal data with left censored outcomes subject to missing resulted from intermittent subject dropout. We adopt an appropriate censored quantile regression technique to handle left censoring and employ the idea of inverse probability weighting to tackle the issue associated with informative intermittent missing mechanism. Asymptotic properties are established for the proposed estimator. We evaluate our method by simulation studies. The proposed method is applied to the Michigan PBB study to investigate the PBB decay profile.

In Chapter 4, we develop a censored quantile regression procedure for longitudinal data with irregular outcome-dependent follow-up. We adopt a proportional intensity model for the follow-up visit process and propose an inverse intensity-ratio weighted least absolute deviation estimator in censored quantile regression model. This approach yields consistent estimates of the quantile regression parameters provided that the model for the follow-up visit process is correctly specified. We evaluate our method by simulation studies. The proposed method is also applied to the Michigan PBB study to investigate the PBB decay profile.

In Chapter 5, we provide a summary of our completed work and plans for future work.

Chapter 2

Accelerated Recurrence Time Analysis of Recurrent Events Data Observed in a Time Window

2.1 Regression Procedures

2.1.1 Data and Model

Let $\mathbf{T} = \{T^{(1)}, T^{(2)}, \ldots\}$ be the recurrent event times. Let $\mathbf{Z}$ be the associated $p \times 1$ vector of covariates. The corresponding counting process of the recurrent event is denoted by $N(t) = \sum_{j=1}^{\infty} I(T^{(j)} \leq t)$. Let $\{L, R\}$ be the first and last follow-up time for $\mathbf{T}$. Denote the window observed counting process by $\tilde{N}(t) = \sum_{j=1}^{\infty} I(L \leq T^{(j)} \leq (R \wedge t))$. The observed data consists of $\{\tilde{N}(\cdot), L, R, \mathbf{Z}\}$. It is assumed that $N(\cdot)$ and $\{L, R\}$ are independent conditionally on $\mathbf{Z}$.

The mean function, defined as $\mu_{\mathbf{Z}}(t) = E(N(t)|\mathbf{Z})$, is of interest. Its inverse function, $\tau_{\mathbf{Z}}(u) = \inf\{t : \mu_{\mathbf{Z}}(t) > u\}$, represents time to expected frequency u. The accelerated recurrence time model takes the form that

$$\tau_{\mathbf{Z}}(u) = \exp\left(\mathbf{X}^T \boldsymbol{\beta}_0(u)\right), \tag{2.1}$$

where $u > 0$ and $\mathbf{X} = (1, \mathbf{Z}^T)^T$.

2.1.2 Two-Stage Estimation

Estimation Procedure

With the observation window starting from zero, Huang and Peng (2009) proposed an estimating equation that

$$E\left[\mathbf{X} I\left(R \geq \exp\left(\mathbf{X}^T \boldsymbol{\beta}_0(u)\right)\right)\left\{N\left(\exp\left(\mathbf{X}^T \boldsymbol{\beta}_0(u)\right)\right) - u\right\}\right] = 0.$$

However, with $L > 0$, $N\left(\exp\left(\mathbf{X}^T \boldsymbol{\beta}_0(u)\right)\right)$ is not always observable because of missing observation before the follow-up starts. If we restrict our analysis on the subsample with $L = 0$, too much information was lost especially when most of subjects have delayed starting time. Therefore, to extend usable sample size, we propose a baseline frequency point strategy. Define a baseline frequency point v, less than u. The definition of $\tau_{\mathbf{Z}}(\cdot)$ implies that $E\left[N\left\{\tau_{\mathbf{Z}}(u)\right\} - N\left\{\tau_{\mathbf{Z}}(v)\right\}\right] = u - v$. Restricting the estimating equation on the subsample with $N\left(\tau_{\mathbf{Z}}(u)\right) - N\left(\tau_{\mathbf{Z}}(v)\right)$ observed, we have

$$\begin{aligned} 0 =& E\Bigg[\mathbf{X} I\left\{L \le \exp\left(\mathbf{X}^T \boldsymbol{\beta}_0(v)\right)\right\} I\left\{\exp\left(\mathbf{X}^T \boldsymbol{\beta}_0(u)\right) \le R\right\} \\ & \times \left\{\sum_{j=1}^{\infty} I\left\{\exp\left(\mathbf{X}^T \boldsymbol{\beta}_0(v)\right) < T^{(j)} \le \exp\left(\mathbf{X}^T \boldsymbol{\beta}(u)\right)\right\} - (u - v)\right\}\Bigg]. \qquad (2.2) \end{aligned}$$

The only problem here is that $\boldsymbol{\beta}_0(v)$ is unknown. We note that $0 = \tau_{\mathbf{Z}}(0) = \exp\left(\mathbf{X}^T \boldsymbol{\beta}_0(0)\right)$ and $v < u$. Hence, we propose obtaining $\hat{\boldsymbol{\beta}}(u)$ by sequentially solving the following estimating equation for $\boldsymbol{\beta}(u)$ plugging in $\hat{\boldsymbol{\beta}}(v)$ as $\boldsymbol{\beta}_0(v)$:

$$\begin{aligned} 0 =& \sqrt{n} \boldsymbol{\Phi}(\boldsymbol{\beta}(u), u, v, \hat{\boldsymbol{\beta}}(v)) \\ =& n^{-1/2} \sum_{i=1}^{n} \mathbf{X}_i I\left\{\exp\left(\mathbf{X}_i^T \boldsymbol{\beta}(u)\right) \le R_i\right\} I\left\{L_i \le \exp\left(\mathbf{X}_i^T \hat{\boldsymbol{\beta}}(v)\right)\right\} \\ & \times \left[\sum_{j=1}^{\infty} I\left\{\exp\left(\mathbf{X}_i^T \hat{\boldsymbol{\beta}}(v)\right) < T_i^{(j)} \le \exp\left(\mathbf{X}_i^T \boldsymbol{\beta}(u)\right)\right\} - (u - v)\right]. \qquad (2.3) \end{aligned}$$

The rule of selecting a baseline point, adopted in the following simulations and real analysis, is that $v(u) = \max\{v_h : v_h < u, h = 1, \ldots, H\}$, where $0 = v_0 < v_1 < \cdots < v_H < U$ is a sequence of equally spaced baseline points. Other choices of $v(u)$ are possible but need to satisfy that $v(u) < u$.

The solution-finding problem in (2.3) is equivalent to locating the minimizer of the following objective function,

$$
\begin{aligned}
&\sqrt{n}\Psi(\boldsymbol{\beta}(u), u, v, \hat{\boldsymbol{\beta}}(v)) \\
=&n^{-1/2}\sum_{i=1}^{n}\left[\sum_{j=1}^{\infty}\left(\mathbf{X}_i^T\beta \wedge \log(R_i) - \log(T_i^{[j]})\right)^{+} I\left\{\exp\left(\mathbf{X}_i^T\hat{\boldsymbol{\beta}}(v)\right) \leq T_i^{[j]} \leq R_i\right\}\right. \\
&\left. - \left(\mathbf{X}_i^T\beta \wedge \log(R_i)\right)(u-v)\right] \times I\left\{L_i \leq \exp\left(\mathbf{X}_i^T\hat{\beta}(v)\right)\right\}.
\end{aligned}
$$

However, some arbitrariness can be involved in the selection of baseline points. A careless choice may lead to low efficiency in estimation and even large bias especially when the sample size is small or moderate. In estimating equation (2.2), time to expected frequency u is compared with time to expected preselected baseline frequency, v. If v is too small, there may be only few subjects satisfying $L \leq \tau_{\mathbf{Z}}(v)$; if v is too large, the time interval between $\tau_{\mathbf{Z}}(v)$ and $\tau_{\mathbf{Z}}(u)$ would be very narrow. Our empirical rule of selecting the equally spaced baseline points is that $v_1 \times \{\# \text{ of subjects with } L = 0\} \approx 10$.

In the second stage, we propose to improve the performance through an augmented estimating equation. The left end of the time interval is broaden, from time to expected baseline frequency, $\tau_{\mathbf{Z}}(v) \geq L$, to the follow-up starting time, L. The stochastic estimating equation, which utilizes the event information between L and $\tau_{\mathbf{Z}}(u)$, is

$$
\begin{aligned}
0 =&\sqrt{n}\boldsymbol{\Delta}(\boldsymbol{\beta}; u) = n^{-1/2}\sum_{i=1}^{n}\left[\mathbf{X}_i I\left(\exp\left(\mathbf{X}_i^T\boldsymbol{\beta}(u)\right) \leq R_i\right)\right. \\
&\left.\times\left\{\sum_{j=1}^{\infty} I\left(L_i \leq T_i^{(j)} \leq \exp\left(\mathbf{X}_i^T\boldsymbol{\beta}(u)\right)\right) - (u - u \wedge \tilde{\mu}_{\mathbf{Z}_i}(L_i; \boldsymbol{\beta}))\right\}\right], \qquad (2.4)
\end{aligned}
$$

where $\tilde{\mu}_{\mathbf{Z}}(L; \boldsymbol{\beta}) = \int_0^{\infty} I(L > \exp(\mathbf{X}^T\boldsymbol{\beta}(u)))du$ represents the expectation of the recurrent events number before the follow-up starts. Note that the parameter $\boldsymbol{\beta}$ here is a function on $(0, U]$, not a single vector.

The problem here is that the equation $\sqrt{n}\mathbf{\Delta}(\boldsymbol{\beta}; u) = 0$ is not easy to solve. That's why we consider a modification

$$\sqrt{n}\tilde{\mathbf{\Delta}}(\mathbf{a}; u, \boldsymbol{\beta}) = n^{-1/2} \sum_{i=1}^{n} \left[\mathbf{X}_i I\left(\exp\left(\mathbf{X}_i^T \mathbf{a}\right) \le R_i\right) \right.$$
$$\left. \times \left\{ \sum_{j=1}^{\infty} I\left(L_i \le T_i^{(j)} \le \exp\left(\mathbf{X}_i^T \mathbf{a}\right)\right) - \left(u - u \wedge \tilde{\mu}_{\mathbf{Z}_i}(L_i; \boldsymbol{\beta})\right) \right\} \right]. \tag{2.5}$$

We propose a cadlag estimator $\hat{\boldsymbol{\beta}}(u)$ jumping only on a equally-spaced grid, $\mathcal{S}_{K(n)} = \{0 = u_0 < u_1 < \cdots < u_{K(n)} = U\}$. So $\tilde{\mu}_{\mathbf{Z}}\left(L; \hat{\boldsymbol{\beta}}\right) = \int_0^\infty I\left(L > \exp\left(\mathbf{X}^T \hat{\boldsymbol{\beta}}(u)\right)\right) du = \sum_{k=0}^{K(n)} \frac{U}{K(n)} I\left(L > \exp\left(\mathbf{X}^T \hat{\boldsymbol{\beta}}(u_k)\right)\right)$. We propose the following iterative algorithm of estimating $\boldsymbol{\beta}_0$.

1. Let $\hat{\boldsymbol{\beta}}_{(0)}$ be the initial estimator from estimating equation (2.3). Estimate the recurrent events number before the follow-up starts $\tilde{\mu}_{\mathbf{Z}}\left(L; \hat{\boldsymbol{\beta}}_{(0)}\right)$ denoted by $\tilde{\mu}_{(0)}$. set $m = 1$;

2. Find $\hat{\boldsymbol{\beta}}_{(m)}$ by solving $\sqrt{n}\tilde{\mathbf{\Delta}}\left(\mathbf{a}; u, \hat{\boldsymbol{\beta}}_{(m-1)}\right) = 0$;

3. Estimate $\tilde{\mu}_{(m)} = \tilde{\mu}\left(L; \hat{\boldsymbol{\beta}}_{(m)}\right)$;

4. Increase m by 1 and go back to step 2 until the convergence criteria, i.e. $\left\|\hat{\boldsymbol{\beta}}_{(m)} - \hat{\boldsymbol{\beta}}_{(m-1)}\right\| < 10^{-10}$ and $\left\|\tilde{\mu}_{(m)} - \tilde{\mu}_{(m-1)}\right\| < 10^{-10}$ is met.

Denote the converged estimator as $\hat{\boldsymbol{\beta}}$. In step 2, the solution to $\sqrt{n}\tilde{\mathbf{\Delta}}(\mathbf{a}; u, \hat{\boldsymbol{\beta}}_{(m-1)}) = 0$ is the same as the minimizer of $\mathbf{a}$ in the following objective function:

$$\sqrt{n}\Theta(\mathbf{a}; u, \hat{\boldsymbol{\beta}}_{(m-1)}) = n^{-1/2} \sum_{i=1}^{n} \left[\sum_{j=1}^{\infty} \left(\mathbf{X}_i^T \mathbf{a} \wedge \log R_i - \log T_i^{(j)}\right)^+ I(L_i \le T_i^{(j)} \le R_i) \right.$$
$$\left. - \left(\mathbf{X}_i^T \mathbf{a} \wedge \log R_i\right) \left\{u - u \wedge \tilde{\mu}_{\mathbf{Z}_i}\left(L_i; \hat{\boldsymbol{\beta}}_{(m-1)}\right)\right\} \right].$$

This algorithm shares similar idea with the EM algorithm. At each iteration, the expectation of recurrent events number before the follow-up starts based on estimates from previous iteration is plugged in to the objective function to obtain the updated estimates.

Asymptotic Results

We can show that $\hat{\boldsymbol{\beta}}_{(0)}$ is consistent. Based on that, we can prove that $\hat{\boldsymbol{\beta}}$ is consistent although the original estiamting equation $\sqrt{n}\boldsymbol{\Delta}(\boldsymbol{\beta}; u) = 0$ may contain inconsistent roots. We also want to point out that $\hat{\boldsymbol{\beta}}_{(m)}$ ieself is a legitimate estimator for any m.

The regularity conditions include:

C 1. For $u \in (0, U]$, $E\left[\mathbf{X}^{\otimes 2} I\left(L \leq \exp(\mathbf{X}^T \boldsymbol{\beta}_0(v(u)))\right) I\left(R > \exp(\mathbf{X}^T \boldsymbol{\beta}_0(u))\right)\right]$ is non-singular.

C 2. $\boldsymbol{\beta}_0(u)$ is a Lipchitz continuous function and in the interior of a compact and bounded space $\mathcal{B}$ for all $u \in (0, U]$.

C 3. $\|\mathbf{Z}\|$ is bounded.

C 4. $\sum_{j=1}^{\infty} I(T^{(j)} \leq R)$ is bounded.

C 5. (L, R), given $\mathbf{Z}$, has a bounded conditional density function $f_{L,R|\mathbf{Z}}(l, r)$ at $\{(l, r) = (\tau_{\mathbf{Z}}(u_1),\tau_{\mathbf{Z}}(u_2)) : u_1, u_2 \in (0, U]\}$, for all $\mathbf{Z}$.

C 6. $\dot{\mu}_{\mathbf{Z}}(t) = d\mu_{\mathbf{Z}}(t)/dt$ is continuous and bounded at $\{\tau_{\mathbf{Z}}(u) : u \in (0, U]\}$ for all $\mathbf{Z}$.

C 7. $\inf_{u \in (0,U]} eigminE\left[\mathbf{X}^{\otimes 2} \dot{\mu}(\tau_{\mathbf{Z}}(u))\tau_{\mathbf{Z}}(u) I\left(L \leq \tau_{\mathbf{Z}}(v(u))\right) I\left(R \geq \tau_{\mathbf{Z}}(u)\right)\right] > 0$.

We establish the uniform consistency and weak convergence of $\hat{\boldsymbol{\beta}}(\tau)$ stated in the following theorems.

Theorem 2.1.1. *Under conditions C1-C6,* $\sup_{u \in (0,U]} \left\|\hat{\boldsymbol{\beta}}_{(0)}(u) - \boldsymbol{\beta}_0(u)\right\| \xrightarrow{p} 0$.

Theorem 2.1.2. *Under conditions C1-C7,* $n^{1/2}\left\{\hat{\boldsymbol{\beta}}_{(0)} - \boldsymbol{\beta}_0\right\}$ *converges weakly to a Gaussian process with mean 0 and covariance matrix* Σ*, where* Σ *is presented in the proof (2.5 Appendix).*

Theorem 2.1.3. *Under conditions C1-C6,* $\sup_{u\in(0,U]}\left\|\hat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}_0(u)\right\| \xrightarrow{p} 0.$

Theorem 2.1.4. *Under conditions C1-C7,* $n^{1/2}\left\{\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right\}$ *converges weakly to a Gaussian process with mean 0 and covariance matrix* Σ*, where* Σ *is presented in the proof (2.5 Appendix).*

The proof of all theorems are presented in section 2.5 Appendix.

Inference

Inference about $\hat{\boldsymbol{\beta}}(u)$ are important for scientific conclusions; however, as seen in proof of Theorem 2.1.4, the covariance matrix of $n^{1/2}\left\{\hat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}_0(u)\right\}$ is complicated and not available. Therefore, we adopt the resampling approach of Jin et al. (2001) to estimate the covariance matrix. Let v_i, $i = 1, \dots, n$ be i.i.d. from a nonnegative distribution of unit mean and unit variance, e.g., an exponential distribution of unit rate. We consider perturbed estimating equations

$$\begin{aligned}
0 =& n^{-1/2}\sum_{i=1}^{n} v_i \mathbf{X}_i I\left\{\exp\left(\mathbf{X}_i^T \boldsymbol{\beta}(u)\right) \le R_i\right\} I\left\{L_i \le \exp\left(\mathbf{X}_i^T \hat{\boldsymbol{\beta}}(v(u))\right)\right\} \\
&\times \left[\sum_{j=1}^{\infty} I\left\{T_i^{(j)} \in \left(\exp\left(\mathbf{X}_i^T \hat{\boldsymbol{\beta}}(v(u))\right), \exp\left(\mathbf{X}_i^T \boldsymbol{\beta}(u)\right)\right]\right\} - (u - v(u))\right],
\end{aligned} \tag{2.6}$$

and

$$\begin{aligned}
0 =& n^{-1/2}\sum_{i=1}^{n} v_i \Bigg[\mathbf{X}_i I\left(\exp\left(\mathbf{X}_i^T \boldsymbol{\beta}(u)\right) \le R_i\right) \\
&\times \left\{\sum_{j=1}^{\infty} I\left(T_i^{(j)} \in \left(L_i, \exp\left(\mathbf{X}_i^T \boldsymbol{\beta}(u)\right)\right]\right) - (u - u \wedge \tilde{\mu}_{\mathbf{Z}_i}(L_i; \boldsymbol{\beta}))\right\}\Bigg].
\end{aligned} \tag{2.7}$$

Solving perturbed estimating equations (2.6) and (2.7) following the same two-stage estimating procedure, we could obtain a new estimator denoted by $\hat{\boldsymbol{\beta}}^*(\cdot)$. It can be shown that the distribution of $n^{1/2}\left\{\hat{\boldsymbol{\beta}}^*(u) - \hat{\boldsymbol{\beta}}(u)\right\}$ conditionally on the data is the same as $n^{1/2}\left\{\hat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}_0(u)\right\}$. Thus, we can approximate the distribution of $n^{1/2}\left\{\hat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}_0(u)\right\}$ by a simulated distribution of $n^{1/2}\left\{\hat{\boldsymbol{\beta}}^*(u) - \hat{\boldsymbol{\beta}}(u)\right\}$. Pointwise confidence interval for $\boldsymbol{\beta}_0(\cdot)$ can be obtained by the Wald method,

$$\hat{\boldsymbol{\beta}}(u) \pm \Phi_{0.975}^{-1} SE\left\{\hat{\boldsymbol{\beta}}^*(u)\right\},$$

where $SE\left\{\hat{\boldsymbol{\beta}}^*(u)\right\}$ is the empirical standard error of $\left\{\hat{\boldsymbol{\beta}}^*(u) - \hat{\boldsymbol{\beta}}(u)\right\}$ and $\Phi_{0.975}^{-1}$ is the 97.5th quantile from standard normal distribution.

2.1.3 Estimator Based on Counting Process

Estimation Procedure

Our key idea to estimate $\boldsymbol{\beta}_0(u)$ is based on the fact that

$$E\Big\{\mathbf{X}(\tilde{N}(\exp(\mathbf{X}^T\boldsymbol{\beta}_0(u))) - \int_0^u I(L \leq \exp(\mathbf{X}^T\boldsymbol{\beta}_0(s) \leq R)ds)\Big\} = 0, \qquad (2.8)$$

where $\mathbf{X} = (1, \mathbf{Z}^T)^T$.

According to (2.8), the estimation of $\boldsymbol{\beta}_0(u)$ only includes $\{\boldsymbol{\beta}_0(s) : s < u\}$. This motivates us a grid-based estimation procedure for $\boldsymbol{\beta}_0(u)$ sequentially from $u = 0$ to the above limit of interest, say U. Define a grid $S_{L(n)} = \{0 = u_0 < u_1 < \cdots < u_{L(n)} = U\}$. Our proposed estimator $\hat{\boldsymbol{\beta}}(u)$ is a right-continuous piecewise-constant function that jumps only at grid $S_{L(n)}$. Note that $0 = \tau_{\mathbf{Z}}(0) = \exp(\mathbf{X}^T\boldsymbol{\beta}_0(0))$; therefore, we always set $\exp(\mathbf{X}^T\hat{\boldsymbol{\beta}}(0)) = 0$. We propose to obtain $\hat{\boldsymbol{\beta}}(u_k)$ $(k = 1, 2, \ldots, L(n))$ by sequentially solving the following monotone estimating equation for $\boldsymbol{\beta}(u_k)$:

$$n^{1/2}\mathbf{S}_n(\boldsymbol{\beta}, u) = n^{-1/2}\sum_{i=1}^{n}\mathbf{X}_i\Big\{\tilde{N}_i(\exp(\mathbf{X}_i^T\boldsymbol{\beta}(u_k))) - \sum_{m=0}^{k-1} I(L_i \le \exp(\mathbf{X}_i^T\hat{\boldsymbol{\beta}}(u_m)) \le R_i)(u_{m+1}-u_m)\Big\} = 0. \tag{2.9}$$

Because (2.9) is not continuous, an exact root may not exist and the proposed estimator $\hat{\boldsymbol{\beta}}(u_k)$ are defined as generalized solutions (Fygenson and Ritov 1994).

The monotonicity of (2.9) greatly facilitates the computation. It implies that all generalized solutions belong to a convex set and that the left side of (2.9) is the gradient of a convex function. The solution-finding problem for (2.9) is equivalent to locating the minimizer of the following L_1-type convex objective function:

$$\begin{aligned} l_j(\mathbf{h}) = &\sum_{i=1}^{n}\sum_{j=1}^{\infty} I(L_i \le T_i^{(j)} \le R_i)\left|\log T_i^{(j)} - \mathbf{X}_i^T\mathbf{h}\right| \\ &+ \left|R^* - (\sum_{i=1}^{n}\sum_{j=1}^{\infty} I(L_i \le T_i^{(j)} \le R_i)(-\mathbf{X}_i))^T\mathbf{h}\right| \\ &+ \left|R^* - (\sum_{i=1}^{n} 2\mathbf{X}_i\sum_{m=0}^{k-1} I(L_i \le \exp(\mathbf{X}_i^T\hat{\boldsymbol{\beta}}(u_m)) \le R_i)(u_{m+1}-u_m))^T\mathbf{h}\right|, \end{aligned}$$

where R^* is a very large number and $j = 1, \ldots, L(n)$. The foregoing minimization problem can be easily solved using the Barrodale-Roberts algorithm (Barrodale and Roberts 1974), the implementation of which is available in standard statistical software, for example, the $l1fit()$ function in S-PLUS or the $rq()$ function in R package $quantreg$.

Asymptotic Results

The proposed estimator $\hat{\boldsymbol{\beta}}(\cdot)$ has the properties of uniform consistency and weak convergence under some regularity conditions. Define $F_{L|\mathbf{Z}}(l) = Pr(L \leq l|\mathbf{Z})$, $F_{R|\mathbf{Z}} = Pr(R \leq r|\mathbf{Z})$, $f_{L|\mathbf{Z}}(l) = dF_{L|\mathbf{Z}}(l)/dl$, and $f_{R|\mathbf{Z}}(r) = dF_{R|\mathbf{Z}}(r)/dr$. The regularity conditions are as follows:

C 1. $\mathcal{Z}$ is compact, i.e., $\sup_i \|\mathbf{Z}\| < \infty$.

C 2. (a) Each component of $E\left[\mathbf{X}\sum_{j=1}^{\infty} I\left(L \leq T^{(j)} \leq \exp\left(\mathbf{X}^T\boldsymbol{\beta}_0(u)\right) \wedge R\right)\right]$ is a Lipschitz function of u, and (b) $f_{L|\mathbf{Z}}(t)$ and $f_{R|\mathbf{Z}}(t)$ are bounded above uniformly in t and $\mathbf{Z}$.

C 3. (a) $E\left[I\left(L \leq \exp\left(\mathbf{X}^T\mathbf{b}\right) \leq R\right)\dot{\mu}_{\mathbf{Z}}\left(\exp\left(\mathbf{X}^T\mathbf{b}\right)\right)|\mathbf{Z}\right] > 0$ for any $\mathbf{b} \in \mathcal{B}(d_0)$, (b) $E\left(\mathbf{Z}^{\otimes 2}\right)$ is positive definite, and (c) each component of $E[\mathbf{X}^{\otimes 2}\exp\left(\mathbf{X}^T\mathbf{b}\right)$ $\{f_{L|\mathbf{Z}}(\exp(\mathbf{X}^T\mathbf{b})) - f_{R|\mathbf{Z}}(\exp(\mathbf{X}^T\mathbf{b}))\}]\{E[\mathbf{X}^{\otimes 2}\exp(\mathbf{X}^T\mathbf{b})I(L \leq \exp(\mathbf{X}^T\mathbf{b}) \leq R)$ $\dot{\mu}_{\mathbf{Z}}(\exp(\mathbf{X}^T\mathbf{b}))]\}^{-1}$ is uniformly bounded in $\mathbf{b} \in \mathcal{B}(d_0)$, where $\mathcal{B}(d_0)$ is a neighborhood containing $\{\boldsymbol{\beta}_0(u), u \in (0, U]\}$, defined in Appendix A.

C 4. $\inf_{u\in[v,U]} eigmin E\{I(L \leq \exp(\mathbf{X}^T\boldsymbol{\beta}_0(u)) \leq R)\dot{\mu}_{\mathbf{Z}}(\exp(\mathbf{X}^T\boldsymbol{\beta}_0(u)))\exp(\mathbf{X}^T\boldsymbol{\beta}_0(u))$ $\mathbf{X}^{\otimes 2}\} > 0$ for any $v \in (0, U]$, where $eigmin(\cdot)$ denotes the minimum eigenvalue of a matrix.

C 5. $N_i(t)$ is bounded above for all $i = 1, \ldots, n$.

We have the following theorems.

Theorem 2.1.5. *Assuming that conditions C1-C4 hold, if* $\lim_{n\to\infty}\|S_{L(n)}\| = 0$, *then* $\sup_{u\in[v,U]}\|\hat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}_0(u)\| \xrightarrow{p} 0$, *where* $0 < v < U$.

Theorem 2.1.6. *Assuming that conditions C1-C6 hold, if* $\lim_{n\to\infty} n^{1/2}\|S_{L(n)}\| = 0$, *then* $n^{1/2}\{\hat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}_0(u)\}$ *converges weakly to a Gaussian process for* $u \in [v, U]$, *where* $0 < v < U$.

The proof of all theorems are presented in section 2.5 Appendix.

Inference

Similarly, we adopt the resampling approach of Jin et al. (2001) to estimate the covariance matrix. Let v_i, $i = 1, \ldots, n$ be i.i.d. from a nonnegative distribution of unit mean and unit variance, e.g., an exponential distribution of unit rate. We consider a perturbed estimating equation

$$n^{-1/2} \sum_{i=1}^{n} v_i \mathbf{X}_i \Big\{ \tilde{N}_i(\exp(\mathbf{X}_i^T \boldsymbol{\beta}(u_k))) - \sum_{m=0}^{k-1} I(L_i \leq \exp(\mathbf{X}_i^T \hat{\boldsymbol{\beta}}(u_m)) \leq R_i)(u_{m+1} - u_m) \Big\} = 0. \qquad (2.10)$$

The perturbed estimating equation (2.10) could be solved by minimizing an objective function using $rq()$ function in R. Denote a new estimator denoted by $\hat{\boldsymbol{\beta}}^*(\cdot)$. We can approximate the distribution of $n^{1/2}\{\hat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}_0(u)\}$ by a simulated distribution of $n^{1/2}\{\hat{\boldsymbol{\beta}}^*(u) - \hat{\boldsymbol{\beta}}(u)\}$. Pointwise confidence interval for $\boldsymbol{\beta}_0(\cdot)$ can be obtained by the Wald method,

$$\hat{\boldsymbol{\beta}}(u) \pm \Phi^{-1}_{0.975} SE\{\hat{\boldsymbol{\beta}}^*(u)\},$$

where $SE\{\hat{\boldsymbol{\beta}}^*(u)\}$ is the empirical standard error of $\{\hat{\boldsymbol{\beta}}^*(u) - \hat{\boldsymbol{\beta}}(u)\}$ and $\Phi^{-1}_{0.975}$ is the 97.5th quantile from standard normal distribution.

2.2 Simulation Studies

Finite-sample performance of the proposed method is evaluated through Monte Carlo simulations similar to Huang and Peng (2009). A Gamma frailty on a standard homogeneous Poisson process was applied to generate recurrent events. Variance of the Gamma frailty, σ^2, determines the level of intra-individual correlation. We considered $\sigma^2 = 0$ and 0.5, where 0 corresponds to zero intra-individual correlation. Two covaraites, Z_1 and Z_2, were generated from $Bernoulli(0.5)$ and $Uniform(-0.5, 0.5)$ respectively. A reccurrent event time sequence was generated by

$$T^{(j)} = \exp\left\{\min\left(1, \frac{T^{*(j)}}{1.5\gamma}\right) Z_1 + Z_2\right\} T^{*(j)}/\gamma, \qquad j = 1, 2, \ldots,$$

where $\{T^{*(j)}, j = 1, 2, \ldots\}$ is a recurrent event sequence generated from a standard homogeneous Poisson process and frailty γ was generated from a Gamma distribution. It can be shown that under this setup,

$$\tau_{\mathbf{Z}}(u) = \exp\left\{\log(u) + \min(1, \frac{u}{1.5}) Z_1 + Z_2\right\}.$$

Covariate Z_1 had an increasing effect and Covariate Z_2 had a constant effect. The first visit time, L, was generated from $w \cdot Unif(0, 1)$, where $w \sim Bernoulli(0.8)$. We included w to ensure L had a probability mass of 0.2 at zero and render a scenario that the low tail of $\boldsymbol{\beta}_0$ was identifiable. The last follow-up visit time, R was generated from distribution $Unif(L, 12)$. This observation window resulted in an average of 4 observed recurrent events.

Under each configuration, we generated 500 datasets of sample size $n = 100$. For interval estimation and inference, the resampling size of 100 was chosen. An equally spaced grid on $u \in (0, 3]$ with size 0.02 was adopted when estimating $\boldsymbol{\beta}_0$. For two-stage estimators, baseline points v's were chosen as $\{0.5k : k = 0, 1, \ldots, 5\}$. The rule of selecting a baseline point is $v(u) = \max\{0.5k : 0.5k < u, k = 0, 1, \ldots, 5\}$. The estimator from estimating equation (2.3) is referred as the initial estimator and the estimator from the iterative algorithm is referred as the iterative estimator.

Simulation results are summarized in figures. Figure 2.1 and Figure 2.2 present simulation results from the set-up with $\sigma^2 = 0$ and the set-up with $\sigma^2 = 0.5$ respectively. In the first row, we plots the empirical bias from the stage-one estimator (initial, doted line), the stage-two estimator (iterative, dashed line), and the estimator based on counting process (sequential, solid line), versus expected frequency u. It shows that the iterative estimator and the counting process based estimator have smaller bias compared to the initial estimator. The plots in the second row depicts the empirical mean squared errors (MSE) versus expected frequency u. The iterative estimator and the counting process based estimator have smaller MSE than the initial estimator. The third row presents the coverage probability of 95% confidence intervals (CI) obtained from resampling of the iterative estimator and the counting process based estimator. The resulting 95% CIs are slightly under-covered and yet have coverage probabilities fairly close to the nominal value. In the last row, we plot the relative efficiency of the counting process based estimator over the iterative estimator. The effciency gain seems to increase with the expected frequency and can be over 40% at some large u.

We also compare the estimation efficiency between the proposed estimator and Huang and Peng (2009)'s estimator assuming that the observation window starts from zero. We set L = 0 while keeping $(T^{(j)}; R; \mathbf{Z})$ generated from the same way. In Figure 2.3, we plot the relative efficiency of the counting process based restimator to Huang and Peng (2009)'s estimator. It is shown that the new proposal for estimating the ART model is always more efficient than the method of Huang and Peng (2009). The efficiency gain seems to increase with the expected frequency and can be over 100% at some large u. This finding is consistent with Koenker (2008)'s empirical results about the efficiency comparison between Peng and Huang (2008)'s and Powell (1984, 1986)'s methods on censored quantile regression.

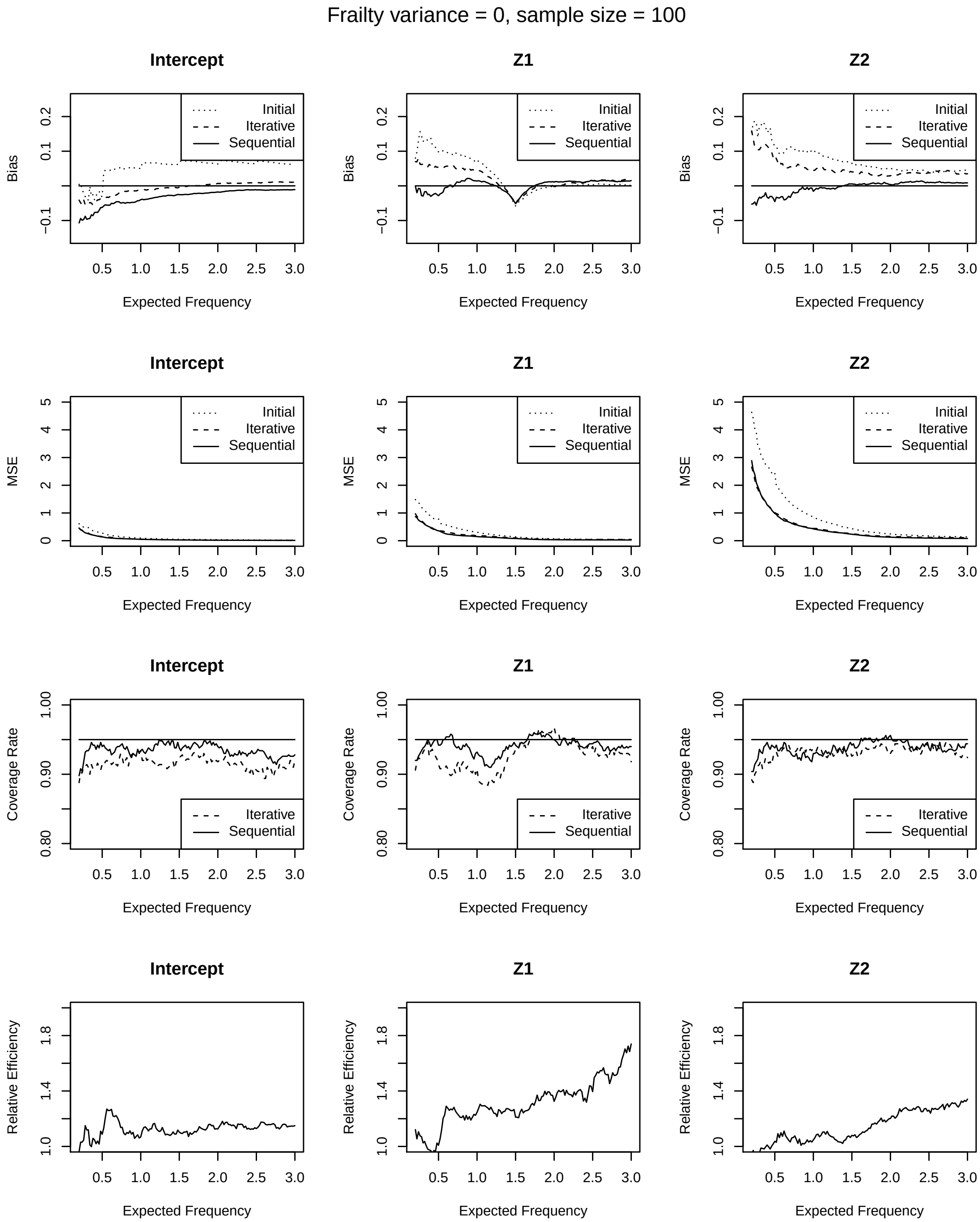

Figure 2.1: Bias, MSE, coverage rate and relative efficiency of the counting process based estimator compared with the stage-two estimator; Gamma frailty = 1; sample size = 100.

