Smoothing Techniques and Semiparametric Regression Models for Recurrent Event Data

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Tianmeng Lyu

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Xianghua Luo, Adviser

May 2018

Acknowledgements

I would like to express my sincere gratitude to my advisor, Professor Xianghua Luo, for her guidance, advice and support in my academic life. She has set up a great example as a researcher for me to follow. I also would like to thank Professors Gongjun Xu, Chiung-Yu Huang, and Yifei Sun for their generous help and insightful advice on my research. I'm very grateful to Professors Lynn Eberly and Eric Lock for serving as my committee members and I really appreciate the opportunities to work with them on several exciting projects.

I thank all the faculty and staff members in Division of Biostatistics for providing such a welcoming and encouraging environment. I would like to thank Professors Susan Telke, Saonli Basu, Julian Wolfson, Wei Pan and Cavan Reilly for their support and guidance during my graduate assistantship. I treasure the rewarding experience working with the faculty in Center for Magnetic Resonance Research and School of Medicine. Special thanks go to Professors Gülin Öz and Alexa Pragman for the inspiring collaborations.

I would like to thank my classmates and friends in the division who made my life much more enjoyable. I thank Qinshu Lian, Jin Jin, Jincheng Zhou, Lifeng Lin, Chong Wu, Chen Gao, Chuyu Deng, Yangqing Deng, Jaron Arbet, Adam Kaplan, Zhenxun Wang and others for their encouragement and support over the years. My graduate life would not have been the same without them.

Finally, I would like to express my heartful gratitude to my family for their love and support. I thank my parents, Ke Lyu and Hongyan Yao, for their endless love, caring and encouragement, and also for setting good examples for me both in life and in work. I express my deepest gratitude and love to my husband, Chao Li, for always being there for me. I would not have completed my dissertation without his support and encouragement.

Abstract

Recurrent event data are frequently encountered in biomedical and clinical studies where the event of interest can happen for multiple times, such as recurrent hospitalizations and recurrent infections. The analysis of recurrent event data can be based on the gap times between consecutive events or on the total time to event. In this dissertation, we improve the estimation and inference procedures of the accelerated failure time model for recurrent gap time data using the induced smoothing technique in the first project, and we focus on regression models on the rate function of the recurrent event process in the second and the third projects.

The semiparametric accelerated failure time (AFT) model is especially appealing in analyzing recurrent gap time data owing to its direct interpretation of covariate effects. In general, estimation of the semiparametric AFT model is challenging because the rank-based estimating function is a non-smooth step function. In the first project, we extend the induced smoothing approach to the AFT model for recurrent gap time data. Our proposed smooth estimating function permits the application of standard numerical methods for both the regression coefficients estimation and the standard error estimation. The proposed method is applied to the data analysis of repeated hospitalizations for patients in the Danish Psychiatric Center Register.

In the second project, we focus on the semiparametric additive rates model where the regression coefficients quantify the absolute difference in the occurrence rate of the recurrent events between different groups. The model estimation requires the values of time-dependent covariates being observed throughout the entire follow-up period. In practice, however, time-dependent covariates are usually only measured at intermittent follow-up visits. To solve this problem, we propose to kernel smooth functions involving time-dependent covariates across subjects in the estimating function. In the third project, we extend the kernel smoothing approach to the additive-multiplicative rates model with intermittently observed time-dependent covariates. The additive-multiplicative rates model allows some covariates to have additive effects and others to have multiplicative effects. The proposed methods are illustrated by analyzing data from an epidemiologic study which aims to evaluate the effect of streptococcal infections on recurrent pharyngitis episodes.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Recurrent event data analysis

Different from the traditional survival data where the event of interest, e.g., death, can only occur once, in recurrent event data each subject can experience the event of interest repeatedly over time. Such data are increasingly collected in clinical and epidemiological studies, such as recurrent infections after transplantations, repeated failures of a medical device, and recurrent hospitalizations of patients with psychiatric disorders. Characterizing the recurrent event process and identifying the potential risk factors could be of significant interest in clinical studies.

Depending on the nature of recurrent events and the research interest, the focus of statistical analysis can be on the *gap times* (illustrated in Figure 1.1) between consecutive events or on *total time to event* (illustrated in Figure 1.2).

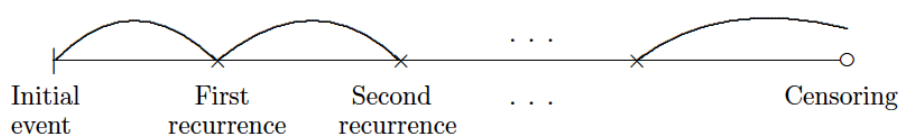

Figure 1.1: Illustration of the recurrent gap time data.

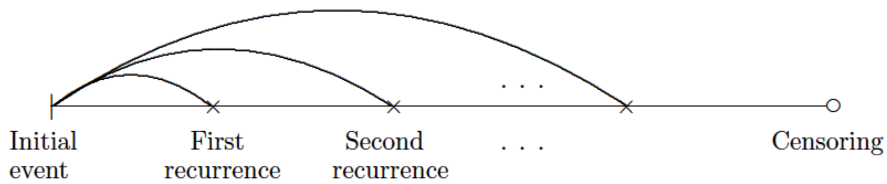

Figure 1.2: Illustration of the recurrent time-to-event data.

For the former, as discussed in Wang and Chang (1999), the unique structure of recurrent gap time data imposes difficulties in model estimation. First, due to the correlation among gap times of the same subject, the recurrent gap times beyond the first gap are subject to "induced informative censoring" even when the total censoring time is completely random. Second, the last censored gap time is expected to be longer than the previous uncensored gap times due to intercept sampling. Last, unlike the clustered survival data where the cluster size is typically assumed to be non-informative, the number of recurrent gap times of a subject is usually informative because subjects who are at a higher risk tend to have more gap times. Therefore, it is not appropriate to naively treat recurrent gap time data as clustered survival data and apply methods for clustered survival data to recurrent gap time data. Several authors, including Wang and Chang (1999) and Peña et al. (2001), have developed nonparametric methods to estimate the distribution of recurrent gap times, whereas others have developed various semiparametric regression models for evaluating covariate effects on the recurrent gap times. Huang and Chen (2003) proposed a marginal proportional hazards (PH) model, and Chang (2004) and Strawderman (2005) studied accelerated failure time (AFT) models for recurrent gap time data. Lu (2005) proposed linear transformation models and Sun et al. (2006) considered additive hazards models for recurrent gap time data. More recently, Luo et al. (2013) studied quantile regression models and Kang et al. (2015) proposed a class of transformed hazards models for recurrent gap time data.

When the time-to-event data are analyzed, the interest mainly focuses on modeling the occurrence intensity or rate of recurrent events over time. Let $N^*(t)$ be the number of recurrent events occurring at or prior to time t. Note that $N^*(t)$ is called the counting process or recurrent event process in literature. The *intensity function* of the recurrent event process $N^*(t)$ is defined as

$$\lambda(t|H(t)) = \lim_{\Delta \to 0^+} \frac{P(N^*(t+\Delta) - N^*(t) > 0|H(t))}{\Delta},$$

where $H(t)$ represents the event history up to and including t. Note that the intensity function $\lambda(t|H(t))$ can be regarded as the instantaneous risk of recurrent event conditioning on the event history. On the other hand, the *rate function* of the recurrent event process is defined as

$$\lambda(t) = \lim_{\Delta \to 0^+} \frac{P(N^*(t+\Delta) - N^*(t) > 0)}{\Delta},$$

namely, the occurrence rate at t unconditionally on the event history. Thus $\lambda(t)$ is often referred to as the marginal rate function. Andersen and Gill (1982) and Prentice et al. (1981a) proposed Cox-type conditional regression models which assume that the effects of covariates are multiplicative on the intensity function of the underlying counting process, while Pepe and Cai (1993) and Lin et al. (2000) considered Cox-type marginal regressions on the rate function. Other semiparametric models, such as additive intensity or rate models, have been considered by Liu and Wu (2011) and Schaubel et al. (2006). For nonparametric methods, Lawless and Nadeau (1995) and Nelson (1995) studied the estimation of the cumulative rate function and Chiang et al. (2005) explored a smoothing technique for estimating the rate function. See Cook and Lawless (2007) for a comprehensive review of the methods for recurrent event analysis.

In this dissertation, we apply smoothing techniques to improve the estimation of different regression models on recurrent event data. Specifically, for the analysis of recurrent gap time data, we apply the induced smoothing method to the AFT model, which overcomes the computational difficulties arising from the existing non-smooth estimating function. For the analysis of time-to-event data, we apply the kernel smoothing technique to the estimating functions of the additive rates model and the additive-multiplicative rates model to deal with the intermittently observed time-dependent covariates.

1.2 Data examples

1.2.1 Hospitalization data of psychiatric patients from the Danish Psychiatric Central Register

The hospitalization data from the Danish Psychiatric Central Register (Munk-Jørgensen and Mortensen, 1997) computerized all admissions to psychiatric hospitals and psychiatric wards in general hospitals in Denmark since 1969. We consider a subset of the data, which was composed of a cohort of 286 individuals who were first admitted to or contacted with Danish psychiatric services between April 1 and December 31, 1970. The maximum follow-up time was set to be 3 years to avoid any potential change in the distributional pattern of recurrent gap times. The details about this cohort have been described elsewhere (Luo and Huang, 2011; Luo et al., 2013). Briefly, among the 286 subjects, 106 (37%) were females, 230 (80%) had schizophrenia onset after 20 years old, 115 (40%) were censored after the initial hospitalization or contact with no records of rehospitalization, 56 (20%) had one rehospitalization, and 115 (40%) had two or more rehospitalization records. The average number of rehospitalization was 1.7. The median disease onset age was 26 with a range of 14 to 88 years old.

Note that 9 of the 286 patients died before the end of the follow-up time, hence, the independent censoring assumption was not expected to be seriously violated. The main scientific interest is to estimate the effect of the disease onset age on the gap time between two successive hospitalizations. We further discuss the data analysis for recurrent gap time data in Chapter 2.

1.2.2 A study of streptococcal infections on the risk of pharyngitis

Pharyngitis is an infection of the pharynx, the back of the throat, which is often due to viruses, but several bacteria including group A streptococcus (GAS) are also a common cause of pharyngitis. GAS pharyngitis, which is also known as strep throat, is more prevalent in children and usually occurs in late winter and early spring. According to a World Health Organization report, there are over 616 million new cases of GAS pharyngitis per year and 550 million among them occur in less developed countries. Besides, other streptococcal infections, including group C, G streptococcus (GCS, GGS) may also cause pharyngitis, so it is of interest to explore the effect of streptococci on the risk of pharyngitis. Between March 2002 and March 2004, 307 school children in a rural area near Vellore, India were recruited in the study of pharyngitis (Jose et al., 2018). During the follow-up time, cases of pharyngitis were identified weekly and the streptococcal infection status was also determined for those with pharyngitis at the time when pharyngitis was diagnosed. In addition, monthly visits were scheduled to identify the streptococcal infections. In this dataset, the occurrence of pharyngitis is the event of interest which may happen recurrently and the streptococcal infections are time-dependent binary variables that were only intermittently observed. The analysis of this dataset is discussed in Chapter 3 and Chapter 4.

1.3 Organizations

The remaining dissertation is organized as follows. In Chapter 2, we first introduce the semiparametric accelerate failure time (AFT) model for recurrent gap time data and discuss the limitations of the current non-smooth estimating function and resampling-based variance estimation methods. Then we propose a smooth estimating function for the AFT model with recurrent gap time data by applying the induced smoothing method (Lyu et al., 2018). Starting from Chapter 3, we focus on the analysis of time-to-event data. We propose a semiparametric estimator for additive rates model for data with intermittently observed time-dependent covariates in Chapter 3. The proposed estimating function is constructed by incorporating kernel smoothed mean covariate process. In Chapter 4, we consider the additive-multiplicative rates model with intermittently observed time-dependent covariates. The additive-multiplicative rates model is more general and includes the additive rates model and the proportional rates model as special cases. Some conclusion remarks are included in Chapter 5.

Chapter 2

Induced smoothing for rank-based regression with recurrent gap time data

2.1 Introduction

Recurrent event data are frequently encountered in clinical and epidemiological studies, where each subject can experience an event of interest repeatedly. Examples of recurrent events include rehospitalizations experienced by patients with psychiatric disorders (Heslin and Weiss, 2015), recurrent infections after hematopoietic cell transplantations (Barker et al., 2005), and many others. Depending on the nature of recurrent events and the research interest, the focus of statistical analysis can be placed on the time-to-event data by modeling the intensity or rate function of the counting process or on the gap times between consecutive events. For the former, various nonparametric and semiparametric methods have been developed in the literature. Some nonparametric methods include the estimation of the cumulative rate function (Lawless and Nadeau, 1995; Nelson, 1995) and techniques for estimating the rate function (Chiang et al., 2005). Several authors (Andersen and Gill, 1982; Pepe and Cai, 1993; Prentice et al., 1981a; Lin et al., 2000) considered Cox-type models which assume that the effects of covariates are multiplicative on the intensity or rate functions of the underlying counting process, whereas others considered additive intensity or rate models (Liu and Wu, 2011; Schaubel et al., 2006).

Alternatively, the focus can be placed on the gap times between recurrent events. As discussed in Wang and Chang (1999), the unique sequential ordering structure of recurrent gap time data generates difficulties in model estimation. First, due to the correlation among gap times of the same subject, the recurrent gap times beyond the first gap are subject to induced informative censoring even when the total censoring time is completely random. Second, the last censored gap time is expected to be longer than the previous uncensored gap times. Lastly, unlike the clustered survival data where the cluster size is typically assumed to be non-informative, the number of recurrent gap times of a subject is usually informative since subjects who are at a higher risk tend to have more gap times. Therefore, it is not appropriate to naively treat recurrent gap time data as independently censored clustered survival data and apply methods for clustered survival data to recurrent gap time data. Several authors (Wang and Chang, 1999; Peña et al., 2001) have developed nonparametric methods to estimate the distribution of recurrent gap times, while others (Du, 2009; Du et al., 2011) studied nonparametric estimation of the gap time hazard function in the presence of covariates. Semiparametric regression models for recurrent gap time data include proportional hazards (PH) models (Huang and Chen, 2003), accelerated failure time (AFT) models (Chang, 2004; Strawderman, 2005), linear transformation models (Lu, 2005), additive hazards models (Sun et al., 2006), and more recently, quantile regression models (Luo et al., 2013) and transformed hazards models (Kang et al., 2015).

Among the various recurrent gap time models, the AFT model is particularly appealing as it provides a direct interpretation of the covariate effects on the (transformed) length of gap times. Nevertheless, similar to the AFT models for univariate survival data (Prentice, 1978; Tsiatis, 1990; Kalbfleisch and Prentice, 2011, and reference therein), the estimation of the AFT model for recurrent gap time data (Chang, 2004) usually relies on rank-based estimating functions which are non-smooth step functions of regression parameters. It is well known that solving non-smooth, rank-based estimating equations could be computationally challenging since the solution to a non-smooth estimating equation typically does not exist. In addition to the difficulties in point estimation, variance estimation for the semiparametric AFT models has also been found challenging. This is because the asymptotic variance depends on the slope of the estimating function which can not be evaluated directly when the estimating function is non-smooth. Popular alternatives for variance estimation include the bootstrap method (Efron and Tibshirani, 1993) and the perturbation method (Parzen et al., 1994; Chang, 2004). However, both methods require solving rank-based estimating equations for numerous times, and hence can be computationally inefficient and unstable since they depend heavily on the point estimation from the non-smooth estimating functions, which is not guaranteed to succeed, for each resampling.

To tackle the difficulties in variance estimation for the AFT models with univariate survival data, Zeng and Lin (2008) proposed new resampling methods which only require evaluating the estimating functions repeatedly rather than solving them. These methods (Zeng and Lin, 2008) can greatly improve the efficiency in computing for the variance estimation; however, the challenge in the point estimation remains unresolved. Alternatively, efforts have been made on improving the point and variance estimation simultaneously by approximating the rank-based estimation function by a continuously differentiable estimating function so that the standard numerical methods can be applied in the inference procedure. In particular, Brown and Wang (2007) proposed the so-called induced smoothing technique for the rank-based estimating function for univariate survival data with Gehan's weight. Later, it was extended to general weights (Chiou et al., 2015a). Similar smoothing techniques have been extended to clustered survival data (Johnson and Strawderman, 2009; Chiou et al., 2015b). To our knowledge, no efforts have been made on improving the estimation of the AFT model with recurrent gap time data in literature. In this chapter we propose to extend the induced smoothing technique to the AFT model for recurrent gap time data.

The rest of this chapter is organized as follows. In Section 2.2, we first introduce the notation and setting of the AFT model for recurrent gap time data. We then briefly introduce the non-smooth rank-based estimating functions. In Section 2.3, we present the proposed induced smoothing method for the recurrent-gap-time AFT model followed by its large-sample properties and an asymptotic variance estimator. In Section 2.4, we conduct simulation studies to compare the proposed induced smoothing method with the existing rank-based estimating function method with various variance estimation methods. A real data analysis using the patient contact data from the Danish Psychiatric Central Register is presented in Section 2.5. Some concluding remarks are provided in Section 2.6.

2.2 The AFT model and rank-based estimating functions

2.2.1 The AFT model for recurrent gap time data

Consider a study with n subjects being recruited after each experienced an initial event and being followed on the recurrence of the event. Let $i = 1, \ldots, n$ index the subjects and $j = 0, 1, \ldots$ index the recurrent events of the ith subject, with $j = 0$ indicating the initial event. Let T_{ij} denote the gap time between the $(j-1)$th event and the jth event for subject i. Among the various regression models for recurrent gap times, the AFT model is of particular interest because of its direct interpretation of covariate effects on the (transformed) gap time variable. Let $\mathbf{Z}_i$ be the $p \times 1$ vector of baseline covariates. We impose the usual linear model for the logarithm-transformed gap times:

$$\log(T_{ij}) = \boldsymbol{\beta}_0^\intercal \mathbf{Z}_i + \epsilon_{ij}, \tag{2.1}$$

where $\boldsymbol{\beta}_0$ is the true $p \times 1$ vector of regression parameters and has the usual interpretation of covariate effects as in linear models. The error terms within each subject, $\epsilon_{ij}, j = 1, 2, \ldots$, are assumed to have an unknown common marginal distribution, and the correlation structure among the error terms is left unspecified. In this way, the correlation between two gap times ϵ_{ij} and $\epsilon_{ij'}$ is allowed to depend on j and j'. Finally, we assume that the error vectors $\epsilon_i = (\epsilon_{i1}, \epsilon_{i2}, \ldots)^\intercal$, $i = 1, \ldots, n$, are independently and identically distributed (i.i.d.) across subjects.

Note that the identical marginal distribution condition assumed for Model (2.1) is weaker than the shared frailty model which assumes that the error terms of the same subject are i.i.d. given a subject-specific frailty variable. Under the shared frailty model, each pair of gap times in the set $\{\log(T_{ij}), j = 1, \ldots\}$ are required to have the same correlation. The identical marginal distribution condition for Model (2.1) leaves the within-subject correlation structure fully unspecified, hence Model (2.1) allows more sophisticated correlation structure in real data, such as the autoregressive (AR) and the unstructured correlation.

In most applications, the observation of recurrent events is subject to right censoring due to loss of follow-up or end of study. Let C_i be the censoring time of the recurrent event process for the ith subject, which is assumed to be independent of $\{T_{ij}; j \geq 1\}$ conditional on $\mathbf{Z}_i$. Let m_i denote the number of observed events so that m_i satisfies $\sum_{j=1}^{m_i} T_{ij} \leq C_i$ and $\sum_{j=1}^{m_i+1} T_{ij} > C_i$, where $\sum_1^0 = 0$. We further define the censoring indicator for the jth event $\delta_{ij} = \mathrm{I}(\sum_{l=1}^{j} T_{il} \leq C_i)$, where $\mathrm{I}(\cdot)$ is an indicator function. Let X_{ij} denote the observed gap time such that $X_{ij} = T_{ij}$ for $j = 1, \ldots, m_i$ and $X_{i,m_i+1} = C_i - \sum_{l=1}^{m_i} X_{il}$. Define the transformed observed gap time $Y_{ij} = \log(X_{ij})$. The observed data of subject i consist of $\{(X_{ij}, \delta_{ij}); j = 1, \ldots, m_i + 1, \mathbf{Z}_i, C_i\}$.

2.2.2 Rank-based estimating function

We begin by considering the simple yet inefficient method that only uses times to first event in model estimation; that is, ignoring gap times of higher orders. Define the residuals $e_{ij}(\boldsymbol{\beta}) = \log(X_{ij}) - \boldsymbol{\beta}^\intercal \mathbf{Z}_i$. Let $N_{ij}(\boldsymbol{\beta}, t) = \delta_{ij} I\{e_{ij}(\boldsymbol{\beta}) \leq t\}$ and $R_{ij}(\boldsymbol{\beta}, t) = I\{e_{ij}(\boldsymbol{\beta}) \geq t\}$ be the counting process and at-risk process on the time scale of the residual, corresponding to subject i's jth gap time. An unbiased weighted rank-based estimating function for $\boldsymbol{\beta}$ based on the time-to-first event data takes the form (Tsiatis, 1990; Wei et al., 1990; Jin et al., 2003):

$$\sum_{i=1}^{n} w(\boldsymbol{\beta}, e_{i1}(\boldsymbol{\beta}))\delta_{i1} \left[\mathbf{Z}_i - \frac{\frac{1}{n}\sum_{l=1}^{n} \mathbf{Z}_l \mathrm{I}\{e_{l1}(\boldsymbol{\beta}) \geq e_{i1}(\boldsymbol{\beta})\}}{\frac{1}{n}\sum_{l=1}^{n} \mathrm{I}\{e_{l1}(\boldsymbol{\beta}) \geq e_{i1}(\boldsymbol{\beta})\}} \right]$$

or, equivalently,

$$\sum_{i=1}^{n} \int_{-\infty}^{\infty} w(\boldsymbol{\beta}, t) \left\{ \mathbf{Z}_i - \frac{S_1(\boldsymbol{\beta}, t)}{S_0(\boldsymbol{\beta}, t)} \right\} dN_{i1}(\boldsymbol{\beta}, t), \tag{2.2}$$

where $S_0(\boldsymbol{\beta}, t) = n^{-1} \sum_{i=1}^{n} R_{i1}(\boldsymbol{\beta}, t)$, $S_1(\boldsymbol{\beta}, t) = n^{-1} \sum_{i=1}^{n} \mathbf{Z}_i R_{i1}(\boldsymbol{\beta}, t)$, and $w(\boldsymbol{\beta}, t)$ is the weight function. Common choices of $w(\boldsymbol{\beta}, t)$ include $w(\boldsymbol{\beta}, t) \equiv 1$ for log-rank (LR) weight (Prentice, 1978) and $w(\boldsymbol{\beta}, t) \equiv S_0(\boldsymbol{\beta}, t)$ for Gehan's weight (Gehan, 1965). Note that the estimating function in (2.2) is constructed based on the linear rank statistic and can be viewed as the sum of the weighted difference between the covariate of a subject with an event (subject i) and the expected covariate among those who are in the "risk set" at the transformed event time of this subject, $\{l : e_{l1}(\boldsymbol{\beta}) \geq e_{i1}(\boldsymbol{\beta})\}$.

To improve the efficiency of estimation, one can make use of information beyond the first gap time. However, as discussed earlier, methods for clustered survival data cannot be directly applied to the recurrent gap time data due to the unique sequential structure of recurrent events. It was demonstrated in Luo and Huang (2011) that, when the underlying recurrent gap times of a subject are exchangeable, the weighted-risk set (WRS) technique can be applied to a reduced dataset to avoid biases in estimation caused by induced informative censoring and the biased sampling of the last censored gap time. Specifically the last censored gap time is not used in the construction of the estimating functions if the number of uncensored gap times of a subject is at least one. For the ease of discussion, we define $m_i^* = \max\{m_i, 1\}$, then $m_i^* = 1$ if subject i has no observed recurrent events and m_i^* equals the number of observed recurrent events m_i if $m_i \geq 1$. Note that $X_{i1} = C_i$ if $m_i = 0$ and $X_{ij} = T_{ij}$ for $j = 1, .., m_i^*$ if $m_i \geq 1$. Thus, the reduced data used in the WRS estimations are $\{(X_{ij}, \delta_{ij}); j = 1, \ldots, m_i^*, \mathbf{Z}_i, C_i\}$ from each subject. The WRS method assigns a weight $1/m_i^*$ to each of the remaining m_i^* gap times of a subject to ensure that overall contribution of each subject to the estimation to be the same to avoid the possible bias caused by informative cluster sizes.

In the same spirit as the WRS method in Luo and Huang (2011), we first define the *averaged counting process* and the *averaged at-risk process* for the AFT model:

$$
\begin{aligned}
N_i^*(\boldsymbol{\beta}, t) &= \frac{1}{m_i^*} \sum_{j=1}^{m_i^*} N_{ij}(\boldsymbol{\beta}, t), \\
R_i^*(\boldsymbol{\beta}, t) &= \frac{1}{m_i^*} \sum_{j=1}^{m_i^*} R_{ij}(\boldsymbol{\beta}, t).
\end{aligned}
$$

Note that these two averaged processes are based on the individual counting processes N_{ij} and R_{ij} defined earlier, which are all on the scale of the residual of the log-transformed gap times. Hence, the two averaged processes $N_i^*(\boldsymbol{\beta}, t)$ and $R_i^*(\boldsymbol{\beta}, t)$ defined here are different than those in Luo and Huang (2011). Let $S_0^*(\boldsymbol{\beta}, t) = n^{-1}\sum_{i=1}^n R_i^*(\boldsymbol{\beta}, t)$ and $S_1^*(\boldsymbol{\beta}, t) = n^{-1}\sum_{i=1}^n \mathbf{Z}_i R_i^*(\boldsymbol{\beta}, t)$. Then, we can replace $n^{-1}\sum_{i=1}^n N_{i1}(\boldsymbol{\beta}, t)$, $n^{-1}\sum_{i=1}^n \mathbf{Z}_i N_{i1}(\boldsymbol{\beta}, t)$, $S_0(\boldsymbol{\beta}, t)$, and $S_1(\boldsymbol{\beta}, t)$ in (2.2) with their respective multivariate counterparts $n^{-1}\sum_{i=1}^n N_i^*(\boldsymbol{\beta}, t)$, $n^{-1}\sum_{i=1}^n \mathbf{Z}_i N_i^*(\boldsymbol{\beta}, t)$, $S_0^*(\boldsymbol{\beta}, t)$, and $S_1^*(\boldsymbol{\beta}, t)$ and construct a new estimating equation:

$$U(\boldsymbol{\beta}) = \sum_{i=1}^n \int_{-\infty}^{\infty} w^*(\boldsymbol{\beta}, t)\left\{\mathbf{Z}_i - \frac{S_1^*(\boldsymbol{\beta}, t)}{S_0^*(\boldsymbol{\beta}, t)}\right\} dN_i^*(\boldsymbol{\beta}, t), \tag{2.3}$$

where the weight function $w^*(\boldsymbol{\beta}, t)$ is required to converge to the same limit as $w(\boldsymbol{\beta}, t)$ as $n \to \infty$. It can be shown that (2.3) is equivalent to

$$\sum_{i=1}^n \frac{1}{m_i^*} \sum_{j=1}^{m_i^*} w^*(\boldsymbol{\beta}, e_{ij}(\boldsymbol{\beta}))\delta_{ij}\left[\mathbf{Z}_i - \frac{S_1^*\{\boldsymbol{\beta}, e_{ij}(\boldsymbol{\beta})\}}{S_0^*\{\boldsymbol{\beta}, e_{ij}(\boldsymbol{\beta})\}}\right]. \tag{2.4}$$

It is easy to show that the empirical processes $n^{-1}\sum_{i=1}^n N_i^*(\boldsymbol{\beta}, t)$, $n^{-1}\sum_{i=1}^n \mathbf{Z}_i N_i^*(\boldsymbol{\beta}, t)$, $S_0^*(\boldsymbol{\beta}, t)$, and $S_1^*(\boldsymbol{\beta}, t)$ converge to the same limits as their respective univariate counterparts and that the mapping defined by U in (2.3) is compactly differentiable with respect to the supremum norm. As a result, we can prove that $U(\boldsymbol{\beta})$ and its univariate counterpart in (2.2) converge weakly to the same limiting distribution and converge uniformly to the same limit. The latter ensures the consistency of the solution, denoted by $\hat{\boldsymbol{\beta}}$, to the estimating equation $U(\boldsymbol{\beta}) = 0$.

Note that, while Chang (2004) was the first to consider the AFT model for recurrent event data, it is worthwhile to point out that the estimating function proposed in Chang (2004) is a special case of (2.4) with the unit or log-rank weight function,

$w^*(\boldsymbol{\beta}, t) = 1$:

$$U_{\mathrm{LR}}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \frac{1}{m_i^*} \sum_{j=1}^{m_i^*} \delta_{ij} \left[\mathbf{Z}_i - \frac{S_1^*\{\boldsymbol{\beta}, e_{ij}(\boldsymbol{\beta})\}}{S_0^*\{\boldsymbol{\beta}, e_{ij}(\boldsymbol{\beta})\}} \right]. \tag{2.5}$$

The existence of a strongly consistent and asymptotically normal sequence of solutions to $U_{\mathrm{LR}}(\boldsymbol{\beta}) = 0$ was established in Chang (2004); however, the involvement of the unknown parameter $\boldsymbol{\beta}$ in the indicator function renders the estimating function in (2.5) a non-smooth step function of $\boldsymbol{\beta}$. Hence, a solution $\hat{\boldsymbol{\beta}}_{\mathrm{LR}}$ such that $U_{\mathrm{LR}}(\hat{\boldsymbol{\beta}}_{\mathrm{LR}}) = 0$ may not exist for a finite sample. An alternative approach is to estimate $\boldsymbol{\beta}$ by minimizing the norm of $U_{\mathrm{LR}}(\boldsymbol{\beta})$, that is $\|U_{\mathrm{LR}}(\boldsymbol{\beta})\| = U_{\mathrm{LR}}(\boldsymbol{\beta})^\intercal U_{\mathrm{LR}}(\boldsymbol{\beta})$. However, because monotonicity in $U_{\mathrm{LR}}(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ is not guaranteed, there may exist multiple solutions to the minimization problem. Therefore, the point estimation based on the non-smooth estimating function in (2.5) could be computationally challenging in applications.

Because the asymptotic variance of the point estimator depends on the slope of the estimating function in (2.5), it is difficult to estimate the variance directly when the estimating function is non-smooth. In the literature, resampling-based methods are commonly used for variance estimation. Among them, the bootstrap method is popular due to the ease of implementation. As an alternative, Chang (2004) adopted the perturbation technique proposed by Parzen et al. (1994) to estimate the variance of $\hat{\boldsymbol{\beta}}_{\mathrm{LR}}$. Briefly, since it has been proved that $n^{-1/2}U_{\mathrm{LR}}(\boldsymbol{\beta})$ converges in distribution to a multivariate normal distribution with mean 0 and covariance $V_{\mathrm{LR}}(\boldsymbol{\beta})$, one can first generate a large number of random vectors R's from a multivariate normal distribution with mean 0 and covariance $\hat{V}_{\mathrm{LR}}(\boldsymbol{\beta})$, where $\hat{V}_{\mathrm{LR}}(\boldsymbol{\beta})$ is a consistent estimator of $V_{\mathrm{LR}}(\boldsymbol{\beta})$. Then, one can solve the equation $U_{\mathrm{LR}}(\boldsymbol{\beta}) = R$ to obtain $\hat{\boldsymbol{\beta}}_{\mathrm{LR}}(R)$ for each R. The variance of $\hat{\boldsymbol{\beta}}_{\mathrm{LR}}$ can be approximated by the sample variance of $\hat{\boldsymbol{\beta}}_{\mathrm{LR}}(R)$'s.