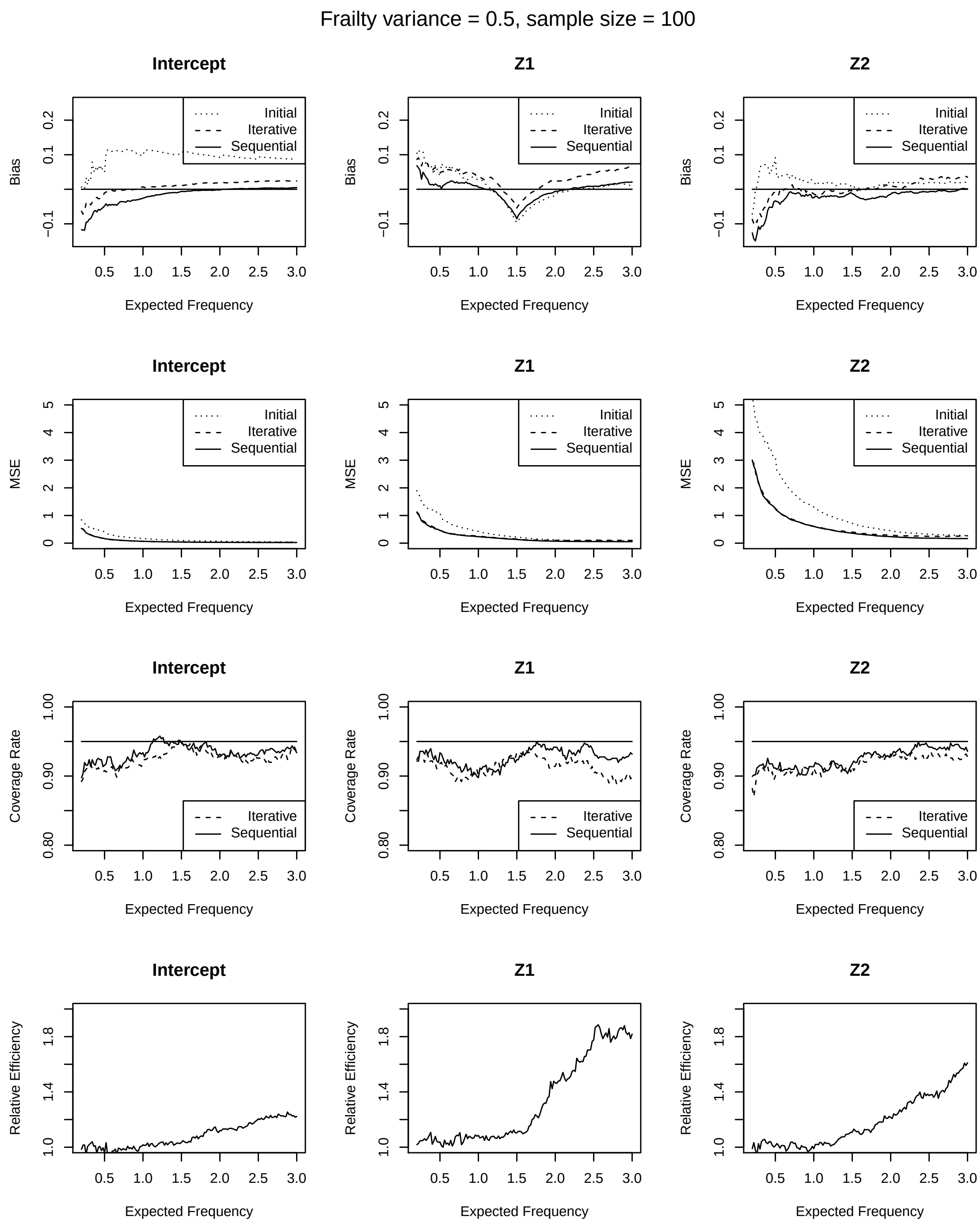

Figure 2.2: Bias, MSE, coverage rate and relative efficiency of the counting process based estimator compared with the stage-two estimator; Gamma frailty variance = 0.5; sample size = 100.

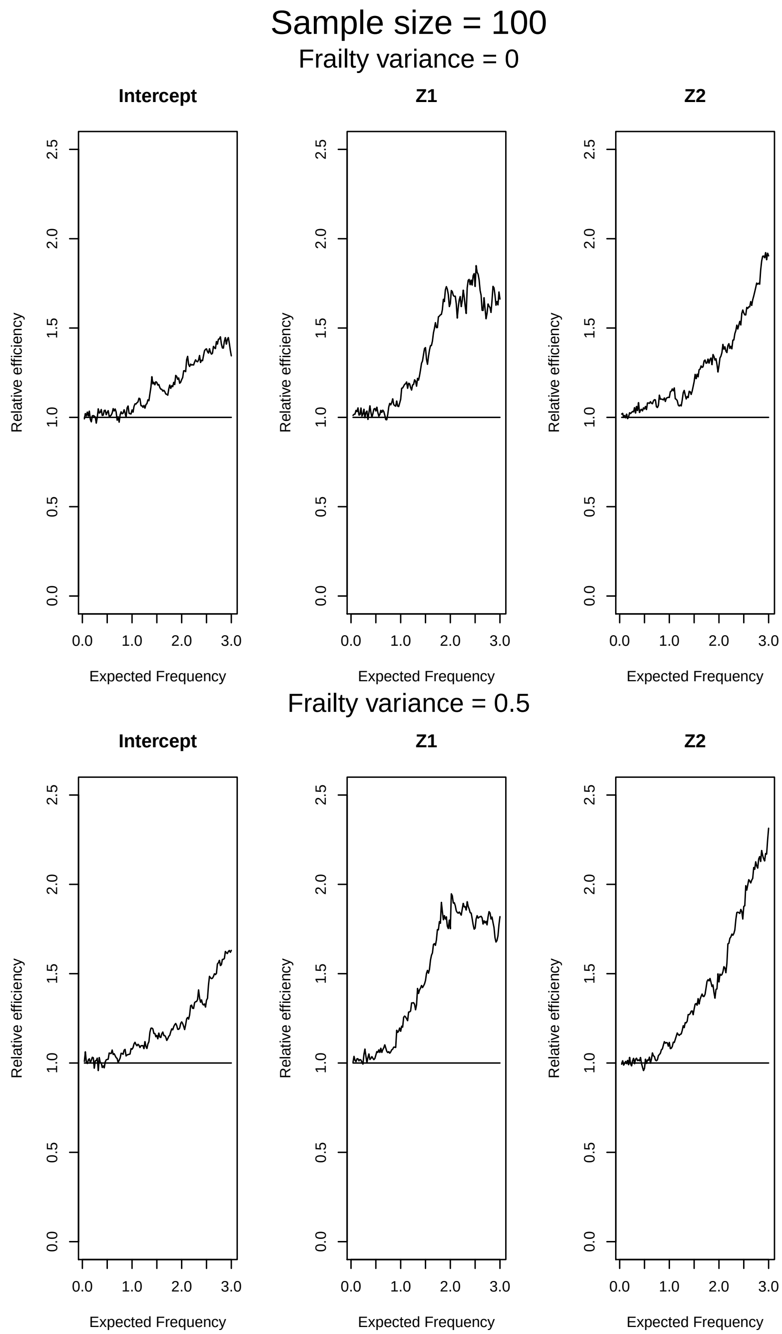

Figure 2.3: Simulation results on the efficiency of the proposed counting process based estimator relative to Huang and Peng (2009)'s estimator

2.3 CFFPR Data Example

Cystic Fibrosis (CF) is a life-limiting genetic disease without known cure yet, affecting about 30,000 people in the United States (Cystic Fibrosis Foundation, 2011). For CF patients, lung infections will result in damaged and lower lung function. Pseudomonas aeruginosa (PA), the most important pathogen that shortens survival of CF patients, infects more than half of people with CF (Cystic Fibrosis Foundation, 2011). Characterizing the timing of PA infections and assessing how it is influenced by potential risk factors can help make treatment decisions and are thus of scientific interests. To address these questions, we utilized the data from 2875 children documented in 1986-2008 CFF Patient Registry (CFFPR). All these children were born in or after 1998, with ΔF508 mutation, and had at least 5 year follow-up in the registry.

We applied the proprosed estimator based on the mean-zero stochastic process to this CFFPR dataset. The recurrent event time $T^{(j)}$ is the age of a CF child when s/he had the jth PA infection. Due to late diagnosis or late entry to the study after diagnosis, some CF children had delayed CFFPR entries after birth. Time from birth to registry entry constitutes the follow-up starting time L in our method framework. In this dataset, age at the first CFFPR visit ranges from 0 to 5.7 years with mean=0.7 years and median=0.4 years. The number of positive PA cultures at CFFPR visits ranges from 0 to 50; the mean and median number of PA infections are 3.9 and 2 respectively. We considered risk factors including sex, patient's CFTR gene classification (I=ΔF508 homozygous, II=ΔF508 heterozygous), meconium ileus (MI), and pancreatic insufficiency. The covariates for a subject are coded as *Female*, 1 if the subject was female and 0 otherwise, *F508/Other*, 1 if the subject was ΔF508 heterozygous and 0 otherwise, *MI*, 1 if the subject was diagnosed by MI and 0 otherwise, and *Pancreat*, 1 if the subject was pancreatic insufficient (defined as never on enzyme) and 0 otherwise.

Figure 2.4 displays the estimated coefficients (solid lines) along with the 95%

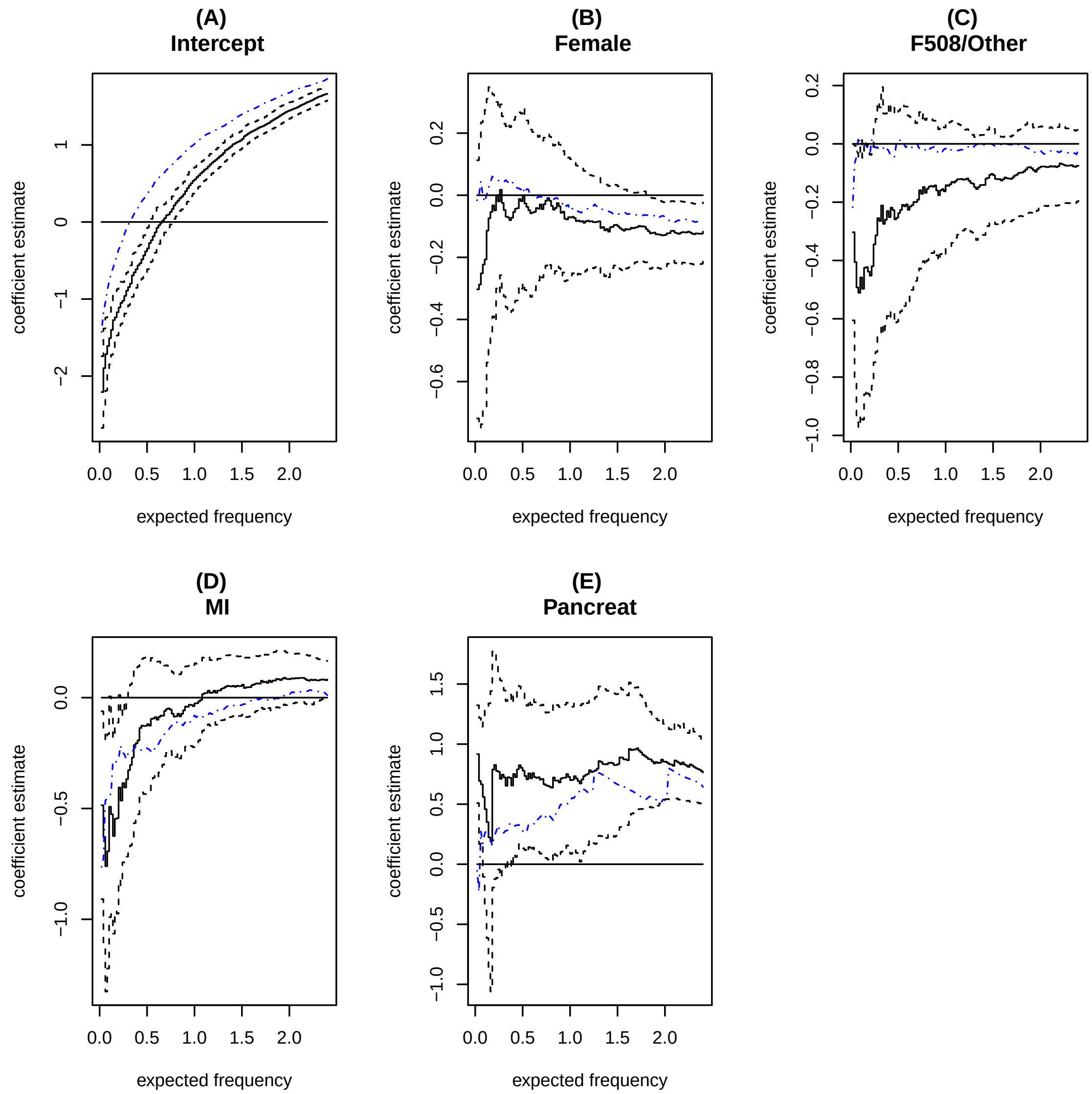

Figure 2.4: Coefficient estimates (solid lines) and 95% pointwise confidence intervals (dotted lines) from the proposed method; the coefficient estimates from Huang and Peng (2009)'s method (dash dotted line)

pointwise confidence intervals (dashed lines). The intercept (panel A) represents the estimated log time to expected frequency of PA infections for the reference group, i.e., CF boys with homozygous F508del mutations, had no MI, and pancreatic insufficient. The non-intercept coefficient estimates (panels B-E) plot the estimated effects of covariates, which are allowed to be frequency-varying. Negative coefficient estimates indicate sooner progression to recurrence of PA infections. To better summarize the varying covariate effect estimates, we present the average covariate effects in the frequency intervals (0.4, 1.4], (1.4, 2.4], and (0.4, 2.4] respectively in Table 2.1. The estimates strongly suggest that CF children with pancreatic insufficiency tend to experience recurrent PA infections at earlier ages than CF children with pancreatic sufficiency. CF girls have marginal increased risk at later recurrent PA infections. The average effect estimates for MI demonstrate a cross-over pattern, changing from -0.37 to 0.74, though not reaching statistical significance in either frequency interval. There is not enough evidence to show a significant difference between F508del homozygous group and F508del heterozygous group. We also conducted constancy tests for each covariate effect. MI has a frequency-dependent effect on the timing of PA recurrence, while other covariates displayedmore constant effects over the frequency of PA infections.

In Figure 2.4, we also plot the coefficient estimate obtained from applying Huang and Peng (2009)'s method assuming that the observation window starts from zero (dashed dotted lines). intercepts estimated by Huang and Peng (2009)'s method are significantly larger than those from the proposed method. This observation conforms to the intuition that naively treating PA infection frequency before registry entry as 0 would lead to over-optimistic estimates for time to expected frequency.

2.4 Remarks

The accelerated recurrent model offers a useful and flexible alternative to current approaches for analyzing recurrent events data. In this project, we propose a two estimation procedures for the accelerated recurrent time model when recurrent events data is only observed in a random time window. In addition, we require L and R to be always observed, which is often true in registry study setting. Both estimators are

Table 2.1: The CFFPR example: average covariate effect estimates along with standard errors (SE) and the corresponding p values, and the p values from constancy tests.

Average Covariate Effect Estimates					
Frequency Interval		*Sex*	*F508/Other*	*MI*	*Pancreat*
(0.4, 1.4]	Estimate	-0.06	-0.17	-0.04	0.74
	SE	0.10	0.12	0.09	0.29
	P value	0.55	0.15	0.69	0.01
(1.4, 2.4]	Estimate	-0.11	-0.10	0.07	0.86
	SE	0.06	0.07	0.06	0.21
	P value	0.05	0.19	0.21	<.001
(0.4, 2.4]	Estimate	-0.09	-0.13	0.02	0.80
	SE	0.08	0.09	0.08	0.24
	P value	0.28	0.16	0.81	< .001
Constancy Tests					
Constancy tests	P value	0.32	0.26	0.06	0.50

consistent and asymptotically normal. The second estimator based on a mean-zero stochastic process is more efficient and easier in implementation, which have been shown by simulation studies.

2.5 Appendix

2.5.1 Proof of Theorems

Lemma 1. Define $\psi(\mathbf{a}, u, v, \mathbf{b}) = E\{\Psi(\mathbf{a}, u, v, \mathbf{b})\}$. Given h, and $u \in (v_h, v_{h+1}]$, $\psi(\mathbf{a}, u, v_h, \boldsymbol{\beta}_0(v_h))$ has a unique minimizer at $\mathbf{a} = \boldsymbol{\beta}_0(u)$, under condition(a).

Proof: Define

$$B(\mathbf{X}, L, R; \mathbf{a}, u, v, \mathbf{b}) = E\Bigg\{ I\left(L \le \exp(\mathbf{X}^T\mathbf{b})\right) \Bigg[\sum_{j=1}^{\infty} (\mathbf{X}^T\mathbf{a} \wedge \log R - \log T^{(j)})^+$$

$$\times I(\mathbf{X}^T\mathbf{b} \le \log T^{(j)} \le \log R) - (\mathbf{X}^T\mathbf{a} \wedge \log R)(u - v) \Bigg] \Bigg| \mathbf{X}, L, R \Bigg\}$$

Since we have

$$
\begin{aligned}
&\psi(\mathbf{a}, u, v_h, \boldsymbol{\beta}_0(v_h)) \\
&= E[B(\mathbf{X}, L, R; \mathbf{a}, u, v_h, \boldsymbol{\beta}_0(v_h)) I(L > \exp(\mathbf{X}^T\boldsymbol{\beta}_0(v_h)))] \\
&+ E[B(\mathbf{X}, L, R; \mathbf{a}, u, v_h, \boldsymbol{\beta}_0(v_h)) I\{L \leq \exp(\mathbf{X}^T\boldsymbol{\beta}_0(v_h)), \min(\mathbf{X}^T\boldsymbol{\beta}_0(u), \mathbf{X}^T\mathbf{a}) \geq \log R\}] \\
&+ E[B(\mathbf{X}, L, R; \mathbf{a}, u, v_h, \boldsymbol{\beta}_0(v_h)) I\{L \leq \exp(\mathbf{X}^T\boldsymbol{\beta}_0(v_h)), \mathbf{X}^T\boldsymbol{\beta}_0(u) \geq \log R > \mathbf{X}^T\mathbf{a}\}] \\
&+ E[B(\mathbf{X}, L, R; \mathbf{a}, u, v_h, \boldsymbol{\beta}_0(v_h)) I\{L \leq \exp(\mathbf{X}^T\boldsymbol{\beta}_0(v_h)), \mathbf{X}^T\boldsymbol{\beta}_0(u) < \log R\}],
\end{aligned}
$$

we could prove it case by case.

- When $L > \exp\left\{\mathbf{X}^T\boldsymbol{\beta}_0(v_h)\right\}$, $B(\mathbf{X}, L, R; \mathbf{a}, u, v_h, \boldsymbol{\beta}_0(v_h)) = B(\mathbf{X}, L, R; \boldsymbol{\beta}_0(u), u, v_h, \boldsymbol{\beta}_0(v_h)) = 0.$
- When $L \leq \exp(\mathbf{X}^T\boldsymbol{\beta}_0(v_h))$,

  1. When $\mathbf{X}^T\boldsymbol{\beta}_0(u) \geq \log R$ and $\mathbf{X}^T\mathbf{a} \geq \log R$,

  $$
  \begin{aligned}
  &B(\mathbf{X}, L, R; \mathbf{a}, u, v, \boldsymbol{\beta}_0(v_h)) = B(\mathbf{X}, L, R; \boldsymbol{\beta}_0(u), u, v, \boldsymbol{\beta}_0(v_h)) \\
  =&E\left\{ \sum_{j=1}^{\infty} (\log R - \log T^{(j)})^+ I(\mathbf{X}^T\boldsymbol{\beta}_0(v_h) \leq \log T^{(j)} \leq \log R) \right. \\
  &\left. - \log R(u - v_h) \middle| \mathbf{X}, L, R \right\}
  \end{aligned}
  $$

  2. When $\mathbf{X}^T\boldsymbol{\beta}_0(u) \geq \log R$ and $\mathbf{X}^T\mathbf{a} < \log R$,consider

  $$
  \begin{aligned}
  f(y; \mathbf{X}, L, R) =&E\left\{ \sum_{j=1}^{\infty} I(\mathbf{X}^T\boldsymbol{\beta}_0(v_h) \leq \log T^{(j)} \leq \log R) \right. \\
  &\left. \times (y \wedge \log R - \log T^{(j)})^+ - (y \wedge \log R)(u - v_h) \middle| \mathbf{X}, L, R \right\}.
  \end{aligned}
  $$

$$
\begin{aligned}
&\frac{\partial f(y;\mathbf{X},L,R)}{\partial y}\\
=&E\left\{\sum_{j=1}^{\infty} I(\mathbf{X}^T\boldsymbol{\beta}_0(v_h)\leq \log T^{(j)}\leq y<\log R)\right.\\
&\left.-I(y<\log R)(u-v_h)\middle|\mathbf{X},L,R\right\}\\
\leq&E\left\{I(y<\log R)\left[\sum_{j=1}^{\infty} I(\mathbf{X}^T\boldsymbol{\beta}_0(v_h)\leq \log T^{(j)}\leq \log R)\right.\right.\\
&\left.\left.-(u-v_h)\right]\middle|\mathbf{X},L,R\right\}\\
\leq&E\left\{I(y<\log R)\left[\sum_{j=1}^{\infty} I(\mathbf{X}^T\boldsymbol{\beta}_0(v_h)\leq \log T^{(j)}\leq \mathbf{X}^T\boldsymbol{\beta}_0(u))\right.\right.\\
&\left.\left.-(u-v_h)\right]\middle|\mathbf{X},L,R\right\}\\
=&0
\end{aligned}
$$

Since $\mathbf{X}^T\mathbf{a} < \mathbf{X}^T\boldsymbol{\beta}_0(u)$, we have $B(\mathbf{X},L,R;\mathbf{a},u,v_h,\boldsymbol{\beta}_0(v_h)) = f(\mathbf{X}^T\mathbf{a})$ $\geq f(\mathbf{X}^T\boldsymbol{\beta}_0(u)) = B(\mathbf{X},L,R;\boldsymbol{\beta}_0(u),u,v_h,\boldsymbol{\beta}_0(v_h))$.

3. When $\mathbf{X}^T\boldsymbol{\beta}_0(u) < \log R$,

$$
\begin{aligned}
&\frac{\partial f(y;\mathbf{X},L,R)}{\partial y}\\
=&E\Bigg\{\sum_{j=1}^{\infty} I(\mathbf{X}^T\boldsymbol{\beta}_0(v_h)\leq \log T^{(j)}\leq y<\log R)\\
&-I(y<\log R)(u-v_h)\Bigg|\mathbf{X},L,R\Bigg\}\\
=&E\Bigg\{I(y<\log R)\Bigg[\sum_{j=1}^{\infty} I(\mathbf{X}^T\boldsymbol{\beta}_0(v_h)\leq \log T^{(j)}\leq y)\\
&-(u-v_h)\Bigg]\Bigg|\mathbf{X},L,R\Bigg\}\\
=&I(y<\log R)\left[I(y\geq \mathbf{X}^T\boldsymbol{\beta}_0(v_h))(\mu(\exp(y))-v_h)-(u-v_h)\right]\\
&\begin{cases}
<0 & \text{if } y<\mathbf{X}^T\boldsymbol{\beta}_0(u);\\
=0 & \text{if } y=\mathbf{X}^T\boldsymbol{\beta}_0(u);\\
>0 & \text{if } \mathbf{X}^T\boldsymbol{\beta}_0(u)<y<\log R.\\
=0 & \text{if } y\geq \log R
\end{cases}
\end{aligned}
$$

So $y=\mathbf{X}^T\boldsymbol{\beta}_0(u)$ is the unique minimizor of $f(y;\mathbf{X},L,R)$ which means

$$
B(\mathbf{X},L,R;\mathbf{a},u,v_h,\boldsymbol{\beta}_0(v_h))\geq B(\mathbf{X},L,R;\boldsymbol{\beta}_0(u),u,v_h,\boldsymbol{\beta}_0(v_h)),
$$

where the equality holds if and only if $\mathbf{X}^T\mathbf{a}=\mathbf{X}^T\boldsymbol{\beta}_0(u)$.

In summary, for any $u\in(v_h,v_{h+1}]$, $\psi(\mathbf{a},u,v_h,\boldsymbol{\beta}_0(v_h))\geq\psi(\boldsymbol{\beta}_0(u),u,v_h,\boldsymbol{\beta}_0(v_h))$. And under condition (a), we have $Pr\{\mathbf{X}^T(\boldsymbol{\beta}_0(u)-\mathbf{a})I(L\leq\exp(\mathbf{X}^T\boldsymbol{\beta}_0(v_h)))I(\log R>\mathbf{X}^T\boldsymbol{\beta}_0(u))\neq 0\}>0$ for any $\mathbf{a}\neq\boldsymbol{\beta}_0(u)$. Thus, strict inequlity holds that

$$
\begin{aligned}
&E\left\{B(\mathbf{X},L,R;\mathbf{a},u,v_h,\boldsymbol{\beta}_0(v_h))I\left(L\leq\exp(\mathbf{X}^T\boldsymbol{\beta}_0(v_h))\right)I\left(\mathbf{X}^T\boldsymbol{\beta}_0(u)<\log R\right)\right\}\\
>&E\left\{B(\mathbf{X},L,R;\boldsymbol{\beta}_0(u),u,v_h,\boldsymbol{\beta}_0(v_h))I\left(L\leq\exp(\mathbf{X}^T\boldsymbol{\beta}_0(v_h))\right)I\left(\mathbf{X}^T\boldsymbol{\beta}_0(u)<\log R\right)\right\}.
\end{aligned}
$$

Hence,$\psi(\mathbf{a}, u, v_h, \boldsymbol{\beta}_0(v_h)) > \psi(\boldsymbol{\beta}_0(u), u, v_h, \boldsymbol{\beta}_0(v_h))$ for any $\mathbf{a} \neq \boldsymbol{\beta}_0(u)$. Lemma 1 is proved.

Proof of *Theorem 2.1.1*: $\boldsymbol{\beta}_0(u)$ is a Lipchitz function, so there exist a finite number C_0 that for any $u_1, u_2 \in (0, U]$, $\|\boldsymbol{\beta}_0(u_1) - \boldsymbol{\beta}_0(u_2)\| \leq C_0|u_1 - u_2|$. If we could prove that $\max_{u_k: k=1,\ldots,K(n)} \|\hat{\boldsymbol{\beta}}(u_k) - \boldsymbol{\beta}_0(u_k)\| \xrightarrow{p} 0$, then

$$
\begin{aligned}
&\sup_{u\in(0,U]} \|\hat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}_0(u)\| \\
&\leq \max_{k=1,\ldots,K(n)} \|\hat{\boldsymbol{\beta}}(u_k) - \boldsymbol{\beta}_0(u_k)\| + \sup_{u\in(0,U]} \|\boldsymbol{\beta}_0(u) - \boldsymbol{\beta}_0(v(u))\| \\
&\leq \max_{k=1,\ldots,K(n)} \|\hat{\boldsymbol{\beta}}(u_k) - \boldsymbol{\beta}_0(u_k)\| + C_0 \sup_{u\in(0,U]} |u - v(u)| \\
&\leq \max_{k=1,\ldots,K(n)} \|\hat{\boldsymbol{\beta}}(u_k) - \boldsymbol{\beta}_0(u_k)\| + C_0 \cdot \frac{U}{K(n)} \\
&\xrightarrow{p} 0
\end{aligned}
$$

as $n \to \infty$ and $K(n) \to \infty$. Therefore, to prove theorem 1, we only need to prove that $\max_{k=1,\ldots,K(n)} \|\hat{\boldsymbol{\beta}}(u_k) - \boldsymbol{\beta}_0(u_k)\| \xrightarrow{p} 0$.

Denote

$$
\begin{aligned}
&\varphi(\mathbf{X}, L, R, \mathbf{T}; \mathbf{a}, u, v, \mathbf{b}) = I(L \leq \exp(\mathbf{X}^T\mathbf{b}))\times \\
&\Bigg\{ \sum_{j=1}^{\infty} \left[\left(\mathbf{X}^T\mathbf{a} - \log T^{(j)}\right)^+ + \left(\log R - \log T^{(j)}\right)^+ \right] (\mathbf{X}^T\mathbf{b} \leq \log T^{(j)} \leq \log R) \\
&- \sum_{j=1}^{\infty} \left(\mathbf{X}^T\mathbf{a} \vee \log R - \log T^j\right)^+ \left(\mathbf{X}^T\mathbf{b} \leq \log T^{(j)} \leq \log R\right) \\
&- (\mathbf{X}^T\mathbf{a} \wedge \log R)(u - v) \Bigg\}
\end{aligned}
$$

Since linear functions, concave and convex functions are Glivenko-Cantelli (G-C) classes and the sum or product of G-C classes are also G-C classes, $\varphi(\mathbf{X}, L, R, \mathbf{T}; \mathbf{a}, u, v, \mathbf{b})$ is a G-C class with index $\mathbf{a}$, u, v, and $\mathbf{b}$ under condition (b). This fact coupled with pointwise convergence by the strong law of large numbers implies the convergence of $\Psi(\mathbf{a}, u, v, \mathbf{b})$ to $\psi(\mathbf{a}, u, v, \mathbf{b})$ uniformly in $\mathbf{a}$, u, v,and $\mathbf{b}$ (Andersen & Gill, 1982)

1. For $u_k \in (0, v_1]$, it can be proved as Theorem 3. in Huang and Peng's paper (2009) that $\max_{u_k \in (0,v_1]} \|\hat{\boldsymbol{\beta}}(u_k) - \boldsymbol{\beta}(u_k)\| \xrightarrow{a.s.} 0$.

2. For $u_k \in (v_h, v_{h+1}]$, $h = 1, 2, \ldots, H$, we need to prove $\max_{u_k \in (v_h,v_{h+1}]} \|\hat{\boldsymbol{\beta}}(u_k) - \boldsymbol{\beta}(u_k)\| \xrightarrow{p} 0$ given $\hat{\boldsymbol{\beta}}(v_h) \xrightarrow{a.s.} \boldsymbol{\beta}_0(v_h)$.

According to Glivenko-Cantelli theorem,

$$\sup_{\mathbf{a} \in \mathcal{B}, u \in (0,U], v \in (0,U], \mathbf{b} \in \mathcal{B}} |\Psi(\mathbf{a}, u, v, \mathbf{b}) - \psi(\mathbf{a}, u, v, \mathbf{b})| \xrightarrow{a.s.} 0$$

which implies

$$\sup_{\boldsymbol{\beta} \in \mathcal{B}, u_k \in (v_h, v_{h+1}], \hat{\boldsymbol{\beta}}(v_h) \in \mathcal{B}} |\Psi(\boldsymbol{\beta}, u_k, v_h, \hat{\boldsymbol{\beta}}(v_h)) - \psi(\boldsymbol{\beta}, u_k, v_h, \hat{\boldsymbol{\beta}}(v_h))| \xrightarrow{a.s.} 0. \tag{2.11}$$

Under condition (c), (d), (e) and (f), given $\hat{\boldsymbol{\beta}}(v_h) \xrightarrow{a.s.} \boldsymbol{\beta}_0(v_h)$,

$$\sup_{\boldsymbol{\beta} \in \mathcal{B}, u_k \in (v_h, v_{h+1}], \hat{\boldsymbol{\beta}}(v_h) \in \mathcal{B}} |\psi(\boldsymbol{\beta}, u_k, v_h, \hat{\boldsymbol{\beta}}(v_h)) - \psi(\boldsymbol{\beta}, u_k, v_h, \boldsymbol{\beta}_0(v_h))| \xrightarrow{a.s.} 0. \tag{2.12}$$

$$
\begin{aligned}
0 \leq & \psi(\hat{\boldsymbol{\beta}}(u_k), u_k, v_h, \boldsymbol{\beta}_0(v_h)) - \psi(\boldsymbol{\beta}_0(u_k), u_k, v_h, \boldsymbol{\beta}_0(v_h)) \\
= & \psi(\hat{\boldsymbol{\beta}}(u_k), u_k, v_h, \boldsymbol{\beta}_0(v_h)) - \psi(\hat{\boldsymbol{\beta}}(u_k), u_k, v_h, \hat{\boldsymbol{\beta}}(v_h)) \\
& + \psi(\hat{\boldsymbol{\beta}}(u_k), u_k, v_h, \hat{\boldsymbol{\beta}}(v_h)) - \psi(\boldsymbol{\beta}_0(u_k), u_k, v_h, \hat{\boldsymbol{\beta}}(v_h)) \\
& + \psi(\boldsymbol{\beta}_0(u_k), u_k, v_h, \hat{\boldsymbol{\beta}}(v_h)) - \psi(\boldsymbol{\beta}_0(u_k), u_k, v_h, \boldsymbol{\beta}_0(v_h)) \\
= & \psi(\hat{\boldsymbol{\beta}}(u_k), u_k, v_h, \boldsymbol{\beta}_0(v_h)) - \psi(\hat{\boldsymbol{\beta}}(u_k), u_k, v_h, \hat{\boldsymbol{\beta}}(v_h)) \qquad (2.13) \\
& + \psi(\hat{\boldsymbol{\beta}}(u_k), u_k, v_h, \hat{\boldsymbol{\beta}}(v_h)) - \Psi(\hat{\boldsymbol{\beta}}(u_k), u_k, v_h, \hat{\boldsymbol{\beta}}(v_h)) \qquad (2.14) \\
& + \Psi(\hat{\boldsymbol{\beta}}(u_k), u_k, v_h, \hat{\boldsymbol{\beta}}(v_h)) - \Psi(\boldsymbol{\beta}_0(u_k), u_k, v_h, \hat{\boldsymbol{\beta}}(v_h)) \qquad (2.15) \\
& + \Psi(\boldsymbol{\beta}_0(u_k), u_k, v_h, \hat{\boldsymbol{\beta}}(v_h)) - \psi(\boldsymbol{\beta}_0(u_k), u_k, v_h, \hat{\boldsymbol{\beta}}(v_h)) \qquad (2.16) \\
& + \psi(\boldsymbol{\beta}_0(u_k), u_k, v_h, \hat{\boldsymbol{\beta}}(v_h)) - \psi(\boldsymbol{\beta}_0(u_k), u_k, v_h, \boldsymbol{\beta}_0(v_h)) \qquad (2.17)
\end{aligned}
$$

Since (2.15) is less than or equal to zero,

$$
\begin{aligned}
& \max_{u_k \in (v_h, v_{h+1}]} |\psi(\hat{\boldsymbol{\beta}}(u_k), u_k, v_h, \boldsymbol{\beta}_0(v_h)) - \psi(\boldsymbol{\beta}_0(u_k), u_k, v_h, \boldsymbol{\beta}_0(v_h))| \\
\leq & \max_{u_k \in (v_h, v_{h+1}]} |(2.13) + (2.14) + (2.16) + (2.17)| \\
\leq & \max_{u_k \in (v_h, v_{h+1}]} |(2.13)| + \max_{u_k \in (v_h, v_{h+1}]} |(2.14)| + \max_{u_k \in (v_h, v_{h+1}]} |(2.16)| + \max_{u_k \in (v_h, v_{h+1}]} |(2.17)|
\end{aligned}
$$

According to (2.12), $\max_{u_k \in (v_h, v_{h+1}]} |(2.13)|$ and $\max_{u_k \in (v_h, v_{h+1}]} |(2.17)|$ converges to zero almost surely. According to (4.10), $\max_{u_k \in (v_h, v_{h+1}]} |(2.14)|$ and $\max_{u_k \in (v_h, v_{h+1}]} |(2.16)|$ converges to zero almost surely. Hence, we have $\max_{u_k \in (v_h, v_{h+1}]} |\psi(\hat{\boldsymbol{\beta}}(u_k), u_k, v_h, \boldsymbol{\beta}_0(v_h)) - \psi(\boldsymbol{\beta}_0(u_k), u_k, v_h, \boldsymbol{\beta}_0(v_h))| \xrightarrow{a.s.} 0$.