Note that both the bootstrap and the perturbation method require solving the estimating equation for a large number of times, which causes the computational burden to increase in a great amount, especially when the estimating function is non-smooth. In addition, the two variance estimation methods rely on the success of each resampling's point estimation whose challenges have been discussed previously.

2.3 The proposed induced smooth estimating function

Since the rank-based estimating functions discussed in Section 2.2.2 are non-smooth, causing difficulties in parameter estimation, we propose a monotonic, smooth estimating function in this section. We want to reemphasize that although Johnson and Strawderman (2009) have proposed a smooth estimating function for the clustered survival data AFT model, their method cannot be directly applied to the recurrent gap time data because of the unique structure of this type of data.

For univariate survival data, it has been proved that, when using Gehan's weight, the estimating function in (2.2) is monotonic and corresponds to a convex objective function (Fygenson and Ritov, 1994). If the parameter is estimated by minimizing the objective function, then the set of minimizers would be convex although the minimizer may not be unique. Later, it was showed that applying an *induced smoothing* technique on the rank-based estimating function with Gehan's weight leads to an estimating function which is both smooth and monotonic, essential for improving the computation for both the point and variance estimation (Brown and Wang, 2007). We now consider extending the induced smoothing technique to the setting of recurrent gap time data. We start with the rank-based estimating function for the recurrent gap time data in (2.4) by using a Gehan-type weight, defined as $w^*(\boldsymbol{\beta}, t) = S_0^*(\boldsymbol{\beta}, t)$, which converges to the same limit as Gehan's weight for univariate survival data

$w(\boldsymbol{\beta}, t) = S_0(\boldsymbol{\beta}, t)$. The estimating function then becomes

$$\begin{aligned} U_{\mathrm{G}}(\boldsymbol{\beta}) &= \sum_{i=1}^{n} \int_{-\infty}^{\infty} S_0^*(\boldsymbol{\beta}, t) \left\{ \mathbf{Z}_i - \frac{S_1^*(\boldsymbol{\beta}, t)}{S_0^*(\boldsymbol{\beta}, t)} \right\} dN_i^*(\boldsymbol{\beta}, t) \\ &= \sum_{i=1}^{n} \frac{1}{m_i^*} \sum_{j=1}^{m_i^*} S_0^*\{\boldsymbol{\beta}, e_{ij}(\boldsymbol{\beta})\}\delta_{ij} \left[\mathbf{Z}_i - \frac{S_1^*\{\boldsymbol{\beta}, e_{ij}(\boldsymbol{\beta})\}}{S_0^*\{\boldsymbol{\beta}, e_{ij}(\boldsymbol{\beta})\}} \right] \\ &= \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i^*} \sum_{l=1}^{n} \sum_{k=1}^{m_l^*} \frac{\delta_{ij}}{m_i^* m_l^*} (\mathbf{Z}_i - \mathbf{Z}_l) \mathrm{I}\left\{ e_{lk}(\boldsymbol{\beta}) \geq e_{ij}(\boldsymbol{\beta}) \right\}. \end{aligned} \tag{2.6}$$

Then, we can apply the induced smoothing technique to the estimating function with the Gehan-type weight in (2.6) as follows. Let W be a $p \times 1$ independent standard normal vector, then a smoothed estimating function can be proposed by replacing $U_{\mathrm{G}}(\boldsymbol{\beta})$ with $\mathrm{E}_W[U_{\mathrm{G}}(\tilde{\boldsymbol{\beta}})]$, where $\tilde{\boldsymbol{\beta}} = \boldsymbol{\beta} + n^{-1/2}W$, and E_W denotes the expectation with respect to W. This leads to a smooth, monotonic estimating function:

$$U_{\mathrm{G}}^{(s)}(\boldsymbol{\beta}) = \mathrm{E}_W \left[U_{\mathrm{G}}(\tilde{\boldsymbol{\beta}}) \right] = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i^*} \sum_{l=1}^{n} \sum_{k=1}^{m_l^*} \frac{\delta_{ij}}{m_i^* m_l^*} (\mathbf{Z}_i - \mathbf{Z}_l) \mathrm{E}_W \left[\mathrm{I}\{ e_{lk}(\tilde{\boldsymbol{\beta}}) \geq e_{ij}(\tilde{\boldsymbol{\beta}}) \} \right].$$

It is easy to show that

$$\begin{aligned} \mathrm{E}_W \left[I\left\{ e_{lk}(\tilde{\boldsymbol{\beta}}) \geq e_{ij}(\tilde{\boldsymbol{\beta}}) \right\} \right] &= \mathrm{E}_W \left[\mathrm{I}\left\{ Y_{lk} - (\boldsymbol{\beta} + n^{-1/2}W)^\intercal \mathbf{Z}_l \geq Y_{ij} - (\boldsymbol{\beta} + n^{-1/2}W)^\intercal \mathbf{Z}_i \right\} \right] \\ &= \mathrm{E}_W \left[\mathrm{I}\left\{ (\boldsymbol{\beta} + n^{-1/2}W)^\intercal (\mathbf{Z}_l - \mathbf{Z}_i) \leq Y_{lk} - Y_{ij} \right\} \right] \\ &= \Phi \left\{ \frac{Y_{lk} - Y_{ij} - \boldsymbol{\beta}^\intercal (\mathbf{Z}_l - \mathbf{Z}_i)}{r_{il}} \right\}, \end{aligned}$$

where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal random variable and $r_{il}^2 = n^{-1}(\mathbf{Z}_l - \mathbf{Z}_i)^\intercal (\mathbf{Z}_l - \mathbf{Z}_i)$. Let $h_{lk,ij}(\boldsymbol{\beta}) = \{Y_{lk} - Y_{ij} - \boldsymbol{\beta}^\intercal (\mathbf{Z}_l - \mathbf{Z}_i)\}/r_{il}$, then we have

$$\mathrm{E}_W\left[I\{e_{lk}(\tilde{\boldsymbol{\beta}}) \geq e_{ij}(\tilde{\boldsymbol{\beta}})\}\right] = \Phi(h_{lk,ij}(\boldsymbol{\beta})).$$

Thus, the resulting smooth estimating function can be expressed as

$$U_{\mathrm{G}}^{(s)}(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i^*}\sum_{l=1}^{n}\sum_{k=1}^{m_l^*}\frac{\delta_{ij}}{m_i^* m_l^*}(\mathbf{Z}_i - \mathbf{Z}_l)\Phi(h_{lk,ij}(\boldsymbol{\beta})). \tag{2.7}$$

Let $\dot{U}_{\mathrm{G}}^{(s)}(\boldsymbol{\beta}) = \partial\left\{\frac{1}{n}U_{\mathrm{G}}^{(s)}(\boldsymbol{\beta})\right\}/\partial\boldsymbol{\beta}$, then

$$\dot{U}_{\mathrm{G}}^{(s)}(\boldsymbol{\beta}) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{m_i^*}\sum_{l=1}^{n}\sum_{k=1}^{m_l^*}\frac{\delta_{ij}}{m_i^* m_l^*}\frac{1}{r_{il}}\phi(h_{lk,ij}(\boldsymbol{\beta}))(\mathbf{Z}_i - \mathbf{Z}_l)(\mathbf{Z}_i - \mathbf{Z}_l)^{\intercal},$$

where $\phi(\cdot)$ is the probability density function of a standard normal random variable. It can be easily shown that the smooth estimating function in (2.7) is the derivative of the convex objective function

$$L_{\mathrm{G}}^{(s)}(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i^*}\sum_{l=1}^{n}\sum_{k=1}^{m_l^*}\frac{\delta_{ij}}{m_i^* m_l^*}\left[\{e_{lk}(\boldsymbol{\beta}) - e_{ij}(\boldsymbol{\beta})\}\Phi(h_{ij,lk}(\boldsymbol{\beta})) + r_{il}\phi(h_{ij,lk}(\boldsymbol{\beta}))\right]. \tag{2.8}$$

The estimator $\hat{\boldsymbol{\beta}}_{\mathrm{G}}^{(s)}$ is obtained by minimizing the objective function $L_{\mathrm{G}}^{(s)}(\boldsymbol{\beta})$. The consistency and asymptotic normality of $\hat{\boldsymbol{\beta}}_{\mathrm{G}}^{(s)}$ are stated in the following theorem and proved in Appendix A.

Theorem 1. Under the regularity conditions in Appendix A, $\hat{\boldsymbol{\beta}}_{\mathrm{G}}^{(s)}$ is a strongly consistent estimator of $\boldsymbol{\beta}_0$ and $\sqrt{n}(\hat{\boldsymbol{\beta}}_{\mathrm{G}}^{(s)} - \boldsymbol{\beta}_0)$ converges in distribution to $N(\mathbf{0}, \Sigma)$, where $\Sigma = A^{-1}V(A^{-1})^{\intercal}$, and $V = \lim_{n\to\infty}\mathrm{Var}\left\{n^{-1/2}U_{\mathrm{G}}(\boldsymbol{\beta}_0)\right\}, A = \partial/\partial\boldsymbol{\beta}\left\{\lim_{n\to\infty} n^{-1}U_{\mathrm{G}}(\boldsymbol{\beta}_0)\right\}$.

Note that the smooth estimator $\hat{\boldsymbol{\beta}}_{\mathrm{G}}^{(s)}$ has the same asymptotic properties as the estimator $\hat{\boldsymbol{\beta}}_{\mathrm{G}}$ (defined in Appendix A) based on the non-smooth estimating function with the Gehan-type weight.

Since the proposed estimating function in (2.7) is smooth and thus differentiable, one can use $\dot{U}_{\mathrm{G}}^{(s)}(\hat{\boldsymbol{\beta}}_{\mathrm{G}}^{(s)})$ to estimate $A_{\mathrm{G}}(\boldsymbol{\beta}_0)$ (Brown and Wang, 2007; Johnson and Strawderman, 2009). Hence, we propose to use $\dot{U}_{\mathrm{G}}^{(s)}(\hat{\boldsymbol{\beta}}_{\mathrm{G}}^{(s)})^{-1}\hat{V}_{\mathrm{G}}\left\{\dot{U}_{\mathrm{G}}^{(s)}(\hat{\boldsymbol{\beta}}_{\mathrm{G}}^{(s)})^{-1}\right\}^{\mathsf{T}}$ to estimate the asymptotic variance of $\sqrt{n}\left(\hat{\boldsymbol{\beta}}_{\mathrm{G}}^{(s)}-\boldsymbol{\beta}_0\right)$, where $\hat{V}_G$ is the sample variance of $\{n^{-1/2}U_{\mathrm{G},b}^{(s)}(\hat{\boldsymbol{\beta}}_{\mathrm{G}}^{(s)}), b=1,\ldots,N_B\}$ and $U_{\mathrm{G},b}^{(s)}(\hat{\boldsymbol{\beta}}_{\mathrm{G}}^{(s)})$ is the smooth estimating function based on the bth bootstrap sample at $\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}_{\mathrm{G}}^{(s)}$.

2.4 Simulation

Simulation studies were conducted to assess the performance of the proposed smooth estimating function as compared to the non-smooth rank-based estimating function with various variance estimation methods. For each simulation scenario, 1000 datasets were generated, each with a sample size of $n=100$ or $n=200$. All resampling sizes (number of bootstraps or perturbations) were set to be 200.

We began by generating the log gap times $\log(T_{ij}), i=1,\ldots,n, j=1,2,\ldots,$ from the AFT model:

$$\log(T_{ij}) = \beta_1 Z_{i1} + \beta_2 Z_{i2} + \epsilon_{ij}, \tag{2.9}$$

where $\beta_1=\beta_2=0.5$, and $\epsilon_{ij}=\alpha_i+\epsilon_{ij}^*$. The covariate Z_1 had a Bernoulli distribution with success probability equal to 0.5 and Z_2 followed a uniform distribution on the interval $[0,1]$. The frailties α_i followed a normal distribution with mean -1 and variance ρ. Two types of distributions of the random errors ϵ_{ij}^* were examined: normal distribution and logistic distribution, and the parameters of the distributions were determined so that ϵ_{ij}^* had mean zero and variance $1-\rho$. Two values of the variance parameter, $\rho = 0.2, 0.4$, were considered to achieve different levels of within-subject correlations. Note that Model (2.9) implies a uniform correlation structure and the within subject correlation is ρ. It is easy to prove that the above shared frailty model satisfies the identical marginal distribution condition assumed in Model (2.1). The censoring times C_i were generated from uniform distributions to yield desirable censoring rates (i.e., percent of subjects without any observed events), $c_p = 25\%$ and 50%.

To show that the proposed method is valid when the data have more complicated correlation structure, we considered scenarios where $\log(T_{ij})$ follow a first-order autoregression or AR(1) model:

$$\log(T_{ij}) = \beta_1 Z_{i1} + \beta_2 Z_{i2} + \alpha + \omega_{ij},$$

where $\alpha = -1$, $\omega_{ij} = \rho\omega_{ij-1} + v_{ij}$ and v_{ij} followed a normal distribution with mean zero and variance $1-\rho^2$ for $j = 2, \ldots$. We started by generating ω_{i1} from a normal distribution with mean zero and variance 1. Two levels of ρ, 0.2 and 0.4, were considered. It can easily be proved that the above AR(1) model also satisfies the identical marginal distribution condition in Model (2.1).

With the simulated data, we first compared the performance of the non-smooth estimating equation with either the log-rank weight in (2.5) (Chang, 2004) or Gehan's weight in (2.6) to the performance of the proposed smooth estimating equation in (2.7). The simulation results for data with an uniform correlation structure in normal or logistic random errors, and data with the AR(1) correlation structure are presented in Table 2.1, Table 2.2 and Table 2.3, respectively. For the point estimates, we report the relative bias (Bias) and the Monte-Carlo empirical standard deviation of the point estimates (SD). For each variance estimation method, we report the average standard errors (ASE) and the coverage percentage (CP) of the 95% confidence intervals (CIs).

The simulation results show that the average point estimates based on the non-smooth and smooth estimating functions are all virtually unbiased. We noticed that under the simulation scenarios that we used, the non-smooth method with the log-rank weight failed to converge for about half a percent of the simulated datasets (the results in the tables are based on the simulated datasets with converged point estimates).

As for the variance estimation, the asymptotic variance estimator of the proposed smooth estimating function gives satisfactory variance estimation with the ASE being close to the Monte-Carlo empirical SD and the bootstrap ASE and the CP being close to its nominal level (95%). The Monte-Carlo SD of the estimates from the proposed smooth estimating function method and the non-smooth method with Gehan's weight are close; and both are smaller than that of the non-smooth method with log-rank weight (Chang, 2004) for the simulated data. It should be noted that since the bootstrap method and the perturbation method (Parzen et al., 1994) for the non-smooth method with log-rank weight need solving the non-smooth estimating equations for numerous times, the variance estimation suffers from the same non-convergence problem as in the point estimation (the ASE and CP in the tables are based on the converged bootstrap samples or perturbed samples only). The computing time of the asymptotic variance estimator based on the proposed smooth estimating function method was substantially shorter than that of the bootstrap or perturbation method of the non-smooth methods as expected.