Next, we followed Huang & Peng (2009)'s proof of theorem 3. If $\max_{u_k \in (v_h, v_{h+1}]} \|\hat{\boldsymbol{\beta}}(u_k) - \boldsymbol{\beta}_0(u_k)\| \stackrel{a.s.}{\not\longrightarrow} 0$, then there exist an $\epsilon > 0$ and a sequence of $\{u, \boldsymbol{\zeta}\}$, satisfying that $\|\psi(\boldsymbol{\zeta}, u, v_h, \boldsymbol{\beta}_0(v_h)) - \psi(\boldsymbol{\beta}_0(u), u, v_h, \boldsymbol{\beta}_0(v_h))\| \to 0$ and $\|\boldsymbol{\zeta} - \boldsymbol{\beta}_0(u)\| > \epsilon$. There must exist a subsequence converging to $\{u^*, \boldsymbol{\zeta}^*\}$. Since $\psi(\mathbf{a}, u, v, \mathbf{b})$ is a continuous function in $\mathbf{a}$ and u and $\boldsymbol{\beta}_0(u)$ is continuous in u, we have $\psi(\boldsymbol{\zeta}^*, u^*, v_h, \boldsymbol{\beta}_0(v_h)) = \psi(\boldsymbol{\beta}_0(u^*), u^*, v_h, \boldsymbol{\beta}_0(v_h))$ and $\boldsymbol{\zeta}^* \neq \boldsymbol{\beta}_0(u^*)$ which contradicts with the fact that $\boldsymbol{\beta}_0(u^*)$ is the unique minimizor.

Since H is a finite number, we could conclude that $\max_{k=1,2,\ldots,K(n)} \|\hat{\boldsymbol{\beta}}(u_k) - \boldsymbol{\beta}_0(u_k)\| \xrightarrow{a.s.} 0$. Proof is completed.

Lemma 2. Denote $\boldsymbol{\phi}(\mathbf{a}, u, v, \mathbf{b}) = E\{\boldsymbol{\Phi}(\mathbf{a}, u, v, \mathbf{b})\}$. Given $\sup_{u\in[v_h,v_{h+1}]} \left\|\hat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}_0(u)\right\| \xrightarrow{p} 0$,

$$\sup_{u\in(v_h,v_{h+1}]} \Bigg\| \sqrt{n} \Bigg\{ \boldsymbol{\Phi}(\hat{\boldsymbol{\beta}}(u), u, v_h, \hat{\boldsymbol{\beta}}(v_h)) - \boldsymbol{\Phi}(\boldsymbol{\beta}_0(u), u, v_h, \boldsymbol{\beta}_0(v_h)) \\ - \boldsymbol{\phi}(\hat{\boldsymbol{\beta}}(u), u, v_h, \hat{\boldsymbol{\beta}}(v_h)) + \boldsymbol{\phi}(\boldsymbol{\beta}_0(u), u, v_h, \boldsymbol{\beta}_0(v_h)) \Bigg\} \Bigg\| \xrightarrow{p} 0.$$

Proof: Denote

$$\begin{aligned} &\mathbf{A}_i(\mathbf{a}, u, v, \mathbf{b}) \\ &= I(L_i \le \exp(\mathbf{X}_i^T \mathbf{b})) I(\mathbf{X}_i^T \mathbf{a} \le \log R_i) \mathbf{X}_i \left[N(\exp(\mathbf{X}_i^T \mathbf{a})) - N(\exp(\mathbf{X}_i^T \mathbf{b})) - (u - v) \right] \end{aligned}$$

Under the conditions (b) - (f), there exist finite numbers M_1, M_2, M_3, M_4, M_5 that

- $\|\mathbf{X}\| \le M_1$;
- $\sum_{j=1}^{\infty} I(T^{(j)} \le R) \le M_2$;
- $\dot{\mu}_{\mathbf{Z}}(t) \le M_3$;
- $f_{R|\mathbf{Z}}(t) \le M_4$ and $f_{L|\mathbf{Z}}(t) \le M_4$;
- $\|\frac{\partial \exp(\mathbf{X}^T \mathbf{a})}{\partial \mathbf{a}}\| \le M_5$ for any $\mathbf{a} \in \mathcal{B}$.

$$
\begin{aligned}
& E\|\mathbf{A}_i(\mathbf{a}_1, u, v, \mathbf{b}) - \mathbf{A}_i(\mathbf{a}_2, u, v, \mathbf{b})\|^2 \\
=& E\Bigg\{ I(L_i \le \exp(\mathbf{X}_i \mathbf{b}))\mathbf{X}_i \\
& \times \Bigg[I\left(\mathbf{X}_i^T\mathbf{a}_1 \le \log R_i, \mathbf{X}_i^T\mathbf{a}_2 \le \log R_i\right) \left(N\left(\exp(\mathbf{X}_i^T\mathbf{a}_1)\right) - N\left(\exp(\mathbf{X}_i^T\mathbf{a}_2)\right)\right) \\
& + I\left(\mathbf{X}_i^T\mathbf{a}_1 \le \log R_i, \mathbf{X}_i^T\mathbf{a}_2 > \log R_i\right) \left(N\left(\exp(\mathbf{X}_i^T\mathbf{a}_1)\right) - N\left(\exp(\mathbf{X}_i^T\mathbf{b})\right) - u + v\right) \\
& - I\left(\mathbf{X}_i^T\mathbf{a}_1 > \log R_i, \mathbf{X}_i^T\mathbf{a}_2 \le \log R_i\right) \left(N\left(\exp(\mathbf{X}_i^T\mathbf{a}_2)\right) - N\left(\exp(\mathbf{X}_i^T\mathbf{b})\right) - u + v\right) \\
& - I\left(\mathbf{X}_i^T\mathbf{a}_1 > \log R_i, \mathbf{X}_i^T\mathbf{a}_2 > \log R_i\right) \cdot 0 \Bigg]\Bigg\}^2 \\
\le& M_1^2 \cdot M_2 \cdot M_3 \cdot M_5 \|\mathbf{a}_1 - \mathbf{a}_2\| \\
& + M_1^2 \cdot M_2^2 \cdot M_4 \cdot M_5 \|\mathbf{a}_1 - \mathbf{a}_2\| \\
& + M_1^2 \cdot M_2^2 \cdot M_4 \cdot M_5 \|\mathbf{a}_1 - \mathbf{a}_2\| \\
=& C_1 \cdot \|\mathbf{a}_1 - \mathbf{a}_2\|
\end{aligned}
$$

where $C_1 = M_1^2 M_2 M_3 M_5 + 2M_1^2 M_2^2 M_4 M_5$.

$$
\begin{aligned}
& E\|\mathbf{A}_i(\mathbf{a}, u, v, \mathbf{b}_1) - \mathbf{A}_i(\mathbf{a}, u, v, \mathbf{b}_2)\|^2 \\
=& E\Bigg[I(R_i \ge \exp(\mathbf{X}_i \mathbf{a}))\mathbf{X}_i \\
& \times \Bigg\{ I\left(L_i \le \min(\exp(\mathbf{X}_i^T\mathbf{b}_1), \exp(\mathbf{X}_i^T\mathbf{b}_2))\right) \left(N\left(\exp(\mathbf{X}_i^T\mathbf{b}_2)\right) - N\left(\exp(\mathbf{X}_i^T\mathbf{b}_1)\right)\right) \\
& + I\left(\exp(\mathbf{X}_i^T\mathbf{b}_2) < L_i \le \exp(\mathbf{X}_i^T\mathbf{b}_1)\right) \left(N\left(\exp(\mathbf{X}_i^T\mathbf{a})\right) - N\left(\exp(\mathbf{X}_i^T\mathbf{b}_1)\right) - u + v\right) \\
& - I\left(\exp(\mathbf{X}_i^T\mathbf{b}_1) < L_i \le \exp(\mathbf{X}_i^T\mathbf{b}_2)\right) \left(N\left(\exp(\mathbf{X}_i^T\mathbf{a})\right) - N\left(\exp(\mathbf{X}_i^T\mathbf{b}_2)\right) - u + v\right) \\
& - I\left(L_i > \max(\exp(\mathbf{X}_i^T\mathbf{b}_1), \exp(\mathbf{X}_i^T\mathbf{b}_2))\right) \cdot 0 \Bigg\}\Bigg]^2 \\
\le& M_1^2 \cdot M_2 \cdot M_3 \cdot M_5 \|\mathbf{b}_1 - \mathbf{b}_2\| \\
& + M_1^2 \cdot M_2^2 \cdot M_4 \cdot M_5 \|\mathbf{b}_1 - \mathbf{b}_2\| \\
& + M_1^2 \cdot M_2^2 \cdot M_4 \cdot M_5 \|\mathbf{b}_1 - \mathbf{b}_2\| \\
=& C_2 \cdot \|\mathbf{b}_1 - \mathbf{b}_2\|
\end{aligned}
$$

where $C_2 = M_1^2 M_2 M_3 M_5 + 2M_1^2 M_2^2 M_4 M_5$.

$$
\begin{aligned}
&\sup_{u\in(v_h,v_{h+1}],\|\mathbf{a}-\boldsymbol{\beta}_0(u)\|\le\epsilon_1,\|\mathbf{b}-\boldsymbol{\beta}_0(v_h)\|\le\epsilon_2} E\|\mathbf{A}_i\left(\mathbf{a},u,v_h,\mathbf{b}\right)-\mathbf{A}_i\left(\boldsymbol{\beta}_0(u),u,v_h,\boldsymbol{\beta}_0(v_h)\right)\|^2 \\
=&\sup_{u\in(v_h,v_{h+1}],\|\mathbf{a}-\boldsymbol{\beta}_0(u)\|\le\epsilon_1,\|\mathbf{b}-\boldsymbol{\beta}_0(v_h)\|\le\epsilon_2} E\Big\|\mathbf{A}_i\left(\mathbf{a},u,v_h,\mathbf{b}\right)-\mathbf{A}_i\left(\mathbf{a},u,v_h,\boldsymbol{\beta}_0(v_h)\right) \\
&\qquad + \mathbf{A}_i\left(\mathbf{a},u,v_h,\boldsymbol{\beta}_0(v_h)\right)-\mathbf{A}_i\left(\boldsymbol{\beta}_0(u),u,v_h,\boldsymbol{\beta}_0(v_h)\right)\Big\|^2 \\
\le&2\sup_{u\in(v_h,v_{h+1}],\|\mathbf{a}-\boldsymbol{\beta}_0(u)\|\le\epsilon_1,\|\mathbf{b}-\boldsymbol{\beta}_0(v_h)\|\le\epsilon_2} E\left\|\mathbf{A}_i\left(\mathbf{a},u,v_h,\mathbf{b}\right)-\mathbf{A}_i\left(\mathbf{a},u,v_h,\boldsymbol{\beta}_0(v_h)\right)\right\|^2 \\
&+2\sup_{u\in(v_h,v_{h+1}],\|\mathbf{a}-\boldsymbol{\beta}_0(u)\|\le\epsilon_1} E\left\|\mathbf{A}_i\left(\mathbf{a},u,v_h,\boldsymbol{\beta}_0(v_h)\right)-\mathbf{A}_i\left(\boldsymbol{\beta}_0(u),u,v_h,\boldsymbol{\beta}_0(v_h)\right)\right\|^2 \\
\le&2C_2\epsilon_2+2C_1\epsilon_1
\end{aligned}
$$

$$
\begin{aligned}
&\sup_{u\in(v_h,v_{h+1}]} Var\bigg\{\mathbf{A}_i(\hat{\boldsymbol\beta}(u),u,v_h,\hat{\boldsymbol\beta}(v_h)) - \mathbf{A}_i(\boldsymbol\beta_0(u),u,v_h,\boldsymbol\beta_0(v_h)) \\
&\qquad - E\left[\mathbf{A}_i(\hat{\boldsymbol\beta}(u),u,v_h,\hat{\boldsymbol\beta}(v_h))\right] + E\left[\mathbf{A}_i(\boldsymbol\beta_0(u),u,v_h,\boldsymbol\beta_0(v_h))\right]\bigg\} \\
&\le \sup_{u\in(v_h,v_{h+1}]} E\Big\|\mathbf{A}_i(\hat{\boldsymbol\beta}(u),u,v_h,\hat{\boldsymbol\beta}(v_h)) - \mathbf{A}_i(\boldsymbol\beta_0(u),u,v_h,\boldsymbol\beta_0(v_h))\Big\|^2 \\
&= \sup_{u\in(v_h,v_{h+1}]} \bigg\{\Big\|\mathbf{A}_i(\hat{\boldsymbol\beta}(u),u,v_h,\hat{\boldsymbol\beta}(v_h)) - \mathbf{A}_i(\boldsymbol\beta_0(u),u,v_h,\boldsymbol\beta_0(v_h))\Big\|^2 \\
&\qquad \times I\left(\sup_{u\in(v_h,v_{h+1}]}\|\hat{\boldsymbol\beta}(u)-\boldsymbol\beta_0(u)\| \le \epsilon_1, \quad \|\hat{\boldsymbol\beta}(v_h)-\boldsymbol\beta_0(v_h)\| \le \epsilon_2\right) \\
&\qquad + \Big\|\mathbf{A}_i(\hat{\boldsymbol\beta}(u),u,v_h,\hat{\boldsymbol\beta}(v_h)) - \mathbf{A}_i(\boldsymbol\beta_0(u),u,v_h,\boldsymbol\beta_0(v_h))\Big\|^2 \\
&\qquad \times I\left(\sup_{u\in(v_h,v_{h+1}]}\|\hat{\boldsymbol\beta}(u)-\boldsymbol\beta_0(u)\| > \epsilon_1 \text{ or } \|\hat{\boldsymbol\beta}(v_h)-\boldsymbol\beta_0(v_h)\| > \epsilon_2\right)\bigg\} \\
&\le \sup_{u\in(v_h,v_{h+1}],\|\mathbf{a}-\boldsymbol\beta_0(u)\|\le\epsilon_1,\|\mathbf{b}-\boldsymbol\beta_0(v_h)\|\le\epsilon_2} E\Big\|\mathbf{A}_i\left(\mathbf{a},u,v_h,\mathbf{b}\right) - \mathbf{A}_i\left(\boldsymbol\beta_0(u),u,v_h,\boldsymbol\beta_0(v_h)\right)\Big\|^2 \\
&+ M_1^2M_2^2 \sup_{u\in(v_h,v_{h+1}]} Pr\left(\sup_{u\in(v_h,v_{h+1}]}\|\hat{\boldsymbol\beta}(u)-\boldsymbol\beta_0(u)\| > \epsilon_1 \text{ or } \|\hat{\boldsymbol\beta}(v_h)-\boldsymbol\beta_0(v_h)\| > \epsilon_2\right) \\
&\le 2C_2\epsilon_2 + 2C_1\epsilon_1 + M_1^2M_2^2 Pr\left(\sup_{u\in(v_h,v_{h+1}]}\|\hat{\boldsymbol\beta}(u)-\boldsymbol\beta_0(u)\| > \epsilon_1\right) \\
&+ M_1^2M_2^2 Pr\left(\|\hat{\boldsymbol\beta}(v_h)-\boldsymbol\beta_0(v_h)\| > \epsilon_2\right)
\end{aligned}
\tag{2.18}
$$

The inequality in (2.18) holds for any $\epsilon_1 > 0$ and $\epsilon_2 > 0$. Since $\hat{\boldsymbol\beta}(u)$ is uniformly consistent to $\boldsymbol\beta_0(u)$, as ϵ_1 and ϵ_2 goes to zero,

$$
\begin{aligned}
&\sup_{u\in(v_h,v_{h+1}]} Var\bigg\{\mathbf{A}_i(\hat{\boldsymbol\beta}(u),u,v_h,\hat{\boldsymbol\beta}(v_h)) - \mathbf{A}_i(\boldsymbol\beta_0(u),u,v_h,\boldsymbol\beta_0(v_h)) \\
&\qquad\qquad - E\left[\mathbf{A}_i(\hat{\boldsymbol\beta}(u),u,v_h,\hat{\boldsymbol\beta}(v_h))\right] + E\left[\mathbf{A}_i(\boldsymbol\beta_0(u),u,v_h,\boldsymbol\beta_0(v_h))\right]\bigg\} \xrightarrow{p} 0.
\end{aligned}
\tag{2.19}
$$

Lemma 2 is proved.

Proof of *Theorem 2.1.2*: Let $o_I(p_n)$ denote a term that converges uniformly to 0 in probability in $u \in I$ after being divided by p_n. According to Lemma 2.,

$$\begin{aligned}
&\sqrt{n}\left[\boldsymbol{\Phi}(\boldsymbol{\beta}_0(u), u, v_h, \boldsymbol{\beta}_0(v_h)) + \boldsymbol{\phi}(\hat{\boldsymbol{\beta}}(u), u, v_h, \hat{\boldsymbol{\beta}}(v_h)) - \boldsymbol{\phi}(\boldsymbol{\beta}_0(u), u, v_h, \boldsymbol{\beta}_0(v_h))\right] \\
=&\sqrt{n}\boldsymbol{\Phi}(\hat{\boldsymbol{\beta}}(u), u, v_h, \hat{\boldsymbol{\beta}}(v_h)) + o_{(v_h, v_{h+1}]}(1) = o_{(v_h, v_{h+1}]}(1) \qquad (2.20)
\end{aligned}$$

Define

$$\begin{aligned}
&\mathbf{D}_{\boldsymbol{\phi}}^{(1)}(\mathbf{a}, u, v, \mathbf{b}) = \frac{\partial \boldsymbol{\phi}(\mathbf{a}, u, v, \mathbf{b})}{\partial \mathbf{a}} \\
=& - E\Bigg[\exp(\mathbf{X}^T\mathbf{a})\mathbf{X}^{\otimes 2}\Big\{\mu_Z(\exp(\mathbf{X}^T\mathbf{a})) - \mu_Z(\exp(\mathbf{X}^T\mathbf{b})) - u + v\Big\} \\
&\qquad \times \int_0^{\exp(\mathbf{X}^T\mathbf{b})} f_{L,R|\mathbf{Z}}(l, \exp(\mathbf{X}^T\mathbf{a}))dl\Bigg] \\
&+ E\left\{\int_{\exp(\mathbf{X}^T\mathbf{a})}^{\infty}\int_0^{\exp(\mathbf{X}^T\mathbf{b})} f_{L,R|Z}(l, r)dldr\mathbf{X}^{\otimes 2}\dot{\mu}(\exp(\mathbf{X}^T\mathbf{a}))\exp(\mathbf{X}^T\mathbf{a})\right\}
\end{aligned}$$

and

$$\begin{aligned}
&\mathbf{D}_{\boldsymbol{\phi}}^{(2)}(\mathbf{a}, u, v, \mathbf{b}) = \frac{\partial \boldsymbol{\phi}(\mathbf{a}, u, v, \mathbf{b})}{\partial \mathbf{b}} \\
&= E\Bigg[\exp(\mathbf{X}^T\mathbf{b})\mathbf{X}^{\otimes 2}\Big\{\mu_Z(\exp(\mathbf{X}^T\mathbf{a})) - \mu_Z(\exp(\mathbf{X}^T\mathbf{b})) - u + v\Big\} \\
&\qquad \times \int_{\exp(\mathbf{X}^T\mathbf{a})}^{\infty} f_{L,R|\mathbf{Z}}(\exp(\mathbf{X}^T\mathbf{b}), r)dr\Bigg] \\
&- E\left[\int_{\exp(\mathbf{X}^T\mathbf{a})}^{\infty}\int_0^{\exp(\mathbf{X}^T\mathbf{b})} f_{L,R|Z}(l, r)dldr\mathbf{X}^{\otimes 2}\dot{\mu}_{\mathbf{Z}}(\exp(\mathbf{X}^T\mathbf{b}))\exp(\mathbf{X}^T\mathbf{b})\right].
\end{aligned}$$

Using Taylor expansion of $\boldsymbol{\phi}(\hat{\boldsymbol{\beta}}(u), u, v_h, \hat{\boldsymbol{\beta}}(v_h))$, we get

$$\begin{aligned}
&\sqrt{n}\left[\boldsymbol{\phi}(\hat{\boldsymbol{\beta}}(u), u, v_h, \hat{\boldsymbol{\beta}}(v_h)) - \boldsymbol{\phi}(\boldsymbol{\beta}_0(u), u, v_h, \boldsymbol{\beta}_0(v_h))\right] \\
&= \mathbf{D}_{\boldsymbol{\phi}}^{(1)}(\boldsymbol{\beta}_0(u), u, v_h, \boldsymbol{\beta}_0(v_h))\sqrt{n}(\hat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}_0(u)) + o\left(\sqrt{n}(\hat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}_0(u))\right) \\
&+ \mathbf{D}_{\boldsymbol{\phi}}^{(2)}(\boldsymbol{\beta}_0(u), u, v_h, \boldsymbol{\beta}_0(v_h))\sqrt{n}(\hat{\boldsymbol{\beta}}(v_h) - \boldsymbol{\beta}_0(v_h)) + o\left(\sqrt{n}(\hat{\boldsymbol{\beta}}(v_h) - \boldsymbol{\beta}_0(v_h))\right)
\end{aligned}$$

where

$$\mathbf{D}_{\boldsymbol{\phi}}^{(1)}(\boldsymbol{\beta}_0(u), u, v_h, \boldsymbol{\beta}_0(v_h)) = E\left[\mathbf{X}^{\otimes 2}\dot{\mu}(\tau_{\mathbf{Z}}(u))\tau_{\mathbf{Z}}(u)I\left(L \leq \tau_{\mathbf{Z}}(v(u))\right)I\left(R \geq \tau_{\mathbf{Z}}(u)\right)\right],$$

and

$$\mathbf{D}_{\boldsymbol{\phi}}^{(2)}(\boldsymbol{\beta}_0(u), u, v_h, \boldsymbol{\beta}_0(v_h)) = -E\left[\mathbf{X}^{\otimes 2}\dot{\mu}(\tau_{\mathbf{Z}}(v_h))\tau_{\mathbf{Z}}(v_h)I\left(L \leq \tau_{\mathbf{Z}}(v(u))\right)I\left(R \geq \tau_{\mathbf{Z}}(u)\right)\right].$$

After replacing $\sqrt{n}\left\{\boldsymbol{\phi}(\hat{\boldsymbol{\beta}}(u), u, v_h, \hat{\boldsymbol{\beta}}(v_h)) - \boldsymbol{\phi}(\boldsymbol{\beta}_0(u), u, v_h, \boldsymbol{\beta}_0(v_h))\right\}$ in the LHS of (2.20) by Taylor expansion, under condition (b*), we get

$$\begin{aligned}
&\sqrt{n}(\hat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}_0(u)) \\
&= -\left[\mathbf{D}_{\boldsymbol{\phi}}^{(1)}(\boldsymbol{\beta}_0(u), u, v_h, \boldsymbol{\beta}_0(v_h)) + o(1)\right]^{-1}\Bigg\{\sqrt{n}\boldsymbol{\Phi}(\boldsymbol{\beta}_0(u), u, v_h, \boldsymbol{\beta}_0(v_h)) + \epsilon(u) \\
&+ \left(\mathbf{D}_{\boldsymbol{\phi}}^{(2)}(\boldsymbol{\beta}_0(u), u, v_h, \boldsymbol{\beta}_0(v_h)) + o(1)\right)\sqrt{n}(\hat{\boldsymbol{\beta}}(v_h) - \boldsymbol{\beta}_0(v_h))\Bigg\}
\end{aligned}$$

Given $\sqrt{n}\left(\hat{\boldsymbol{\beta}}(v_h) - \boldsymbol{\beta}_0(v_h) - \frac{1}{n}\sum_{i=1}^n \boldsymbol{\xi}_i(v_h)\right) \xrightarrow{p} 0$, we have

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}_0(u) - \frac{1}{n}\sum_{i=1}^n \boldsymbol{\xi}_i(u)\right) \xrightarrow{p} 0,$$

where $\boldsymbol{\xi}_i(u) = -\left[\mathbf{D}_{\boldsymbol{\phi}}^{(1)}(\boldsymbol{\beta}_0(u), u, v_h, \boldsymbol{\beta}_0(v_h))\right]^{-1}$

$$\times\left\{\mathbf{A}_i(\boldsymbol{\beta}_0(u), u, v_h, \boldsymbol{\beta}_0(v_h)) + \mathbf{D}_{\boldsymbol{\phi}}^{(2)}(\boldsymbol{\beta}_0(u), u, v_h, \boldsymbol{\beta}_0(v_h))\xi_i(v_h)\right\}.$$

Based on this relationship and the fact $\boldsymbol{\xi}_i(0) = 0$, we can get $\boldsymbol{\xi}_i(v_h)$ denoted by $\boldsymbol{\xi}_{h,i}$ sequentially for $h = 1, 2, \ldots, H$. After determining $\boldsymbol{\xi}_{h,i}$'s, for $u \in (v_h, v_{h+1}]$, we can get $\boldsymbol{\xi}_i(u) = -\Big[\mathbf{D}_{\boldsymbol{\phi}}^{(1)}(\boldsymbol{\beta}_0(u), u, v_h, \boldsymbol{\beta}_0(v_h))\Big]^{-1}$

$$\times\Big\{\mathbf{A}_i(\boldsymbol{\beta}_0(u), u, v_h, \boldsymbol{\beta}_0(v_h)) + \mathbf{D}_{\boldsymbol{\phi}}^{(2)}(\boldsymbol{\beta}_0(u), u, v_h, \boldsymbol{\beta}_0(v_h))\boldsymbol{\xi}_{h,i}\Big\}.$$

Note, $\{\boldsymbol{\xi}(u), u \in (0, U]\}$ is a Donsker class, which implies that

$$\sqrt{n}\left\{\hat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}_0(u)\right\} \xrightarrow{D} G(u)$$

with mean 0 and covariance $\boldsymbol{\Sigma}(s,t) = E\{\boldsymbol{\xi}_1(s), \boldsymbol{\xi}_1(t)\}$.

Lemma 3. Let $\theta(u, \mathbf{a}, \beta) = E\left[\Theta(u, \mathbf{a}, \beta)\right]$. Given u, $\theta(u, \mathbf{a}, \boldsymbol{\beta}_0)$ has a unique minimizer at $\mathbf{a} = \boldsymbol{\beta}_0(u)$, under the same condition as in lemma 1.

Proof: Denote

$$\begin{aligned} H(\mathbf{X}, L, R; u, \mathbf{a}, \mathbf{b}) = E\Big[&\sum_{j=1}^{\infty}\left(\mathbf{X}^T\mathbf{a} \wedge \log R - \log T^{(j)}\right)^{+} I(L \le T^{(j)} \le R) \\ &- (\mathbf{X}^T\mathbf{a} \wedge \log R) \cdot (u - \tilde{\mu}_{\mathbf{Z}}(L; \mathbf{b}) \wedge u)\Big|\mathbf{X}, L, R\Big] \end{aligned}$$

Since we have

$$\begin{aligned} &\theta(u, \mathbf{a}, \boldsymbol{\beta}_0) \\ =&E\left\{H(\mathbf{X}, L, R; u, \mathbf{a}, \boldsymbol{\beta}_0) I(\mathbf{X}^T\boldsymbol{\beta}_0(u) \ge \log R\ ,\ \mathbf{X}^T\mathbf{a} \ge \log R)\right\} \\ &+ E\left\{H(\mathbf{X}, L, R; u, \mathbf{a}, \boldsymbol{\beta}_0) I(\mathbf{X}^T\mathbf{a} < \log R \le \mathbf{X}^T\boldsymbol{\beta}_0(u))\right\} \\ &+ E\left\{H(\mathbf{X}, L, R; u, \mathbf{a}, \boldsymbol{\beta}_0) I(L \le \exp(\mathbf{X}^T\boldsymbol{\beta}_0(u)) \le R)\right\} \\ &+ E\left\{H(\mathbf{X}, L, R; u, \mathbf{a}, \boldsymbol{\beta}_0) I(\exp(\mathbf{X}^T\boldsymbol{\beta}_0(u)) < L)\right\}, \end{aligned}$$

we could prove it case by case.

- When $\mathbf{X}^T\boldsymbol{\beta}_0(u) \geq \log R$, $\mathbf{X}^T\mathbf{a} \geq \log R$,

$$H(\mathbf{X}, L, R; u, \boldsymbol{\beta}_0(u), \boldsymbol{\beta}_0) = H(\mathbf{X}, L, R; u, \mathbf{a}, \boldsymbol{\beta}_0)$$

- When $\mathbf{X}^T\mathbf{a} < \log R \leq \mathbf{X}^T\boldsymbol{\beta}_0(u)$, $\Rightarrow \mu_{\mathbf{Z}}(L) \leq u$,

  consider $f(y; \mathbf{X}, L, R) = E\Big[\sum_{j=1}^{\infty}\big(y - \log T^{(j)}\big)^{+} I(L \leq T^{(j)} \leq R) - y(u - \mu_{\mathbf{Z}}(L))\Big|\mathbf{X}, L, R\Big]$,

$$\begin{aligned}\frac{df(y; \mathbf{X}, L, R)}{dy} &= E\Big[\sum_{j=1}^{\infty} I(y \geq \log T^{(j)}) I(L \leq T^{(j)} \leq R) - (u - \mu_{\mathbf{Z}}(L))\Big|\mathbf{X}, L, R\Big] \\ &= I(L \leq \exp(y))(\mu_{\mathbf{Z}}(\exp(y) \wedge R) - \mu_{\mathbf{Z}}(L)) - (u - \mu_{\mathbf{Z}}(L)) \\ &\leq 0,\end{aligned}$$

  because $\mu_{\mathbf{Z}}(\exp(y) \wedge R) \leq u$ and $u \geq \mu_{\mathbf{Z}}(L)$.

  $\Rightarrow$

$$H(\mathbf{X}, L, R; u, \boldsymbol{\beta}_0(u), \boldsymbol{\beta}_0) = f(\log R) \leq f(\mathbf{X}^T\mathbf{a}) = H(\mathbf{X}, L, R; u, \mathbf{a}, \boldsymbol{\beta}_0).$$

- When $L \leq \exp(\mathbf{X}^T\boldsymbol{\beta}_0(u)) \leq R$,

$$\begin{aligned}\frac{df(y; \mathbf{X}, L, R)}{dy} &= I(L \leq \exp(y))(\mu_{\mathbf{Z}}(\exp(y) \wedge R) - \mu_{\mathbf{Z}}(L)) - (u - \mu_{\mathbf{Z}}(L)) \\ &\begin{cases} < 0 & \text{if } y < \mathbf{X}^T\boldsymbol{\beta}_0(u); \\ = 0 & \text{if } y = \mathbf{X}^T\boldsymbol{\beta}_0(u); \\ > 0 & \text{if } \mathbf{X}^T\boldsymbol{\beta}_0(u) < y \leq \log R. \end{cases}\end{aligned}$$

  So $y = \mathbf{X}^T\boldsymbol{\beta}_0(u)$ is the unique minimizor of $f(y; \mathbf{X}, L, R)$ which means

$$H(\mathbf{X}, L, R; \mathbf{a}, u, \boldsymbol{\beta}_0) \geq H(\mathbf{X}, L, R; \boldsymbol{\beta}_0(u), u, \boldsymbol{\beta}_0),$$

where the equality holds if and only if $\mathbf{X}^T\mathbf{a} = \mathbf{X}^T\boldsymbol{\beta}_0(u)$.

- when $\exp(\mathbf{X}^T\boldsymbol{\beta}_0(u)) < L$,

$$
\begin{aligned}
H(\mathbf{X}, L, R; u, \mathbf{a}, \boldsymbol{\beta}_0) &= E\Big[\sum_{j=1}^{\infty}\left(\mathbf{X}^T\mathbf{b} \wedge \log R - \log T^{(j)}\right)^+ I(L \leq T^{(j)} \leq R)\Big] \\
&\geq 0 = H(\mathbf{X}, L, R; u, \boldsymbol{\beta}_0(u), \boldsymbol{\beta}_0).
\end{aligned}
$$

In summary, for any $u \in (0, U]$, $\theta(\mathbf{a}, u, \boldsymbol{\beta}_0) \geq \theta(\boldsymbol{\beta}_0(u), u, \boldsymbol{\beta}_0)$. Under condition (a), we have $Pr\left\{[\boldsymbol{\beta}_0(u) - \mathbf{a}]\,\mathbf{X}I(L \leq \exp(\mathbf{X}^T\boldsymbol{\beta}_0(v_h)))I(\log R > \mathbf{X}^T\boldsymbol{\beta}_0(u)) \neq 0\right\} > 0$ for any $\mathbf{a} \neq \boldsymbol{\beta}_0(u)$. Since $\exp(\mathbf{X}^T\boldsymbol{\beta}_0(v_h)) < \exp(\mathbf{X}^T\boldsymbol{\beta}_0(u))$,

$$
\begin{aligned}
&Pr\left\{[\boldsymbol{\beta}_0(u) - \mathbf{a}]\,\mathbf{X}I(L \leq \exp(\mathbf{X}^T\boldsymbol{\beta}_0(u)) \leq R) \neq 0\right\} \\
\geq &Pr\left\{[\boldsymbol{\beta}_0(u) - \mathbf{a}]\,\mathbf{X}I(L \leq \exp(\mathbf{X}^T\boldsymbol{\beta}_0(v_h)))I(\log R > \mathbf{X}^T\boldsymbol{\beta}_0(u)) \neq 0\right\} \\
> &0.
\end{aligned}
$$

Then stict inequality holds that

$$
\begin{aligned}
&E\{H(\mathbf{X}, L, R; u, \mathbf{a}, \boldsymbol{\beta}_0)I(L \leq \exp(\mathbf{X}^T\boldsymbol{\beta}_0(u)) \leq R)\} \\
> &E\{H(\mathbf{X}, L, R; u, \boldsymbol{\beta}_0(u), \boldsymbol{\beta}_0)I(L \leq \exp(\mathbf{X}^T\boldsymbol{\beta}_0(u)) \leq R)\}.
\end{aligned}
$$

Hence, $\theta(\mathbf{a}, u, \boldsymbol{\beta}_0)$ has a unique minimizer at $\mathbf{a} = \boldsymbol{\beta}_0(u)$ given u. End of proof.