For comparison, we also applied two existing methods to recurrent gap time data: (1) analyzing the time to first event only data with the induced smoothing method for univariate survival data (Brown and Wang, 2007), and (2) applying the induced smoothing method for clustered survival data (Johnson and Strawderman, 2009) to the recurrent gap time data, by ignoring their sequential structure. The results are

Table 2.1: Results for simulated data with uniform correlation structure and normal random error. The standard error for the non-smooth models, with the log-rank weight and Gehan's weight, is estimated by the bootstrap method or the perturbation method (Parzen et al.); the standard error for the proposed method is estimated by the bootstrap method and the asymptotic variance (ASV) estimator; n is the sample size; c_p is the percent of subjects without events; ρ is the within-subject correlation; $\bar{m}$ is the average number of gap times, observed or censored per subject; Bias is the relative bias computed as the difference of the mean estimated parameter and the true value divided by the true value; SD is the Monte-Carlo standard deviation; ASE is the mean standard error; CP is the proportion of the 95% confidence intervals covering the true value.

					Non-smooth, log-rank weight						Non-smooth, Gehan's weight			
							Bootstrap		Perturbation				Bootstrap	
n	c_p	ρ	$\bar{m}$		Bias	SD	ASE	CP	ASE	CP	Bias	SD	ASE	CP
100	0.25	0.2	3.41	β_1	-0.005	0.211	0.210	0.943	0.212	0.943	0.004	0.184	0.186	0.949
				β_2	-0.026	0.378	0.361	0.927	0.363	0.926	-0.022	0.321	0.323	0.946
		0.4	3.80	β_1	0.005	0.213	0.215	0.958	0.218	0.964	0.005	0.193	0.193	0.958
				β_2	0.017	0.372	0.373	0.938	0.376	0.937	0.021	0.321	0.336	0.952
	0.50	0.2	1.88	β_1	0.004	0.239	0.250	0.958	0.267	0.965	0.006	0.217	0.235	0.964
				β_2	-0.037	0.424	0.436	0.956	0.449	0.957	-0.006	0.394	0.409	0.960
		0.4	2.00	β_1	0.023	0.248	0.248	0.942	0.264	0.958	0.017	0.231	0.233	0.954
				β_2	-0.000	0.438	0.436	0.950	0.447	0.947	0.004	0.404	0.406	0.945
200	0.25	0.2	3.41	β_1	0.002	0.143	0.149	0.956	0.150	0.957	0.009	0.123	0.130	0.956
				β_2	-0.018	0.276	0.257	0.926	0.258	0.931	-0.004	0.231	0.225	0.937
		0.4	3.80	β_1	0.004	0.154	0.152	0.940	0.153	0.944	0.001	0.133	0.134	0.946
				β_2	-0.001	0.261	0.264	0.942	0.264	0.947	-0.002	0.225	0.233	0.960
	0.50	0.2	1.88	β_1	-0.005	0.169	0.173	0.944	0.177	0.951	-0.007	0.153	0.162	0.959
				β_2	-0.009	0.302	0.303	0.943	0.308	0.949	0.003	0.275	0.281	0.953
		0.4	2.00	β_1	0.002	0.169	0.174	0.956	0.177	0.962	0.002	0.158	0.162	0.956
				β_2	0.003	0.301	0.304	0.946	0.310	0.945	0.010	0.280	0.281	0.939

					Proposed, smooth						Naive			
							Bootstrap		ASV		Univariate		Clustered	
n	c_p	ρ	$\bar{m}$		Bias	SD	ASE	CP	ASE	CP	Bias	SD	Bias	SD
100	0.25	0.2	3.41	β_1	0.005	0.184	0.186	0.949	0.182	0.945	0.003	0.217	0.039	0.176
				β_2	-0.022	0.321	0.323	0.944	0.315	0.939	-0.004	0.378	0.017	0.291
		0.4	3.80	β_1	0.005	0.193	0.193	0.956	0.189	0.949	0.010	0.221	0.066	0.234
				β_2	0.022	0.321	0.336	0.950	0.326	0.949	0.024	0.371	0.058	0.400
	0.50	0.2	1.88	β_1	0.008	0.218	0.235	0.964	0.228	0.950	0.016	0.233	0.070	0.220
				β_2	-0.006	0.394	0.408	0.961	0.394	0.954	-0.005	0.423	0.022	0.377
		0.4	2.00	β_1	0.020	0.231	0.233	0.952	0.227	0.943	0.015	0.244	0.120	0.269
				β_2	0.005	0.404	0.406	0.943	0.393	0.934	0.003	0.426	0.056	0.469
200	0.25	0.2	3.41	β_1	0.009	0.123	0.130	0.955	0.129	0.953	0.009	0.148	0.049	0.122
				β_2	-0.004	0.231	0.225	0.936	0.222	0.936	0.018	0.274	0.033	0.211
		0.4	3.80	β_1	0.001	0.133	0.134	0.946	0.133	0.939	0.001	0.155	0.062	0.163
				β_2	-0.002	0.225	0.233	0.960	0.230	0.953	0.002	0.260	0.034	0.280
	0.50	0.2	1.88	β_1	-0.005	0.154	0.162	0.958	0.160	0.956	-0.000	0.163	0.056	0.157
				β_2	0.004	0.275	0.281	0.953	0.277	0.949	0.003	0.295	0.048	0.271
		0.4	2.00	β_1	0.003	0.158	0.162	0.957	0.160	0.952	0.002	0.166	0.120	0.192
				β_2	0.011	0.280	0.281	0.939	0.277	0.934	0.010	0.292	0.109	0.333

Table 2.2: Results for simulated data with uniform correlation structure and logistic random error. The standard error for the non-smooth models, with the log-rank weight and Gehan's weight, is estimated by the bootstrap method or the perturbation method (Parzen et al.); the standard error for the proposed method is estimated by the bootstrap method and the asymptotic variance (ASV) estimator; n is the sample size; c_p is the percent of subjects without events; ρ is the within-subject correlation; $\bar{m}$ is the average number of gap times, observed or censored per subject; Bias is the relative bias computed as the difference of the mean estimated parameter and the true value divided by the true value; SD is the Monte-Carlo standard deviation; ASE is the mean standard error; CP is the proportion of the 95% confidence intervals covering the true value.

					Non-smooth, log-rank weight						Non-smooth, Gehan's weight			
							Bootstrap		Perturbation				Bootstrap	
n	c_p	ρ	$\bar{m}$		Bias	SD	ASE	CP	ASE	CP	Bias	SD	ASE	CP
100	0.25	0.2	3.39	β_1	0.002	0.214	0.204	0.939	0.206	0.942	0.018	0.182	0.181	0.945
				β_2	-0.018	0.369	0.354	0.938	0.355	0.930	-0.006	0.313	0.314	0.946
		0.4	3.80	β_1	0.014	0.210	0.211	0.950	0.213	0.955	0.020	0.189	0.189	0.946
				β_2	0.003	0.372	0.367	0.939	0.369	0.940	0.011	0.329	0.329	0.952
	0.50	0.2	1.88	β_1	0.016	0.235	0.237	0.952	0.257	0.958	0.022	0.224	0.226	0.948
				β_2	0.039	0.405	0.412	0.955	0.421	0.951	0.055	0.381	0.389	0.954
		0.4	2.00	β_1	0.000	0.246	0.241	0.949	0.257	0.955	0.007	0.229	0.229	0.948
				β_2	-0.006	0.433	0.423	0.932	0.433	0.931	-0.005	0.398	0.397	0.949
200	0.25	0.2	3.40	β_1	0.013	0.152	0.145	0.942	0.146	0.944	0.018	0.127	0.126	0.948
				β_2	-0.012	0.262	0.252	0.944	0.253	0.945	-0.002	0.221	0.218	0.953
		0.4	3.80	β_1	0.006	0.150	0.150	0.949	0.150	0.947	0.007	0.135	0.132	0.949
				β_2	0.012	0.262	0.261	0.954	0.262	0.951	0.017	0.228	0.230	0.957
	0.50	0.2	1.88	β_1	0.016	0.158	0.165	0.957	0.168	0.958	0.016	0.149	0.156	0.965
				β_2	0.024	0.286	0.289	0.948	0.294	0.947	0.029	0.270	0.270	0.951
		0.4	1.99	β_1	0.005	0.175	0.170	0.931	0.173	0.934	0.008	0.160	0.160	0.944
				β_2	0.013	0.306	0.296	0.943	0.301	0.943	0.004	0.274	0.275	0.949

					Proposed, smooth						Naive			
							Bootstrap		ASV		Univariate		Clustered	
n	c_p	ρ	$\bar{m}$		Bias	SD	ASE	CP	ASE	CP	Bias	SD	Bias	SD
100	0.25	0.2	3.39	β_1	0.019	0.182	0.181	0.943	0.177	0.942	0.013	0.216	0.047	0.172
				β_2	-0.005	0.314	0.314	0.945	0.306	0.941	-0.005	0.377	0.035	0.287
		0.4	3.80	β_1	0.020	0.189	0.189	0.947	0.186	0.941	0.007	0.219	0.072	0.226
				β_2	0.012	0.329	0.329	0.953	0.320	0.947	-0.001	0.371	0.070	0.392
	0.50	0.2	1.88	β_1	0.024	0.225	0.226	0.947	0.219	0.936	0.027	0.240	0.089	0.226
				β_2	0.056	0.381	0.389	0.955	0.376	0.938	0.051	0.400	0.108	0.373
		0.4	2.00	β_1	0.009	0.229	0.229	0.948	0.223	0.942	0.014	0.241	0.130	0.275
				β_2	-0.003	0.398	0.397	0.948	0.384	0.941	-0.002	0.425	0.081	0.446
200	0.25	0.2	3.40	β_1	0.018	0.127	0.126	0.949	0.125	0.946	0.010	0.151	0.049	0.123
				β_2	-0.001	0.221	0.218	0.953	0.215	0.950	-0.007	0.264	0.037	0.211
		0.4	3.80	β_1	0.008	0.135	0.132	0.949	0.131	0.946	0.000	0.152	0.069	0.168
				β_2	0.018	0.227	0.230	0.958	0.228	0.955	0.014	0.256	0.065	0.288
	0.50	0.2	1.88	β_1	0.017	0.149	0.156	0.967	0.154	0.964	0.020	0.161	0.080	0.152
				β_2	0.030	0.270	0.270	0.951	0.265	0.944	0.025	0.284	0.074	0.262
		0.4	1.99	β_1	0.009	0.160	0.160	0.943	0.158	0.938	0.015	0.166	0.125	0.193
				β_2	0.005	0.274	0.275	0.949	0.271	0.948	0.007	0.291	0.087	0.323

Table 2.3: Results for simulated data with AR(1) correlation structure. The standard error for the non-smooth models, with the log-rank weight and Gehan's weight, is estimated by the bootstrap method or the perturbation method (Parzen et al.); the standard error for the proposed method is estimated by the bootstrap method and the asymptotic variance (ASV) estimator; n is the sample size; c_p is the percent of subjects without events; ρ is the correlation parameter in the AR(1) correlation structure; $\bar{m}$ is the average number of gap times, observed or censored per subject; Bias is the relative bias computed as the difference of the mean estimated parameter and the true value divided by the true value; SD is the Monte-Carlo standard deviation; ASE is the mean standard error; CP is the proportion of the 95% confidence intervals covering the true value.

					Non-smooth, log-rank weight						Non-smooth, Gehan's weight			
							Bootstrap		Perturbation				Bootstrap	
n	c_p	ρ	$\bar{m}$		Bias	SD	ASE	CP	ASE	CP	Bias	SD	ASE	CP
100	0.25	0.2	3.22	β_1	-0.028	0.213	0.209	0.943	0.211	0.944	-0.009	0.183	0.185	0.954
				β_2	-0.031	0.367	0.365	0.934	0.367	0.938	-0.038	0.319	0.324	0.933
		0.4	3.38	β_1	-0.029	0.207	0.209	0.943	0.211	0.945	-0.014	0.182	0.187	0.948
				β_2	-0.021	0.373	0.363	0.939	0.364	0.942	-0.010	0.325	0.326	0.942
	0.50	0.2	1.85	β_1	0.008	0.251	0.249	0.951	0.265	0.957	0.010	0.230	0.233	0.957
				β_2	-0.027	0.427	0.433	0.950	0.445	0.947	0.007	0.388	0.403	0.958
		0.4	1.92	β_1	0.012	0.261	0.247	0.929	0.265	0.940	0.015	0.241	0.232	0.934
				β_2	-0.021	0.430	0.433	0.948	0.445	0.952	0.006	0.394	0.402	0.945
200	0.25	0.2	3.23	β_1	-0.021	0.150	0.147	0.935	0.148	0.940	-0.010	0.128	0.128	0.941
				β_2	-0.024	0.259	0.257	0.936	0.257	0.942	-0.029	0.221	0.223	0.950
		0.4	3.38	β_1	-0.019	0.149	0.148	0.948	0.149	0.949	-0.012	0.126	0.131	0.947
				β_2	-0.030	0.263	0.256	0.938	0.257	0.934	-0.032	0.225	0.226	0.947
	0.50	0.2	1.85	β_1	-0.022	0.168	0.174	0.946	0.177	0.951	-0.012	0.155	0.161	0.963
				β_2	-0.027	0.299	0.303	0.952	0.308	0.950	0.000	0.273	0.278	0.950
		0.4	1.92	β_1	-0.018	0.177	0.173	0.941	0.176	0.944	-0.014	0.162	0.161	0.942
				β_2	-0.008	0.291	0.302	0.954	0.307	0.955	0.006	0.266	0.278	0.963

					Proposed, smooth						Naive			
							Bootstrap		ASV		Univariate		Clustered	
n	c_p	ρ	$\bar{m}$		Bias	SD	ASE	CP	ASE	CP	Bias	SD	Bias	SD
100	0.25	0.2	3.22	β_1	-0.009	0.183	0.185	0.955	0.181	0.950	0.007	0.218	-0.023	0.143
				β_2	-0.038	0.320	0.324	0.934	0.315	0.931	-0.021	0.385	-0.041	0.252
		0.4	3.38	β_1	-0.013	0.182	0.187	0.948	0.183	0.942	0.012	0.216	-0.030	0.163
				β_2	-0.009	0.325	0.325	0.940	0.317	0.933	-0.002	0.387	-0.034	0.282
	0.50	0.2	1.85	β_1	0.012	0.231	0.233	0.955	0.227	0.947	0.013	0.244	0.018	0.210
				β_2	0.007	0.388	0.402	0.958	0.388	0.951	0.004	0.415	0.017	0.336
		0.4	1.92	β_1	0.017	0.242	0.232	0.935	0.226	0.922	0.028	0.257	0.030	0.236
				β_2	0.007	0.395	0.402	0.943	0.389	0.935	0.003	0.423	0.021	0.366
200	0.25	0.2	3.23	β_1	-0.010	0.128	0.128	0.941	0.127	0.939	0.002	0.153	-0.025	0.101
				β_2	-0.028	0.221	0.223	0.950	0.220	0.946	-0.022	0.269	-0.038	0.174
		0.4	3.38	β_1	-0.012	0.126	0.131	0.946	0.129	0.946	0.012	0.151	-0.032	0.111
				β_2	-0.032	0.225	0.225	0.947	0.223	0.948	-0.012	0.268	-0.053	0.192
	0.50	0.2	1.85	β_1	-0.011	0.155	0.161	0.964	0.159	0.959	-0.009	0.167	-0.002	0.140
				β_2	0.001	0.273	0.278	0.950	0.274	0.946	0.002	0.293	0.020	0.244
		0.4	1.92	β_1	-0.013	0.162	0.161	0.943	0.159	0.939	-0.004	0.172	0.012	0.160
				β_2	0.006	0.266	0.278	0.964	0.275	0.957	0.009	0.285	0.027	0.256

shown in the lower-right panel of Table 2.1-Table 2.3. Whereas the point estimates of the univariate method are satisfactory, this method is obviously less efficient (i.e., larger SDs) than the proposed method. As expected, the point estimates of the clustered survival data method are biased, and the biases increase with the within-subject correlation, which demonstrates that naively applying methods for clustered survival data in the analysis of recurrent gap times can yield substantial bias.