Proof of *Theorem 2.1.3*: Let

$$
\begin{aligned}
\vartheta(\mathbf{X}, L, R, \mathbf{T}; u, \mathbf{a}, \mathbf{b}) = &\sum_{j=1}^{\infty}\left[(\mathbf{X}^T\mathbf{a} - \log T^(j))^+ + (\log R - \log T^{(j)})^+\right] I(L \leq T^{(j)} \leq R) \\
&- \sum_{j=1}^{\infty}(\mathbf{X}^T\mathbf{a} \vee \log R - \log T^{(j)})^+ I(L \leq T^{(j)} \leq R) \\
&- (\mathbf{X}^T\mathbf{a} \wedge \log R)(u - u \wedge \tilde{\mu}_{\mathbf{Z}}(L; \mathbf{b}))
\end{aligned}
$$

Since linear functions, and concave or convex functions are Glivenko-Cantelli classes and the sum or product of G-C classes are still G-C class, $\vartheta(\mathbf{X}, L, R, \mathbf{T}; u, \mathbf{a}, \mathbf{b})$ is a G-C class with index $u, \mathbf{a}, \mathbf{b}$ under condition (b). This fact coupled with pointwise convergence by the strong law of large numbers implies the convergence of $\tilde{\Theta}(u, \mathbf{a}, \boldsymbol{\beta})$ to $\theta(u, \mathbf{a}, \boldsymbol{\beta})$ uniformly in u, $\mathbf{a}$, and $\boldsymbol{\beta}$ (Andersen & Gill, 1982).

$$\sup_{u\in(0,U],\mathbf{a}\in\mathcal{B},\boldsymbol{\beta}\in\mathcal{B}} |\Theta(u, \mathbf{a}, \boldsymbol{\beta}) - \theta(u, \mathbf{a}, \boldsymbol{\beta})| \xrightarrow{p} 0 \tag{2.21}$$

$$\begin{aligned}
&\sup_{u\in(0,U],\mathbf{a}\in\mathcal{B}} |\theta(u, \mathbf{a}, \boldsymbol{\beta}_0) - \theta(u, \mathbf{a}, \hat{\boldsymbol{\beta}}_{(m-1)})| \\
=&\sup_{u\in(0,U],\mathbf{a}\in\mathcal{B}} \left| E\left[(\mathbf{X}^T\mathbf{a} \wedge \log R) \left(\tilde{\mu}_{\mathbf{Z}}(L; \boldsymbol{\beta}_0) \wedge u - \tilde{\mu}_{\mathbf{Z}}(L; \hat{\boldsymbol{\beta}}_{(m-1)}) \wedge u \right) \right] \right| \\
=&\sup_{u\in(0,U],\mathbf{a}\in\mathcal{B}} E\left| (\mathbf{X}^T\mathbf{a} \wedge \log R) \int_0^u (I(L \geq \exp(\mathbf{X}^T\boldsymbol{\beta}_0(v))) \right. \\
&\left. \quad - I(L \geq \exp(\mathbf{X}^T\hat{\boldsymbol{\beta}}_{(m-1)}(v))))dv \right| \\
\leq&\sup_{u\in(0,U],\mathbf{a}\in\mathcal{B}} E\left[(\mathbf{X}^T\mathbf{a} \wedge \log R) u M_4 M_5 \sup_{v\in(0,u]} \|\hat{\boldsymbol{\beta}}_{(m-1)}(v) - \boldsymbol{\beta}_0(v)\| \right] \\
\xrightarrow{p}& 0
\end{aligned} \tag{2.22}$$

$$\begin{aligned}
0 \leq& \theta(u, \hat{\boldsymbol{\beta}}_{(m)}(u), \boldsymbol{\beta}_0) - \theta(u, \boldsymbol{\beta}_0(u), \boldsymbol{\beta}_0) \\
=& \theta(u, \hat{\boldsymbol{\beta}}_{(m)}(u), \boldsymbol{\beta}_0) - \theta(u, \hat{\boldsymbol{\beta}}_{(m)}(u), \hat{\boldsymbol{\beta}}_{(m-1)}) && (2.23)\\
&+ \theta(u, \hat{\boldsymbol{\beta}}_{(m)}(u), \hat{\boldsymbol{\beta}}_{(m-1)}) - \Theta(u, \hat{\boldsymbol{\beta}}_{(m)}(u), \hat{\boldsymbol{\beta}}_{(m-1)}) && (2.24)\\
&+ \Theta(u, \hat{\boldsymbol{\beta}}_{(m)}(u), \hat{\boldsymbol{\beta}}_{(m-1)}) - \Theta(u, \boldsymbol{\beta}_0(u), \hat{\boldsymbol{\beta}}_{(m-1)}) && (2.25)\\
&+ \Theta(u, \boldsymbol{\beta}_0(u), \hat{\boldsymbol{\beta}}_{(m-1)}) - \theta(u, \boldsymbol{\beta}_0(u), \hat{\boldsymbol{\beta}}_{(m-1)}) && (2.26)\\
&+ \theta(u, \boldsymbol{\beta}_0(u), \hat{\boldsymbol{\beta}}_{(m-1)}) - \theta(u, \boldsymbol{\beta}_0(u), \boldsymbol{\beta}_0) && (2.27)
\end{aligned}$$

Since (2.25) is less than or equal to zero,

$$\begin{aligned}
&\sup_{u\in(0,U]} |\theta(u,\hat{\boldsymbol{\beta}}_{(m)}(u),\boldsymbol{\beta}_0) - \theta(u,\boldsymbol{\beta}_0(u),\boldsymbol{\beta}_0)| \\
\leq &\sup_{u\in(0,U]} |(2.23)+(2.24)+(2.26)+(2.27)| \\
\leq &\sup_{u\in(0,U]} |(2.23)| + \sup_{u\in(0,U]} |(2.24)| + \sup_{u\in(0,U]} |(2.26)| + \sup_{u\in(0,U]} |(2.27)|
\end{aligned}$$

According to (2.21), $\sup_{u\in(0,U]} |(2.24)|$ and $\sup_{u\in(0,U]} |(2.26)|$ converges to zero in probability. According to (2.22), $\sup_{u\in(0,U]} |(2.23)|$ and $\sup_{u\in(0,U]} |(2.27)|$ converges to zero in probability. Hence, we have $\sup_{u\in(0,U]} \left|\theta(u,\hat{\boldsymbol{\beta}}_{(m)}(u),\boldsymbol{\beta}_0) - \theta(u,\boldsymbol{\beta}_0(u),\boldsymbol{\beta}_0)\right| \xrightarrow{p} 0$. Following similar arguments to those in Huang and Peng's (2009) proof of theorem, $\sup_{u\in(0,U]} |\hat{\boldsymbol{\beta}}_{(m)}(u) - \boldsymbol{\beta}_0(u)| \xrightarrow{p} 0$.

Lemma 4. Given $\hat{\boldsymbol{\beta}}_{(m-1)}$ and $\hat{\boldsymbol{\beta}}_{(m)}$ are uniformly consistent,

$$\begin{aligned}
\sup_{u\in(0,U]} \|\sqrt{n}\Big\{ &\tilde{\boldsymbol{\Delta}}(\hat{\boldsymbol{\beta}}_{(m)}(u);u,\hat{\boldsymbol{\beta}}_{(m-1)}) - \tilde{\boldsymbol{\Delta}}(\boldsymbol{\beta}_0(u);u,\boldsymbol{\beta}_0) \\
&- \tilde{\boldsymbol{\delta}}(\hat{\boldsymbol{\beta}}_{(m)}(u);u,\hat{\boldsymbol{\beta}}_{(m-1)}) + \tilde{\boldsymbol{\delta}}(\boldsymbol{\beta}_0(u);u,\boldsymbol{\beta}_0)\Big\}\| \xrightarrow{p} 0,
\end{aligned}$$

where $\tilde{\boldsymbol{\delta}}(\mathbf{a};u,\mathbf{b}) = E[\tilde{\boldsymbol{\Delta}}(\mathbf{a};u,\mathbf{b})]$.

Proof: Denote $\mathbf{D}_i(\mathbf{a};u,\mathbf{b}) = \mathbf{X}_i I(\mathbf{X}_i^T\mathbf{a} \leq \log R_i)[N(\exp(\mathbf{X}_i^T\mathbf{a})) \vee N(L_i) - N(L_i) -$

$u + u \wedge \tilde{\mu}_{\mathbf{Z}_i}(L_i;\mathbf{b})]$.

$$
\begin{aligned}
& E[\mathbf{D}_i(\mathbf{a}_1;u,\mathbf{b}) - \mathbf{D}_i(\mathbf{a}_2;u,\mathbf{b})]^2 \\
=& E\Big\{\mathbf{X}_i I(\max(\mathbf{X}_i^T\mathbf{a}_1, \mathbf{X}_i^T\mathbf{a}_2) \le \log R_i) \\
& \qquad\qquad \times [N(\exp(\mathbf{X}_i^T\mathbf{a}_1)) \vee N(L_i) - N(\exp(\mathbf{X}_i^T\mathbf{a}_2)) \vee N(L_i)] \\
& + \mathbf{X}_i I(\mathbf{X}_i^T\mathbf{a}_1 \le \log R_i < \mathbf{X}_i^T\mathbf{a}_2)[N(\exp(\mathbf{X}_i^T\mathbf{a}_1)) \vee N(L_i)N(L_i) - u + u \wedge \tilde{\mu}_{\mathbf{Z}_i}(L_i;\mathbf{b})] \\
& - \mathbf{X}_i I(\mathbf{X}_i^T\mathbf{a}_2 \le \log R_i < \mathbf{X}_i^T\mathbf{a}_1)[N(\exp(\mathbf{X}_i^T\mathbf{a}_2)) \vee N(L_i)N(L_i) - u + u \wedge \tilde{\mu}_{\mathbf{Z}_i}(L_i;\mathbf{b})] \\
& + \mathbf{X}_i I(\min(\mathbf{X}_i^T\mathbf{a}_1, \mathbf{X}_i^T\mathbf{a}_2) > \log R_i) \times 0\Big\}^2 \\
\le& M_1^2 \cdot M_2 \cdot M_3 \cdot M_5 \cdot \|\mathbf{a}_1 - \mathbf{a}_2\| \\
& + M_1^2 \cdot M_2^2 \cdot M_4 \cdot M_5 \cdot \|\mathbf{a}_1 - \mathbf{a}_2\| \\
& + M_1^2 \cdot M_2^2 \cdot M_4 \cdot M_5 \cdot \|\mathbf{a}_1 - \mathbf{a}_2\| \\
=& C_3\|\mathbf{a}_1 - \mathbf{a}_2\| \qquad (2.28)
\end{aligned}
$$

where $C_3 = M_1^2 M_2 M_3 M_5 + 2M_1^2 M_2^2 M_4 M_5$.

$$
\begin{aligned}
& E[\mathbf{D}_i(\mathbf{a};u,\mathbf{b}_1) - \mathbf{D}_i(\mathbf{a};u,\mathbf{b}_2)]^2 \\
=& E\left\{\mathbf{X}_i I(\mathbf{X}_i^T\mathbf{a} \le \log R_i)[u \wedge \tilde{\mu}_{\mathbf{Z}_i}(L_i;\mathbf{b}_1) - u \wedge \tilde{\mu}_{\mathbf{Z}_i}(L_i;\mathbf{b}_2)]\right\}^2 \\
\le& E\left\{\mathbf{X}_i\left[\int_0^U (I(L \ge \exp(\mathbf{X}_i^T\mathbf{b}_1(u))) - I(L \ge \exp(\mathbf{X}_i^T\mathbf{b}_2(u))))du\right]\right\} \\
\le& M_1^2 \cdot U \cdot M_4 \cdot M_5 \cdot \sup_{u\in(0,U]} \|\mathbf{b}_1(u) - \mathbf{b}_2(u)\| \\
=& C_4 \sup_{u'\in(0,U]} \|\mathbf{b}_1(u') - \mathbf{b}_2(u')\| \qquad (2.29)
\end{aligned}
$$

where $C_4 = M_1^2 M_4 M_5 U$.

$$
\begin{aligned}
&\sup_{u\in(0,U],\|\mathbf{a}-\boldsymbol{\beta}_0(u)\|\le\epsilon_1,\sup_{w\in(0,U]}\|\mathbf{b}(w)-(\boldsymbol{\beta}_0(w)\|\le\epsilon_2} E\left[\mathbf{D}_i(\mathbf{a};u,\mathbf{b})-\mathbf{D}_i(\boldsymbol{\beta}_0(u);u,\boldsymbol{\beta}_0)\right]^2\\
\le&2\times\sup_{u\in(0,U],\|\mathbf{a}-\boldsymbol{\beta}_0(u)\|\le\epsilon_1,\sup_{w\in(0,U]}\|\mathbf{b}(w)-(\boldsymbol{\beta}_0(w)\|\le\epsilon_2} E\left[\mathbf{D}_i(\mathbf{a};u,\mathbf{b})-\mathbf{D}_i(\mathbf{a};u,\boldsymbol{\beta}_0)\right]^2\\
&+2\times\sup_{u\in(0,U],\|\mathbf{a}-\boldsymbol{\beta}_0(u)\|\le\epsilon_1} E\left[\mathbf{D}_i(\mathbf{a};u,\boldsymbol{\beta}_0)-\mathbf{D}_i(\boldsymbol{\beta}_0(u);u,\boldsymbol{\beta}_0)\right]^2\\
\le&2C_4\epsilon_2+2C_3\epsilon_1
\end{aligned}
$$

$$
\begin{aligned}
&\sup_{u\in(0,U]} Var\Big\{\mathbf{D}_i(\hat{\boldsymbol{\beta}}_{(m)}(u),u,\hat{\boldsymbol{\beta}}_{(m-1)})-\mathbf{D}_i(\boldsymbol{\beta}_0(u),u,\boldsymbol{\beta}_0)\\
&\qquad -E\left[\mathbf{D}_i(\hat{\boldsymbol{\beta}}_{(m)}(u),u,\hat{\boldsymbol{\beta}}_{(m-1)})\right]+E\left[\mathbf{D}_i(\boldsymbol{\beta}_0(u),u,\boldsymbol{\beta}_0)\right]\Big\}\\
&\le\sup_{u\in(0,U]} E\left\|\mathbf{D}_i(\hat{\boldsymbol{\beta}}_{(m)}(u),u,\hat{\boldsymbol{\beta}}_{(m-1)})-\mathbf{D}_i(\boldsymbol{\beta}_0(u),u,\boldsymbol{\beta}_0)\right\|^2\\
&=\sup_{u\in(0,U]}\Big\{\left\|\mathbf{D}_i(\hat{\boldsymbol{\beta}}_{(m)}(u),u,\hat{\boldsymbol{\beta}}_{(m-1)})-\mathbf{D}_i(\boldsymbol{\beta}_0(u),u,\boldsymbol{\beta}_0)\right\|^2\\
&\qquad\times I\left(\|\hat{\boldsymbol{\beta}}_{(m)}(u)-\boldsymbol{\beta}_0(u)\|\le\epsilon_1,\quad\sup_{w\in(0,U]}\|\hat{\boldsymbol{\beta}}_{(m-1)}(w)-\boldsymbol{\beta}_0(w)\|\le\epsilon_2\right)\\
&\qquad+\left\|\mathbf{D}_i(\hat{\boldsymbol{\beta}}_{(m)}(u),u,\hat{\boldsymbol{\beta}}_{(m-1)})-\mathbf{D}_i(\boldsymbol{\beta}_0(u),u,\boldsymbol{\beta}_0)\right\|^2\\
&\qquad\times I\left(\|\hat{\boldsymbol{\beta}}_{(m)}(u)-\boldsymbol{\beta}_0(u)\|>\epsilon_1\text{ or }\sup_{w\in(0,U]}\|\hat{\boldsymbol{\beta}}_{(m-1)}(w)-\boldsymbol{\beta}_0(w)\|>\epsilon_2\right)\Big\}\\
&\le\sup_{u\in(0,U],\|\mathbf{a}-\boldsymbol{\beta}_0(u)\|\le\epsilon_1,\sup_{w\in(0,U]}\|\mathbf{b}(w)-\boldsymbol{\beta}_0(w)\|\le\epsilon_2} E\left\|\mathbf{D}_i\left(\mathbf{a},u,\mathbf{b}\right)-\mathbf{D}_i\left(\boldsymbol{\beta}_0(u),u,\boldsymbol{\beta}_0\right)\right\|^2\\
&+M_1^2M_2^2\sup_{u\in(0,U]} Pr\left(\|\hat{\boldsymbol{\beta}}_{(m)}(u)-\boldsymbol{\beta}_0(u)\|>\epsilon_1\text{ or }\sup_{w\in(0,U]}\|\hat{\boldsymbol{\beta}}_{(m-1)}(w)-\boldsymbol{\beta}_0(w)\|>\epsilon_2\right)\\
&\le 2C_2\epsilon_2+2C_1\epsilon_1+M_1^2M_2^2\sup_{u\in(0,U]} Pr\left(\|\hat{\boldsymbol{\beta}}_{(m)}(u)-\boldsymbol{\beta}_0(u)\|>\epsilon_1\right)\\
&+M_1^2M_2^2 Pr\left(\sup_{w\in(0,U]}\|\hat{\boldsymbol{\beta}}_{(m-1)}(w)-\boldsymbol{\beta}_0(w)\|>\epsilon_2\right)
\end{aligned}
\tag{2.30}
$$

The inequality in (2.30) holds for any $\epsilon_1 > 0$ and $\epsilon_2 > 0$. Since $\hat{\boldsymbol{\beta}}_{(m-1)}$ and $\hat{\boldsymbol{\beta}}_{(m)}$ uniformly consistent to $\boldsymbol{\beta}_0$, as ϵ_1 and ϵ_2 goes to zero,

$$\sup_{u\in(0,U]} Var\Big\{\mathbf{D}_i(\hat{\boldsymbol{\beta}}_{(m)}(u),u,\hat{\boldsymbol{\beta}}_{(m-1)}) - \mathbf{D}_i(\boldsymbol{\beta}_0(u),u,\boldsymbol{\beta}_0) \\ - E\Big[\mathbf{D}_i(\hat{\boldsymbol{\beta}}_{(m)}(u),u,\hat{\boldsymbol{\beta}}_{(m-1)})\Big] + E\left[\mathbf{D}_i(\boldsymbol{\beta}_0(u),u,\boldsymbol{\beta}_0)\right]\Big\} \xrightarrow{p} 0. \qquad (2.31)$$

Lemma 4 is proved.

Proof of *Theorem 2.1.4*: According to Lemma 4,

$$\Big\|\sqrt{n}\Big\{\tilde{\Delta}(\hat{\boldsymbol{\beta}}_{(m)}(u);u,\hat{\boldsymbol{\beta}}_{(m-1)}) - \tilde{\Delta}(\boldsymbol{\beta}_0(u);u,\boldsymbol{\beta}_0) \\ - \tilde{\delta}(\hat{\boldsymbol{\beta}}_{(m)}(u);u,\hat{\boldsymbol{\beta}}_{(m-1)}) + \tilde{\delta}(\boldsymbol{\beta}_0(u);u,\boldsymbol{\beta}_0)\Big\}\Big\| = o_{(0,U]}(1). \qquad (2.32)$$

Suppose that $\sqrt{n}(\hat{\boldsymbol{\beta}}_{(m-1)} - \boldsymbol{\beta}_0) = n^{-1/2}\sum_{i=1}^n \boldsymbol{\xi}_i$ is a tight Gaussian process on $(0,U]$, and $\hat{\boldsymbol{\beta}}_{(m)}$ is uniformly consistent to $\boldsymbol{\beta}_0$. We want to prove that $\sqrt{n}(\hat{\boldsymbol{\beta}}_{(m)} - \boldsymbol{\beta}_0) = n^{-1/2}\sum_{i=1}^n \boldsymbol{\xi}_i^*$ is also a Gaussian process.

First, we can see that

$$\begin{aligned}\tilde{\boldsymbol{\delta}}(\mathbf{a};u,\mathbf{b}) = E\Big\{&\mathbf{X}I(R \geq \exp(\mathbf{X}^T\mathbf{a})) \\ &\times\Big[N(\exp(\mathbf{X}^T\mathbf{a})\vee L) - N(L) - u + \int_0^u I(L > \exp(\mathbf{X}^T\mathbf{b}(w)))dw\Big]\Big\} \\ =E\Big\{&\mathbf{X}I(R \geq \exp(\mathbf{X}^T\mathbf{a}))\Big[N(\exp(\mathbf{X}^T\mathbf{a})\vee L) - N(L) - u\Big]\Big\} \\ &+ \int_0^u E\Big\{\mathbf{X}I(R \geq \exp(\mathbf{X}^T\mathbf{a}))I(L > \exp(\mathbf{X}^T\mathbf{b}(w)))\Big\}dw \qquad (2.33)\end{aligned}$$

Denote

$$\begin{aligned}\tilde{\boldsymbol{\delta}}_1(\boldsymbol{\beta}_0(u);u,\boldsymbol{\beta}_0) &= \frac{d\delta(\mathbf{a};u,\mathbf{b})}{d\mathbf{a}}\Big|_{\mathbf{a}=\boldsymbol{\beta}_0(u),\mathbf{b}=\boldsymbol{\beta}_0} \\ &=E\left[\mathbf{X}^{\otimes 2}\dot{\mu}(\tau_{\mathbf{Z}}(u))\tau_{\mathbf{Z}}(u)I\left(L \leq \tau_{\mathbf{Z}}(v(u))\right)I\left(R \geq \tau_{\mathbf{Z}}(u)\right)\right]\end{aligned}$$

and $\boldsymbol{\rho}(u,w) = \frac{dE\left\{\mathbf{X}I(R\geq\tau_{\mathbf{Z}}(u))I(L>\exp(\mathbf{X}^T\mathbf{a}))\right\}}{d\mathbf{a}}\Big|_{\mathbf{a}=\boldsymbol{\beta}_0(w)}$. Using Taylor expansion, we get

$$\begin{aligned}
&\sqrt{n}[\tilde{\delta}(\hat{\boldsymbol{\beta}}_{(m)}(u);u,\hat{\boldsymbol{\beta}}_{(m-1)}) - \tilde{\delta}(\boldsymbol{\beta}_0(u);u,\boldsymbol{\beta}_0)]\\
=&\tilde{\boldsymbol{\delta}}_1(\boldsymbol{\beta}_0(u);u,\boldsymbol{\beta}_0)\sqrt{n}(\hat{\boldsymbol{\beta}}_{(m)}(u)-\boldsymbol{\beta}_0(u)) + o_p\left(\sqrt{n}(\hat{\boldsymbol{\beta}}_{(m)}(u)-\boldsymbol{\beta}_0(u))\right)\\
&+\int_0^u \boldsymbol{\rho}(u,w)\sqrt{n}[\hat{\boldsymbol{\beta}}_{(m-1)}(w)-\boldsymbol{\beta}_0(w)]dw + o_p(\sup_{w\in(0,U]}\sqrt{n}[\hat{\boldsymbol{\beta}}_{(m-1)}(w)-\boldsymbol{\beta}_0(w)])
\end{aligned} \tag{2.34}$$

Since $\sqrt{n}\left[\hat{\boldsymbol{\beta}}_{(m-1)}(u)-\boldsymbol{\beta}_0(u)\right]$ is a tight Gaussian process on $u\in(0,U]$,

$$o_p\left(\sup_{w\in(0,U]}\sqrt{n}(\hat{\boldsymbol{\beta}}_{(m-1)}(w)-\boldsymbol{\beta}_0(w))\right) = o_p(1).$$

Plug equation (2.34) into equation (2.32),

$$\begin{aligned}
&\sqrt{n}(\hat{\boldsymbol{\beta}}_{(m)}(u)-\boldsymbol{\beta}_0(u))\\
=&-\left[\tilde{\boldsymbol{\delta}}_1(\boldsymbol{\beta}_0(u);u,\boldsymbol{\beta}_0)+o_p(1)\right]^{-1}\left\{\sqrt{n}\tilde{\Delta}(\boldsymbol{\beta}_0(u);u,\boldsymbol{\beta}_0)+o_{(0,U]}(1)\right.\\
&\qquad\left.+\int_0^u\boldsymbol{\rho}(u,w)\sqrt{n}[\hat{\boldsymbol{\beta}}_{(m-1)}(w)-\boldsymbol{\beta}_0(w)]dw+o_p(1)\right\} &(2.35)\\
=&n^{-1/2}\sum_{i=1}^n\left\{-\tilde{\boldsymbol{\delta}}_1(\boldsymbol{\beta}_0(u);u,\boldsymbol{\beta}_0)^{-1}\left[D_i(\boldsymbol{\beta}_0(u);u,\boldsymbol{\beta}_0)\right.\right.\\
&\qquad\left.\left.+\int_0^u\boldsymbol{\rho}(u,w)\boldsymbol{\xi}_i(w)dw\right]\right\}+o_{(0,U]}(1) &(2.36)
\end{aligned}$$

Define $\boldsymbol{\xi}_i^* = -\tilde{\boldsymbol{\delta}}_1(\boldsymbol{\beta}_0(u);u,\boldsymbol{\beta}_0)^{-1}\left[D_i(\boldsymbol{\beta}_0(u);u,\boldsymbol{\beta}_0)+\int_0^u\boldsymbol{\rho}(u,w)\boldsymbol{\xi}_i(w)dw\right]$. Since $\boldsymbol{\xi}_i^*$ is a Donsker class, $\sqrt{n}(\hat{\boldsymbol{\beta}}_{(m)}-\boldsymbol{\beta}_0)$ asymptotically converges weakly to a Gaussian process with mean zero and covariance $\boldsymbol{\Sigma}(s,t)=E\{\boldsymbol{\xi}_1^*(s),\boldsymbol{\xi}_1^*(t)\}$.

Note that $\hat{\boldsymbol{\beta}}$ is a consistent solution to the stochastic estimating equation

$\sqrt{n}\Delta(\boldsymbol{\beta};u)=0$. Equation (2.35) implies that

$$\begin{aligned}
&\sqrt{n}(\hat{\boldsymbol{\beta}}(u)-\boldsymbol{\beta}_0(u))\\
=&-\left[\tilde{\boldsymbol{\delta}}_1(\boldsymbol{\beta}_0(u);u,\boldsymbol{\beta}_0)\right]^{-1}\Big\{\sqrt{n}\tilde{\Delta}(\boldsymbol{\beta}_0(u);u,\boldsymbol{\beta}_0)\\
&\qquad+\int_0^u\boldsymbol{\rho}(u,w)\sqrt{n}[\hat{\boldsymbol{\beta}}(w)-\boldsymbol{\beta}_0(w)]dw\Big\}+o_{(0,U]}(1). \qquad (2.37)
\end{aligned}$$

By viewing the foregoing stochastic differential equation (2.37) for $n^{1/2}\left[\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right]$, and using the theory for linear Volterra equations of the second kind (theorem 3.1 and 3.3, Peter Linz 1985), we get

$$\begin{aligned}
n^{1/2}\left[\hat{\boldsymbol{\beta}}(u)-\boldsymbol{\beta}_0(u)\right]&=-\tilde{\boldsymbol{\delta}}_1(\boldsymbol{\beta}_0(u);u,\boldsymbol{\beta}_0)^{-1}\sqrt{n}\tilde{\boldsymbol{\Delta}}(\boldsymbol{\beta}_0(u);u,\boldsymbol{\beta}_0)\\
&\quad+\int_0^u\Gamma(u,w)\left\{-\tilde{\boldsymbol{\delta}}_1(\boldsymbol{\beta}_0(w);w,\boldsymbol{\beta}_0)^{-1}\sqrt{n}\tilde{\boldsymbol{\Delta}}(\boldsymbol{\beta}_0(w);w,\boldsymbol{\beta}_0)\right\}dw\\
&=n^{-1/2}\sum_{i=1}^n\Big[-\tilde{\boldsymbol{\delta}}_1(\boldsymbol{\beta}_0(u);u,\boldsymbol{\beta}_0)^{-1}\mathbf{D}_i(\boldsymbol{\beta}_0(u);u,\boldsymbol{\beta}_0)\\
&\quad+\int_0^u\Gamma(u,w)\left\{-\tilde{\boldsymbol{\delta}}_1(\boldsymbol{\beta}_0(w);w,\boldsymbol{\beta}_0)^{-1}\mathbf{D}_i(\boldsymbol{\beta}_0(w);w,\boldsymbol{\beta}_0)\right\}dw\Big]
\end{aligned}$$

where $\Gamma(u,w)=\sum_{j=1}^{\infty}\mathbf{k}_j(u,w)$ with $\mathbf{k}_1(u,w)=-\tilde{\boldsymbol{\delta}}_1(\boldsymbol{\beta}_0(u);u,\boldsymbol{\beta}_0)^{-1}\boldsymbol{\rho}(u,w)$ and $\mathbf{k}_n(u,w)=\int_w^u\mathbf{k}_1(u,v)\mathbf{k}_{n-1}(v,w)dv$. Define

$$\begin{aligned}
\boldsymbol{\xi}_i^{**}=&-\tilde{\boldsymbol{\delta}}_1(\boldsymbol{\beta}_0(u);u,\boldsymbol{\beta}_0)^{-1}\mathbf{D}_i(\boldsymbol{\beta}_0(u);u,\boldsymbol{\beta}_0)\\
&+\int_0^u\Gamma(u,w)\left\{-\tilde{\boldsymbol{\delta}}_1(\boldsymbol{\beta}_0(w);w,\boldsymbol{\beta}_0)^{-1}\mathbf{D}_i(\boldsymbol{\beta}_0(w);w,\boldsymbol{\beta}_0)\right\}dw.
\end{aligned}$$

Since $\boldsymbol{\xi}_i^{**}$ is a Donsker class, $\sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)$ asymptotically converges weakly to a Gaussian process with mean 0 and covariance matrix $\boldsymbol{\Sigma}(s,t)=E\{\boldsymbol{\xi}_1^{**}(s),\boldsymbol{\xi}_1^{**}(t)\}$.

Proof of *Theorem 2.1.5*: Define $\mathbf{A}(\mathbf{b})=E[X\sum_{j=1}^{\infty}I(L\le T^{(j)}\le\exp(\mathbf{X}^T\mathbf{b})\wedge R)]$ and $\mathbf{B}(\mathbf{b})=E[\mathbf{X}^{\otimes 2}\exp(\mathbf{X}^T\mathbf{b})I(L\le\exp(\mathbf{X}^T\mathbf{b})\le R)\dot{\mu}_{\mathbf{Z}}(\exp(\mathbf{X}^T\mathbf{b}))]$, and $\mathbf{v}_n(\mathbf{b}) = n^{-1}\sum_{i=1}^n \mathbf{X}_i \sum_{j=1}^\infty I(L_i \le T_i^{(j)} \le \exp(\mathbf{X}_i^T\mathbf{b}) \wedge R) - \mathbf{A}(\mathbf{b})$. Define $\tilde{\mathbf{A}}(\mathbf{b}) = E[\mathbf{X}I(L \le \exp(\mathbf{X}^T\mathbf{b}) \le R)]$ and $\mathbf{J}(\mathbf{b}) = E[\mathbf{X}^{\otimes 2}\exp(\mathbf{X}^T\mathbf{b})\{f_{L|\mathbf{Z}}(\exp(\mathbf{X}^T\mathbf{b})) - f_{R|\mathbf{Z}}(\exp(\mathbf{X}^T\mathbf{b}))\}]$, and $\tilde{\mathbf{v}}_n(\mathbf{b}) = n^{-1}\sum_{i=1}^n \mathbf{X}_i I(L_i \le \exp(\mathbf{X}_i^T\mathbf{b}) \le R_i) - \tilde{\mathbf{A}}(\mathbf{b})$.

For $d > 0$, define $\mathcal{B}(d) = \{\mathbf{b} \in R^{p+1} : \inf_{u\in(0,U]} \|\mathbf{A}(\mathbf{b}) - \mathbf{A}(\boldsymbol{\beta}_0(u))\| \le d\}$. Let $a_n = \|S_{L(n)}\|$; then $L(n) = U/a_n$. Let $\mathcal{A}(d) = \{\mathbf{A}(\mathbf{b}) : \mathbf{b} \in \mathcal{B}(d)\}$. Under condition C3, $\mathbf{A}(\cdot)$ is a one-to-one map from $\mathcal{B}(d_0)$ to $\mathcal{A}(d_0)$, since $(\mathbf{b}' - \mathbf{b})(\mathbf{A}(\mathbf{b}') - \mathbf{A}(\mathbf{b})) = E((\mathbf{X}^T\mathbf{b}' - \mathbf{X}^T\mathbf{b})(\tilde{N}(\exp(\mathbf{X}^T\mathbf{b}')) - \tilde{N}(\exp(\mathbf{X}^T\mathbf{b})))) \ge 0$ and the equation only holds when $\mathbf{b} = \mathbf{b}'$ under condition C3. So the inverse function of $\mathbf{A}(\cdot)$ exists, denoted by $\boldsymbol{\kappa}(\cdot)$, such that $\boldsymbol{\kappa}(\mathbf{A}(\mathbf{b})) = \mathbf{b}$ for any $\mathbf{b} \in \mathcal{B}(d_0)$.

Since $\hat{\boldsymbol{\beta}}(u_k)$ is the generalized solution to $n^{1/2}\mathbf{S}_n(\boldsymbol{\beta}, u_k)$, we have

$$\zeta_{n,k} = n^{-1}\sum_{i=1}^n \mathbf{X}_i\Big\{\sum_{j=1}^\infty I(L_i \le T_i^{(j)} \le \exp(\mathbf{X}_i^T\hat{\boldsymbol{\beta}}(u_k)) \wedge R_i) \\ - I(L_i \le \exp(\mathbf{X}_i^T\hat{\boldsymbol{\beta}}(s)) \le R_i)ds\Big\}$$

for $k = 1, 2, \ldots, L(n)$. Here, by the definition of a generalized solution, $\max_{k=1,\ldots,L(n)} \|\zeta_{n,k}\| \le \sup_i \|\mathbf{X}_i\|/n$.