2.5 Data analysis

We applied the proposed method to the hospitalization data from the Danish Psychiatric Central Register (Munk-Jørgensen and Mortensen, 1997) which computerized all admissions to psychiatric hospitals and psychiatric wards in general hospitals in Denmark since 1969. In this chapter, we only considered a subset of the published data, which was composed of a cohort of 286 individuals who were first admitted to or contacted with Danish psychiatric services between April 1 and December 31, 1970. The maximum follow-up time was set to be 3 years to avoid any potential change in the distributional pattern of recurrent gap times. The details about this cohort have been described elsewhere (Luo and Huang, 2011; Luo et al., 2013). Briefly, among the 286 subjects, 106 (37%) were females, 230 (80%) had schizophrenia onset after 20 years old, 115 (40%) were censored after the initial hospitalization or contact with no records of rehospitalization, 56 (20%) had one rehospitalization, and 115 (40%) had two or more rehospitalization records. The average number of rehospitalization was 1.7. The median disease onset age was 26 with a range of 14 to 88 years old. Note that 9 of the 286 patients died before the end of the follow-up time, hence, the independent censoring assumption was not expected to be seriously violated.

Our main interest was to estimate the effect of the disease onset age on the gap time between two successive hospitalizations. We fitted the AFT model to the data

Table 2.4: Regression results for schizophrenia data. SE is standard error estimate; CI is confidence interval.

	Non-smooth						Smooth		
	Log-rank weight			Gehan's weight					
	Estimate	SE	95% CI	Estimate	SE	95% CI	Estimate	SE	95% CI
Log(onset age)	1.444	0.298	(0.860, 2.028)	1.295	0.242	(0.822,1.769)	1.295	0.241	(0.822,1.768)
Gender	0.095	0.276	(-0.445, 0.636)	0.123	0.233	(-0.334,0.580)	0.125	0.235	(-0.335,0.585)

with two covariates, the logarithm-transformed onset age and gender. We applied both the proposed smooth method and non-smooth methods with log-rank or Gehan's weight. The variance for the non-smooth and smooth methods was estimated by the bootstrap and the asymptotic methods, respectively.

As shown in Table 2.4, the point estimates of the effects of log onset age and gender from the non-smooth and smooth estimating functions are similar, while the CIs from the proposed method and the non-smooth method with Gehan's weight are narrower than the non-smooth method with log-rank weight (Chang, 2004), similar to the findings from the simulation study. All methods show that the effect of onset age was significantly associated with gap times between recurrent hospitalization while gender did not have a significant effect, which is in line with the previous findings in literature (Luo and Huang, 2011; Luo et al., 2013).

2.6 Discussion

Despite its appealing direct interpretation, the AFT model (Chang, 2004) has not been widely used in recurrent event data analysis possibly due to the lack of reliable and efficient computing programs. In this chapter, we have introduced an induced smoothing technique to improve the performance of the rank-based AFT model for recurrent gap time data. With simulations and a real data analysis, we have shown that the proposed smooth estimating function method provides similar but more computational stable point and variance estimates as compared to the existing non-smooth estimating function method in Chang (2004). The proposed induced smoothing method also has been shown to be more computationally efficient than the non-smooth methods. Hence we recommend to use the proposed induced smoothing method with the asymptotic variance estimator for data analysis.

In this chapter, we adopted a Gehan-type weight for the induced smoothing method in order to achieve a more tractable objective function. However, the induced smoothing method is applicable to other weight functions such as the log-rank weight or a general weight function. Note that estimating functions with general weights may not be monotonic. In that case, by following similar techniques in Chiou et al. (2015a), one can use an iterative procedure and within each iteration, reweight a monotonic estimating function in the same form as (2.6) to approximate the estimating function with a general weight. We note that, like many correlated-data methods, the proposed induced smoothing method for the recurrent gap time AFT model is robust in the sense that its validity does not depend on the correct specification of the correlation structure. A possible future research direction is to improve the efficiency of estimation by incorporating the correlation structure in the estimating function, such as using the generalized method of moments estimation studied by Li and Yin (2009) for clustered survival data.

Chapter 3

Additive rates model for recurrent event data with intermittently observed time-dependent covariates

3.1 Introduction

Recurrent event data are frequently encountered in clinical and epidemiological studies, where each subject can experience events of interest repeatedly. Examples of recurrent events include infections after hematopoietic cell transplantations (Barker et al., 2005), repeated cardiovascular events (Kohli et al., 2013), and rehospitalizations of patients with psychiatric disorders (Eaton et al., 1992). In such studies, data on time-dependent covariates are often collected during the course of follow-up. Regression methods that can handle time-dependent covariates have been an important tool as investigators are often interested in evaluating the effect of variables that are evolving over time such as in studies of personalized medicine. Researchers may also be interested in utilizing updated information on risk factors during follow-up in dynamic prediction of event risk.

The motivating example of this research is an observational study conducted in

India between 2002 and 2004 which aimed to evaluate the effects of time-varying streptococcal infections, including group A, C, G streptococcus, on the risk of the recurrent pharyngitis (Jose et al., 2018). In this study, participants were examined weekly for the symptoms of pharyngitis and throat swabs were obtained to identify the status of streptococcal infections in symptomatic patients. In addition, monthly visits were scheduled to determine the carriage rate of streptococcal infections in this population. In the analysis, the infections of streptococcal groups were regarded as time-varying risk factors of the recurrent pharyngitis occurrence.

Regression methods for recurrent events are usually formulated based on either the conditional *intensity* or the marginal *rate* function of the counting process of recurrent events. Andersen and Gill (1982) and Prentice et al. (1981a) proposed a proportional intensity model, which postulates a multiplicative covariate effect on the intensity function of the underlying counting process, that is the instantaneous risk of recurrent event conditional on the event history and covariate history. Alternatively, Pepe and Cai (1993) and Lin et al. (2000) proposed proportional rates models which are based on the marginal rate function. Although the proportional intensity or rate models have gained great popularity in applications, they require the covariates to have multiplicative effects on the recurrent event risk. In applications, it is possible that the covariate effects add to, instead of multiplying, the baseline event risk. In this case, it would be more appropriate to use additive models such as the semiparametric additive rates model proposed by Schaubel et al. (2006) and the additive intensity model considered by Liu and Wu (2011). Moreover, the additive models can provide the risk difference estimates which may be desired by epidemiologists.

Although the aforementioned recurrent event models can naturally accommodate time-dependent covariates, their model estimation procedures require the values of time-dependent covariates to be continuously observed throughout the entire follow-up period for all subjects. In practice, however, the time-dependent covariates are often intermittently measured, rendering the existing model estimation procedures not readily applicable. A number of approaches to handle intermittently measured covariates have been discussed and reviewed by Andersen and Liestøl (2003). The first type of methods is a two-stage approach where the values of time-dependent covariates are estimated in the first stage and then the estimated covariate values are used in the regression model in the second stage. Simple methods for the first stage include carrying forward the last observed value or imputing the missing values between two observation times by linear interpolation. More complex methods such as parametric or non-parametric smoothing techniques (Raboud et al., 1993), random effects model (Dafni and Tsiatis, 1998), and stochastic models (Boscardin et al., 1998; Bycott and Taylor, 1998) have also been considered.

The second type of methods involves jointly modeling longitudinally measured covariates and event times. When the event time is univariate, various joint methods have been proposed including selection models, pattern mixture models, and shared parameter models. Readers are referred to Tsiatis and Davidian (2004) and Rizopoulos (2012) for comprehensive reviews. When the event time is recurrent, Henderson et al. (2000) modeled the covariate and recurrent event processes jointly via a latent bivariate Gaussian process, while Li (2016) considered a joint model of the recurrent event process and the longitudinal process for binary covariate specifically. Others considered joint models in the presence of a terminal event (Liu and Huang, 2009; Kim et al., 2012; Cai et al., 2017b). The estimation of the joint models could be computationally intensive, especially when the longitudinal covariates are multi-dimensional or include categorical variables. In addition, the validity of joint modeling relies on certain assumptions about the covariate model and the dependence structure of the repeatedly measured covariates, which may be difficult to verify. Misspecification of the model for longitudinal measurements will result in biased estimation of the event time model.

Recently, Cao et al. (2015) and Li et al. (2016) proposed kernel-weighted estimation procedures for the proportional hazards/rates models with time-dependent covariates. Specifically, Cao et al. (2015) considered the case where data on covariates are not collected at failure times and proposed to smooth the partial likelihood to derive a consistent estimator with a convergence rate slower than root-n. Li et al. (2016) focused on the setting of recurrent event data where, in addition to regular follow-up visits, covariate values are usually collected when an event occurs. As pointed out by the authors, measurements at event visits give a biased representation of the underlying covariate process of an individual. Hence, in the construction of kernel-smoothed pseudo-partial score functions, only covariate values measured at regular visits, whose timing is noninformative of the underlying recurrent event risk, are used to estimate the expected covariate value of individuals in a risk set. The estimated score function gives a consistent estimator with a root-n convergence rate.

In this chapter, we propose a semiparametric estimator for the additive rates model with intermittently observed time-dependent covariates. Specifically, we kernel smooth functions of time-dependent covariates across subjects instead of smoothing individual covariate trajectories. Our proposed method is demonstrated to have better performance than simple covariate imputation methods such as the last covariate carried forward (LCCF) method through simulation studies. We also discuss a few practical issues including the situation when both time-dependent and time-independent covariates are present and the case when different time-dependent covariates are measured at different times.

The remainder of this chapter is organized as follows. In Section 3.2, we first review the additive rates model and the estimation procedure (Schaubel et al., 2006) in the ideal case where time-dependent covariates are monitored continuously, then we present the proposed kernel smoothed estimator for the case where covariates are time-dependent and intermittently observed. Some extensions of the proposed method are discussed in Section 3.3. Section 3.4 compares the performance of the proposed estimator to the two simple approaches including the LCCF and linear interpolation methods with simulation studies. In Section 3.5, we present a real data analysis using the Indian pharyngitis data. Some concluding remarks are included in Section 3.6.

3.2 Model and the proposed estimator

Let $i = 1, \ldots, n$ index the n subjects in a study. Let $N_i^*(t)$ denote the number of events that subject i has experienced at or prior to time t in the absence of censoring. Denote by $\mathbf{Z}_i(t) = (Z_{i1}(t), \ldots, Z_{ip}(t))^\intercal$ a $p \times 1$ vector of possibly time-dependent covariates. The semiparametric additive rates model assumes that the rate function of $N_i^*(t)$ conditioning on the covariates at time t is

$$\lambda\{t \mid \mathbf{Z}_i(t)\} = \lambda_0(t) + \boldsymbol{\beta}^\intercal \mathbf{Z}_i(t),$$

where $\lambda_0(t)$ is an unspecified baseline rate function and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\intercal$ is a $p \times 1$ vector of regression parameters, whose jth component β_j is interpreted as the rate difference associated with one unit difference in $Z_{ij}(t)$. Let C_i denote the follow-up time for subject i and define $Y_i(t) = I(C_i \geq t)$. Let $N_i(t) = N_i^*(t \wedge C_i)$, where $t \wedge C_i = \min(t, C_i)$, be the number of observed events up to C_i. Let τ be a pre-specified time point such that the recurrent event process could potentially be observed beyond τ with a non-zero probability. The observed data $\{N_i(\cdot), \mathbf{Z}_i(\cdot), Y_i(\cdot)\}, i = 1, \ldots, n$, are assumed to be independent and identically distributed.

For model estimation, following Lin and Ying (1994), Schaubel et al. (2006) considered the estimating function

$$
\begin{aligned}
U(\boldsymbol{\beta}) =& \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\{\mathbf{Z}_i(t)-\bar{\mathbf{Z}}(t)\}\{dN_i(t)-Y_i(t)\boldsymbol{\beta}^{\intercal}\mathbf{Z}_i(t)dt\} \\
=& \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\{\mathbf{Z}_i(t)-\bar{\mathbf{Z}}(t)\}dN_i(t)-\left[\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\left\{\mathbf{Z}_i(t)-\bar{\mathbf{Z}}(t)\right\}^{\otimes 2}Y_i(t)dt\right]\boldsymbol{\beta},
\end{aligned}
\tag{3.1}
$$

where $\bar{\mathbf{Z}}(t)=\frac{\frac{1}{n}\sum_{i=1}^{n}Y_i(t)\mathbf{Z}_i(t)}{\frac{1}{n}\sum_{i=1}^{n}Y_i(t)}$, $\mathbf{Z}^{\otimes 0}=1$, $\mathbf{Z}^{\otimes 1}=z$, $\mathbf{Z}^{\otimes 2}=\mathbf{Z}\mathbf{Z}^{\intercal}$. Solving $U(\boldsymbol{\beta})=0$ gives a simple closed-form solution

$$
\widehat{\boldsymbol{\beta}}=\left[\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\{\mathbf{Z}_i(t)-\bar{\mathbf{Z}}(t)\}^{\otimes 2}Y_i(t)dt\right]^{-1}\left[\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\{\mathbf{Z}_i(t)-\bar{\mathbf{Z}}(t)\}dN_i(t)\right]. \tag{3.2}
$$

It is easy to see that the evaluation of the estimator in (3.2), in particular the denominator, requires the time-dependent covariates to be continuously observed throughout the follow-up period. In practice, however, time-dependent covariates are often only observed intermittently. For example, in the pharyngitis data that motivates this research, the bacterial infection status of patients was only identified monthly. In such case, the estimator in (3.2) is not evaluable with the observed data.

A simple method for handling intermittently observed time-dependent covariates is to impute unobserved values using the LCCF approach. Under LCCF, the last known value of the covariate of a subject is used forward in time until a new value is measured or the observation of this subject is censored. This method has been shown to yield biased estimation under the proportional rates model (Li et al., 2016). Another simple approach is to use linear interpolation to estimate covariate values between two observations within each subject. Both simple methods are expected to result in biased estimations, especially when the covariates are binary, under the additive rates model. Instead of imputing missing values in the individual covariate trajectories, we propose a method focusing on smoothing the estimating function using the observed covariate information across subjects.

We first consider the simple case where all covariates in the model are time-dependent and observed at the same regular visits. More general cases such as when both time-dependent and -independent covariates are present in the model or when multiple time-dependent covariates are measured at different regular visits are discussed in Section 3.3. Let $O(\cdot)$ denote the counting process for the regular visits, where regular visits are referred to as pre-scheduled follow-up visits, and when a regular visit occurs, $O(\cdot)$ jumps by 1. In the Indian pharyngitis study, $O(t)$ is a function with unit steps at the monthly visits. Since the participants may be sick at a regular visit, we allow $O(\cdot)$ and $N(\cdot)$ to jump at the same time. We assume that the process $O(\cdot)$ is independent of $Z(\cdot)$ and C. The rate function of $O(\cdot)$ is denoted by $m(t)$, that is, $\mathrm{E}\{dO(t)\} = m(t)dt$.

Let $S^{(k)}(t) = n^{-1}\sum_{i=1}^{n} Y_i(t)\mathbf{Z}_i(t)^{\otimes k}, k = 0, 1, 2$. It is easy to show that the estimating function in (3.1) can be re-expressed as

$$\begin{aligned} U(\boldsymbol{\beta}) =& \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\mathbf{Z}_i(t)dN_i(t) - \int_0^{\tau}\frac{S^{(1)}(t)}{S^{(0)}(t)}\left\{\frac{1}{n}\sum_{i=1}^{n}dN_i(t)\right\} \\ & - \left(\int_0^{\tau}\left[\frac{1}{n}\sum_{i=1}^{n}Y_i(t)\frac{S^{(2)}(t)}{S^{(0)}(t)} - \frac{1}{n}\sum_{i=1}^{n}Y_i(t)\left\{\frac{S^{(1)}(t)}{S^{(0)}(t)}\right\}^{\otimes 2}\right]dt\right)\boldsymbol{\beta}. \end{aligned} \tag{3.3}$$

Thus, $\widehat{\boldsymbol{\beta}}$ can be expressed as

$$\begin{aligned} \widehat{\boldsymbol{\beta}} =& \left(\int_0^{\tau}\left[\frac{1}{n}\sum_{i=1}^{n}Y_i(t)\frac{S^{(2)}(t)}{S^{(0)}(t)} - \frac{1}{n}\sum_{i=1}^{n}Y_i(t)\left\{\frac{S^{(1)}(t)}{S^{(0)}(t)}\right\}^{\otimes 2}\right]dt\right)^{-1} \\ & \left[\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\mathbf{Z}_i(t)dN_i(t) - \int_0^{\tau}\frac{S^{(1)}(t)}{S^{(0)}(t)}\left\{\frac{1}{n}\sum_{i=1}^{n}dN_i(t)\right\}\right]. \end{aligned} \tag{3.4}$$

As can be seen from (3.4), the estimator $\widehat{\boldsymbol{\beta}}$ is a functional of the empirical processes

$n^{-1}\sum_{i=1}^{n}\mathbf{Z}_i(t)dN_i(t)$, $n^{-1}\sum_{i=1}^{n}dN_i(t)$, and $S^{(k)}(t), k=0,1,2$. We assume that the covariates of subject i, $\mathbf{Z}_i(t)$, are observed at this subject's event times (that is, where $N_i(t)$ jumps), which is typically satisfied in recurrent event data, such as in the Indian pharyngitis data example. Hence, the empirical processes $n^{-1}\sum_{i=1}^{n}\mathbf{Z}_i(t)dN_i(t)$ and $n^{-1}\sum_{i=1}^{n}dN_i(t)$ can be computed based on the observed data. However, we note that the processes $S^{(k)}(t)/S^{(0)}(t), k=1,2$, cannot be evaluated when the time-dependent covariates are not continuously observed.

In what follows, we show how to approximate the ratios $S^{(k)}(t)/S^{(0)}(t), k=1,2$, using intermittently observed time-dependent covariate data. Let $s^{(k)}(t)$ denote the expectation of $S^{(k)}(t)$: $s^{(k)}(t)\equiv \mathrm{E}\{S^{(k)}(t)\}=\mathrm{E}\{Y_i(t)\mathbf{Z}_i(t)^{\otimes k}\}$. We aim to find a consistent estimator of $s^{(k)}(t)/s^{(0)}(t)$ to approximate $S^{(k)}(t)/S^{(0)}(t)$. We propose to apply the kernel smoothing method to estimate $s^{(k)}(t)/s^{(0)}(t)$ as follows. Define the kernel smoothed process

$$\widehat{S}_h^{(k)}(t)=n^{-1}\sum_{i=1}^{n}\int_0^{\tau}K_h(t-u)Y_i(u)\mathbf{Z}_i(u)^{\otimes k}dO_i(u), t\in[h,\tau-h],$$

for $k=0,1,2$, where $K_h(t)=K(t/h)/h$, h is the bandwidth, $0<h<\tau/2$, and $K(t)$ is a second order kernel function with support $[-1,1]$. In order to avoid the bias in the boundary region, we let $\widehat{S}_h^{(k)}(t)=\widehat{S}_h^{(k)}(h)$ for $t\in[0,h)$, $\widehat{S}_h^{(k)}(t)=\widehat{S}_h^{(k)}(\tau-h)$ for $t\in(\tau-h,\tau]$. One can prove that $\widehat{S}_h^{(k)}(t)$ converges in probability to the limit $s^{(k)}(t)\cdot m(t)$, where $m(t)=\mathrm{E}\{dO_i(t)\}/dt$ is the rate function of the observation process and usually considered as a nuisance. Therefore, by kernel smoothing all $S^{(k)}(t)$, including $S^{(0)}(t)$, we can construct $\widehat{\xi}_h^{(k)}(t)\equiv\frac{\widehat{S}_h^{(k)}(t)}{\widehat{S}_h^{(0)}(t)}$, which converges in probability to $\frac{s^{(k)}(t)\cdot m(t)}{s^{(0)}(t)\cdot m(t)}=\frac{s^{(k)}(t)}{s^{(0)}(t)}$ as $n\to\infty$. Note that although $S^{(0)}(t)$ can be calculated directly from the observed data, we apply the same kernel smoothing technique on it as for $S^{(1)}(t)$ and $S^{(2)}(t)$ to circumvent the estimation of the nuisance $m(t)$. We also note that, in the construction of the kernel smoothed estimator $\widehat{\xi}_h^{(k)}(t)$, we only utilize the covariate values measured at regular visits (through $O_i(t)$). The covariate values measured at the event times (i.e., when $dN_i(t) = 1$) are only used in the evaluation of $n^{-1}\sum_{i=1}^n \mathbf{Z}_i(t)dN_i(t)$ in the estimating function (3.3).

Now, we replace $S^{(k)}(t)/S^{(0)}(t)$ with $\widehat{\xi}_h^{(k)}(t)$ in Equation (3.3) to obtain the following kernel estimating function:

$$\begin{aligned}\widehat{U}_h(\boldsymbol{\beta}) =& \frac{1}{n}\sum_{i=1}^n \int_0^\tau \mathbf{Z}_i(t)dN_i(t) - \int_0^\tau \widehat{\xi}_h^{(1)}(t)\left\{\frac{1}{n}\sum_{i=1}^n dN_i(t)\right\} \\ &- \left[\int_0^\tau \left\{\frac{1}{n}\sum_{i=1}^n Y_i(t)\widehat{\xi}_h^{(2)}(t) - \frac{1}{n}\sum_{i=1}^n Y_i(t)\widehat{\xi}_h^{(1)}(t)^{\otimes 2}\right\} dt\right]\boldsymbol{\beta}.\end{aligned} \tag{3.5}$$

Solving $\widehat{U}_h(\boldsymbol{\beta}) = 0$ leads to the proposed estimator of $\boldsymbol{\beta}$:

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_h = &\left[\int_0^\tau \left\{\frac{1}{n}\sum_{i=1}^n Y_i(t)\widehat{\xi}_h^{(2)}(t) - \frac{1}{n}\sum_{i=1}^n Y_i(t)\widehat{\xi}_h^{(1)}(t)^{\otimes 2}\right\} dt\right]^{-1} \\ &\left[\frac{1}{n}\sum_{i=1}^n \int_0^\tau \mathbf{Z}_i(t)dN_i(t) - \int_0^\tau \widehat{\xi}_h^{(1)}(t)\left\{\frac{1}{n}\sum_{i=1}^n dN_i(t)\right\}\right].\end{aligned}$$

We summarize the large sample property of $\widehat{\boldsymbol{\beta}}_h$ in the following theorem. The detailed proof is given in Appendix B.

Theorem 2. Let $\boldsymbol{\beta}_0$ denote the true parameter. Under regularity conditions 1-9 in Appendix B, as $n \to \infty$, $\sqrt{n}(\widehat{\boldsymbol{\beta}}_h - \boldsymbol{\beta}_0)$ converges in distribution to a normal random variable with zero mean and variance Σ, where Σ is defined in Appendix B, on the condition that $h = O(n^{-v})$, $1/4 < v < 1/2$.

For the estimation of the baseline mean function $\mu_0(t) = \int_0^t \lambda_0(u)du$, following Schaubel et al. (2006), we have

$$\widehat{\mu}_0(t, \boldsymbol{\beta}) = \int_0^t \frac{\sum_{i=1}^n Y_i(u)\{dN_i(u) - \boldsymbol{\beta}^\intercal \mathbf{Z}_i(u)du\}}{\sum_{i=1}^n Y_i(u)}$$

$$= \sum_{i=1}^{n} \int_0^t \frac{1}{Y.(t)} Y_i(u) dN_i(u) - \int_0^t \frac{S^{(1)}(u)}{S^{(0)}(u)} du \boldsymbol{\beta},$$

where $Y.(t) = \sum_{i=1}^n Y_i(t)$. As discussed before, $\frac{S^{(1)}(t)}{S^{(0)}(t)}$ can not be evaluated directly with observed data. We consider the following estimator

$$\widehat{\mu}_{0,h}(t, \widehat{\boldsymbol{\beta}}_h) = \sum_{i=1}^{n} \int_0^t \frac{1}{Y.(t)} Y_i(u) dN_i(u) - \int_0^t \widehat{\xi}_h^{(1)}(u) du \widehat{\boldsymbol{\beta}}_h.$$

Note that $\widehat{\mu}_{0,h}(t, \widehat{\boldsymbol{\beta}}_h)$ may not give a nondecreasing function because the increment could be negative, especially for the time interval without observed recurrent events. To ensure monotonicity, we propose to estimate the baseline mean function by $\widetilde{\mu}_{0,h}(t, \widehat{\boldsymbol{\beta}}_h)$ $= \max_{0 \le u \le t} \widehat{\mu}_{0,h}(u, \widehat{\boldsymbol{\beta}}_h)$.

3.3 Extensions of the proposed estimator

3.3.1 Estimator for binary covariates with no time trend

In this subsection, we consider a special case when $\mathbf{Z}_i(t)$ is composed of only binary covariates with no time trend at the population level. We show that the regression coefficient estimator has a simpler form than the proposed estimator $\widehat{\boldsymbol{\beta}}_h$ for the general case. First, we assume that the binary covariate $\mathbf{Z}_i(t)$ has no time trend at the population level in the sense that $\mathrm{E}\{\mathbf{Z}_i(t)\} = \boldsymbol{\mu}_1$ and $\mathrm{E}\{\mathbf{Z}_i(t)^{\otimes 2}\} = \boldsymbol{\mu}_2$, where $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are constants. Note that when $Z_i(t)$ is scalar valued, $\mathrm{E}\{Z_i(t)\} = \mathrm{E}\{Z_i(t)^2\} = \mu_1$. We note that, at individual level, $\mathbf{Z}_i(t)$ is allowed to vary over time. In addition, we assume that the censoring time C_i is independent of the covariates. Let $R(t) = \mathrm{E}\{Y_i(t)\}$, then we can prove that

$$\bar{\mathbf{Z}}(t) = \frac{n^{-1} \sum_{i=1}^n Y_i(t) \mathbf{Z}_i(t)}{n^{-1} \sum_{i=1}^n Y_i(t)} \to \frac{R(t)\boldsymbol{\mu}_1}{R(t)} = \boldsymbol{\mu}_1, \tag{3.6}$$

in probability as $n \to \infty$. The estimating function in (3.1) can be written as

$$\begin{aligned}U(\boldsymbol{\beta}) =&\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\{\mathbf{Z}_i(t)-\bar{\mathbf{Z}}(t)\}\{dN_i(t)-Y_i(t)\boldsymbol{\beta}^{\intercal}\mathbf{Z}_i(t)dt\}\\ =&\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\mathbf{Z}_i(t)dN_i(t)-\left\{\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}Y_i(t)\mathbf{Z}_i(t)^{\otimes 2}dt\right\}\boldsymbol{\beta}\\ &-\int_0^{\tau}\bar{\mathbf{Z}}(t)\frac{1}{n}\sum_{i=1}^{n}dN_i(t)+\int_0^{\tau}\bar{\mathbf{Z}}(t)\frac{1}{n}\sum_{i=1}^{n}Y_i(t)\boldsymbol{\beta}^{\intercal}\mathbf{Z}_i(t)dt.\end{aligned}$$

Then using the result in (3.6), one can prove that $U(\boldsymbol{\beta}) = U_s(\boldsymbol{\beta}) + o_p(1)$, where

$$\begin{aligned}U_s(\boldsymbol{\beta}) =&\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\mathbf{Z}_i(t)dN_i(t)-\left\{\boldsymbol{\mu}_2\int_0^{\tau}R(t)dt\right\}\boldsymbol{\beta}-\boldsymbol{\mu}_1\int_0^{\tau}\frac{1}{n}\sum_{i=1}^{n}dN_i(t)\\ &+\left\{\boldsymbol{\mu}_1^{\otimes 2}\int_0^{\tau}R(t)dt\right\}\boldsymbol{\beta}.\end{aligned}$$

The solution of $U_s(\boldsymbol{\beta}) = 0$ can then be expressed as

$$\boldsymbol{\beta}_s = \left\{\int_0^{\tau}R(t)dt\right\}^{-1}\{\boldsymbol{\mu}_2-\boldsymbol{\mu}_1^{\otimes 2}\}^{-1}\left\{\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\mathbf{Z}_i(t)dN_i(t)-\boldsymbol{\mu}_1\int_0^{\tau}\frac{1}{n}\sum_{i=1}^{n}dN_i(t)\right\}.$$

Therefore, an estimator of $\boldsymbol{\beta}_s$, denoted by $\widehat{\boldsymbol{\beta}}_s$, can be obtained by replacing $\int_0^{\tau}R(t)dt$ with $n^{-1}\sum_{i=1}^{n}\min(C_i,\tau)$, and replacing $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ with $\widehat{\boldsymbol{\mu}}_1 = \sum_{i=1}^{n}\sum_{j=1}^{M_i}\mathbf{Z}_{ij}/\sum_{i=1}^{n}M_i$ and $\widehat{\boldsymbol{\mu}}_2 = \sum_{i=1}^{n}\sum_{j=1}^{M_i}\mathbf{Z}_{ij}^{\otimes 2}/\sum_{i=1}^{n}M_i$, respectively, where M_i denotes the number of regular visits of subject i including the baseline and $\mathbf{Z}_{ij}$ denotes the covariates of subject i measured at the j^{th} regular visit.

Compared to the proposed estimator $\widehat{\boldsymbol{\beta}}_h$ for the general case, the estimator $\widehat{\boldsymbol{\beta}}_s$ for the special case of binary covariates with no time trend has a few advantages. First, $\widehat{\boldsymbol{\beta}}_s$ is expected to be more computationally efficient and easier to implement because it does not involve kernel smoothing or integration, and all terms in $\widehat{\boldsymbol{\beta}}_s$ can be expressed as finite summations. Second, when only baseline covariate values, $\mathbf{Z}_i(0)$, are available, $\widehat{\boldsymbol{\beta}}_s$ can still be applied by using $\widehat{\boldsymbol{\mu}}_1 = n^{-1}\sum_{i=1}^n \mathbf{Z}_i(0)$ and $\widehat{\boldsymbol{\mu}}_2 = n^{-1}\sum_{i=1}^n \mathbf{Z}_i(0)^{\otimes 2}$.

3.3.2 Estimation when both time-dependent and -independent covariates are present

Thus far, our discussions focus on the estimation of models with time-dependent covariates only. In practice, however, it is common to collect data on both time-dependent and time-independent covariates. One may be interested in the effect of a time-dependent covariate adjusting for baseline variables or vice versa, e.g., the effect of a time-varying biomarker adjusting for sex or the effect of a randomized treatment adjusting for a time-varying adjuvant treatment. Note that the proposed method in Section 3.2 can be applied to the scenario where both time-dependent and -independent covariates are present. However, instead of kernel smoothing, it is more natural to estimate the mean covariate processes that only involve time-independent covariates with their simple empirical averages. In this subsection, we present a more appropriate method to deal with the two types of covariates.

Let $\mathbf{Z}_i(t) = (Z_{i1}(t), \ldots, Z_{ip}(t))^\intercal$ denote the vector of time-dependent covariates and $\mathbf{W}_i = (W_{i1}, \ldots, W_{iq})^\intercal$ the vector of time-independent covariates. Then the additive rates model can be expressed as

$$\lambda\{t|\mathbf{Z}_i(t), \mathbf{W}_i\} = \lambda_0(t) + \boldsymbol{\beta}^\intercal \mathbf{Z}_i(t) + \boldsymbol{\gamma}^\intercal \mathbf{W}_i,$$

where $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are $p \times 1$ and $q \times 1$ vectors of parameters for the time-dependent covariates and the time-independent covariates, respectively.