Simple algebra shows that

$$\mathbf{A}\{\hat{\boldsymbol{\beta}}(u_k)\} - \mathbf{A}\{\boldsymbol{\beta}_0(u_k)\} - \sum_{l=0}^{k-1}\int_{u_l}^{u_{l+1}} \{\tilde{\mathbf{A}}(\hat{\boldsymbol{\beta}}(s)) - \tilde{\mathbf{A}}(\boldsymbol{\beta}_0(s))\}ds \\ = -\mathbf{v}_n\{\hat{\boldsymbol{\beta}}(u_k)\} + \int_0^{u_k} \tilde{\mathbf{v}}_n\{\hat{\boldsymbol{\beta}}(u_k)\}ds + \zeta_{n,k}.$$

Following similar arguments as in Peng and Huang (2008), it can be shown that under condition C1-C3, $\sup_{u\in(0,U]} \|\mathbf{A}(\hat{\boldsymbol{\beta}}(u)) - \mathbf{A}(\boldsymbol{\beta}_0(u))\| \xrightarrow{p} 0$.

Using Taylor expansion of $\boldsymbol{\kappa}(\mathbf{A}(\hat{\boldsymbol{\beta}}(u)))$ around $\mathbf{A}(\boldsymbol{\beta}_0(u))$ for $u \in [v, U]$, from condition C4, we get that

$$\begin{aligned}\|\hat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}_0(u)\| &\leq \|\mathbf{B}(\boldsymbol{\beta}_0(u))^{-1}(\mathbf{A}(\hat{\boldsymbol{\beta}}(u)) - \mathbf{A}(\boldsymbol{\beta}_0(u)))\| + \|\epsilon_n^*(u)\| \\ &\leq C_6\|\mathbf{A}(\hat{\boldsymbol{\beta}}(u)) - \mathbf{A}(\boldsymbol{\beta}_0(u))\| + \|\epsilon_n^*(u)\|,\end{aligned}$$

where $\sup_{u\in[v,U]} \|\epsilon_n^*(u)\| \xrightarrow{p} 0$ and $C_6 > 0$ does not depend on u. This completes the proof.

Proof of *Theorem 2.1.6*: Following the proof of Lemma B.1. in Peng and Huang (2008), we can show that given $\sup_{u\in(0,U]} \|\mathbf{A}(\hat{\boldsymbol{\beta}}(u)) - \mathbf{A}(\boldsymbol{\beta}_0(u))\| \xrightarrow{p} 0$,

$$\begin{aligned}\sup_{u\in(0,U]} \Bigg\| n^{-1/2} \sum_{i=1}^{n} \mathbf{X}_i(\tilde{N}_i(\exp(\mathbf{X}_i^T\hat{\boldsymbol{\beta}}(u))) - \tilde{N}_i(\exp(\mathbf{X}_i^T\boldsymbol{\beta}_0(u)))) \\ - n^{-1/2}(\mathbf{A}(\hat{\boldsymbol{\beta}}(u)) - \mathbf{A}(\boldsymbol{\beta}_0(u)))\Bigg\| \xrightarrow{p} 0. \qquad (2.38)\end{aligned}$$

and

$$\begin{aligned}\sup_{u\in(0,U]} \Bigg\| n^{-1/2} \sum_{i=1}^{n} \mathbf{X}_i(I(L_i \leq \exp(\mathbf{X}_i^T\hat{\boldsymbol{\beta}}(u)) \leq R_i) - I(L_i \leq \exp(\mathbf{X}_i^T\boldsymbol{\beta}_0(u)) \leq R_i)) \\ - n^{-1/2}(\tilde{\mathbf{A}}(\tilde{\boldsymbol{\beta}}(u)) - \tilde{\mathbf{A}}(\boldsymbol{\beta}_0(u)))\Bigg\| \xrightarrow{p} 0. \qquad (2.39)\end{aligned}$$

Let $o_I(a_n)$ denote a term that converges uniformly to 0 in probability in $u \in I$ after being divided by a_n. Because $n^{1/2}\|S_{L(n)}\| \to 0$, similar arguments as in Peng and Huang (2008) show that $n^{1/2}\mathbf{S}_n(\hat{\boldsymbol{\beta}}, u) = o_{(0,U]}(1)$, a.s. Then (2.38) and (2.39), coupled with the fact that $\mathbf{A}(\hat{\boldsymbol{\beta}}(u))$ converges uniformly to $\mathbf{A}(\boldsymbol{\beta}_0(u))$ for $u \in (0, U]$, imply that

$$
\begin{aligned}
&- n^{1/2}\mathbf{S}_n(\boldsymbol{\beta}_0, u) \\
=& n^{1/2}[\mathbf{A}(\hat{\boldsymbol{\beta}}(u)) - \mathbf{A}(\boldsymbol{\beta}_0(u))] - \int_0^u n^{1/2}[\tilde{\mathbf{A}}(\hat{\boldsymbol{\beta}}(s)) - \tilde{\mathbf{A}}(\boldsymbol{\beta}_0(s))]ds + o_{(0,U]}(1) \\
=& n^{1/2}[\mathbf{A}(\hat{\boldsymbol{\beta}}(u)) - \mathbf{A}(\boldsymbol{\beta}_0(u))] \\
&- \int_0^u (\mathbf{J}(\boldsymbol{\beta}_0(s))\mathbf{B}(\boldsymbol{\beta}_0(s))^{-1} + o_{(0,U]}(1))n^{1/2}(\mathbf{A}(\hat{\boldsymbol{\beta}}(u)) - \mathbf{A}(\boldsymbol{\beta}_0(u)))ds + o_{(0,U]}(1)
\end{aligned}
$$

According to the production integration theory (Gill and Johansen 1990; Andersen et al. 1998, II.6), we get

$$
n^{1/2}(\mathbf{A}(\hat{\boldsymbol{\beta}}(u)) - \mathbf{A}(\boldsymbol{\beta}_0(u))) = \boldsymbol{\phi}\{-n^{1/2}\mathbf{S}_n(\boldsymbol{\beta}_0(u), u)\} + o_{(0,U]}(1), \tag{2.40}
$$

where $\boldsymbol{\phi}(\cdot)$ is a linear operator from $\mathcal{F}$ to $\mathcal{F}$ (see Peng and Huang, 2008).

By the Donsker theorem, $-n^{1/2}\mathbf{S}_n(\boldsymbol{\beta}_0, u)$ converges weakly to a tight Gaussian process $\mathbf{G}(u)$ for $u \in (0, U]$. Hence, $n^{1/2}\{\mathbf{A}(\hat{\boldsymbol{\beta}}(u)) - \mathbf{A}(\boldsymbol{\beta}_0(u))\}$ converges weakly to $\boldsymbol{\phi}(\mathbf{G}(u))$ which is also a Gaussian process since $\boldsymbol{\phi}$ is a linear operator. Using Taylor expansions we have $n^{1/2}\{\hat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}_0(u)\}$ converges weakly to a Gaussian process $\mathbf{B}(\boldsymbol{\beta}_0(u))^{-1}\boldsymbol{\phi}(\mathbf{G}(u))$, which is also a Gaussian.

Chapter 3

Censored Quantile Regression Analysis of Longitudinal Data with an Informative Intermittent Missing Pattern

3.1 Regression Procedures

3.1.1 Data and Model

Suppose that N individuals are expected to be measured at K occasions. Let $\mathbf{y}_i^* = (y_{i1}^*, y_{i2}^*, \ldots, y_{iK}^*)'$, where y_{ij}^* denotes the jth outcome of the ith subject subject to left censoring by a fixed known constant c, and let $\mathbf{z}_i = (\mathbf{z}_{i1}', \mathbf{z}_{i2}', \ldots, \mathbf{z}_{iK}')'$, where $\mathbf{z}_{ij}$ denote the corresponding $p \times 1$ covariate vector, $i = 1, \ldots, N$ and $j = 1, \ldots, K$. Define the observed $y_{ij} = \max(c, y_{ij}^*)$ and $\eta_{ij} = I(y_{ij}^* > c)$, where $I(\cdot)$ is an indicator function. Let $\delta_{ij} = 1$ if $(y_{ij}, \eta_{ij}, \mathbf{z}_{ij})$ is available and 0 if missing. Observed data include $\{(y_{ij}, \eta_{ij}, \mathbf{z}_{ij}) : \delta_{ij} = 1\}$. Without loss of generality, the conditional quantiles of y_{ij}^* given $\mathbf{z}_{ij}$ is defined as $Q_{y_{ij}^*}(\tau|\mathbf{z}_{ij}) = \inf\{y : Pr(y_{ij}^* \leq y|\mathbf{z}_{ij}) \geq \tau\}$ for $\tau \in [0, 1]$. A quantile regression model may linearly link $Q_{y_{ij}^*}(\tau|\mathbf{z}_{ij})$ to $\mathbf{z}_{ij}$ as follows:

$$Q_{y_{ij}^*}(\tau|\mathbf{x}_{ij}) = \mathbf{x}_{ij}^T \boldsymbol{\beta}_0(\tau), \tag{3.1}$$

where $\mathbf{x}_{ij} = (1, \mathbf{z}_{ij})$ and $\boldsymbol{\beta}_0(\tau)$ is a vector of unknown regression coefficients, representing the effects of covariates on the τth conditional quantile of y_{ij}^* and may change with τ. For the observed y_{ij}, the corresponding conditional quantile satisfies

$$Q_{y_{ij}}(\tau|\mathbf{x}_{ij}) = \max(c, \mathbf{x}_{ij}^T \boldsymbol{\beta}_0(\tau)). \tag{3.2}$$

This fact serves as the base for the proposed estimation of $\boldsymbol{\beta}_0(\tau)$.

3.1.2 Estimation Procedure and Inference

With complete data $\{(y_{ij}, \eta_{ij}, \mathbf{z}_{ij}) : i = 1, \dots, N, j = 1, \dots, K\}$, the estimating equation for censored longitudinal data is

$$0 = \sum_{i=1}^{N} \sum_{j=1}^{K} \mathbf{x}_{ij} \left\{ \tau - I(y_{ij} < \max(c, \mathbf{x}_{ij}^T \boldsymbol{\beta})) \right\}. \tag{3.3}$$

Missingness in the data may cause problem in estimating equation (3.3). When the missingness is completely at random, the estimating equation (3.3) is still working. However, when the missingness is informative about the response variable, estimators resulting from equation (3.3) would be biased.

Our strategy is to adopt the inverse probability weighting method and assume the responses are missing at random (MAR). Here are assumptions needed.

Assumptions:

(1) There is a $q \times 1$ random vector $\mathbf{v}_{ij}$, such that

$$\pi_{ij} = P(\delta_{ij} = 1 | \mathbf{y}_i^*, \mathbf{z}_i) = P(\delta_{ij} = 1 | \mathbf{v}_{ij}) = p(\mathbf{v}_{ij}, \boldsymbol{\alpha}),$$

where $\boldsymbol{\alpha}$ denotes some parameters.

(2) For all $\mathbf{v} \in \mathbf{V} \subset \mathbf{R}^q$, $p(\mathbf{v}) > 0$, where $\mathbf{V}$ is a set of all possible $\mathbf{v}_{ij}$.

(3) $\mathbf{v}_{ij}$ is observed whenever $\delta_{ij} = 1$.

We propose to weight the available data on $(y_{ij}, \mathbf{z}_{ij}, \eta_{ij})$ by the inverse probability of $\delta_{ij} = 1$. Specifically, the proposed estimating equation for $\boldsymbol{\beta}_0(\tau)$ takes the form,

$$0 = \mathbf{U}_N^{\pi}(\boldsymbol{\beta}) = N^{-1/2} \sum_{i=1}^{N} \sum_{j=1}^{K} \frac{I(\delta_{ij} = 1)}{\pi_{ij}} \mathbf{x}_{ij} \left\{ \tau - I(y_{ij} < \max(c, \mathbf{x}_{ij}^T \boldsymbol{\beta})) \right\}. \tag{3.4}$$

The probability π_{ij} is usually unknown. We need to conduct additional modeling for the missing mechanism. Note that δ_{ij} is a binary variable. So regression models for binary variable, such as logistic regression, could be adopted depending on the missingness structure present in the data. Define an estimator, such as maximum likelihood estimator, of $\boldsymbol{\alpha}$ as $\hat{\boldsymbol{\alpha}}$ and an estimator of π_{ij} as $\hat{\pi}_{ij} = p(\mathbf{v}_{ij}, \hat{\boldsymbol{\alpha}})$. Plugging in π_{ij} in equation (3.4) by $\hat{\pi}_{ij} = p(\mathbf{v}_{ij}, \hat{\boldsymbol{\alpha}})$, we get a new estimating equation that is

$$\mathbf{U}_N(\boldsymbol{\beta}) = N^{-1/2} \sum_{i=1}^{N} \sum_{j=1}^{K} \frac{I(\delta_{ij} = 1)}{\hat{\pi}_{ij}} \mathbf{x}_{ij} \left\{ \tau - I(y_{ij} < \max(c, \mathbf{x}_{ij}^T \boldsymbol{\beta})) \right\}. \tag{3.5}$$

It can be proved that if $\hat{\boldsymbol{\alpha}}$ and $p(\mathbf{v}, \boldsymbol{\alpha})$ satisfy some condition (C1 in section 3.1.3 Asymptotical Results), estimator $\hat{\boldsymbol{\beta}}(\tau)$ resulting from estimating equation (3.5) would be uniformly consistent and asymptotical normal. Fortunately, the condition C1 is not hard to satisfy. Examples satisfying condition C1 are present in 3.5 Appendix.

The solution-finding problem to (3.5) is equivalent to locating the minimizer of the following objective function,

$$\sum_{i=1}^{N} \sum_{j=1}^{K} \frac{I(\delta_{ij} = 1)}{\hat{\pi}_{ij}} \rho_\tau \left\{ y_{ij} - \max\left(c, \mathbf{x}_{ij}^T \boldsymbol{\beta}\right) \right\}, \tag{3.6}$$

where $\rho_\tau(u) = u\tau I(u \geq 0) - u(1-\tau)I(u < 0)$. The minimizer to (3.6) is denoted by $\hat{\boldsymbol{\beta}}(\tau)$. The minimization problem of (3.6) can be easily solved using the Barrodale-Roberts algorithms ((Barrodale and Roberts, 1974)), the implementation of which is available in standard statistical software, for example, the $l1fit()$ function in S-PLUS or the $rq()$ function in R package *quantreg*.

The covariance matrix of $n^{1/2} \left\{ \hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau) \right\}$ depends on some unknown value, such as the probability density function of the outcome variable. Therefore, we adopted bootstrapping method to obtain the confidence interval of $\boldsymbol{\beta}_0(\tau)$.

3.1.3 Asymptotic Results

We can show that for $\tau \in (0,1)$, $\hat{\boldsymbol{\beta}}(\tau)$ is uniformly consistent and $n^{1/2}\left\{\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\right\}$ converges weakly to a Gaussian process. The regularity conditions include:

C 1. Let $\boldsymbol{\alpha}$ be the parameters in the missing model. Denote $\hat{\boldsymbol{\alpha}}$ as the estimator and $\boldsymbol{\alpha}_0$ as the true value. Define $\hat{\pi}_{ij} = p(\mathbf{v}_{ij}, \hat{\boldsymbol{\alpha}})$.

- (a) $\hat{\boldsymbol{\alpha}}$ is consistent to $\boldsymbol{\alpha}_0$;
- (b) There exist $\boldsymbol{\xi}_{1,j}$, such that $\left\|\sqrt{n}\,(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) - n^{-1/2}\sum_{j=1}^{n}\boldsymbol{\xi}_{1,j}\right\| \xrightarrow{p} 0$;
- (c) $p(\mathbf{v}_{ij}, \boldsymbol{\alpha})$ has continuous derivatives of $\boldsymbol{\alpha}$ around $\boldsymbol{\alpha}_0$, for all $\mathbf{v}_{ij} \in \mathbf{V}$.

C 2. $\mathbf{z}$ is bounded.

C 3.
- (a) $\boldsymbol{\beta}_0(\tau)$ is a Lipschitz continuous for $\tau \in (0,1)$;
- (b) The density function of Y_{ik}, $f_k(y|\mathbf{z}_i)$ is bounded above uniformly in $\mathbf{z}_i$.

C 4. For some $\rho_0 > 0$ and $c_0 > 0$, $\inf_{\mathbf{b}\in\mathcal{B}(\rho_0)} eigmin\mathbf{A}(\mathbf{b}) \leq -c_0$, where $\mathcal{B}(\rho) = \{\mathbf{b} \in R^{p+1} : \inf_{\tau\in(0,1)} \|\mathbf{b} - \boldsymbol{\beta}_0(\tau)\| \leq \rho\}$ and

$$\mathbf{A}(\mathbf{b}) = \frac{d}{d\mathbf{b}} E\left[-\sum_{k=1}^{K}\frac{I(\delta_{ik}=1)}{\pi_{ik}}\mathbf{x}_{ik} I\{y_{ik} \leq \max(c, \mathbf{x}_{ik}^T\mathbf{b})\}\right].$$

We establish the uniform consistency and weak convergence of $\hat{\boldsymbol{\beta}}(\tau)$ stated in the following theorems.

Theorem 3.1.1. *Under conditions C1-C4,* $\sup_{\tau\in(0,1)} \|\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\| \xrightarrow{p} 0$.

Theorem 3.1.2. *Under conditions C1-C4,* $\left\{n^{1/2}\left(\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\right) : \tau \in (0,1)\right\}$ *converges weakly to a Gaussian process with mean 0 and covariance matrix* Σ*, where* Σ *is presented in proof.*

Condition C 1 is usually satisfied if only we have a correct model for missingness (examples see 3.5 Appendix). The proofs of Theorems 3.1.2 – 3.1.2 are presented in 3.5 Appendix.

3.2 Simulation Studies

Finite-sample performance of the proposed method was evaluated through Monte Carlo simulations. We considered the scenario that 200 individuals are expected to have repeated measurements at four time points, 0,5,11, and 18. We call the measurement at time zero as the baseline measurement. Let x_{ij} be the jth visit time of the ith subject. The latent response variable y^*_{ij} is generated from the following models:

- Case 1 (normal random effect and constant slope):

$$y^*_{ij} = 1.5 + a_i - 0.01x_{ij} + e_{ij}, \qquad i = 1, \dots, 200,\ j = 1, \dots, 4,$$

  where random effect $a_i \sim N(0, 1)$ and $e_{ij} \sim N(0, 0.09)$. In this setup, $Q_{y^*_{ij}}(\tau|x_{ij})$ $= 1.5 + \sqrt{(1.09)} \times \Phi^{-1}(\tau) - 0.01x_{ij}$.

- Case 2 (skewed random effect and constant slope):

$$y^*_{ij} = 1 + a_i - 0.01x_{ij} + e_{ij}, \qquad i = 1, \dots, 200,\ j = 1, \dots, 4,$$

  where random effect $a_i \sim \Gamma(3, 0.5) - 1.5$ and $e_{ij} \sim N(0, 0.09)$. In this setup, $Q_{y^*_{ij}}(\tau|x_{ij}) = 1 + F^{-1}_{a+e}(\tau) - 0.01x_{ij}$, where F_{a+e} is the distribution function of $a + e$ and $F^{-1}_{a+e}(\tau)$ is gained from Monte Carlo simulation.

- Case 3 (normal random effect and changing slopes):

$$y_{ij}^* = 1.5 + a_{i0} - (0.01 + a_{i1})x_{ij} - a_{i2}\sqrt{x_{ij}} + e_{ij}, \qquad i = 1, \ldots, 200,\ j = 1, \ldots, 4,$$

where $a_{i0} \sim N(0, 0.91)$, $a_{i1} \sim N(0, 10^{-4})$, $a_{i2} \sim N(0, 0.02)$, and $e_{ij} \sim N(0, 0.09)$. We can show that y_{ij}^* given x_{ij} follows a normal distribution, $N(1.5 - 0.01x_{ij}, (1 + 0.01x)^2)$, that its quantile is linear in x_{ij}. That is, $Q_{y_{ij}^*}(\tau|x_{ij}) = 1.5 + \Phi^{-1}(\tau) - 0.01(1 - \Phi^{-1}(\tau))x_{ij}$.

We left censored y_{ij}^* at zero, i.e. the observed $y_{ij} = \max(y_{ij}^*, 0)$. About 10% of response variables are left censored at zero. In our simulation, y_{i1} is always observed and let $v_{ij} = y_{i1}$, for $j = 2, 3, 4$. The probability of y_{ij} being available in the other three time points follow a logistic regression model:

$$\pi_{ij} \;=\; \frac{\exp(2 - y_{i1})}{1 + \exp(2 - y_{i1})}, \qquad i = 1, \ldots, 200,\ j = 2, 3, 4.$$

The average of π_{ij} for $j = 2, 3, 4$ in three cases are 40%, 30%, and 40% respectively. Bootstrapping size for inference is 500. Each setup is repeated 500 times.

We fitted the model by using the proposed estimating equation (3.4) (referred to as Weighted QR). We also applied Wang and Fygenson (2009)'s method which does not account for informative missingness (referred to as Unweighted QR). Coefficient estimates from both unweighted and weighted estimating equations of the 25th, 50th, and 75th quantiles are summarized in Table 3.1. Without considering informative missingness, unweighted estimators appear to be biased. The coverage rate of 95% confidence interval from bootstrapping seems to deviate from the nominal value. In contrast, the proposed estimator is virtually unbiased. The coverage rate of 95% confidence interval from bootstrapping is close to the nominal value. Table 3.1 also shows that the empirical standard error (Emp SE) agree quite well with average bootstrapping-based standard error estimates.

In summary, our simulation study suggests that the proposed weighted estimator is a good estimator for quantile regression coefficients when outcome is under left censoring and missing at random. It is also demonstrated that bootstrapping is an appropriate approach for inference.

Table 3.1: Comparison of the unweighted censored quantile regression estimator (Unweighted QR) and our proposed weighted estimator with logistic missingness model(Weighted QR): Cov95 – coverage rate of the bootstrapping 95% confidence interval; Emp SE – empirical standard error; Avg SE – average bootstrapping-based standard error estimates.

			Unweighted QR		Weighted QR			
		TRUE	Bias	Cov95	Bias	Emp SE	Avg SE	Cov95
				$\tau = 0.25$				
Case 1	$\hat{\beta}^{(1)}$	.796	-.077	.88	-.006	.098	.104	.94
	$\hat{\beta}^{(2)}$	-.010	-.014	.63	-.000	.005	.006	.97
Case 2	$\hat{\beta}^{(1)}$	.344	-.015	.95	.004	.063	.063	.95
	$\hat{\beta}^{(2)}$	-.010	-.005	.93	.000	.004	.004	.95
Case 3	$\hat{\beta}^{(1)}$	.826	-.054	.91	.006	.099	.100	.94
	$\hat{\beta}^{(2)}$	-.017	-.015	.79	-.000	.007	.007	.94
				$\tau = 0.5$				
Case 1	$\hat{\beta}^{(1)}$	1.50	-.088	.82	-.002	.100	.101	.93
	$\hat{\beta}^{(2)}$	-.010	-.017	.99	.001	.006	.007	.96
Case 2	$\hat{\beta}^{(1)}$	.856	-.049	.89	.001	.076	.078	.95
	$\hat{\beta}^{(2)}$	-.010	-.008	.98	.000	.005	.005	.96
Case 3	$\hat{\beta}^{(1)}$	1.50	-.065	.87	.006	.092	.096	.95
	$\hat{\beta}^{(2)}$	-.010	-.017	.85	.000	.007	.008	.96
				$\tau = 0.75$				
Case 1	$\hat{\beta}^{(1)}$	2.20	-.096	.82	.002	.115	.120	.94
	$\hat{\beta}^{(2)}$	-.010	-.020	.13	.001	.010	.010	.97
Case 2	$\hat{\beta}^{(1)}$	1.50	-.087	.82	-.005	.113	.117	.94
	$\hat{\beta}^{(2)}$	-.010	-.012	.41	.000	.008	.008	.96
Case 3	$\hat{\beta}^{(1)}$	2.17	-.080	.84	.003	.111	.114	.95
	$\hat{\beta}^{(2)}$	-.033	-.019	.35	.001	.010	.011	.95

3.3 PBB Data Example

In this project, we use the same data set as in Terrell et al. (2008) except for six women whose first visit is more than 13 years later since the exposure (defined ad July 1, 1973). These six women were excluded because they would contribute little information for studying PBB change in a time window not far from exposure. The data were collected during 1976 - 1994. Specifically, data were collected in 12 years: 1977, 1978, 1979, 1980, 1981, 1982, 1983, 1988, 1989, 1992, 1993, and 1994. The outcome y^*_{ij} is the underlying PBB level of the ith subject at the jth year in the 12 years listed above. The maximum visit number in the data is 7, which means many y^*_{ij}'s are not available. Knowing that the missingness pattern is related with the observed measurements, we first need to model the missing pattern before conducting quantile regression. We assume that the missingness pattern in the follow-up visit is dependent on the observed measurement at the first visit which is referred to as initial PBB measurement in the following. We categorized the initial PBB measurement (after logarithm transformation) into six groups: $[0, 1)$, $[1, 2)$, $[2, 3)$, $[3, 4)$, $[4, 5)$, and $[5, 7)$. Let $p^{(k)}_j$ be the probability of a woman with first measurement in the range $[k-1, k)$ or $[5, 7)$ for $k = 6$ to have a PBB concentration measurement in the jth year. We plot the empirical sample proportions in Figure 3.1. It shows that high PBB concentration tends to have more measurements. As explained by epidemiological investigator of this study, this occurred because in the late 1980's those with PBB > 10 were contacted by project staff for re-tests. They chose 10 ppb because they noted that it would be easier to see a decrease in PBB if the level was around 10 ppb. In the 1990's, there was an updated health questionnaire done for the entire cohort, and at this time all participants were able to have their PBB levels measured again. Figure 3.1 also demonstrates that (a) the visiting pattern of each group in each year is quite different from the pattern in another year; (b) the probability of being observed does not seem to be linearly correlated

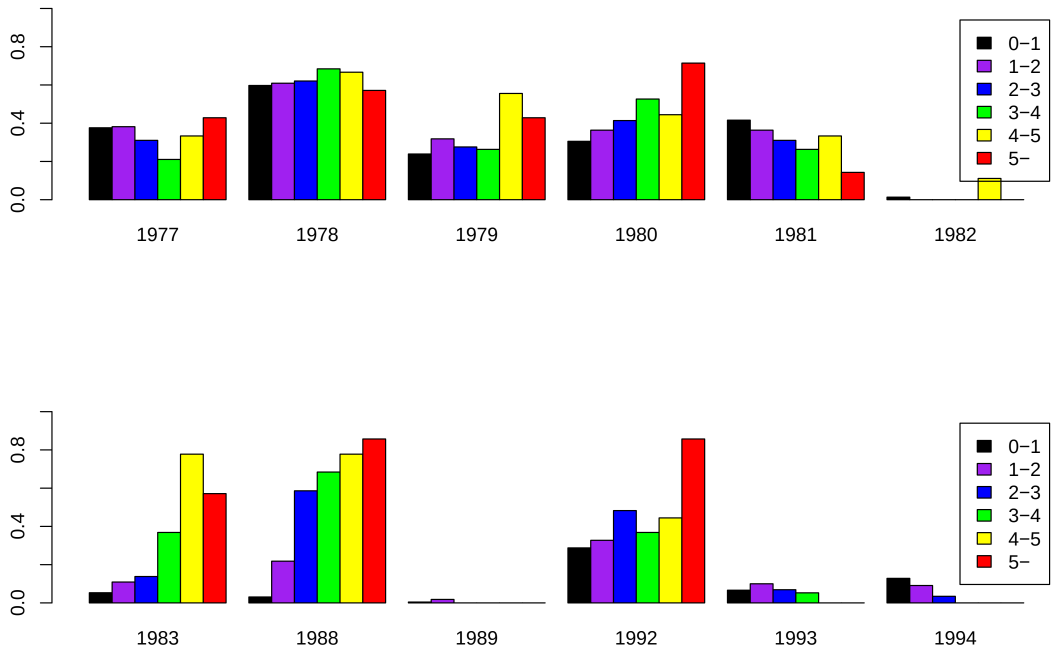

Figure 3.1: Empirical sample proportion of being available for each group in every year

with the continuous initial PBB outcome. Based on these observations, we use the sample proportion of being followed up in each year as the estimator of π_{ij}, $\hat{\pi}_{ij} = \sum_{k=1}^{6} p_j^{(k)} I(\text{ith subject is in kth group according to the initial PBB measurement})$, avoiding strong parametric assumptions.

We fit model (3.1) with x_{ij} being the jth visit time (referred to as Time) of the ith subject since PBB exposure. We fitted the model by using the proposed estimating equation (3.4) (referred to as Weighted QR). We also applied Wang and Fygenson (2009)'s method which does not account for the informative missing pattern (referred to as Unweighted QR). The estimates of quantile regression coefficients $\boldsymbol{\beta}(\tau)$ are summarized in Table 1. From the weighted estimating equation, the negative coefficients for Time conform to our intuition that PBB concentration generally decreases over time. The increasing pattern in the magnitude of the Time coefficient with τ further demonstrates that a faster decaying profile for high PBB quantiles compared to low PBB quantiles.

The decay rates for the 25th, 50th, and 75th percentiles of PBB concentration are not significant, while the estimated decay rates for the 85th, 90th, and 95th quantiles are significantly lower than zero. Specifically, the 85th percentile decreases 3 percents per year (95% confidence interval: 0 - 5 percents per year). The 90th percentile decreases 4 percents per year (95% confidence interval: 1 - 7 percents per year). The 95th percentile decreases 6 percents per year (95% confidence interval: 3 - 11 percents per year).

On the other hand, results from the unweighted quantile regression appear to underestimate the decay rate. In fact, the unweighted quantile regression results in positive slope estimates for time. This result is not right, because the PBB serum concentration is impossible to increase in body. This real example further demonstrates that failing to handle informative missingness, the resulting estimates would be substantially biased and conclusions would be incorrect.

3.4 Remarks

Left censoring and missing measurements are often simultaneously present in longitudinal studies. Failing to handle these data features, as shown in our simulation study and data example, can lead to considerably biased estimation and consequently misleading scientific conclusions. The new regression method developed in this work

Table 3.2: Parameter estimates and 95% confidence interval for 25th quantile, 50th quantile, 75th quantile, 85th quantile, 90th quantile, and 95th quantile

		Unweighted QR		Weighted QR	
Quantile	Effect	Estimate	95% CI	Estimate	95% CI
25th	Intercept	0.159	(0.134, 1.109)	0.182	(0.127, 2.538)
	Time	0.006	(-0.119, 0.013)	-0.000	(-0.595, 0.014)
50th	Intercept	0.851	(0.519, 0.968)	0.773	(0.556, 1.030)
	Time	0.005	(-0.007, 0.022)	-0.008	(-0.026, 0.008)
75th	Intercept	1.533	(1.340, 1.859)	1.476	(1.289, 1.798)
	Time	0.025	(0.000, 0.047)	-0.006	(-0.029, 0.002)
85th	Intercept	2.298	(1.838, 2.924)	2.208	(1.706, 2.695)
	Time	0.019	(-0.012, 0.071)	-0.026	(-0.054, -1e-4)
90th	Intercept	2.973	(2.331, 3.617)	2.897	(2.168, 3.272)
	Time	0.035	(-0.014, 0.085)	-0.043	(-0.069, -0.006)
95th	Intercept	3.872	(3.204, 4.852)	3.917	(3.032, 4.422)
	Time	0.043	(-0.012, 0.075)	-0.060	(-0.113, -0.026)

appropriately account for these complications. Furthermore, by adopting quantile regression modeling, we may offer a more comprehensive view of the profile of longitudinal outcomes as well as its relationship with covariates, as compared to some traditional modeling, such as linear mixed models.

Here we need to emphasize that our proposed weighted estimating equation is sensitive to the probability weight. For example, the unweighted quantile regression without considering informative missingness is a kind of model without appropriate weight. Good probability estimate for missingness is key to gain a good estimate in quantile regression models.

3.5 Appendix

3.5.1 Examples of Estimators Satisfying Condition C 1

Condition C 1 is usually satisfied if only we have a correct model for missingness.

- Example 1: Empirical sample proportion:

  Empirical sample proportion of an event is a consistent estimator for true probability of the event. Define $\boldsymbol{\xi}_{1,j} = I(\text{event happens for the jth subject}) - \boldsymbol{\alpha}_0$. It is easy to see that $\|\sqrt{n}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) - n^{-1/2}\sum_{j=1}^{n}\boldsymbol{\xi}_{1,j}\| \xrightarrow{p} 0$. Also, $p(\mathbf{v}, \boldsymbol{\alpha}) = \boldsymbol{\alpha}$ has continuous derivatives in $\boldsymbol{\alpha}$.

- Example 2: Maximum likelihood estimator (MLE):

  First, MLE is a consistent estimator. Denote the score function as $S(\boldsymbol{\alpha}) = \frac{1}{n}\sum_{i=1}^{n} S_i(\boldsymbol{\alpha}) = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial l_i(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}$, where $l_i(\alpha)$ is the log-likelihood function for the ith subject. Define $s(\boldsymbol{\alpha}) = E(S_i(\boldsymbol{\alpha}))$. Briefly, we have $\|\sqrt{n}\{S(\hat{\boldsymbol{\alpha}}) - S(\boldsymbol{\alpha}_0) - s(\hat{\boldsymbol{\alpha}}) + s(\boldsymbol{\alpha}_0)\}\| \xrightarrow{p} 0$, $S(\hat{\boldsymbol{\alpha}}) = 0$, and $s(\hat{\boldsymbol{\alpha}}) - s(\boldsymbol{\alpha}_0) = (s'(\boldsymbol{\alpha}_0) + o_p(1))(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0)$. Define $\boldsymbol{\xi}_{1,j} = -s'(\boldsymbol{\alpha}_0)^{-1}S_j(\boldsymbol{\alpha}_0)$. It can be proven that $\|\sqrt{n}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) - n^{-1/2}\sum_{j=1}^{n}\boldsymbol{\xi}_{1,j}\| \xrightarrow{p} 0$. Also, $p(\mathbf{v}, \boldsymbol{\alpha}) = \frac{\exp(\mathbf{v}^T\boldsymbol{\alpha})}{1+\exp(\mathbf{v}^T\boldsymbol{\alpha})}$ has continuous derivatives in $\boldsymbol{\alpha}$.