The estimating function for $(\boldsymbol{\beta},\boldsymbol{\gamma})$ is

$$\begin{aligned}U(\boldsymbol{\beta},\boldsymbol{\gamma}) =& \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}(\mathbf{Z}_i(t),\mathbf{W}_i)^{\intercal}dN_i(t) - \int_0^{\tau}(\bar{\mathbf{Z}}(t),\bar{\mathbf{W}}(t))^{\intercal}\left\{\frac{1}{n}\sum_{i=1}^{n}dN_i(t)\right\}\\ &-\left[\int_0^{\tau}\frac{1}{n}\sum_{i=1}^{n}Y_i(t)\left\{\frac{\frac{1}{n}\sum_{i=1}^{n}Y_i(t)(\mathbf{Z}_i(t),\mathbf{W}_i)^{\intercal\otimes 2}}{S^{(0)}(t)}\right.\right.\\ &\left.\left.-\left(\bar{\mathbf{Z}}(t),\bar{\mathbf{W}}(t)\right)^{\intercal\otimes 2}\right\}dt\right](\boldsymbol{\beta},\boldsymbol{\gamma})^{\intercal},\end{aligned}$$

where $\bar{\mathbf{W}}(t) = \left\{\frac{1}{n}\sum_{i=1}^{n}Y_i(t)\mathbf{W}_i\right\}\left\{\frac{1}{n}\sum_{i=1}^{n}Y_i(t)\right\}^{-1}$. We define $S_z^{(k)}(t) = \frac{1}{n}\sum_{i=1}^{n}Y_i(t)$ $\mathbf{Z}_i(t)^{\otimes k}$, $S_w^{(k)}(t) = \frac{1}{n}\sum_{i=1}^{n}Y_i(t)\mathbf{W}_i^{\otimes k}$ for $k=1,2$, and $S^{(2)}(t) = \begin{pmatrix} S_z^{(2)}(t) & S_{zw}^{(2)}(t)\\ S_{wz}^{(2)}(t) & S_w^{(2)}(t)\end{pmatrix}$, where $S_{zw}^{(2)}(t) = \frac{1}{n}\sum_{i=1}^{n}Y_i(t)\mathbf{Z}_i(t)\mathbf{W}_i^{\intercal}$, and $S_{wz}^{(2)}(t) = \frac{1}{n}\sum_{i=1}^{n}Y_i(t)\mathbf{W}_i\mathbf{Z}_i(t)^{\intercal}$. The estimating function can further be written as

$$\begin{aligned}U(\boldsymbol{\beta},\boldsymbol{\gamma}) =& \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}(\mathbf{Z}_i(t),\mathbf{W}_i)^{\intercal}dN_i(t) - \int_0^{\tau}\left(\frac{S_z^{(1)}(t)}{S^{(0)}(t)},\bar{\mathbf{W}}(t)\right)^{\intercal}\left\{\frac{1}{n}\sum_{i=1}^{n}dN_i(t)\right\}\\ &-\left[\int_0^{\tau}\frac{1}{n}\sum_{i=1}^{n}Y_i(t)\left\{\frac{S^{(2)}(t)}{S^{(0)}(t)} - \left(\frac{S_z^{(1)}(t)}{S^{(0)}(t)},\bar{\mathbf{W}}(t)\right)^{\intercal\otimes 2}\right\}dt\right](\boldsymbol{\beta},\boldsymbol{\gamma})^{\intercal}.\end{aligned} \tag{3.7}$$

Note that when the time-dependent covariates $\mathbf{Z}_i(t)$ are observed intermittently, a few quantities in (3.7) are not evaluable: $S_z^{(k)}(t)/S^{(0)}(t), k=1,2$, $S_{wz}^{(2)}(t)/S^{(0)}$ and $S_{zw}^{(2)}(t)/S^{(0)}$, whereas the values of $S_w^{(k)}(t), k=1,2$, and $\bar{\mathbf{W}}(t)$ are known for all time t since $\mathbf{W}_i$ are time-independent. We can use the kernel smoothed processes $\widehat{S}_h^{(k)}(t)$ defined in Section 3.2 to replace $S_z^{(k)}(t)$ for $k=1,2$ and the same $\widehat{S}_h^{(0)}(t)$ to replace $S^{(0)}(t)$. Further, we propose the kernel smoothed processes $\widehat{S}_{zw,h}^{(2)}(t) = n^{-1}\sum_{i=1}^{n}\int_0^{\tau}K_h(t-u)Y_i(u)\mathbf{Z}_i(u)\mathbf{W}_i^{\intercal}dO_i(u)$ and $\widehat{S}_{wz,h}^{(2)}(t) = n^{-1}\sum_{i=1}^{n}\int_0^{\tau}K_h(t-u)Y_i(u)$ $\mathbf{W}_i\mathbf{Z}_i(u)^{\intercal}dO_i(u)$, $t\in[h,\tau-h]$. Similar boundary corrections as described in Section 3.2 are applied. Then, we propose the estimating function, $\widehat{U}_h(\boldsymbol{\beta}, \boldsymbol{\gamma})$, by replacing the non-observable quantities in (3.7) specified above with their kernel smoothed counterparts. A simulation study showed that the proposed method in this subsection is more efficient in the estimation of the regression coefficients of the time-independent covariates than the method forcing the time-independent covariates in the kernel smoothed functions (results not shown).

3.3.3 Estimation when multiple time-dependent covariates are measured on different schedules

In the previous sections, we assume that multiple time-dependent covariates are observed simultaneously at each visit (i.e., synchronized). In practice, it is possible that the time-dependent covariates are measured on different schedules. For example, in social behavioral studies, in order to prevent survey fatigue and maintain a high retention rate, different surveys may be delivered at different visits. In this subsection, we discuss how to extend the proposed method to accommodate multiple time-dependent covariates measured on different time schedules. For ease of discussion, we assume that there are two time-dependent covariates ($p = 2$), $Z_{i1}(t)$ and $Z_{i2}(t)$. The proposed method can be easily extended to the case where $p > 2$.

Let $O_{i1}(t)$ and $O_{i2}(t)$ denote the bivariate observation process that counts the cumulative number of measurements of $Z_{i1}(t)$ and $Z_{i2}(t)$, respectively. We assume that $\{O_1(\cdot), O_2(\cdot)\}$ is independent of $\{\mathbf{Z}(\cdot), C\}$, $\mathrm{E}\{dO_{i1}(u)dO_{i2}(w)\} = m_{12}(u,w)dudw$, $\mathrm{E}\{dO_{ig}(u)\} = m_g(u)du$ for $g = 1, 2$. For $k = 0, 1, 2$, define $\widehat{S}^{(k)}_{g,h}(t) = n^{-1}\sum_{i=1}^{n}\int_0^\tau K_h(t-u)Y_i(u)Z_{ig}(u)^k dO_{ig}(u)$, which consistently estimate $\mathrm{E}\{Y(t)Z_g^k(t)\}m_g(t)$. It is easy to see that $S^{(1)}(t)/S^{(0)}(t)$ in Equation (3.4) can be replaced by $\widetilde{S}^{(1)}(t)/\widetilde{S}^{(0)}(t) = \left(\widehat{S}^{(1)}_{1,h}(t)/\widehat{S}^{(0)}_{1,h}(t), \widehat{S}^{(1)}_{2,h}(t)/\widehat{S}^{(0)}_{2,h}(t)\right)^{\intercal}$. Moreover, the matrix $S^{(2)}(t)/S^{(0)}(t)$ in Equation

(3.4) is

$$\begin{pmatrix} \sum_{i=1}^n Y_i(t)Z_{i1}(t)^2/\sum_{i=1}^n Y_i(t) & \sum_{i=1}^n Y_i(t)Z_{i1}(t)Z_{i2}(t)/\sum_{i=1}^n Y_i(t) \\ \sum_{i=1}^n Y_i(t)Z_{i1}(t)Z_{i2}(t)/\sum_{i=1}^n Y_i(t) & \sum_{i=1}^n Y_i(t)Z_{i2}(t)^2/\sum_{i=1}^n Y_i(t) \end{pmatrix}.$$

As before, for $g = 1, 2$, the diagonal entries $\sum_{i=1}^n Y_i(t)Z_{ig}(t)^2/\sum_{i=1}^n Y_i(t)$ can be replaced by the kernel type estimators $\widehat{S}_{g,h}^{(2)}(t)/\widehat{S}_{g,h}^{(0)}(t)$. The off-diagonal entries involve both $Z_1(t)$ and $Z_2(t)$, thus we consider the following bivariate kernel type estimator $\widehat{S}_{12,h}^{(2)}(t)/\widehat{S}_{12,h}^{(0)}(t)$, where

$$\widehat{S}_{12,h}^{(2)}(t) = n^{-1}\sum_{i=1}^n \int_0^\tau \int_0^\tau K_h(t-u)K_h(t-w)Y_i(u \vee w)Z_{i1}(u)Z_{i2}(w)dO_{i1}(u)dO_{i2}(w),$$

and

$$\widehat{S}_{12,h}^{(0)}(t) = n^{-1}\sum_{i=1}^n \int_0^\tau \int_0^\tau K_h(t-u)K_h(t-w)Y_i(u \vee w)dO_{i1}(u)dO_{i2}(w).$$

We note that $\widehat{S}_{12,h}^{(2)}(t)$ consistently estimates $\mathrm{E}\{Y(t)Z_1(t)Z_2(t)\}m_{12}(t,t)$ and $\widehat{S}_{12,h}^{(0)}(t)$ consistently estimates $\mathrm{E}\{Y(t)\}m_{12}(t,t)$. Thus the off-diagonal entries can be replaced by $\widehat{S}_{12,h}^{(2)}(t)/\widehat{S}_{12,h}^{(0)}(t)$, which consistently estimates the population level quantity $\mathrm{E}\{Y(t)Z_1(t)Z_2(t)\}/\mathrm{E}\{Y(t)\}$. To sum up, define $\widetilde{S}^{(2)}(t)/\widetilde{S}^{(0)}(t) = \begin{pmatrix} \widehat{S}_{1,h}^{(2)}(t)/\widehat{S}_{1,h}^{(0)}(t) & \widehat{S}_{12,h}^{(2)}(t)/\widehat{S}_{12,h}^{(0)}(t) \\ \widehat{S}_{12,h}^{(2)}(t)/\widehat{S}_{12,h}^{(0)}(t) & \widehat{S}_{2,h}^{(2)}(t)/\widehat{S}_{2,h}^{(0)}(t) \end{pmatrix}$, then $\boldsymbol{\beta}$ can be consistently estimated by

$$\widehat{\boldsymbol{\beta}}_h = \left(\int_0^\tau \left[\frac{1}{n}\sum_{i=1}^n Y_i(t)\frac{\widetilde{S}^{(2)}(t)}{\widetilde{S}^{(0)}(t)} - \frac{1}{n}\sum_{i=1}^n Y_i(t)\left\{\frac{\widetilde{S}^{(1)}(t)}{\widetilde{S}^{(0)}(t)}\right\}^{\otimes 2}\right]dt\right)^{-1}$$
$$\left[\frac{1}{n}\sum_{i=1}^n \int_0^\tau \mathbf{Z}_i(t)dN_i(t) - \int_0^\tau \frac{\widetilde{S}^{(1)}(t)}{\widetilde{S}^{(0)}(t)}\left\{\frac{1}{n}\sum_{i=1}^n dN_i(t)\right\}\right].$$

3.4 Simulation

We conducted simulation studies to evaluate the performance of the proposed method. Under each simulation scenario, we generated 1000 data replicates with sample size 300 and 600. The resampling size was set to be 100 in the bootstrap method for variance estimation. The recurrent events were generated based on the following additive intensity model where the intensity of the recurrent event process for subject i is

$$\lambda\{t|Z_i(t), \gamma_i\} = \lambda_0(t) + \beta Z_i(t) + \gamma_i. \tag{3.8}$$

The frailty variable γ_i was generated from a gamma distribution with mean 0.02 and variance 0.004. The baseline intensity function $\lambda_0(t) = 0.1\mathrm{I}(t \leq 10) + 0.3\mathrm{I}(10 < t \leq 20)$. Note that the intensity model in (3.8) implies the additive rates model $\lambda\{t|Z_i(t)\} = \lambda_0^*(t) + \beta Z_i(t)$, where the baseline rate function $\lambda_0^*(t) = 0.02 + \lambda_0(t)$.

In the first set of simulations, we considered a continuous time-dependent covariate defined by $Z_i(t) = b_{0i} + b_{1i}t$, where the random intercept b_{0i} was generated from a normal distribution with mean 1.5 and variance 0.05. The random slope b_{1i} was generated from either a zero mean or a non-zero mean (-0.05) normal distribution with variance $5 \cdot 10^{-4}$. The two cases are referred to as *without time trend* and *with time trend*, respectively. The regression coefficient β is set at 0.2.

In the second set of simulations, we considered a binary time-dependent covariate. First, we generated the baseline value, $Z_i(0)$ from a Bernoulli distribution with probability 0.2. Then the binary covariate process was generated from a multistate process which consists of two states, 0 and 1. The duration of state 0 of subject i was generated from an exponential distribution with rate function $1/\{\xi_i g(t)\}$, and the duration of state 1 was generated from an exponential distribution with rate $1/\xi_i$, where the subject-specific random effect ξ_i followed a gamma distribution with mean

1 and variance 0.25 and the function $g(t)$ was set such that the covariate was either *with a time trend*: $g(t) = 4\mathrm{I}(t \leq 10) + 6\mathrm{I}(10 < t \leq 20)$ or *without a time trend*: $g(t) = 4$ for $t \in [0, 20]$). The regression coefficient is set at $\beta = 0.5$.

In all settings, we let the covariates of a subject be observed at its own event times and each subject has a baseline visit at time 0. For each subject, the time of 20 follow-up visits (if there is no censoring) was generated based on a uniform distribution within each of 20 unit time intervals, $(0, 1], (1, 2], \ldots, (19, 20]$. We allowed each visit to have a certain probability to be missing, p_m = 0%, 20%, 40%, and 60%. The censoring time was simulated from a uniform distribution on the interval $[0, 20]$.

We applied the proposed method and two simple approaches, the LCCF method and the linear interpolation method, to the simulated data. For the proposed method, we used the Epanechnikov kernel function and a bandwidth selection procedure as follows. First, we define the averaged squared error as $ASE(h) = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau Y_i(u)$ $\left\{\widehat{\xi}_h^{(1)}(u) - \xi^{(1)}(u)\right\}^2 dO_i(u)$. Since $ASE(h)$ involves the unknown quantity $\xi^{(1)}(\cdot)$, we define $CV(h)$ with the leave-one-out estimator as $CV(h) = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau Y_i(u)$ $\left\{Z_i(u) - \widehat{\xi}_{h,-i}^{(1)}(u)\right\}^2 dO_i(u)$. It is easy to show that

$$\begin{aligned} CV(h) =& \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau Y_i(u)\left\{Z_i(u) - \xi^{(1)}(u)\right\}^2 dO_i(u) \\ &+ \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau Y_i(u)\left\{\widehat{\xi}_{h,-i}^{(1)}(u) - \xi^{(1)}(u)\right\}^2 dO_i(u) \\ &- \frac{2}{n}\sum_{i=1}^{n}\int_0^\tau Y_i(u)\left\{Z_i(u) - \xi^{(1)}(u)\right\}\left\{\widehat{\xi}_{h,-i}^{(1)}(u) - \xi^{(1)}(u)\right\} dO_i(u). \end{aligned}$$

Since the first item on the right hand side does not involve h and the expectation of the third item is zero, minimizing $ASE(h)$ is on average equivalent to minimizing $CV(h)$. Using similar techniques as those in Chiang et al. (2005), it can be shown that the ASE converges to $O(h^4) + O(1/(nh))$, where the first term corresponds to squared bias and the second corresponds to variance. Thus, we can show that the optimal nonparametric convergence rate is $Cn^{-1/5}$ by following the same argument as in Newey et al. (2004). We then determine the constant C by minimizing $CV(h)$ with $h = Cn^{-1/5}$. In Appendix B we show that the range of the bandwidth for $\widehat{\boldsymbol{\beta}}_h$ is $h = O(n^{-v})$, where $1/4 < v < 1/2$, so after choosing the constant C in the first step, we use $h = Cn^{-1/3}$ for the estimation of $\boldsymbol{\beta}$.