3.5.2 Proof of Theorems

Proof of *Theorem 3.1.1* Define $\boldsymbol{\mu}(\boldsymbol{\beta}, \tau) = E\left\{n^{-1/2}U_n^{\pi}(\boldsymbol{\beta}, \tau)\right\}$. Condition C1 coupled with C2 implies that

$$\sup_{\tau\in(0,1),\boldsymbol{\beta}\in R^{p+1}} \|n^{-1/2}\mathbf{U}_n(\boldsymbol{\beta}, \tau) - n^{-1/2}\mathbf{U}_n^{\pi}(\boldsymbol{\beta}, \tau)\| = o(1), \quad a.s.$$

Define $\mathcal{F} = \sum_{k=1}^{K}\frac{I(\delta_{ik}=1)}{\pi_{ik}}\mathbf{x}_{ik}\left[\tau - I\left\{y_{ik} \le \max(c, \mathbf{x}_{ik}^T\boldsymbol{\beta})\right\}\right]$, $\boldsymbol{\beta} \in R^{p+1}$, $\tau \in (0,1)$. The function class $\mathcal{F}$ is Donsker and thus Glivenko-Cantelli because the class of indicator functions is Donsker. It then follows the Glivenko-Cantelli theorem that $\sup_{\boldsymbol{\beta}\in R^{p+1},\tau\in(0,1)}\|n^{-1/2}\mathbf{U}_n^{\pi}(\boldsymbol{\beta}, \tau) - \boldsymbol{\mu}(\boldsymbol{\beta}, \tau)\| = o(1)$, a.s. Therefore,

$$\sup_{\boldsymbol{\beta}\in R^{p+1},\tau\in(0,1)} \|n^{-1/2}\mathbf{U}_n(\boldsymbol{\beta}, \tau) - \boldsymbol{\mu}(\boldsymbol{\beta}, \tau)\| = o(1), \quad a.s. \tag{3.7}$$

Given $\boldsymbol{\mu}(\boldsymbol{\beta}_0(\tau),\tau) = 0$, we have $\sup_{\tau\in(0,1)} \|n^{-1/2}\mathbf{U}_n(\boldsymbol{\beta}_0(\tau),\tau)\| = o(1), \quad a.s.$. And according to our estimating method, $\mathbf{U}_n(\hat{\boldsymbol{\beta}}(\tau),\tau) = o(1)$. Simple manipulation shows that

$$
\begin{aligned}
&\sup_{\tau\in(0,1)} \|\boldsymbol{\mu}(\hat{\boldsymbol{\beta}}(\tau),\tau) - \boldsymbol{\mu}(\boldsymbol{\beta}_0(\tau),\tau)\| \\
&\le \sup_{\tau\in(0,1)} \|\boldsymbol{\mu}(\hat{\boldsymbol{\beta}}(\tau),\tau) - n^{-1/2}\mathbf{U}_n(\hat{\boldsymbol{\beta}}(\tau),\tau)\| + \sup_{\tau\in(0,1)} \|n^{-1/2}\mathbf{U}_n(\hat{\boldsymbol{\beta}}(\tau),\tau)\| \\
&\qquad + \sup_{\tau\in(0,1)} \|n^{-1/2}\mathbf{U}_n(\boldsymbol{\beta}_0(\tau),\tau)\| + \sup_{\tau\in(0,1)} \|n^{-1/2}\mathbf{U}_n(\boldsymbol{\beta}_0(\tau),\tau) - \boldsymbol{\mu}(\boldsymbol{\beta}_0(\tau),\tau)\| \\
&= o(1), \quad a.s. \qquad (3.8)
\end{aligned}
$$

Now we have $\sup_{\tau\in(0,1)} \|\boldsymbol{\mu}(\hat{\boldsymbol{\beta}}(\tau),\tau)\| \xrightarrow{a.s.} 0$. If $\sup_{\tau\in(0,1)} \|\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\| \overset{a.s.}{\not\longrightarrow} 0$, then there must exist $\rho_0 \ge \epsilon > 0$ and a sequence $\{\tau, \boldsymbol{\zeta}\}$ satisfying $\|\boldsymbol{\zeta} - \boldsymbol{\beta}_0(\tau)\| > \epsilon$ such that $\|\boldsymbol{\mu}(\boldsymbol{\zeta},\tau) - \boldsymbol{\mu}(\boldsymbol{\beta}_0(\tau),\tau)\| \to 0$. However, since $\mathbf{u}^T\boldsymbol{\mu}(\boldsymbol{\beta}_0(\tau) + \mathbf{u}\delta,\tau)$ is a decreasing function in δ for any $\mathbf{u} \in R^{p+1}$ satisfying $\|\mathbf{u}\|^2 = 1$ and condition C4, we have

$$
\begin{aligned}
&\|\boldsymbol{\mu}(\boldsymbol{\zeta},\tau) - \boldsymbol{\mu}(\boldsymbol{\beta}_0(\tau),\tau)\|^2 \cdot \|\mathbf{v}\|^2 \\
&\ge \left[\mathbf{v}^T\left\{\boldsymbol{\mu}(\boldsymbol{\zeta},\tau) - \boldsymbol{\mu}(\boldsymbol{\beta}_0(\tau),\tau)\right\}\right]^2 \\
&\ge \left[\mathbf{v}^T\left\{\boldsymbol{\mu}(\boldsymbol{\beta}_0(\tau) + \epsilon\mathbf{v},\tau) - \boldsymbol{\mu}(\boldsymbol{\beta}_0(\tau),\tau)\right\}\right]^2 \\
&\ge c_0^2\epsilon^2 \qquad (3.9)
\end{aligned}
$$

where $\mathbf{v} = \frac{\boldsymbol{\zeta}-\boldsymbol{\beta}_0(\tau)}{\|\boldsymbol{\zeta}-\boldsymbol{\beta}_0(\tau)\|}$. Inequality in (3.9) contradicts with $\|\boldsymbol{\mu}(\boldsymbol{\zeta},\tau) - \boldsymbol{\mu}(\boldsymbol{\beta}_0(\tau),\tau)\| \to 0$. Therefore, we proved that $\sup_{\tau\in(0,1)} \|\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\| \xrightarrow{a.s.} 0$.

Lemma 1. For any positive sequence $\{d_n\}_{n=1}^{\infty}$ satisfying $d_n \to 0$,

$$
\sup_{\mathbf{b},\mathbf{b}'\in\mathcal{B}(\rho_0),\|\mathbf{b}-\mathbf{b}'\|\le d_n,\tau\in(0,1)} \left\|\mathbf{U}_n^{\pi}(\mathbf{b},\tau) - \mathbf{U}_n^{\pi}(\mathbf{b}',\tau) - n^{1/2}\{\boldsymbol{\mu}(\mathbf{b},\tau) - \boldsymbol{\mu}(\mathbf{b}',\tau)\}\right\| \xrightarrow{a.s.} 0.
$$

Proof. This lemma can be proved by using the results in Alexander (1984) and the arguments for theorem 1 of Lai and Ying (1988). The crucial step is to show that there exists $G_0 > 0$ such that

$$var\left[\sum_{k=1}^{K}\frac{I(\delta_{ik}=1)}{\pi_{ik}}\mathbf{x}_{ik}\left[I\{y_{ik}\leq \max(c,\mathbf{x}_{ik}^T\mathbf{b}')\} - \{I(y_{ik}\leq \max(c,\mathbf{x}_{ik}^T\mathbf{b}))\}\right]\right]$$
$$\leq G_0\|\mathbf{b}-\mathbf{b}'\|.$$

This follows from the uniform boundedness of $f_k(y|\mathbf{z})$ and the boundedness of $\mathbf{Z}$ and $\mathcal{B}(\rho_0)$, which are implied by C2 and C3.

Proof of *Theorem 3.1.2* Denote

$$\mathbf{w}(\mathbf{b},\tau) = E\left[\sum_{k=1}^{K}\frac{1}{\pi_{ik}^2}I(\delta_{ik}=1)\mathbf{x}_{ik}[\tau - I\{y_{ik}\leq \max(c,\mathbf{x}_{ik}^T\mathbf{b})\}]\mathbf{h}_{ik}\right],$$

where $\mathbf{h}_{ik} = \frac{\partial p(\mathbf{v}_{ik},\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}}\Big|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}$. Using similar empirical process arguments for $\mathcal{F}$, we can show that $n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K}\frac{1}{\pi_{ik}^2}I(\delta_{ik}=1)\mathbf{x}_{ik}[\tau - I\{y_{ik}\leq \max(c,\mathbf{x}_{ik}^T\mathbf{b})\}]\mathbf{h}_{ik}$ converges to $\mathbf{w}(\mathbf{b},\tau)$ uniformly in both $\mathbf{b}\in R^{p+1}$ and $\tau\in(0,1)$. Denote

$$\boldsymbol{\xi}_{2,i}(\tau) = \sum_{k=1}^{K}\frac{I(\delta_{ik}=1)}{\pi_{ik}}\mathbf{x}_{ik}\left[\tau - I\left\{y_{ik}\leq \max(c,\mathbf{x}_{ik}^T\boldsymbol{\beta}_0(\tau))\right\}\right].$$

It follows from standard asymptotic arguments that

$$
\begin{aligned}
&\mathbf{U}_n\{\boldsymbol{\beta}_0(\tau),\tau\} \\
=&\mathbf{U}_n^{\pi}\{\boldsymbol{\beta}_0(\tau),\tau\} + [\mathbf{U}_n\{\boldsymbol{\beta}_0(\tau),\tau\} - \mathbf{U}_n^{\pi}\{\boldsymbol{\beta}_0(\tau),\tau\}] \\
=&n^{-1/2}\sum_{i=1}^{n}\boldsymbol{\xi}_{2,i}(\tau) \\
&\quad + n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K} I(\delta_{ik}=1)\mathbf{x}_{ik}[\tau - I\{y_{ik}\le \max(c,\mathbf{x}_{ik}^T\boldsymbol{\beta}_0(\tau))\}]\left\{\frac{1}{\hat{\boldsymbol{\pi}}_{ik}} - \frac{1}{\boldsymbol{\pi}_{ik}}\right\} \\
\approx&n^{-1/2}\sum_{i=1}^{n}\boldsymbol{\xi}_{2,i}(\tau) \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K} I(\delta_{ik}=1)\mathbf{x}_{ik}[\tau - I\{y_{ik}\le \max(c,\mathbf{x}_{ik}^T\boldsymbol{\beta}_0(\tau))\}]\frac{1}{\pi_{ik}^2}\sqrt{n}\sum_{j=1}^{n}\mathbf{h}_{ik}\boldsymbol{\xi}_{1,j} \\
=&n^{-1/2}\sum_{i=1}^{n}\boldsymbol{\xi}_{2,i}(\tau) \\
&\quad + n^{-1/2}\sum_{j=1}^{n}\boldsymbol{\xi}_{1,j}\left[\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}\frac{1}{\pi_{ik}^2} I(\delta_{ik}=1)\mathbf{x}_{ik}[\tau - I\{y_{ik}\le \max(c,\mathbf{x}_{ik}^T\boldsymbol{\beta}_0(\tau))\}]\mathbf{h}_{ik}\right] \\
\approx&n^{-1/2}\sum_{i=1}^{n}\boldsymbol{\xi}_{2,i}(\tau) + n^{-1/2}\sum_{j=1}^{n}\boldsymbol{\xi}_{1,j}\mathbf{w}(\boldsymbol{\beta}_0(\tau),\tau) \\
=&n^{-1/2}\sum_{i=1}^{n}\left(\boldsymbol{\xi}_{2,i}(\tau) + \mathbf{w}(\boldsymbol{\beta}_0(\tau),\tau)\boldsymbol{\xi}_{1,i}\right) \qquad (3.10)
\end{aligned}
$$

where $\approx$ denotes asymptotic equivalence uniformly in $\tau \in (0,1)$.

We claim that $\mathcal{F}^* = \{\boldsymbol{\xi}_{2,i}, \tau \in (0,1)\}$ and $\mathcal{F}^{**} = \{\mathbf{w}(\boldsymbol{\beta}_0(\tau),\tau)\boldsymbol{\xi}_{1,i}, \tau \in (0,1)\}$ are Donsker. First, given $\boldsymbol{\beta}_0(\tau)$ is continuous in τ and $\mathbf{w}(\mathbf{b},\tau)$ is monotone in $\mathbf{b}$ and τ, we can show $\mathbf{w}(\boldsymbol{\beta}_0(\tau),\tau)$ is Donsker. Since Donsker property perserves under product, $\mathcal{F}^{**}$ is a Donsker. Similar arguments as $\mathbf{w}(\mathbf{b},\tau)$, we can show that $\mathcal{F}^*$ is also a Donsker.

Next, simple algebraic manipulations show that $\mathbf{U}_n\{\hat{\boldsymbol{\beta}}(\tau),\tau\} - \mathbf{U}_n\{\boldsymbol{\beta}_0(\tau),\tau\} = (I) +$

(II), where

$$(I) = n^{-1/2} \sum_{i=1}^{n} \sum_{k=1}^{K} \frac{I(\delta_{ik} = 1)}{\pi_{ik}} \mathbf{x}_{ik} \Big[I\{y_{ik} \leq \max(c, \mathbf{x}_{ik}^T \hat{\boldsymbol{\beta}}(\tau))\} \\ - I\{y_{ik} \leq \max(c, \mathbf{x}_{ik}^T \boldsymbol{\beta}_0(\tau))\} \Big]$$

and

$$(II) = n^{-1/2} \sum_{i=1}^{n} \sum_{k=1}^{K} I(\delta_{ik} = 1) \mathbf{x}_{ik} \Big[I\{y_{ik} \leq \max(c, \mathbf{x}_{ik}^T \hat{\boldsymbol{\beta}}(\tau))\} \\ - I\{y_{ik} \leq \max(c, \mathbf{x}_{ik}^T \boldsymbol{\beta}_0(\tau))\} \Big] \left(\frac{1}{\hat{\pi}_{ik}} - \frac{1}{\pi_{ik}} \right).$$

From Lemma 1 and the uniform consistency of $\hat{\boldsymbol{\beta}}(\tau)$, we have

$$(I) \approx n^{1/2} \Big[\boldsymbol{\mu}\{\hat{\boldsymbol{\beta}}(\tau), \tau\} - \boldsymbol{\mu}\{\boldsymbol{\beta}_0(\tau), \tau\} \Big].$$

Since $\sup_{i,k} \left\{ \hat{\pi}_{ik}^{-1} - \pi_{ik}^{-1} \right\} = o(1)$, it is easy to see that $\mathbf{U}_n\{\hat{\boldsymbol{\beta}}(\tau), \tau\} - \mathbf{U}_n\{\boldsymbol{\beta}_0(\tau), \tau\}$ is dominated by (I). Taylor expansion of $\boldsymbol{\mu}(\mathbf{b}, \tau)$ around $\mathbf{b} = \boldsymbol{\beta}_0(\tau)$, along with the fact that $\hat{\boldsymbol{\beta}}(\tau)$ uniformly converges to $\boldsymbol{\beta}_0(\tau)$, gives that

$$\mathbf{U}_n\{\hat{\boldsymbol{\beta}}(\tau), \tau\} - \mathbf{U}_n\{\boldsymbol{\beta}_0(\tau), \tau\} = [\mathbf{A}\{\boldsymbol{\beta}_0(\tau) + \varepsilon_n(\tau)\}] \cdot n^{1/2}\{\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\},$$

where $\sup_\tau \|\varepsilon_n(\tau)\| \to 0$. Given $\mathbf{U}_n\{\hat{\boldsymbol{\beta}}(\tau), \tau\} = 0$, this further implies that

$$n^{1/2}\{\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\} = -\mathbf{A}\{\boldsymbol{\beta}_0(\tau)\}^{-1} \mathbf{U}_n\{\boldsymbol{\beta}_0(\tau)\} + \varepsilon_n^*(\tau),$$

where $\sup_\tau \|\varepsilon_n^*(\tau)\| \to 0$.It follows that

$$n^{1/2}\{\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\} \approx n^{-1/2} \sum_{i=1}^{n} \mathbf{A}\{\boldsymbol{\beta}_0(\tau)\}^{-1} \left(\boldsymbol{\xi}_{2,i}(\tau) + \mathbf{w}(\boldsymbol{\beta}_0(\tau), \tau) \boldsymbol{\xi}_{1,i} \right). \qquad (3.11)$$

Weak convergence of $n^{1/2}\{\hat{\boldsymbol{\beta}}(\tau)-\boldsymbol{\beta}_0(\tau)\}$ follows since $\mathcal{F}^*$ and $\mathcal{F}^{**}$ are Donsker classes and the Donsker property preserves under addition.

Chapter 4

Censored Quantile Regression Analysis of Longitudinal Data with Irregular Outcome-Dependent Follow-Up

4.1 Regression Procedures

4.1.1 Data and Model

Let $Y_i^*(t)$ denote the outcome process of interest and $\mathbf{Z}_i(t)$ denote a vector of external covariate processes for the ith subject. Let $[L_i, R_i]$ be a time interval indicating when the ith subject is under study. We assume that $\{L_i, R_i\}$ is independent of the outcome process $Y_i^*(\cdot)$ given covariates $\mathbf{Z}_i(\cdot)$. The outcome process $Y_i^*(\cdot)$ and $\mathbf{Z}_i(\cdot)$ are only observed at L_i when a subject enters the study and at a sequence of follow-up visit times $\{t_i^{(j)} : j = 1, 2, \ldots, m_i\}$ within $(L_i, R_i]$, where m_i is the total number of follow-up visits. Define a counting process on study entry as $N_i^L(t) = I(L_i \leq t)$ and a counting process for follow-up visits as $N_i(t) = \sum_{j=1}^{m_i} I(t_i^{(j)} \leq t)$. We allow $N_i(t)$ to be dependent on previsous outcome measurements in order to accommodate for outcome-dependent follow-up. Observed outcome is left censored at a fixed constant c. Denote $Y_i(t) = \max(c, Y_i^*(t))$. Hence, the observed data consists of n i.i.d. replicates, denoted by $\{L_i,\ \mathbf{Z}_i(L_i),\ Y_i(L_i),\ t_i^{(j)},\ \mathbf{Z}_i(t_i^{(j)}),\ Y_i(t_i^{(j)}),\ R_i :\ j = 1, 2, \ldots, m_i;\ i = 1, 2, \ldots, n\}$.

Define a conditional τth quantile of a random variable Y given $\mathbf{Z}$ as $Q_Y(\tau|\mathbf{Z}) = \inf\{y : Pr(Y \leq y|\mathbf{Z}) \geq \tau\}$. We assume that the conditional quantile of the longitudinal outcome at time t, $Q_{Y_i^*(t)}(\tau|\mathbf{Z}_i)$, follows the marginal regression model

$$Q_{Y_i^*(t)}(\tau|\mathbf{Z}_i(t)) = \mathbf{X}_i(t)^\top \boldsymbol{\beta}_0(\tau), \qquad \text{for all } t > 0 \tag{4.1}$$

where $\mathbf{X}_i(t) = (1, \mathbf{Z}_i(t)^\top)^\top$ and $\boldsymbol{\beta}_0(\tau)$ is a vector of unknown regression coefficients. External covariates $\mathbf{Z}_i(t)$ can include the observation time. For example, in the motivating PBB data where the primary goal is to characterize the change in outcome over time, we let $Q_{Y_i^*(t)}(\tau) = \beta_0(\tau) + \beta_1(\tau)t$, where $\beta_0(\tau)$ represents the baseline τth quantile of outcome at time 0 and $\beta_1(\tau)$ stands for the change rate of the τth quantile over time. This model could also be used with any monotone transformation of $Y_i(t)$. Under model (4.1), quantiles of the left censored outcome $Y_i(t)$ follows a censored quantile regression model

$$Q_{Y_i(t)}(\tau|\mathbf{Z}_i(t)) = \max(c, \mathbf{X}_i(t)^\top \boldsymbol{\beta}_0(\tau)).$$

To have a good estimate in the quantile regression model, we also need to model the visit process. The initial visit time, L_i, is not necessary to be fixed but is assumed to be conditional outcome-independent that $L_i \perp Y_i^*(\cdot)|\mathbf{Z}_i(\cdot)$. This assumption is reasonable in most cases. In the PBB data, participants had little knowledge about how much they were exposed to PBB until they received results from their initial visits. Since conditional outcome-independent visit time will not bias the estimation, it is not needed to specify the distribution of the initial visit time.

On the other hand, follow-up after the initial visit is outcome-dependent. In the PBB data, subjects with high PBB levels are more likely to have frequent follow-up visits. Define a history function $\mathcal{H}_i(t)$ as all observed data before time t of the ith subject. The follow-up visit process are assumed to follow a proportional intensity model (Andersen and Gill, 1982) that

$$\begin{aligned} \lambda(t|\mathcal{H}_i(t)) &= \lim_{\Delta t \to 0} \frac{1}{\Delta t} P\{N_i(t+\Delta t) - N_i(t) = 1|\mathcal{H}_i(t)\} \\ &= I(L_i < t \leq R_i)\lambda_0(t)\exp\left(\mathbf{h}_i(t)^\top \boldsymbol{\alpha}_0\right), \end{aligned} \tag{4.2}$$

where $\mathbf{h}_i(t)$ is a time-dependent covariates belong to $\mathcal{H}_i(t)$ and $\boldsymbol{\alpha}_0$ is a vector of unknown coefficients. Suppose that $\mathbf{h}_i(t)$ contains the initial outcome $Y_i(L_i)$, then the corresponding coefficient α_0 represents how the follow-up visit process depends on the initial outcome. A positive α_0 means that subjects with large initial outcome tend to have more visits while a negative coefficient means the opposite. And $\alpha_0 = 0$

implies that follow-up is independent of the initial outcome. The choice of $\mathbf{h}_i(t)$ could contains initial outcome measurements or previous outcome measurements. The validity of the inverse intensity-ratio weighting appraoch requires a stronger assumption that $dN_i(t) \perp \{Y_i^*(t), \mathbf{Z}_i(t)\}|\mathcal{H}_i(t)$; in words, $dN_i(t)$ is independent of current outcome and covariates given the history.

4.1.2 Estimation Procedure

Without outcome-dependent follow-up, we can follow Powell (1986)'s method to estimate $\boldsymbol{\beta}_0(\tau)$ through minimizing an objective function

$$n^{-1/2}\sum_{i=1}^{n}\left[\int_0^\infty \rho_\tau\left\{Y_i(t) - \max\left(c, \mathbf{X}_i(t)^\top\boldsymbol{\beta}\right)\right\}\left(dN_i^L(t) + dN_i(t)\right)\right], \tag{4.3}$$

where $\rho_\tau(u) = u \cdot \{\tau - I(u < 0)\}$ is the quantile loss function.

However, as discussed in introduction, if follow-up is outcome-dependent, there would be bias in estimates by minimizing the objective function (4.3). Suppose $\boldsymbol{\alpha}_0$ is known. The inverse intensity-ratio weighting approach weights each observation by the reciprocal of its intensity-ratio function compared to a reference group in model (4.2). Specifically, the objective function (4.3) consists of two parts: one is about the measurement at study entry L_i; the other is about measurements at follow-up visits. Study entry time L_i is conditional independent with outcome, hence we do not need to weight the part about $Y_i(L_i)$; it is equivalent to that its weight $= 1$. Since follow-up visits are outcome-dependent, we weight follow-up visits by the reciprocal of their corresponding intensity ratio defined as $w_i(t; \boldsymbol{\alpha}_0)$, where $w_i(t; \boldsymbol{\alpha}) = \exp\left(\mathbf{h}_i(t)^\top\boldsymbol{\alpha}\right)$. To balance the weights between these two parts, 1 versus $w_i(t_i^{(j)}, \boldsymbol{\alpha}_0)$'s, covariates in model (4.2) are centered at their average. For example, suppose $g_i(t)$ is a covariate belong to $\mathbf{h}_i(t)$. Define a centered covariate $g_i^*(t)$ equal to $g_i(t) - \bar{g}$, where $\bar{g} = \sum_{i=1}^n \sum_{j=1}^{m_i} g_i(t_i^{(j)}) / \sum_{i=1}^n m_i$. By centering covariates, we can ensure that at the "average level", its inverse intensity-ratio weights is 1.

Since $\boldsymbol{\alpha}_0$ is unknown, the estimation procedure consists of two steps. The first step is to estimate $\boldsymbol{\alpha}_0$ in model (4.2) through maximizing a partial likelihood function (Andersen and Gill, 1982). With consistent estimates $\hat{\boldsymbol{\alpha}}$, we can estimate the intensity-ratio weights by plugging $\hat{\boldsymbol{\alpha}}$ into $w_i(t; \boldsymbol{\alpha})$. In the second step, we estimate $\boldsymbol{\beta}_0(\tau)$ by minimizing inverse intensity-ratio weighted objective function $\Psi(\boldsymbol{\beta}; \tau, \hat{\boldsymbol{\alpha}})$, where

$$\begin{aligned}\Psi_\tau(\boldsymbol{\beta}; \boldsymbol{\alpha}) = n^{-1/2} \sum_{i=1}^n \Big[\int_0^\infty & \rho_\tau \left\{ Y_i(t) - \max\left(c, \mathbf{X}_i(t)^\top \boldsymbol{\beta}\right) \right\} \\ & \times \left(dN_i^L(t) + \frac{1}{w_i(t; \boldsymbol{\alpha})} dN_i(t) \right) \Big] \end{aligned} \tag{4.4}$$

The minimizer to the objective function $\Psi(\boldsymbol{\beta}; \tau, \hat{\boldsymbol{\alpha}})$ is also a solution to

$$\mathbf{U}_\tau\{\boldsymbol{\beta}; \hat{\boldsymbol{\alpha}}\} = 0, \tag{4.5}$$

where

$$\begin{aligned}\mathbf{U}_\tau(\boldsymbol{\beta}; \boldsymbol{\alpha}) = n^{-1/2} \sum_{i=1}^n \Big[\int_0^\infty & \mathbf{X}_i(t) I\left(\mathbf{X}_i(t)^\top \boldsymbol{\beta} > c\right) \\ & \times \left\{ I\left(Y_i(t) \le \mathbf{X}_i(t)^\top \boldsymbol{\beta}\right) - \tau \right\} \left(dN_i^L(t) + \frac{1}{w_i(t; \boldsymbol{\alpha})} dN_i(t) \right) \Big]. \end{aligned}$$

This fact will help to prove the asymptotical normality of $\sqrt{n}(\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau))$.

Applying the objective function $\Psi_\tau(\boldsymbol{\beta};\hat{\boldsymbol{\alpha}})$ to a sample, it is equivalent to

$$n^{-1/2}\sum_{i=1}^{n}\Bigg[\rho_\tau\left\{Y_i(L_i)-\max\left(c,\mathbf{X}_i(L_i)^\top\boldsymbol{\beta}\right)\right\} \\ +\sum_{j=1}^{m_i}\frac{1}{w_i\left(t_i^{(j)};\hat{\boldsymbol{\alpha}}\right)}\rho_\tau\left\{Y_i(t_i^{(j)})-\max\left(c,\mathbf{X}_i\left(t_i^{(j)}\right)^\top\boldsymbol{\beta}\right)\right\}\Bigg], \quad (4.6)$$

By treating each observed $Y_i(t)$ as independent and specifying weights as the reciprocal of estimated intensity-ratio, the minimization problem can be implemented by standard statistical software, such as the *crq()* function in R package *quantreg*.

4.1.3 Asymptotic Properties

First, we need to show good asymptotic properties of $\hat{\boldsymbol{\alpha}}$. By imposing stronger conditions as in Andersen and Gill (1982), we can follow their arguments to show that $\hat{\boldsymbol{\alpha}}$ converges to $\boldsymbol{\alpha}_0$ almost surely and $\sqrt{n}(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0)+n^{-1/2}\mathbf{J}(\boldsymbol{\alpha}_0)^{-1}\sum_{i=1}^{n}\boldsymbol{\iota}_i(\boldsymbol{\alpha}_0)\xrightarrow{d}0$, where

$$\mathbf{J}(\boldsymbol{\alpha})=-E\Bigg(\frac{1}{n}\sum_{i=1}^{n}\int_0^\infty\Bigg[\frac{\sum_{j=1}^{n}I(L_j<t\le R_j)\mathbf{h}_j(t)^{\otimes 2}e^{\mathbf{h}_j(t)^\top\boldsymbol{\alpha}}}{\sum_{j=1}^{n}I(L_j<t\le R_j)e^{\mathbf{h}_j(t)^\top\boldsymbol{\alpha}}} \\ -\left\{\frac{\sum_{j=1}^{n}I(L_j<t\le R_j)\mathbf{h}_j(t)e^{\mathbf{h}_j(t)^\top\boldsymbol{\alpha}}}{\sum_{j=1}^{n}I(L_j<t\le R_j)e^{\mathbf{h}_j(t)^\top\boldsymbol{\alpha}}}\right\}^{\otimes 2}\Bigg]dN_i(t)\Bigg)$$

and

$$\boldsymbol{\iota}_i(\boldsymbol{\alpha})=\int_0^\infty\left\{\mathbf{h}_i(t)-\frac{\sum_{j=1}^{n}I(L_j<t\le R_j)\mathbf{h}_j(t)e^{\mathbf{h}_j(t)^\top\boldsymbol{\alpha}}}{\sum_{j=1}^{n}I(L_j<t\le R_j)e^{\mathbf{h}_j(t)^\top\boldsymbol{\alpha}}}\right\} \\ \times\left\{dN_i(t)-I(L_i<t\le R_i)\lambda_0(t)e^{\mathbf{h}_i(t)^\top\boldsymbol{\alpha}}dt\right\}.$$

Since a good estimation of $\boldsymbol{\beta}_0(\tau)$ relies on a good estimation of $\boldsymbol{\alpha}_0$, in the following discussion of $\hat{\boldsymbol{\beta}}(\tau)$ it is always assumed to be true that $\hat{\boldsymbol{\alpha}}\xrightarrow{a.s.}\boldsymbol{\alpha}_0$ and

$\sqrt{n}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) + n^{-1/2}\mathbf{J}(\boldsymbol{\alpha}_0)^{-1}\sum_{i=1}^{n}\boldsymbol{\iota}_i(\boldsymbol{\alpha}_0) \xrightarrow{d} 0.$

Define $\zeta_i^{\tau}(\boldsymbol{\beta};\boldsymbol{\alpha}) = \int_0^{\infty}\rho_{\tau}[Y_i(t) - \max\left\{c, \mathbf{X}_i(t)^{\top}\boldsymbol{\beta}\right\}][dN_i^L(t) + \exp\{-\mathbf{h}_i(t)^{\top}\boldsymbol{\alpha}\}dN_i(t)]$ and $\psi_{\tau}(\boldsymbol{\beta};\boldsymbol{\alpha}) = E\{n^{-1/2}\Psi_{\tau}(\boldsymbol{\beta};\boldsymbol{\alpha})\}$. Define $\mathbf{l}_i^{\tau}(\boldsymbol{\beta};\boldsymbol{\alpha}) = \int_0^{\infty}\mathbf{X}_i(t)I\{\mathbf{X}_i(t)^{\top}\boldsymbol{\beta} > c\}[I\{Y_i(t) < \mathbf{X}_i(t)^{\top}\boldsymbol{\beta}\} - \tau][dN_i^L(t) + \exp\{-\mathbf{h}_i(t)^{\top}\boldsymbol{\alpha}\}dN_i(t)]$ and $\boldsymbol{\mu}_{\tau}(\boldsymbol{\beta};\boldsymbol{\alpha}) = E\left\{n^{-1/2}\mathbf{U}_{\tau}(\boldsymbol{\beta};\boldsymbol{\alpha})\right\}$. To guarantee good asymptotic properties of $\hat{\boldsymbol{\beta}}(\tau)$, we further need some regularity conditions:

C1. (a) There exists $\gamma > 0$ such that $E[\int_0^{\infty} I\{\mathbf{X}(t)^{\top}\boldsymbol{\beta}(\gamma) > c\}\mathbf{X}(t)^{\otimes 2}\{dN^L(t) + I(L < t \leq R)\lambda_0(t)dt\}]$ is positive definite;

(b) The conditional density function of $Y(t)$ given $\mathbf{Z}(t)$, $f_{Y(t)}\{y|\mathbf{Z}(t)\}$, is continuous and greater than zero at $y = \mathbf{X}(t)^{\top}\boldsymbol{\beta}_0(\tau)$ for any $\tau \in [\gamma, 1)$.

C2. $\boldsymbol{\beta}_0(\tau)$ is in the interior of a compact space $\mathcal{B}$ for all $\tau \in [\gamma, 1)$.

C3. There exists a small circle of $\boldsymbol{\alpha}$ centered at $\boldsymbol{\alpha}_0$, denoted by $\mathcal{A}$, such that $\frac{\partial\psi_{\tau}(\boldsymbol{\beta};\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}} = E\Big(\int_0^{\infty}\rho_{\tau}\left[Y(t) - \max\left\{c, \mathbf{X}(t)^{\top}\boldsymbol{\beta}\right\}\right]\mathbf{h}(t)\exp\{-\mathbf{h}(t)^{\top}\boldsymbol{\alpha}\}dN(t)\Big)$ is bounded uniformly in $\boldsymbol{\beta} \in \mathcal{B}$, $\boldsymbol{\alpha} \in \mathcal{A}$, and $\tau \in [\gamma, 1)$.

C4. $\zeta_i^{\tau}(\boldsymbol{\beta};\boldsymbol{\alpha})$ has finite first and second moments for any $\boldsymbol{\beta} \in \mathcal{B}$, $\boldsymbol{\alpha} \in \mathcal{A}$, and $\tau \in [\gamma, 1)$;

C5. (a) The covariate space $\mathcal{Z}$ is compact, that is, $\sup_{i,t}\|\mathbf{Z}_i(t)\| < \infty$, where $\|\cdot\|$ stands for Euclidean norm;

(b) $\int_0^{\infty}\exp\left\{-\mathbf{h}_i(t)^{\top}\boldsymbol{\alpha}\right\}dN_i(t)$ is uniformly bounded for any $\boldsymbol{\alpha} \in \mathcal{A}$;

(c) $f_{Y(t)}\left\{\mathbf{X}(t)^{\top}\boldsymbol{\beta}_0(\tau)|\mathbf{Z}(t)\right\}$ is uniformly bounded for any $\mathbf{Z}(t) \in \mathcal{Z}$ and $\tau \in [\gamma, 1)$;

(d) For any $d \geq 0$, there exists a positive constant M^+ such that

$$\sup_{\tau \in [\gamma,1)} E \bigg| \int_0^\infty I\left\{|\mathbf{X}(t)^\top \boldsymbol{\beta}_0(\tau) - c| \leq \|\mathbf{X}(t)\| d\right\} \\ \times \left\{dN^L(t) + I(L < t \leq R)\lambda_0(t)dt\right\} \bigg| \leq M^+ \cdot d;$$

(e) $E\left[\int_0^\infty \mathbf{h}(t) \exp\left\{-\mathbf{h}(t)^\top \boldsymbol{\alpha}\right\} dN(t)\right]$ is uniformly bounded for $\boldsymbol{\alpha} \in \mathcal{A}$.

C6. Define

$$\mathbf{B}_\tau(\boldsymbol{\beta}; \boldsymbol{\alpha}_0) = \frac{\partial \boldsymbol{\mu}_\tau(\boldsymbol{\beta}; \boldsymbol{\alpha}_0)}{\partial \boldsymbol{\beta}} = E\bigg[\int_0^\infty \mathbf{X}(t)^{\otimes 2} I\left\{\mathbf{X}(t)^\top \boldsymbol{\beta} > c\right\} \\ \times f_{Y(t)}\left\{\mathbf{X}(t)^\top \boldsymbol{\beta} | \mathbf{X}(t)\right\} \left\{dN^L(t) + I(L < t \leq R)\lambda_0(t)dt\right\}\bigg].$$

$\inf_{\tau \in [\gamma,1)} eigmin \mathbf{B}_\tau(\boldsymbol{\beta}_0(\tau); \boldsymbol{\alpha}_0) > 0$, where $eigmin(\cdot)$ denotes minimum eigenvalue of a matrix.

We establish the asymptotic properties of $\hat{\boldsymbol{\beta}}(\tau)$ stated in the following theorems.

Theorem 4.1.1. *Under conditions C1-C4,* $\sup_{\tau \in [\gamma,1)} \|\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\| \to 0, \quad a.s..$

Theorem 4.1.2. *Under conditions C1-C6,* $\left\{n^{1/2}\left[\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\right] : \tau \in [\gamma, 1)\right\}$ *converges weakly to a Gaussian process with mean 0 and covariance matrix* Σ*, where* Σ *is presented in* (4.15) *in Appendix.*

Condition C1 is very standard for identifiability of $\boldsymbol{\beta}_0(\tau)$ in censored quantile regression. The boundedness of parameter space, the derivative function, and the moments of the objective function imposed by conditions C2-C4 and the boundedness requirement of covariates, the inverse intensity weighted number of measurements and the density function associated with the outcome imposed in condition C5 are easy to met in practice. The condition C5 (d) is a condition to rule out the situation that there exists a subpopulation with positive probability satisfying that the τth quantile of their outcome is at c. Condition C6 is to ensure the finite asymptotical variance of $\sqrt{n}\left\{\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\right\}$.

4.1.4 Inference

To make inference on $\boldsymbol{\beta}_0(\tau)$, bootstrapping procedures can be used. Each time, randomly select n subjects with replacement. Based on every new bootstrapped sample, repeat the estimation procedure and obtain a new estimator, denoted by $\hat{\boldsymbol{\beta}}(\tau)^*$. The asymptotic distribution of $\sqrt{n}(\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau))$ can be approximated by the distribution of $\sqrt{n}(\hat{\boldsymbol{\beta}}^*(\tau) - \hat{\boldsymbol{\beta}}(\tau))$.

We also developed a sample-based inference procedure. The reason we can not directly estimate the asymptotic variance matrix is that the unknown density function $f_{Y_i(t)}\left(\mathbf{X}_i^\top \boldsymbol{\beta}_0(\tau)|\mathbf{X}_i\right)$ is involved in matrix $\mathbf{B}_\tau(\boldsymbol{\beta}_0(\tau);\boldsymbol{\alpha}_0)$. Huang (2002) and Peng and Fine (2009) proposed a novel way to estimate a derivative matrix, such as matrix $\mathbf{B}_\tau(\boldsymbol{\beta}_0(\tau);\boldsymbol{\alpha}_0)$, by adding a "little-o" perturbation $\mathbf{E}$ to the primitive function which is $\mathbf{U}_\tau(\boldsymbol{\beta}_0(\tau);\boldsymbol{\alpha}_0)$ here. The idea is that if we could find two $(p+1)\times(p+1)$ matrices $\boldsymbol{\beta}^* \equiv (\boldsymbol{\beta}_1^*, \ldots, \boldsymbol{\beta}_(^*p+1))$ and $\mathbf{e} \equiv (\mathbf{e}_1, \ldots, \mathbf{e}_{(p+1)})$ such that $\mathbf{e}_j$ is relatively goes to zero compared to $\mathbf{U}_\tau(\boldsymbol{\beta}_0(\tau);\boldsymbol{\alpha}_0)$ as n increases and $\mathbf{e}_j = \mathbf{U}_\tau(\boldsymbol{\beta}_j^*;\boldsymbol{\alpha}_0) - \mathbf{U}_\tau(\boldsymbol{\beta}_0(\tau);\boldsymbol{\alpha}_0) \approx \sqrt{n}\mathbf{B}_\tau(\boldsymbol{\beta}_0(\tau);\boldsymbol{\alpha}_0)(\boldsymbol{\beta}_j^* - \boldsymbol{\beta}_0(\tau))$ for $j = 1, \ldots, (p+1)$, we can estimate matrix $\mathbf{B}_\tau(\boldsymbol{\beta}_0(\tau);\boldsymbol{\alpha}_0)$ by $n^{-1/2}\mathbf{E}\mathbf{D}^{-1}$, where $\mathbf{D} = (\boldsymbol{\beta}_1^* - \boldsymbol{\beta}_0(\tau), \ldots, \boldsymbol{\beta}_{p+1}^* - \boldsymbol{\beta}_0(\tau))$. The specific procedure follows:

1. Define $\mathbf{l}_j^\tau(\boldsymbol{\beta};\boldsymbol{\alpha}) = \int_0^\infty \mathbf{X}_j(t) I\left(\mathbf{X}_j(t)^\top \boldsymbol{\beta} > 0\right)\left\{I\left(Y_j(t) \leq \mathbf{X}_j(t)^\top \boldsymbol{\beta}\right) - \tau\right\}$ $\times\left(dN_j^L(t) + \frac{1}{w_j(t;\boldsymbol{\alpha})} dN_j(t)\right)$, and $\boldsymbol{\Omega}(\tau) = n^{-1}\sum_{j=1}^n \left\{\mathbf{l}_j^\tau(\hat{\boldsymbol{\beta}}(\tau);\hat{\boldsymbol{\alpha}})\right\}^{\otimes 2}$, where $\mathbf{v}^{\otimes 2} = \mathbf{v}\mathbf{v}^\top$. Find a symmetric and nonsingular $(p+1)\times(p+1)$ matrix $\mathbf{E}$ such that $\boldsymbol{\Omega}(\tau) = \mathbf{E}^2$.

2. Solve the equation

$$\mathbf{U}_\tau(\mathbf{b}; \hat{\boldsymbol{\alpha}}, \hat{\gamma}) = \mathbf{U}_\tau(\hat{\boldsymbol{\beta}}(\tau); \hat{\boldsymbol{\alpha}}, \hat{\gamma}) + \mathbf{e}_j \tag{4.7}$$

   for $\mathbf{b}$, and denote the solution by $\boldsymbol{\beta}_j^*$ $(j = 1, \ldots, p+1)$.

3. Calculate $\mathbf{D} = \Big(\mathbf{b}_1 - \hat{\boldsymbol{\beta}}(\tau), \ldots, \mathbf{b}_{p+1} - \hat{\boldsymbol{\beta}}(\tau)\Big)$.

4. Compute $n^{-1/2}\mathbf{E}\mathbf{D}^{-1}$, which provide consistent estimate for $\mathbf{B}_\tau(\boldsymbol{\beta}_0(\tau); \boldsymbol{\alpha}_0)$.

It is not very straightforward to solve the estimating equation in step 2. In order to take advantage of existing software package, we convert the estimating equation as a new solution-finding problem:

$$\sum_{i=1}^{n} \int_0^\infty \mathbf{X}_i(t) I(\mathbf{X}_i(t)^\top \mathbf{b} > 0) \left\{ I(Y_i(t) \leq \mathbf{X}_i(t)^\top \mathbf{b}) - \tau \right\} \left(dN_i^L(t) + \frac{1}{w_i(t; \boldsymbol{\alpha})} dN_i(t) \right)$$
$$+ I({X^*}^\top \mathbf{b} > 0) X^* \left\{ I(0 \leq {X^*}^\top \mathbf{b}) - \tau \right\} = 0, \tag{4.8}$$

plus a condition that $X^{*\top}\mathbf{b} > 0$, where $X^* = -n^{1/2}\mathbf{e}_j(\tau)/(1-\tau)$. The former equation (4.8), ignoring the condition $X^{*\top}\mathbf{b} > 0$, can be solved using *crq()* function. Denote the estimate by $\mathbf{b}_j$. If it satisfies that $X^{*\top}\mathbf{b}_j > 0$, then $\boldsymbol{\beta}_j^* = \mathbf{b}_j$. Else if $X^{*\top}\mathbf{b}_j < 0$, we replace $\mathbf{e}_j(\tau)$ by $-\mathbf{e}_j(\tau)$. Since the change in the direction of $\mathbf{e}_j$ does not change its desired asymptotic order, our procedure is still valid. Then we have a new ${X^*}' = -X^*$ and repeat step 2 to obtain a new vector $\mathbf{b}_j'$. Since $\mathbf{b}_j$ and $\mathbf{b}_j'$ are both close to $\hat{\boldsymbol{\beta}}(\tau)$, $X'^{*\top}\mathbf{b}_j'$ which is close to $-X^{*\top}\mathbf{b}_j$ should be greater than zero. Then define $\boldsymbol{\beta}_j^* = \mathbf{b}_j'$. Note that if the direction of $\mathbf{e}_j(\tau)$ is changed, the direction of $\mathbf{e}_j$ in matrix $\mathbf{E}$ needs to be changed as well.

Denote the estimators of $\mathbf{B}_\tau(\boldsymbol{\beta}_0(\tau); \boldsymbol{\alpha}_0)$ by $\hat{\mathbf{B}}(\tau)$. A consistent sample-based covariance estimator may be given by

$$n^{-1}\sum_{i=1}^{n}\left[-\hat{\mathbf{B}}(\tau)^{-1}\left\{\mathbf{l}_i^{\tau}(\hat{\boldsymbol{\beta}}(\tau);\hat{\boldsymbol{\alpha}})-\hat{\mathbf{A}}(\tau)\hat{\mathbf{J}}^{-1}\boldsymbol{\iota}_i(\hat{\boldsymbol{\alpha}})\right\}\right]$$
$$\times\left[-\hat{\mathbf{B}}(\tau)^{-1}\left\{\mathbf{l}_i^{\tau}(\hat{\boldsymbol{\beta}}(\tau);\hat{\boldsymbol{\alpha}})-\hat{\mathbf{A}}(\tau)\hat{\mathbf{J}}^{-1}\boldsymbol{\iota}_i(\hat{\boldsymbol{\alpha}})\right\}\right]^{\top}.$$

4.2 Simulation Studies

Simulation studies were conducted to assess finite-sample performance of the proposed method. We also compare the proposed estimator with censored quantile regression for longitudinal data without adjustment for outcome-dependent follow-up (Wang and Fygenson, 2009). Two covariates are considered: $Z_{i1} \sim Uniform(0,1)$ and $Z_{i2} \sim Bernoulli(0.5)$. Two distributions are adopted: Normal distribution and Gamma distribution. Specifically, we have the following two scenarios:
Case 1: $Y_i(t) = \max(0,\ \ 4.5 + a_i - Z_{i1} + Z_{i2} - t + \varepsilon_i(t))$,
where $a_i \sim N(0,\ \frac{1}{4}\left\{(Z_{i1}+Z_{i2}+1)^2 - \frac{1}{2}\right\})$ and $\varepsilon_i(t) \sim N(0,\ \frac{1}{8})$ which are mutually independent. It follows a quantile regression model that

$$Q_{Y_i(t)}(\tau|Z_{i1},Z_{i2}) = \max\left[0, 4.5+\Phi^{-1}(\tau)+\left\{-1+\Phi^{-1}(\tau)\right\}Z_{i1}+\{1+\Phi^{-1}(\tau)\}Z_{i,2}-t\right]$$

Case 2: $Y_i(t) = \max(0,\ \ 3.5 + a_i - 2Z_{i1} - t + \varepsilon_i(t))$,
where $a_i \sim Gamma(3,\ \frac{1}{4}(Z_{i1}+Z_{i2}+1))$ and $\varepsilon_i(t) \sim Gamma(1,\ \frac{1}{4}(Z_{i1}+Z_{i2}+1))$ which are mutually independent. It follows a quantile regression model that

$$Q_{Y_i(t)}(\tau|Z_{i1},Z_{i2}) = \max\Big(0,\ \ 3.5 + F^{-1}_{Gamma(4,1)}(\tau)$$
$$+\left(-2+F^{-1}_{Gamma(4,1)}(\tau)\right)Z_{i1} + F^{-1}_{Gamma(4,1)}(\tau)Z_{i2} - t\Big).$$

The study entry time L_i were generated from $Uniform(0,1)$; the end of follow-up

R_i were generated from $Uniform(4, 5)$. Follow-up visits were generated according to a proportional intensity model:

$$P\{dN_i(t) = 1|H_i(t)\} = I(L_i \leq t \leq R_i)0.2t \exp\left(0.2Y(t^-)\right) dt,$$

where $Y(t^-)$ represents the last observed outcome before time t. A positive coefficient corresponding to $Y(t^-)$ indicates that subjects with large previous outcome have higher intensity of follow-up visits. Under these setups, the average number of visits are 4.4 and the average left censoring rate is 10% in both case 1 and case 2.

For each scenario, we generated 1000 data sets of sample size $n = 200$. Two methods were applied to estimate covariate effects on three outcome quantiles (25th, 50th, and 75th): the censored quantile regression for longitudinal data without adjustment for irregular outcome-dependnet follow-up (Naive); the proposed inverse intensity-ratio weighted approach (Proposed). Bias and empirical standard deviation of estimators from two methods are present in table 1. The proposed estimator has much smaller bias compared to the naive estimator. Considering the empirical standard deviation, the proposed estimator is virtually unbiased while the naive estimator is biased in the presence of outcome-dependent follow-up.

Both bootstrapping and sample-based inference were applied for the proposed estimator. The bootstrapping size is 500. The Emp SD in table 1 stands for the empirical standard deviation of 1000 $\hat{\boldsymbol{\beta}}$'s and the Avg SD is the average of estimated standard deviation from bootstrapping or sample-based inferences. Cov95 in table 1 stands for the coverage rate of 95% confidence intervals covering the true value. Generally, both standard deviation estimates from two inference procedures are acceptably close to the empirical standard deviation. In particular, the SD estimates from bootstrapping method is closer to the empirical SD compared to

Table 4.1: Comparison of the censored quantile regression estimator for longitudinal data ignoring outcome-dependent follow-up (Naive) and the proposed inverse intensity-ratio weighted estimator (Proposed): Emp SD – empirical standard deviation; Avg SD – the average of standard deviation estimates; Cov95 – the coverage rate of a 95% confidence interval.

		Naive		Proposed					
						Bootstrapping		Sample-based	
Effect	True	Bias	EmpSD	Bias	EmpSD	AvgSD	Cov95	AvgSD	Cov95
				Case 1					
				$\tau = 0.25$					
Intercept	4.163	-0.077	0.157	-0.011	0.162	0.172	0.97	0.194	0.95
Z_1	-1.337	0.149	0.284	0.028	0.307	0.323	0.96	0.348	0.94
Z_2	0.663	0.131	0.175	0.011	0.190	0.189	0.95	0.201	0.94
t	-1	0.033	0.032	0.0007	0.035	0.041	0.97	0.054	0.96
				$\tau = 0.5$					
Intercept	4.5	-0.066	0.145	-0.007	0.144	0.153	0.95	0.162	0.95
Z_1	-1	0.134	0.269	0.015	0.269	0.287	0.96	0.301	0.95
Z_2	1	0.130	0.169	0.002	0.171	0.173	0.95	0.174	0.93
t	-1	0.028	0.027	-0.0006	0.028	0.031	0.97	0.036	0.97
				$\tau = 0.75$					
Intercept	4.837	-0.061	0.155	0.002	0.149	0.159	0.96	0.165	0.95
Z_1	-0.663	0.117	0.300	-0.010	0.278	0.298	0.96	0.300	0.94
Z_2	1.337	0.142	0.191	0.0008	0.178	0.185	0.96	0.184	0.94
t	-1	0.025	0.027	-0.002	0.027	0.030	0.97	0.033	0.96
				Case 2					
				$\tau = 0.25$					
Intercept	4.134	-0.029	0.121	0.006	0.121	0.125	0.94	0.135	0.93
Z_1	-1.366	0.062	0.218	-0.003	0.211	0.225	0.96	0.235	0.95
Z_2	0.634	0.065	0.127	-0.001	0.129	0.132	0.95	0.135	0.94
t	-1	0.022	0.026	0.001	0.027	0.031	0.97	0.038	0.97
				$\tau = 0.5$					
Intercept	4.418	-0.043	0.142	0.006	0.136	0.146	0.96	0.155	0.94
Z_1	-1.082	0.097	0.261	-0.005	0.244	0.263	0.96	0.275	0.95
Z_2	0.918	0.093	0.166	-0.007	0.154	0.155	0.94	0.157	0.93
t	-1	0.027	0.028	0.001	0.028	0.031	0.97	0.035	0.96
				$\tau = 0.75$					
Intercept	4.777	-0.060	0.195	0.002	0.178	0.190	0.96	0.199	0.95
Z_1	-0.723	0.134	0.383	0.001	0.327	0.339	0.94	0.352	0.93
Z_2	1.277	0.151	0.241	-0.008	0.209	0.210	0.94	0.205	0.92
t	-1	0.033	0.035	0.0004	0.034	0.037	0.97	0.040	0.96

the sample-based inference procedure. Both procedures have acceptable coverage rates that are close to the nominal value. There is no trend of being over-covered or under-covered.

We take a close examination of the boostrapping SD estimates and sample-based SD estimates. Take the intercept in case 1 for example. Figure 3. contains histogram plots of 1000 SD estimates from two inference procedures respectively. The solid upright line stands for the empirical SD = 0.121, which could be viewed as a true value. The bootstrapping SD estimates are distributed closer to the true value. We conclude that the bootstrapping SD estimates is more accurate than the sample-based SD estimates. We also notice an interesting trend that the difference between the average of SD estimates from two inference procedure gets smaller when τ increases, especially in case 1. Since the effective sample size that $I(\mathbf{X}_i^\top \boldsymbol{\beta}_0(\tau) > c)$ increases as τ increases, we guess that as the sample size increases, the performance of the sample-based inference procedure improves. Besides accuracy, we also compared the computation time. The computation of the sample-based approach is about 50 times as fast as the bootstrapping procedure. Summarily, bootstrapping performs better than the sample-based method while the latter one saves much computation time.

In our simulation, we encounter a "feasible" problem with estimation in lower quantiles. It is straightforward in censored quantile regression that if left censoring rate in a sample is larger than, e.g., 25%, it is very likely that estimation of the 25%th

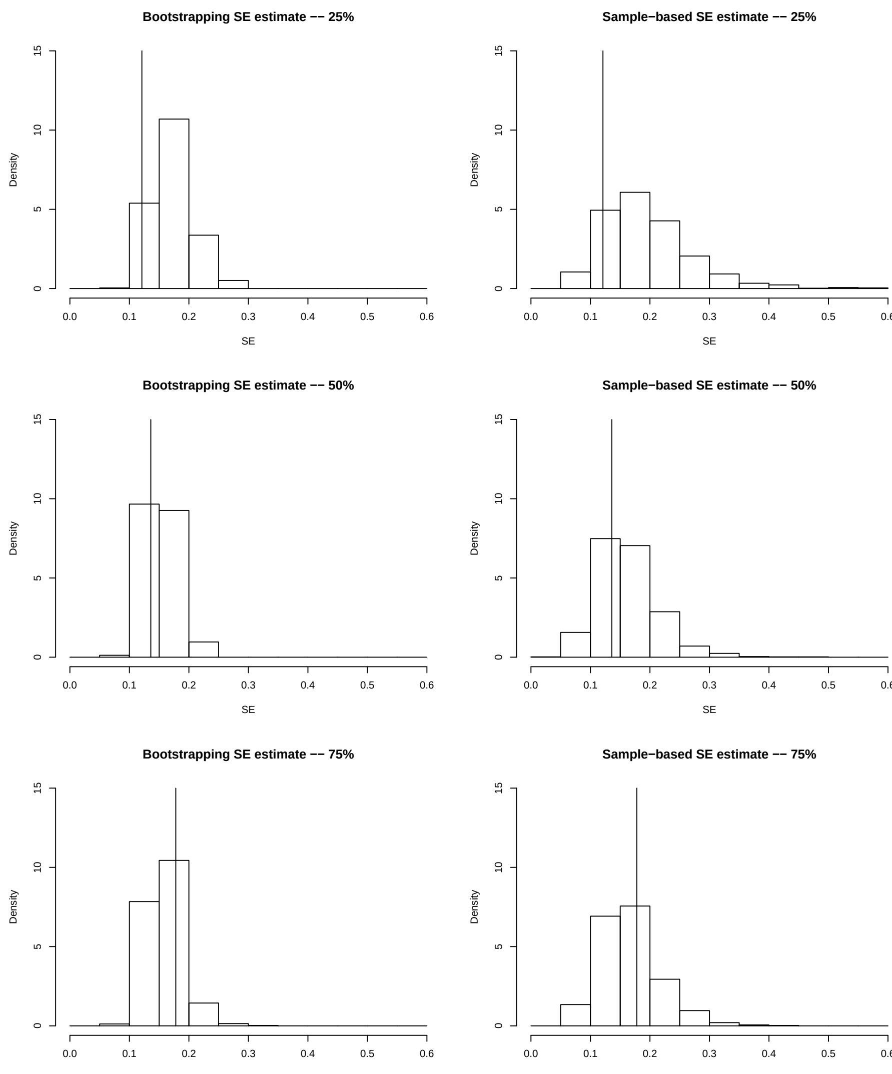

Figure 4.1: Comparision between the SE estimates from bootstrapping and sample-based appraoch; Intercept in case 1 for 25%th quantile, 50%th quantile and 75%th quantile

quantile is not feasible. The reason can be expressed in math that for any $\boldsymbol{\beta}$,

$$
\begin{aligned}
&\sum[I(Y \leq \max(c, \mathbf{X}^T\boldsymbol{\beta})) - \tau] \\
&\geq \sum I(Y = c) \cdot (1 - \tau) - \sum I(Y > c) \cdot \tau \\
&> n \cdot \tau \cdot (1 - \tau) - n \cdot (1 - \tau) \cdot \tau = 0.
\end{aligned}
$$

However, for weighted censored quantile regression estimating equation, the story is a little bit different that the requirement for a feasible estimate also depends on the weight. Since $\sum w[I(Y \leq \max(c, \mathbf{X}^T\boldsymbol{\beta})) - \tau] \geq \sum wI(Y = c)[1 - \tau] - \sum wI(Y > c)\tau$, a necessary condition is that $\sum wI(Y = c)[1 - \tau] - \sum wI(Y > c)\tau \leq 0$. In our simulation setups and the PBB data example, subjects with lower outcome have fewer visits thus have larger weight. In other words, weights of $I(Y = c)$ are larger than weights of $I(Y > c)$; therefore, even that the left censoring rate is less than τ, it is still possible that the estimation of $\boldsymbol{\beta}_0(\tau)$ is not feasible.

In summary, simulation studies demonstrate that ignoring outcome-dependent follow-up would result in biased estimation in censored quantile regression while our proposed weighted estimator is unbiased. Two inference procedures are valid. SD estimate from bootstrapping approach is more accurate while the sample-based inference procedure saves computation time.

4.3 PBB Data Example

Polybrominated biphenyls (PBB's) are manufactured chemicals added as flame retardants to electrical devices, plastics, and various textiles. A widespread contamination with PBB's occurred in Michigan during 1973 - 1974 when Fire master FF-1, instead of NutriMaster, was accidentally mixed with animal feed. Residents on

Michigan farms and neighboring communities were exposed to PBB's by consuming contaminated animal food products. The Michigan Department of Public (now Community) Health (MDCH), in collaboration with the US Health Service, established a registry of individuals exposed to the contaminated food products. Since the initial enrollment period (1976 - 1978), the MDCH has periodically contacted cohort members to obtain additional serum samples. Serum samples from cohort members were analyzed for PBB during 1976C1993.

Our analysis is focused on understanding the decay rate over time of serum PBB concentration in females. PBB's are stable, persistent halogenated organic pollutants with extremely long half-lives. Participants continued to have measurable PBB levels in serum after more than 20 years. PBB exposure is of special concern to the fetus and neonate because it can cross the placenta and is concentrated in breast milk. Therefore, we are interested in the elimination rate of PBB in serum among female subjects.

We include females who were born before the contamination incident (July, 1973) if they had at least two serum PBB measurements, an initial serum PBB measurement of ≥ 2 parts per billion (p.p.b.) and after age 16 and if the time between any two consecutive measurements was at least 6 months apart. We required an initial serum PBB measurement of at least 2 ppb to ensure that their levels were above the limit of detection of 1 pbb. We exclude females who were younger than age 16 at initial measurement because childhood growth could potentially affect the compartment mobility and thus the equilibrium of serum PBB concentration levels. We also excluded measurements taken during pregnancy or during any period of breast-feeding because of the potential mobilization of PBB into the bloodstream during these times.

Our analysis includes 386 women. The longitudinal data set was arranged with one observation per serum PBB measurement, yielding 2 - 7 measurements per woman. Initial PBB concentration level ranges from 1 to 559.80 p.p.b. (mean=11.44, median=2.40). Outcome Y is defined as $log(PBB)$. Time origin is set as the exposure time to PBB which is defined at July 1, 1973 (t=0). Study enter time L is defined to be the first visit time and R is the end of study, December 31, 1993 (t=20.5).

When we model follow-up visits, the entire study were divided into three periods (1976-1981, 1982-1989, and 1990-1993) according to the study design. During 1982-1989, a substudy focused on high PBB levels was established since these participants were more severely exposed and it may be easier to see a decrease in PBB. After 1990, all participants were contacted again for serum samples. Converting three periods as time since the time origin, time intervals are $(2.5, 8.5]$, $(8.5, 16.5]$, and $(16.5, 20.5]$. We also assume that follow-up are dependent on the initial measurement $log(PBB)$. Hence, the proportional intensity model of follow-up visits is specified as

$$
\begin{aligned}
&P(dN_i(t)|H_i(t)) = I(L_i < t \leq R_i)\lambda_0(t) \\
&\times \exp\left(\alpha_1 I(t \leq 8.5) \cdot Y_i(L_i) + \alpha_2 \cdot I(8.5 < t \leq 16.5) Y_i(L_i) + \alpha_3 \cdot I(t > 16.5) Y_i(L_i)\right),
\end{aligned}
$$

where α_1, α_2 and α_3 represent the effects of the initial outcome on the follow-up visit process in three time intervals respectively. Table 4.2 presents the estimates of coefficients in the recurrent event model. All coefficient estimates are positive, which means that subjects with high initial log(PBB) levels tend to have more follow-up visits. The estimate of α_2 is much larger than α_1 and α_3 and is significantly larger than zero (p value < 0.01). The outcome-dependent follow-up pattern is especially strong during 1982-1989. An increase of 1 in initial $log(PBB)$ would result in a 1.79 time increase in intensity of follow-up visits during year 1982-1990.

Table 4.2: Parameter estimates of proportional intensity model

Coeff	Estimate	exp(Estimate)	p-value
α_1	0.027	1.03	0.61
α_2	0.584	1.79	< 0.01
α_3	0.039	1.04	0.52

We model the outcome log(PBB) by marginal quantile regression models that

$$Q_{Y_i(t)}(\tau) = \beta_0(\tau) + \beta_1(\tau) \times t, \qquad 0 < \tau < 1, \quad t > 0, \tag{4.9}$$

where $\beta_0(\tau)$ represents the τth quantile of $log(PBB)$ level at time origin and $\beta_1(\tau)$ represents the decay rate of the population τth quantile of $log(PBB)$ level over time. Both censored quantile regression for longitudinal data ignoring outcome-dependent follow-up (Naive) and the inverse intensity-ratio weighted estimator (Proposed) are applied. Confidence interval is calculated through bootstrapping. Bootstrapping size is 500. Table 4.3 presents point and interval estimates for six quantiles (25th, 50th, 75th, 85th, 90th, and 95th). All $\beta_1(\tau)$ estimates from Naive method are positive, which means that the distribution of $log(PBB)$ shifts up as time goes. It contradicts with the biologic fact that human bodies can not produce chemical PBB's.

On the other hand, decay rate $\beta_1(\tau)$ estimates from our proposed approach are all negative except for the $\beta_1(\tau)$ estimate of the 25th quantile which is really close to zero. Negative coefficient estimates for time demonstrates that PBB concentration generally decreases over time. Due to that fewer low outcomes are observed in follow-up visits, the empirical quantile is higher than the population quantile in follow-up time. By weighting more in the observed low outcomes, the weighted sample quantile is pulled towards the actual population quantile. Furthermore, there is an increasing pattern in the magnitude of $\beta_1(\tau)$ estimates as τ increases. It is consistent to our conjecture that upper quantiles decay faster than lower quantiles. This varying covariate effect pattern is not capturable in fixed effect models. Decay rates of lower percentiles are not significant while the decay rate of the 95th quantile is significantly less than zero. For example, the estimated decay rate for the 90th quantile is 2.6% $(= 1 - \exp(-0.052))$ per year with 95% confidence interval between -0.5% and 5.4%. The estimated decay rate for the 95th quantile is 5.1% $(= 1 - \exp(-0.052))$ per year with 95% confidence interval between 0.7% and 9.2%.

Table 4.3: Parameter estimates and 95% confidence interval for 25th quantile, 50th quantile, 75th quantile, 85th quantile, 90th quantile, and 95th quantile

		Naive		Proposed	
Quantile	Effect	Estimate	95% CI	Estimate	95% CI
25th quantile	Intercept	0.150	(0.092, 0.208)	0.182	(0.155, 0.210)
	Time	0.009	(2E-4, 0.018)	6E-17	(-0.008, 0.008)
50th quantile	Intercept	0.852	(0.707, 0.997)	0.904	(0.609, 1.199)
	Time	0.006	(-0.007, 0.018)	-0.009	(-0.024, 0.007)
75th quantile	Intercept	1.496	(1.246, 1.745)	1.435	(1.171, 1.699)
	Time	0.028	(0.001, 0.054)	-4E-17	(-0.012, 0.012)
85th quantile	Intercept	2.298	(1.763, 2.833)	2.057	(1.635, 2.479)
	Time	0.019	(-0.027, 0.064)	-8E-4	(-0.024, 0.022)
90th quantile	Intercept	2.829	(2.182, 3.475)	2.956	(2.357, 3.555)
	Time	0.036	(-0.017, 0.090)	-0.026	(-0.056, 0.005)
95th quantile	Intercept	3.813	(3.112, 4.514)	4.047	(3.379, 4.716)
	Time	0.046	(-0.003, 0.096)	-0.052	(-0.097, -0.007)

In summary, the PBB concentration distribution shifts down slowly over time. Upper quantiles decease faster than lower quantiles. Significant decrease at 95th quantile has been shown by the data. Ignoring outcome-dependent follow-up would result in very biased estimates of $\beta_1(\tau)$ with opposite sign and opposite changing trend as τ increases.

4.4 Remarks

Quantile regression analysis is a very robust and flexible approach for longitudinal data with skewed outcome and varying covariate effects. Irregular outcome-dependent follow-up is a special missing pattern which is easy to be overlooked since there is no obvious blank cell in data. However, similar to regular outcome-dependent missing data, it can result in biased estimation in marginal regression methods, such as quantile regression. The proposed inverse intensity-ratio weighted estimator can correct the bias due to irregular outcome-dependent follow-up, given the model specification for the visit process is correct.

4.5 Appendix

Proof of *Theorem 1.*

First, we want to prove that $\psi_\tau(\boldsymbol{\beta};\boldsymbol{\alpha}_0)$ has a unique minimizer at $\boldsymbol{\beta} = \boldsymbol{\beta}_0(\tau)$. Define $\nu_\tau\{\boldsymbol{\beta};\mathbf{Z}(t)\} = E\left(\rho_\tau\left[Y(t) - \max\left\{c, \mathbf{X}(t)^\top\boldsymbol{\beta}\right\}\right] \middle| \mathbf{Z}(t)\right)$. We can show that $\nu_\tau\{\boldsymbol{\beta};\mathbf{Z}(t)\} \geq \nu_\tau\{\boldsymbol{\beta}_0(\tau);\mathbf{Z}(t)\}$ for any given $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0(\tau)$ in any possible situations.

(1) When $\mathbf{X}(t)^\top\boldsymbol{\beta}_0(\tau) \leq c$ and $\mathbf{X}(t)^\top\boldsymbol{\beta} \leq c$, $\nu_\tau\{\boldsymbol{\beta}_0(\tau);\mathbf{Z}(t)\} = \nu_\tau\{\boldsymbol{\beta};\mathbf{Z}(t)\}$.

(2) When $\mathbf{X}(t)^\top\boldsymbol{\beta}_0(\tau) \leq c$ and $\mathbf{X}(t)^\top\boldsymbol{\beta} > c$,

$$
\begin{aligned}
&\nu_\tau\{\boldsymbol{\beta}_0(\tau);\mathbf{Z}(t)\} - \nu_\tau\{\boldsymbol{\beta};\mathbf{Z}(t)\} \\
&= E\left[I\{Y(t) = c\}(\tau - 1)\left\{\mathbf{X}(t)^\top\boldsymbol{\beta} - c\right\}\middle|\mathbf{Z}(t)\right] \\
&\quad + E\left(I\left\{c < Y(t) \leq \mathbf{X}(t)^\top\boldsymbol{\beta}\right\}\left[\tau\left\{\mathbf{X}(t)^\top\boldsymbol{\beta} - c\right\} + Y(t) - \mathbf{X}(t)^\top\boldsymbol{\beta}\right]\middle|\mathbf{Z}(t)\right) \\
&\quad + E\left[I\left\{Y(t) > \mathbf{X}(t)^\top\boldsymbol{\beta}\right\}\tau\left\{\mathbf{X}(t)^\top\boldsymbol{\beta} - \tau\right\}\middle|\mathbf{Z}(t)\right] \\
&\leq E\left([I\{Y(t) = c\}(\tau - 1) + \tau I\{Y(t) > c\}]\middle|\mathbf{Z}(t)\right)\left\{\mathbf{X}(t)^\top\boldsymbol{\beta} - c\right\}.
\end{aligned}
$$

Since $\mathbf{X}(t)^\top\boldsymbol{\beta}_0(\tau) \leq c$, we have that $E[I\{Y(t) = c\}] \geq \tau$ and $E[I\{Y(t) > c\}] \leq 1 - \tau$. Therefore, $\nu_\tau\{\boldsymbol{\beta}_0(\tau);\mathbf{Z}(t)\} - \nu_\tau\{\boldsymbol{\beta};\mathbf{Z}(t)\} \leq 0$.

(3) When $\mathbf{X}(t)^\top\boldsymbol{\beta}_0(\tau) > c$,

$$
\begin{aligned}
&\nu_\tau\{\boldsymbol{\beta}_0(\tau);\mathbf{Z}(t)\} - \nu_\tau\{\boldsymbol{\beta};\mathbf{Z}(t)\}\\
&= (1-\tau)P\left\{Y(t) \le \mathbf{X}(t)^\top\boldsymbol{\beta}_0(\tau)\right\}\left[\mathbf{X}(t)^\top\boldsymbol{\beta}_0(\tau) - \max\left\{c, \mathbf{X}(t)^\top\boldsymbol{\beta}\right\}\right]\\
&\quad - \tau P\left\{Y(t) > \mathbf{X}(t)^\top\boldsymbol{\beta}_0(\tau)\right\}\left[\mathbf{X}(t)^\top\boldsymbol{\beta}_0(\tau) - \max\left\{c, \mathbf{X}(t)^\top\boldsymbol{\beta}\right\}\right]\\
&\quad + E\left(\int_{\max\{c,\mathbf{X}(t)^\top\boldsymbol{\beta}\}}^{\mathbf{X}(t)^\top\boldsymbol{\beta}_0(\tau)}\left[y - \max\left\{c, \mathbf{X}(t)^\top\boldsymbol{\beta}\right\}\right] f_{Y(t)}\{y|\mathbf{Z}(t)\}dy\bigg|\mathbf{Z}(t)\right)\\
&= E\left(\int_{\max\{c,\mathbf{X}(t)^\top\boldsymbol{\beta}\}}^{\mathbf{X}(t)^\top\boldsymbol{\beta}_0(\tau)}\left[\max\left\{c, \mathbf{X}(t)^\top\boldsymbol{\beta}\right\} - y\right] f_{Y(t)}\{y|\mathbf{Z}(t)\}dy\bigg|\mathbf{Z}(t)\right)\\
&\le 0
\end{aligned}
\tag{4.10}
$$

Under condition C1 (b), there exists a region around $\mathbf{X}(t)^\top\boldsymbol{\beta}_0(\tau)$ satisfying that $f_{Y(t)}\{y|\mathbf{Z}(t)\} > 0$; thus, $\left[\max\left\{c, \mathbf{X}(t)^\top\boldsymbol{\beta}\right\} - y\right] f_{Y(t)}\{y|\mathbf{Z}(t)\} < 0$. So if $\mathbf{X}(t)^\top\boldsymbol{\beta} \neq \mathbf{X}(t)^\top\boldsymbol{\beta}_0(\tau)$, then $\nu_\tau\{\boldsymbol{\beta}_0(\tau);\mathbf{Z}(t)\} < \nu_\tau\{\boldsymbol{\beta};\mathbf{Z}(t)\}$.

By condition C1 (a), we have that for any $\tau \in [\gamma, 1)$ and $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0(\tau)$,

$$
\begin{aligned}
E\Big[\int_0^\infty & I\left\{\mathbf{X}(t)^\top\boldsymbol{\beta}_0(\tau) > c\right\}\left\{\mathbf{X}(t)^\top\boldsymbol{\beta} - \mathbf{X}(t)^\top\boldsymbol{\beta}_0(\tau)\right\}^2\\
&\times\left\{dN^L(t) + I(L < t \le R)\lambda_0(t)dt\right\}\Big] > 0.
\end{aligned}
$$

Hence, for any $\tau \in [\gamma, 1)$ and $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0(\tau)$,

$$
\begin{aligned}
&E\Big[\int_0^\infty I\left\{\mathbf{X}(t)^\top\boldsymbol{\beta}_0(\tau) > c\right\}[\nu_\tau\{\boldsymbol{\beta};\mathbf{Z}(t)\}.\\
&- \nu_\tau\{\boldsymbol{\beta}_0(\tau);\mathbf{Z}(t)\}]\left\{dN^L(t) + I(L < t \le R)\lambda_0(t)dt\right\}\Big] > 0.
\end{aligned}
$$

Since for any $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0(\tau)$,

$$
\begin{aligned}
&\psi_\tau(\boldsymbol{\beta};\boldsymbol{\alpha}_0) - \psi_\tau\left\{\boldsymbol{\beta}_0(\tau);\boldsymbol{\alpha}_0\right\} \\
&= E\bigg(\int_0^\infty I\left\{\mathbf{X}(t)^\top\boldsymbol{\beta}_0(\tau) \leq c\right\} I\left\{\mathbf{X}(t)^\top\boldsymbol{\beta} \leq c\right\}\left[\nu_\tau\{\boldsymbol{\beta};\mathbf{Z}(t)\} - \nu_\tau\left\{\boldsymbol{\beta}_0(\tau);\mathbf{Z}(t)\right\}\right] \\
&\qquad\qquad \times \left\{dN^L(t) + I(L < t \leq R)\lambda_0(t)dt\right\}\bigg) \\
&\quad + E\bigg(\int_0^\infty I\left\{\mathbf{X}(t)^\top\boldsymbol{\beta}_0(\tau) \leq c\right\} I\left\{\mathbf{X}(t)^\top\boldsymbol{\beta} > c\right\}\left[\nu_\tau\{\boldsymbol{\beta};\mathbf{Z}(t)\} - \nu_\tau\left\{\boldsymbol{\beta}_0(\tau);\mathbf{Z}(t)\right\}\right] \\
&\qquad\qquad \times \left\{dN^L(t) + I(L < t \leq R)\lambda_0(t)dt\right\}\bigg) \\
&\quad + E\bigg(\int_0^\infty I\left\{\mathbf{X}(t)^\top\boldsymbol{\beta}_0(\tau) > c\right\}\left[\nu_\tau\{\boldsymbol{\beta};\mathbf{Z}(t)\} - \nu_\tau\left\{\boldsymbol{\beta}_0(\tau);\mathbf{Z}(t)\right\}\right] \\
&\qquad\qquad \times \left\{dN^L(t) + I(L < t \leq R)\lambda_0(t)dt\right\}\bigg) \\
&> 0.
\end{aligned}
$$

Therefore, under condition C1, we have proved that $\boldsymbol{\beta}_0(\tau)$ is a unique minimizer of $\psi_\tau(\boldsymbol{\beta};\boldsymbol{\alpha}_0)$.

Given $\hat{\boldsymbol{\alpha}} \xrightarrow{a.s.} \boldsymbol{\alpha}_0$, under condition C3, we have

$$
\sup_{\tau\in[\gamma,1),\ \boldsymbol{\beta}\in\mathcal{B}} \left|\psi_\tau(\boldsymbol{\beta};\hat{\boldsymbol{\alpha}}) - \psi_\tau(\boldsymbol{\beta};\boldsymbol{\alpha}_0)\right| \xrightarrow{a.s.} 0. \tag{4.11}
$$

Note that the objective function

$$
\begin{aligned}
\zeta_i^\tau(\boldsymbol{\beta};\boldsymbol{\alpha}) = \int_0^\infty \Big(&\tau\left[Y_i(t) - \max\left\{c, \mathbf{X}_i(t)^\top\boldsymbol{\beta}\right\}\right] \\
&- \tau I\left[Y_i(t) \leq \max\left\{c, \mathbf{X}_i(t)^\top\boldsymbol{\beta}\right\}\right]\left[Y_i(t) - \max\left\{c, \mathbf{X}_i(t)^\top\boldsymbol{\beta}\right\}\right] \\
&+ (1-\tau) I\left[Y_i(t) \leq \max\left\{c, \mathbf{X}_i(t)^\top\boldsymbol{\beta}\right\}\right]\left[\max\left\{c, \mathbf{X}_i(t)^\top\boldsymbol{\beta}\right\} - Y_i(t)\right]\Big) \\
&\times \left[dN_i^L(t) + \exp\left\{-\mathbf{h}_i(t)^\top\boldsymbol{\alpha}\right\} dN_i(t)\right]
\end{aligned}
$$

These three terms in the parenthesis $(\cdot)$ are either concave or convex functions of $\boldsymbol{\beta}$ and linear in τ, and $\exp\left\{-\mathbf{h}_i(t)^\top \boldsymbol{\alpha}\right\}$ is an either concave or convex function of $\boldsymbol{\alpha}$. This fact coupled with pointwise convergence by the strong law of large numbers given condition C4, implies the uniform convergence of $n^{-1/2}\Psi_\tau(\boldsymbol{\beta};\boldsymbol{\alpha})$ (Rockafellar, 1970 (Theorem 10.8)) that is

$$\sup_{\tau\in[\gamma,1),\ \boldsymbol{\beta}\in\mathcal{B},\ \boldsymbol{\alpha}\in\mathcal{A}} |n^{-1/2}\Psi_\tau(\boldsymbol{\beta};\boldsymbol{\alpha}) - \psi_\tau(\boldsymbol{\beta};\boldsymbol{\alpha})| \xrightarrow{a.s.} 0.$$

Coupled with (4.11), we will have that

$$\sup_{\boldsymbol{\beta}\in\mathcal{B},\ \tau\in[\gamma,1)} |n^{-1/2}\Psi_\tau(\boldsymbol{\beta};\hat{\boldsymbol{\alpha}}) - \psi_\tau(\boldsymbol{\beta};\boldsymbol{\alpha}_0)| \xrightarrow{a.s.} 0. \tag{4.12}$$

Given $\psi_\tau\left\{\boldsymbol{\beta}_0(\tau);\boldsymbol{\alpha}_0\right\} = 0$ and $\Psi_\tau\left\{\hat{\boldsymbol{\beta}}(\tau);\hat{\boldsymbol{\alpha}}\right\} = 0$, simple manipulation shows that

$$\sup_{\tau\in[\gamma,1)} \left|\psi_\tau\left\{\hat{\boldsymbol{\beta}}(\tau);\boldsymbol{\alpha}_0\right\} - \psi_\tau\left\{\boldsymbol{\beta}_0(\tau);\boldsymbol{\alpha}_0\right\}\right| \le \sup_{\tau\in[\gamma,1)} \left|\psi\left\{\hat{\boldsymbol{\beta}}(\tau);\boldsymbol{\alpha}_0\right\} - n^{-1/2}\Psi_\tau\left\{\hat{\boldsymbol{\beta}}(\tau);\hat{\boldsymbol{\alpha}}\right\}\right|.$$

By (4.12), we have that

$$\sup_{\tau\in[\gamma,1)} \left|\psi_\tau\left\{\hat{\boldsymbol{\beta}}(\tau);\boldsymbol{\alpha}_0\right\} - \psi_\tau\left\{\boldsymbol{\beta}_0(\tau);\boldsymbol{\alpha}_0\right\}\right| \xrightarrow{a.s.} 0. \tag{4.13}$$

Based on (4.13), we can prove uniform strong convergency of $\hat{\boldsymbol{\beta}}(\tau)$ by following similar arguments in the proof of theorem 3 in Huang and Peng (2009). Specifically, we need to prove that for any $\epsilon > 0$, there exists $\delta > 0$ such that if $\sup_{\tau\in[\gamma,1)} |\psi_\tau\left\{\boldsymbol{\beta}(\tau);\boldsymbol{\alpha}_0\right\} - \psi_\tau\left\{\boldsymbol{\beta}_0(\tau);\boldsymbol{\alpha}_0\right\}| < \delta$, then $\sup_{\tau\in[\gamma,1)} \|\boldsymbol{\beta}(\tau) - \boldsymbol{\beta}_0(\tau)\| < \epsilon$. Suppose that this is not true. Then, there must exist a constant $\epsilon^* > 0$. For any $\left\{\frac{1}{k} : k = 1, 2, \ldots\right\}$, there exists $(\boldsymbol{\beta}_k, \tau_k)$ such that $|\psi_{\tau_k}\left\{\boldsymbol{\beta}_k;\boldsymbol{\alpha}_0\right\} - \psi_{\tau_k}\left\{\boldsymbol{\beta}_0(\tau_k);\boldsymbol{\alpha}_0\right\}| < \frac{1}{k}$ but $\|\boldsymbol{\beta}_k - \boldsymbol{\beta}_0(\tau_k)\| > \epsilon^*$. Since $\mathcal{B}$ is a compact space, there exists a subsequence of $(\boldsymbol{\beta}_k, \tau_k)$ that converges to, say, $(\boldsymbol{\beta}^*, \tau^*)$. Then, we have that $\psi_{\tau^*}(\boldsymbol{\beta}^*;\boldsymbol{\alpha}_0) =$

$\psi_{\tau^*}\{\boldsymbol{\beta}_0(\tau^*);\boldsymbol{\alpha}_0\}$ but $\|\boldsymbol{\beta}^*-\boldsymbol{\beta}_0(\tau^*)\| \geq \epsilon^*$. This contradicts that $\boldsymbol{\beta}_0(\tau^*)$ is a unique minimizer of $\psi_{\tau^*}(\boldsymbol{\beta};\boldsymbol{\alpha}_0)$. Therefore, it is proved that for any $\epsilon > 0$, there exists $\delta > 0$ such that if $\sup_{\tau\in[\gamma,1)} |\psi_\tau\{\boldsymbol{\beta}(\tau);\boldsymbol{\alpha}_0\} - \psi_\tau\{\boldsymbol{\beta}_0(\tau);\boldsymbol{\alpha}_0\}| < \delta$, then $\sup_{\tau\in[\gamma,1)} \|\boldsymbol{\beta}(\tau)-\boldsymbol{\beta}_0(\tau)\| < \epsilon$. Consequently, given $\sup_{\tau\in[\gamma,1)} \left|\psi_\tau\left\{\hat{\boldsymbol{\beta}}(\tau);\boldsymbol{\alpha}_0\right\} - \psi_\tau\{\boldsymbol{\beta}_0(\tau);\boldsymbol{\alpha}_0\}\right| \xrightarrow{a.s.} 0$, it is easy to prove that $\sup_{\tau\in[\gamma,1)} \|\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\| \xrightarrow{a.s.} 0$.

Lemma 1.

$$\sup_{\tau\in[\gamma,1)} \left\|\mathbf{U}_\tau\left\{\hat{\boldsymbol{\beta}}(\tau);\hat{\boldsymbol{\alpha}}\right\} - \mathbf{U}_\tau\{\boldsymbol{\beta}_0(\tau);\boldsymbol{\alpha}_0\} - n^{1/2}\left[\boldsymbol{\mu}_\tau\left\{\hat{\boldsymbol{\beta}}(\tau);\hat{\boldsymbol{\alpha}}\right\} - \boldsymbol{\mu}_\tau\{\boldsymbol{\beta}_0(\tau);\boldsymbol{\alpha}_0\}\right]\right\| \xrightarrow{p} 0.$$

Proof of Lemma 1.

This lemma can be proved by using the results in Alexander (1984) and the arguments for theorem 1 of Lai and Ying (1988). The crucial step is to show that

$$\sup_{\tau\in[\gamma,1)} Var\left[\mathbf{l}_i^\tau\left\{\hat{\boldsymbol{\beta}}(\tau);\hat{\boldsymbol{\alpha}}\right\} - \mathbf{l}_i^\tau\{\boldsymbol{\beta}_0(\tau);\boldsymbol{\alpha}_0\}\right] \xrightarrow{p} 0. \tag{4.14}$$

Under condion C5 (a) and (b), there exists a finite number M_1 such that when $\hat{\boldsymbol{\alpha}} \in \mathcal{A}$,

$$
\begin{aligned}
&\sup_{\tau\in[\gamma,1)} Var\left[\mathbf{l}_i^{\tau}\left\{\hat{\boldsymbol{\beta}}(\tau);\hat{\boldsymbol{\alpha}}\right\}-\mathbf{l}_i^{\tau}\{\boldsymbol{\beta}_0(\tau);\boldsymbol{\alpha}_0\}\right]\\
&\le E\left[\mathbf{l}_i^{\tau}\left\{\hat{\boldsymbol{\beta}}(\tau);\hat{\boldsymbol{\alpha}}\right\}-\mathbf{l}_i^{\tau}\{\boldsymbol{\beta}_0(\tau);\boldsymbol{\alpha}_0\}\right]^2\\
&\le M_1\cdot\sup_{\tau\in[\gamma,1)} E\|\mathbf{l}_i^{\tau}\left\{\hat{\boldsymbol{\beta}}(\tau);\hat{\boldsymbol{\alpha}}\right\}-\mathbf{l}_i^{\tau}\{\boldsymbol{\beta}_0(\tau);\boldsymbol{\alpha}_0\}\|\\
&\le M_1\cdot\sup_{\tau\in[\gamma,1)} E\Bigg\|\int_0^\infty \mathbf{X}_i(t)I\left\{\mathbf{X}_i(t)^\top\hat{\boldsymbol{\beta}}(\tau)>c\right\}I\left\{\mathbf{X}_i(t)^\top\boldsymbol{\beta}_0(\tau)>c\right\}\\
&\quad\times\left[I\left\{Y_i(t)\le\mathbf{X}_i(t)^\top\hat{\boldsymbol{\beta}}(\tau)\right\}-I\left\{Y_i(t)\le\mathbf{X}_i(t)^\top\boldsymbol{\beta}_0(\tau)\right\}\right]\\
&\quad\times\left\{dN_i^L(t)+I(L_i<t\le R_i)\lambda_0(t)dt\right\}\Bigg\|\\
&+M_1\cdot\sup_{\tau\in[\gamma,1)} E\Bigg\|\int_0^\infty \mathbf{X}_i(t)I\left\{\mathbf{X}_i(t)^\top\hat{\boldsymbol{\beta}}(\tau)>c\right\}I\left\{\mathbf{X}_i(t)^\top\boldsymbol{\beta}_0(\tau)\le c\right\}\\
&\quad\times\left[I\left\{Y_i(t)\le\mathbf{X}_i(t)^\top\hat{\boldsymbol{\beta}}(\tau)\right\}-\tau\right]\left\{dN_i^L(t)+I(L_i<t\le R_i)\lambda_0(t)dt\right\}\Bigg\|\\
&+M_1\cdot\sup_{\tau\in[\gamma,1)} E\Bigg\|\int_0^\infty \mathbf{X}_i(t)\left\{\mathbf{X}_i(t)^\top\hat{\boldsymbol{\beta}}(\tau)\le c\right\}I\left\{\mathbf{X}_i(t)^\top\boldsymbol{\beta}_0(\tau)>c\right\}\\
&\quad\times\left[I\left\{Y_i(t)\le\mathbf{X}_i(t)^\top\boldsymbol{\beta}_0(\tau)\right\}-\tau\right]\Bigg]\left\{dN_i^L(t)+I(L_i<t\le R_i)\lambda_0(t)dt\right\}\Bigg\|\\
&+M_1\cdot\sup_{\tau\in[\gamma,1)} E\Bigg\|\int_0^\infty \mathbf{X}_i(t)I\left\{\mathbf{X}_i(t)^\top\hat{\boldsymbol{\beta}}(\tau)>c\right\}\left[I\left\{Y_i(t)\le\mathbf{X}_i(t)^\top\hat{\boldsymbol{\beta}}(\tau)\right\}-\tau\right]\\
&\qquad\qquad\qquad\times\left[\exp\left\{-\mathbf{h}_i(t)^\top\boldsymbol{\alpha}_0\right\}-\exp\left\{-\mathbf{h}_i(t)^\top\hat{\boldsymbol{\alpha}}\right\}\right]dN_i(t)\Bigg\|\\
&=(I)+(II)+(III)+(IV)
\end{aligned}
$$

Under condition C5 (a) - (c) and theorem 1,

$$
\begin{aligned}
&(I)\le M_1\cdot\sup_{\tau\in[\gamma,1)} E\Bigg\|\Bigg[\int_0^\infty \mathbf{X}_i(t)^{\otimes 2}I\left\{\mathbf{X}_i(t)^\top\boldsymbol{\beta}>c\right\}I\left\{\mathbf{X}_i(t)^\top\boldsymbol{\beta}_0(\tau)>c\right\}\\
&\times f_{Y_i(t)}\{\mathbf{X}_i(t)^\top\boldsymbol{\beta}_0(\tau)|\mathbf{Z}_i(t)\}\left\{dN_i^L(t)+I(L_i<t\le R_i)\lambda_0(t)dt\right\}+o_p(1)\Bigg]\\
&\times\left\{\hat{\boldsymbol{\beta}}(\tau)-\boldsymbol{\beta}_0(\tau)\right\}\Bigg\|\\
&\xrightarrow{p} 0
\end{aligned}
$$

When $\left\{\mathbf{X}_i(t)^\top\hat{\boldsymbol{\beta}}(\tau)-c\right\}\left\{\mathbf{X}_i(t)^\top\boldsymbol{\beta}_0(\tau)-c\right\} \leq 0$, it is easy to see that $|\mathbf{X}_i(t)^\top\boldsymbol{\beta}_0(\tau)-c| \leq |\mathbf{X}_i(t)^\top\boldsymbol{\beta}_0(\tau)-\mathbf{X}_i(t)^\top\hat{\boldsymbol{\beta}}(\tau)| \leq \|\mathbf{X}_i(t)\|\|\boldsymbol{\beta}_0(\tau)-\hat{\boldsymbol{\beta}}(\tau)\|$. Under condition C5 (a), (b), and (d) and theorem 1,

$$
\begin{aligned}
(II) \leq \sup_{\tau\in[\gamma,1)} E\Bigg\| \int_0^\infty & \mathbf{X}_i(t) I\left\{|\mathbf{X}_i(t)^\top\boldsymbol{\beta}_0(\tau)-c| \leq \|\mathbf{X}_i(t)\|\|\boldsymbol{\beta}_0(\tau)-\hat{\boldsymbol{\beta}}(\tau)\|\right\} \\
& \times \left[I\left\{Y_i(t) \leq \mathbf{X}_i(t)^\top\hat{\boldsymbol{\beta}}(\tau)\right\}-\tau\right]\left\{dN_i^L(t)+I(L_i<t\leq R_i)\lambda_0(t)dN_i(t)\right\}\Bigg\| \\
\xrightarrow{p} 0. &
\end{aligned}
$$

Similarly, it can be proved that $(III) \xrightarrow{p} 0$.

Under condition C5 (a) and (e) and the consistency of $\hat{\boldsymbol{\alpha}}$,

$$
\begin{aligned}
(IV) \leq M_1 \cdot \sup_{\tau\in[\gamma,1)} & \left\| \int_0^\infty \mathbf{X}_i(t) I\left\{\mathbf{X}_i(t)^\top\hat{\boldsymbol{\beta}}(\tau)>c\right\}\left[I\left\{Y_i(t)\leq \mathbf{X}_i(t)^\top\hat{\boldsymbol{\beta}}(\tau)\right\}-\tau\right] \right. \\
& \left. \times \mathbf{h}_i(t)\exp(-\mathbf{h}_i(t)^\top\boldsymbol{\alpha}_0)dN_i(t)\right\| \|\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0\| \\
\xrightarrow{p} 0 &
\end{aligned}
$$

Therefore, (4.14) has been proved and Lemma 1 is a direct consequence.

Proof of Theorem 2.

According to Lemma 1 and $\mathbf{U}_\tau\left\{\hat{\boldsymbol{\beta}}(\tau);\hat{\boldsymbol{\alpha}}\right\}=\mathbf{0}$, we have

$$
\begin{aligned}
& -\mathbf{U}_\tau\left\{\boldsymbol{\beta}_0(\tau);\boldsymbol{\alpha}_0\right\} \\
& = n^{1/2}\left\{\boldsymbol{\mu}_\tau\left(\hat{\boldsymbol{\beta}};\hat{\boldsymbol{\alpha}}\right)-\boldsymbol{\mu}_\tau\left(\boldsymbol{\beta}_0;\boldsymbol{\alpha}_0\right)\right\}+o_{p:\tau\in[\gamma,1)}(1) \\
& = \left[\mathbf{B}_\tau\left\{\boldsymbol{\beta}_0(\tau);\boldsymbol{\alpha}_0\right\}+o_p(1)\right]\cdot n^{1/2}\left\{\hat{\boldsymbol{\beta}}(\tau)-\boldsymbol{\beta}_0(\tau)\right\} \\
& \quad + \mathbf{A}_\tau\left\{\boldsymbol{\beta}_0(\tau);\boldsymbol{\alpha}_0\right\}\cdot n^{1/2}\left\{\hat{\boldsymbol{\alpha}}(\tau)-\boldsymbol{\alpha}_0(\tau)\right\}+o_{p:\tau\in[\gamma,1)}(1)
\end{aligned}
$$

where $\mathbf{A}_\tau(\boldsymbol{\beta};\boldsymbol{\alpha}) = \frac{\partial \boldsymbol{\mu}_\tau(\boldsymbol{\beta};\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} = -\int_0^\infty \mathbf{X}_i(t) I\left\{\mathbf{X}_i(t)^\top \boldsymbol{\beta} > c\right\}$

$$\times \left[I\left\{Y_i(t) \le \mathbf{X}_i(t)^\top \boldsymbol{\beta}\right\} - \tau\right] \mathbf{h}_i(t)^\top \exp\left\{-\mathbf{h}_i(t)^\top \boldsymbol{\alpha}\right\} dN_i(t),$$

and $o_{p:\tau\in[\gamma,1)}(1)$ means uniform convergence in probability to zero over $\tau \in [\gamma, 1)$.

Under condition C6,

$$\begin{aligned} n^{1/2}\left\{\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\right\} &= -\mathbf{B}_\tau\left\{\boldsymbol{\beta}_0(\tau);\boldsymbol{\alpha}_0\right\}^{-1} \\ &\times \left[\mathbf{U}_\tau\left\{\boldsymbol{\beta}_0(\tau);\boldsymbol{\alpha}_0\right\} + \mathbf{A}_\tau\left\{\boldsymbol{\beta}_0(\tau);\boldsymbol{\alpha}_0\right\} \cdot n^{1/2}\left\{\hat{\boldsymbol{\alpha}}(\tau) - \boldsymbol{\alpha}_0(\tau)\right\}\right] + o_{p:\tau\in[\gamma,1)}(1) \end{aligned}$$

Therefore,

$$\begin{aligned} n^{1/2}\{\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\} = n^{-1/2}\sum_{i=1}^{n}\Big[&-\mathbf{B}_\tau\left\{\boldsymbol{\beta}_0(\tau);\boldsymbol{\alpha}_0\right\}^{-1} \mathbf{l}_i^\tau\left\{\boldsymbol{\beta}_0(\tau);\boldsymbol{\alpha}_0\right\} \\ &+ \mathbf{B}_\tau\left\{\boldsymbol{\beta}_0(\tau);\boldsymbol{\alpha}_0\right\}^{-1} \mathbf{A}_\tau\left\{\boldsymbol{\beta}_0(\tau);\boldsymbol{\alpha}_0\right\} \mathbf{J}(\boldsymbol{\alpha}_0)^{-1} \boldsymbol{\iota}_i(\boldsymbol{\alpha}_0)\Big] + o_{p:\tau\in[\gamma,1)}(1). \end{aligned}$$

According to the definition of quantile and the quantile regression model assumption, $\mathbf{X}_i(t)^\top \boldsymbol{\beta}_0(\tau)$ increases in τ. Since $\int_0^\infty \tau \mathbf{X}_i(t) I\{\mathbf{X}_i(t)^\top \boldsymbol{\beta}_0(\tau) > c\}\left[dN_i^L(t) + \exp\{-\mathbf{h}_i(t)^\top \boldsymbol{\alpha}_0\} dN_i(t)\right]$ and $\int_0^\infty \mathbf{X}_i(t) I\{\mathbf{X}_i(t)^\top \boldsymbol{\beta}_0(\tau) > c\} I\{Y_i(t) \le \mathbf{X}_i(t)^\top \boldsymbol{\beta}_0(\tau)\}\left[dN_i^L(t) + \exp\{-\mathbf{h}_i(t)^\top \boldsymbol{\alpha}_0\} dN_i(t)\right]$ are bounded and monotone functions on $\tau \in [\gamma, 1)$, $\{\mathbf{l}_i^\tau(\boldsymbol{\beta}_0(\tau);\boldsymbol{\alpha}_0) : \tau \in [\gamma, 1)\}$ is a Donsker class. By Donsker theorem and pointwise central limit theory, $n^{1/2}\left\{\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\right\}$ converges weakly to a Gaussian process with covariance matrix $\boldsymbol{\Sigma}(\tau_1, \tau_2)$ for $\tau \in [\gamma, 1)$, where

$$\boldsymbol{\Sigma}(\tau_1, \tau_2) = E\left\{\boldsymbol{\xi}_i(\tau_1)\boldsymbol{\xi}_i(\tau_2)^\top\right\} \tag{4.15}$$

with $\boldsymbol{\xi}_i(\tau) = -\mathbf{B}_\tau\left\{\boldsymbol{\beta}_0(\tau);\boldsymbol{\alpha}_0\right\}^{-1} \mathbf{l}_i^\tau\left\{\boldsymbol{\beta}_0(\tau);\boldsymbol{\alpha}_0\right\} + \mathbf{B}_\tau\left\{\boldsymbol{\beta}_0(\tau);\boldsymbol{\alpha}_0\right\}^{-1} \mathbf{A}_\tau\left\{\boldsymbol{\beta}_0(\tau);\boldsymbol{\alpha}_0\right\} \mathbf{J}(\boldsymbol{\alpha}_0)^{-1} \boldsymbol{\iota}_i(\boldsymbol{\alpha}_0)$.

Chapter 5

Summary and Future Works

5.1 Summary

In my dissertation we focused on three complex data scenarios often encountered in biomedical observational studies. We proposed flexible semiparametric regression methods that properly handle realistic data features including window observation of recurrent events, outcome-dependent follow-up of longitudinal measurements, and etc.

For recurrent event data subject to window observation, we developed a two-stage estimation procedure as well as a novel counting process based estimation procedure under the accelerated recurrence time model. The counting process based estimation procedure is more efficient and simpler in computation. The resulting estimators are shown to be uniformly consistent and converge weakly to a Gaussian. Resampling method of Jin et al. (2001) and a new sample-based procedure are employed for inference. Simulation studies show satisfactory performance of our methods with moderate sample size. An application to the CFFPR data demonstrates practical utility of our proposals.

We have also investigated quantile regression for longitudinal data with outcomes subject to left censoring and follow-up pattern being outcome-dependent. While left censoring is handled by censored quantile regression technique, we investigate two different strategies to account for the dependency between outcomes and follow-up visit times. In both strategies, the idea of inverse probability weighting is the essential to deal with the relevant statistical issues. We evaluate our proposals by simulation studies. The proposed methods are applied to the Michigan PBB study to investigate the PBB decay profile. Both simulations and the real data example demonstrate that failing to handle these data feature can lead to considerably biased estimation and consequently misleading scientific conclusions.

5.2 Future Works

In this subsection we discuss work to be done in near future and possible extensions of this dissertation work.

First, we will conduct more sensitivity analysis simulations for our second and third projects to obtain a better understanding about how a misspecified model of missing pattern would influence the quantile regression estimator. Specifically, we can consider situations when a significant covariate is not included in the model of missingness or the whole model assumption is not correct.

It is worthwhile to consider the double-robust estimating method for the second and the third projects. Double-robust approach could also be considered as a diagnosis approach. If the double-robust estimates are quite different from the proposed estimates, either the quantile regression model or the model of missing data is likely to be misspecified.

For the first project, our proposed estimators of $\boldsymbol{\beta}_0(u)$ requires the ART model assumptions for all $0 < v < u$. It would be best if we can find a competative estimator under a local model with only ART model assumption for expected frequency equal to u.

Bibliography

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

Barrodale, I. and Roberts, F. D. K. (1974). Solution of an overdetermined system of equations in the l_1 norm. *Communications of the ACM* **17,** 319–320.

Buzkova, P. and Lumley, T. (2007). Longitudinal data analysis for generalized linear models with follow-up dependent on outcome-related variables. *The Canadian Journal of Statistics* **35,** 485–500.

Chiang, C.-T. and Wang, M.-C. (2009). Varying-coefficient model for the occurrence rate function of recurrent events. *Annals of the Institute of Statistical Mathematics* **61,** 197–213.

Fine, J. P., Yan, J., and Kosorok, M. R. (2004). Temporal process regression. *Biometrika* **91,** 683–703.

Fitzmaurice, G. M., Lipsitz, S. R., Ibrahim, J. G., Gelber, R., and Lipshultz, S. (2006). Estimation in regression models for longitudinal binary data with outcome-dependent follow-up. *Biostatistics* **7,** 469–485.

Geraci, M. and Bottai, M. (2007). Quantile regression for longitudinal data using the asymmetric laplace distribution. *Biostatistics* **8,** 140–154.

Huang, Y. (2002). Censored regression with the multistate accelerated sojourn times model. *J. R. Statist. Soc. B* **64,** 17–29.

Huang, Y. and Peng, L. (2009). Accelerated recurrence time models. *Scandinavian Journal of Statistics* **36,** 636–648.

Ibrahim, J. G. and Molenberghs, G. (2009). Missing data methods in longitudinal studies: a review. *Test* **18,** 1–43.

Jin, Z., Ying, Z., and Wei, L. J. (2001). A simple resampling method by perturbing the minimand. *Biometrika* **88,** 381–390.

Jung, S.-H. (1996). Quasi-likelihood for median regression models. *Journal of the American Statistical Association* **91,** 251–257.

Koenker, R. (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis* **91,** 74–89.

Koenker, R. (2008). Censored quantile regression redux. *Journal of Statistical Software* **27,**.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **88,** 963–974.

Lawless, J. F. and Nadeau, C. (1995). Some simple robust methods for the analysis of recurrent events. *Technometrics* **37,** 158–168.

Lee, M. and Kong, L. (2013). Quantile regression for longitudinal biomarker data subject to left censoring and dropouts. *Communications in Statistics - Theory and Methods* .

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73,** 13–22.

Lin, D. Y., Wei, L. J., Yang, I., and Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **62,** 711–730.

Lin, D. Y., Wei, L. J., and Ying, Z. (1998). Accelerated failure time models for counting process. *Biometrika* **85,** 605–618.

Lin, D. Y., Wei, L. J., and Ying, Z. (2001). Semiparametric transformation models for point processes. *Journal of the American Statistical Association* **96,** 620–628.

Lin, H., Scharfstein, D. O., and Rosenheck, R. A. (2004). Analysis of longitudinal data with irregular, outcome-dependent follow-up. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* **66,** 791–813.

Lipsitz, S. R., Fitzmaurice, G. M., Ibrabim, J. G., Gelder, R., and Lipshultz, S. (2002). Parameter estimation in longitudinal studies with outcome-dependent follow-up. *Biometrics* **58,** 621–630.

Lipsitz, S. R., Fitzmaurice, G. M., Molenberghs, G., and Zhao, L. P. (1997). Quantile regression methods for longitudinal data with drop-outs: application to cd4 cell counts of patients infected with the human immunodeficiency virus. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **46,** 463–476.

Little, R. J. A. and Rubin, D. B. (1987). *Statistical analysis with missing data*. Wiley, New york.

Nielsen, G. G., Gill, R. D., Andersen, P. K., and Sorenson, T. I. A. (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian Journal of Statistics* **19(1),** 25–43.

Oakes, D. (1992). Frailty models for multiple event times. *In survival Analysis: State* *of the Art (J.P. Klein and P.K. Goel eds.), Kluwer Academic, Amsterdam* pages 371–379.

Peng, L. and Fine, J. (2009). Competing risks quantile regression. *Journal of the American Statistical Association* **104,** 1440–1453.

Peng, L. and Huang, Y. (2008). Survival analysis with quantile regression models. *Journal of American Statistical Association* **103,** 637 – 649.

Pepe, M. S. and Cai, J. (1993). Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates. *Journal of the American Statistical Association* **88,** 811–820.

Powell, J. L. (1984). Least absolute deviations estimation for the censored regression model. *Journal of Econometrics* **25,** 303–325.

Powell, J. L. (1986). Censored regression quantiles. *Journal of Econometrics* **32,** 143–155.

Prentice, R. L., Williams, B. J., and Peterson, A. V. (1981). On the regression analysis of multivaraite failure time data. *Biometrika* **68,** 373–379.

Reid, N. (1994). A conversation with sir david cox. *Statist. Sci.* **9,** 439–455.

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association* **90,** 106–121.

Rubin, D. B. (1976). Inference and missing data. *Biometrika* **63,** 581–592.

Ryu, D., Sinha, D., Mallick, B., Lipsitz, S. R., and Lipshultz, S. E. (2007). Longitudinal studies with outcome-dependent follow-up. *Journal of the American Statistical Association* **102,** 952–961.

Schaubel, D. E. and Cai, J. (2004). Regression methods for gap time hazard functions of sequentially ordered multivariate failure time data. *Biometrika* **91,** 291–303.

Terrell, M. L., Manatunga, A. K., Small, C. M., Cameron, L. L., Wirth, J., Blanck, H. M., Lyles, R. H., and Marcus, M. (2008). A decay model for assessing polybrominated biphenyl exposure among women in the michigan long-term pbb study. *Journal of Exposure Science and Environmental Epidemiology* **18,** 410–420.

Wang, H. J. and Fygenson, M. (2009). Inference for censored quantile regression models in longitudinal studies. *The Annals of Statistics* **37,** 756–781.

Ware, J. H. (1985). Linear models for analysis of longitudinal studies. *Am. Statistician* **39,** 95–101.

Yi, G. Y. and He, W. (2009). Median regression models for longitudinal data with dropouts. *Biometrics* **65,** 618–625.

Yuan, Y. and Yin, G. (2010). Bayesian quantile regression for longitudinal studies with nonignorable missing data. *Biometrics* **66,** 105–114.