In the simulation tables, we report the relative bias (Bias) and the Monte-Carlo empirical standard deviation of the point estimates (SD). For the proposed method, we also report the average standard errors (ASE) estimated by the bootstrap method and the coverage percentage (CP) of the 95% confidence intervals. Table 3.1 shows the simulation results when Z_i is continuous. For the scenarios with no time trend in the covariate, the LCCF method gives biased point estimates and the bias increases as the missing probability increases. The linear interpolation method and the proposed method give virtually unbiased point estimates. For the scenarios with time trend in the covariate, both the LCCF and linear interpolation methods give biased estimates, while the proposed method still provides virtually unbiased estimates. For the variance estimation of the proposed method, the ASEs are all close to the Monte Carlo SDs, and the coverage percentages are all close to 95%. As expected, the SDs (and ASEs) of the proposed method decrease as the sample size increases and increase as the missing rate increases.

Table 3.2 shows the results when the time-dependent covariate is binary. The two simple methods provide biased estimations regardless whether there is time trend or not in the covariate. The proposed estimator gives virtually unbiased estimates for all scenarios. The simplified estimator for binary covariate without time trend as proposed in Section 3.3.1 was applied to the scenarios without time trend; the results show that it provides almost unbiased estimates and the SD is slightly smaller than the SD for the proposed kernel estimator. We also applied the simplified estimator

Table 3.1: Simulation results for the model with a continuous time-dependent covariate: p_m is the missing probability of the covariate values at regular visits; Bias is the relative bias computed by dividing the difference of the mean of the 1000 estimated parameters and the true value by the true value; SD is the standard deviation of the 1000 estimated values; ASE is the mean of the 1000 estimated standard errors by bootstrap method; CP is the proportion of 95% confidence intervals covering the true value.

	LCCF		Linear		Proposed			
p_m	Bias	SD	Bias	SD	Bias	SD	ASE	CP
n=300								
Without time trend								
0%	0.037	0.049	0.005	0.047	-0.001	0.048	0.049	0.940
20%	0.043	0.049	0.005	0.047	-0.004	0.051	0.051	0.945
40%	0.051	0.050	0.006	0.047	0.003	0.055	0.056	0.946
60%	0.060	0.050	0.007	0.047	0.001	0.063	0.064	0.952
With time trend								
0%	-0.567	0.047	-0.020	0.045	0.016	0.046	0.044	0.943
20%	-0.728	0.049	-0.038	0.045	0.019	0.048	0.047	0.941
40%	-0.952	0.050	-0.071	0.045	0.024	0.051	0.051	0.948
60%	-1.277	0.053	-0.134	0.045	0.036	0.058	0.058	0.950
n=600								
Without time trend								
0%	0.036	0.036	0.003	0.035	0.001	0.035	0.034	0.943
20%	0.042	0.036	0.004	0.035	0.003	0.037	0.036	0.945
40%	0.050	0.036	0.005	0.035	0.005	0.040	0.039	0.938
60%	0.059	0.037	0.006	0.035	0.002	0.046	0.045	0.944
With time trend								
0%	-0.565	0.033	-0.022	0.031	0.008	0.032	0.031	0.942
20%	-0.724	0.034	-0.040	0.031	0.013	0.033	0.033	0.935
40%	-0.946	0.035	-0.072	0.031	0.016	0.035	0.036	0.947
60%	-1.271	0.037	-0.136	0.031	0.016	0.041	0.041	0.942

Table 3.2: Simulation results for the model with a binary time-dependent covariate: p_m is the missing probability of the covariate values at regular visits; Bias is the relative bias computed by dividing the difference of the mean of the 1000 estimated parameters and the true value by the true value; SD is the standard deviation of the 1000 estimated values; ASE is the mean of the 1000 estimated standard errors by bootstrap method; CP is the proportion of 95% confidence intervals covering the true value; – means non-applicable.

	LCCF		Linear		Proposed (Simple)		Proposed			
p_m	Bias	SD	Bias	SD	Bias	SD	Bias	SD	ASE	CP
n=300										
Without time trend										
0%	-0.095	0.033	0.153	0.042	0.002	0.040	0.003	0.041	0.039	0.933
20%	-0.141	0.033	0.107	0.041	0.002	0.042	0.004	0.043	0.042	0.941
40%	-0.206	0.032	0.031	0.039	0.003	0.044	0.006	0.046	0.046	0.949
60%	-0.299	0.031	-0.094	0.037	0.004	0.050	0.008	0.053	0.053	0.945
With time trend										
0%	-0.108	0.033	0.148	0.041	-	-	-0.006	0.040	0.040	0.944
20%	-0.155	0.032	0.104	0.040	-	-	-0.004	0.042	0.042	0.952
40%	-0.221	0.031	0.031	0.038	-	-	-0.003	0.045	0.046	0.953
60%	-0.319	0.031	-0.096	0.038	-	-	-0.002	0.054	0.053	0.938
n=600										
Without time trend										
0%	-0.096	0.023	0.151	0.029	0.001	0.027	0.001	0.028	0.028	0.933
20%	-0.142	0.023	0.106	0.029	0.001	0.029	0.002	0.031	0.029	0.930
40%	-0.206	0.022	0.029	0.027	0.001	0.031	0.003	0.032	0.032	0.940
60%	-0.300	0.021	-0.095	0.026	0.001	0.034	0.004	0.037	0.036	0.941
With time trend										
0%	-0.106	0.024	0.150	0.029	-	-	-0.003	0.028	0.028	0.941
20%	-0.154	0.023	0.106	0.028	-	-	-0.002	0.029	0.030	0.945
40%	-0.222	0.022	0.030	0.027	-	-	-0.001	0.031	0.032	0.958
60%	-0.319	0.022	-0.096	0.026	-	-	-0.001	0.037	0.037	0.944

to the case where only the measurements at baseline and event times are available (i.e., $p_m = 100\%$ for the regular visits, not shown in the table). When $n = 300$, the bias is 0.019 (SD = 0.092); when $n = 600$, the bias is 0.009 (SD = 0.063). The results demonstrate that when by design, there are only measurements of $\mathbf{Z}(t)$ at baseline (and event times), the simplified estimator is a valid and efficient method if it is believed that there is no time trend in the binary time-dependent covariate(s).

We also conducted simulation studies to evaluate the performance of the two extensions of the proposed method described in Section 3.3.2 and Section 3.3.3, namely, (1) when both time-dependent and -independent covariates are present in the model and (2) when time-dependent covariates are measured on different time schedules. For the first extension, we simulated data with one continuous time-independent covariate W_i from a normal distribution with mean 1.5 and variance 0.05 and one binary time-dependent covariate $Z_i(t)$, in the same way as for the binary covariate with time trend described before. For the second extension, we simulated data with two time-dependent covariates, one binary and one continuous, following the same way as before, except that the measuring times of the two covariates were simulated separately. We explored situations where each covariate was either with or without time trend. The simulation results of the two extensions are presented in Table 3.3. It is shown that the extensions of the proposed method perform well under all scenarios.

Table 3.3: Simulation results for the extensions of the proposed method: (a) both time-dependent and time-independent covariates are present, where β is the coefficient for the binary, time-dependent covariate Z with time trend, and γ is for the continuous, time-independent covariate W; (b) two time-dependent covariates with different observation time schedules, where β_1 is the coefficient for the continuous covariate Z_1 and β_2 is for the binary covariate Z_2. Notations: p_m is the missing probability of the covariate values at regular visits; Bias is the relative bias computed by dividing the difference of the mean of the 1000 estimated parameters and the true value by the true value; SD is the standard deviation of the 1000 estimated values; ASE is the mean of the 1000 estimated standard errors by bootstrap method; CP is the proportion of 95% confidence intervals covering the true value.

(a)

	β				γ			
p_m	Bias	SD	ASE	CP	Bias	SD	ASE	CP
n=300								
0%	-0.004	0.050	0.052	0.949	-0.015	0.066	0.065	0.952
20%	-0.000	0.055	0.056	0.949	-0.016	0.066	0.066	0.951
40%	-0.000	0.061	0.063	0.944	-0.017	0.068	0.067	0.952
60%	0.006	0.076	0.075	0.940	-0.013	0.070	0.070	0.951
n=600								
0%	-0.002	0.036	0.036	0.942	0.000	0.047	0.046	0.944
20%	-0.002	0.039	0.039	0.946	0.000	0.047	0.046	0.944
40%	-0.001	0.043	0.044	0.947	-0.000	0.048	0.047	0.940
60%	0.004	0.053	0.052	0.945	0.002	0.049	0.049	0.948

(b)

Time trend		β_1				β_2			
Z_1	Z_2	Bias	SD	ASE	CP	Bias	SD	ASE	CP
n=300									
No	No	0.013	0.058	0.056	0.944	-0.000	0.052	0.052	0.962
Yes	No	0.016	0.057	0.059	0.952	0.011	0.049	0.048	0.943
No	Yes	0.014	0.056	0.055	0.949	-0.010	0.055	0.053	0.928
Yes	Yes	0.030	0.056	0.058	0.950	0.006	0.050	0.049	0.938
n=600									
No	No	-0.003	0.040	0.039	0.941	0.002	0.036	0.036	0.946
Yes	No	0.009	0.044	0.045	0.960	0.007	0.034	0.034	0.950
No	Yes	0.016	0.040	0.039	0.936	-0.008	0.036	0.037	0.952
Yes	Yes	0.001	0.043	0.044	0.958	0.003	0.036	0.035	0.929

3.5 Real data analysis

We applied the proposed method to a study which investigated the effect of streptococci on the risk of pharyngitis. Pharyngitis is an infection of the pharynx, the back of the throat, which is often due to viruses, but several bacteria which include group A streptococcus (GAS) are also a common cause of pharyngitis. Specifically, the pharyngitis caused by GAS is also known as strep throat and is prevalent in children and usually occurs in late winter and early spring. Besides, bacteria of other streptococcal groups including GCS and GGS may also cause pharyngitis, and thus it is of clinical interest to investigate the effect of these bacteria on the risk of pharyngitis. Between March 2002 and March 2004, 307 school children in a rural area near Vellore, India were recruited. During the follow-up time, cases of pharyngitis were identified weekly (referred to as 'event visits') and the streptococci status was also determined for those with pharyngitis at the time when pharyngitis was diagnosed. In addition, monthly visits were scheduled to monitor the streptococci status (referred to as 'regular visits'). The detailed design of this study can be found in Jose et al. (2018). Although the regular visits were scheduled on a monthly basis, the actual observation times were irregularly spaced across subjects to balance the workload. It is reasonable to assume that the regular observation process $O_i(t)$ is independent of the covariate processes $\mathbf{Z}_i(t)$ and the censoring time C_i.

The start time of the study, March 11, 2002, is used as the time origin of the recurrent event process of the occurrence of pharyngitis. By choosing calendar time as the time scale, we can avoid modeling the confounding effect of season which is a nuisance in this study. Note that 74 (out of 307) school children were recruited in the second year after June 15, 2003, for whom, the at-risk indicator $Y_i(t)$ is modified to reflect whether subject i has been enrolled in the study prior to time t and remained under observation at time t. During the two-year follow-up, 640 pharyngitis occurrences were identified and 2827 regular visits were recorded. Among throat cultures collected in the regular visits, about 11.43% of them were positive for GAS, 2.90% were positive for GCS and 15.32% were positive for GGS. Among the cultures collected in the event visits, about 17.19% of them were positive for GAS, 4.69% were positive for GCS and 17.66% were positive for GGS. Since GAS, GCS and GGS all belong to the Streptococcus genus family, they are likely to be correlated. We applied McNemar's test for pairwise comparison using the measurements in the first regular visit of each child to test if these bacterial infections were correlated with each other. The results show that GCS was significantly correlated with both GAS and GGS but no significant correlation was observed between GAS and GGS, thus to avoid collinearity, we fit the additive rates model with only GAS and GGS to explore their relationship with the occurrence of pharyngitis. The estimated rate difference for the time-dependent GAS and GGS status based on the proposed kernel method are 0.067 and 0.020, respectively, and their corresponding 95% confidence intervals are (0.028, 0.106) and (-0.013, 0.053). Thus we conclude that positive GAS was associated with a higher risk of pharyngitis, while the GGS infection status was not significantly associated with the risk of pharyngitis. Figure 3.1 shows the estimated baseline mean function with pointwise 95% confidence bands.

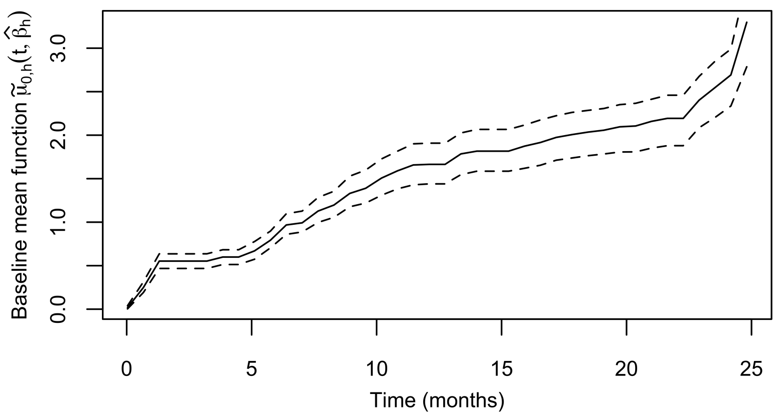

Figure 3.1: Estimation of the baseline mean function for Indian pharyngitis data. Time 0 is the start time of the study, March 11, 2002. The dashed lines are the 95% pointwise confidence bands based on the bootstrap samples.

3.6 Discussion

In this chapter, we propose a kernel smoothed estimating function method to deal with intermittently measured time-dependent covariates in the additive rates model. Compared to the Cox-type models, the additive model is more appealing to practitioners when the rate difference is of primary interest or the proportional rates assumption is violated. In relation to the recent works on Cox-type model (Cao et al., 2015; Li et al., 2016), the proposed work offers an alternative tool and overcomes the unique technical difficulties arising in additive models. Moreover, when multiple time-dependent covariates are in presence and measured on different schedules, the methods in Cao et al. (2015) and Li et al. (2016) cannot be directly applied. In this case, we further extend our method by using multivariate kernels to obtain consistent estimates.

The important feature that distinguishes the proposed method from its competitors lies in that it does not require modeling the underlying covariate process. Popular methods, such as the joint modeling approaches, requires a complete specification of the joint distribution of the recurrent event process and the covariate process, which is a challenging task when both continuous and binary covariates are present. Instead, we apply nonparametric kernel smoothing method to approximate the mean covariate process to obtain consistent estimates, and hence is more robust against model misspecifications.

In the motivating example, the covariates were measured at both event visits and regular visits, which is typical in recurrent event data since the subjects are still at risk after an event occurs. If the covariates are not observed at the time of events, a double kernel approach similar to what was proposed for the proportional rates model by Cao et al. (2015) can be extended to the additive rates model, but the convergence rate of the resulting estimator would be slower than the regular root-n rate. A less computationally intensive and simpler method is to carry forward the last observed value to replace the missing observation at event times in $\mathbf{Z}_i(t)dN_i(t)$ in Equation (3.5) and keep the rest of the terms in the estimating function which involve kernel smoothing the same. The performance of these two approaches will be evaluated in future research.

As another future direction, we can apply the kernel smoothing method to deal with intermittently measured covariates in additive-multiplicative rates model. It is also of interest to investigate model checking procedures to determine whether a covariate has an additive or multiplicative effect. For example, Lin et al. (2000) have proposed a standardized score-type process to check the multiplicative assumption for recurrent event data with continuously monitored time-dependent covariates. Research on checking the additive or multiplicative assumption for intermittently observed time-dependent covariates is warranted.

Chapter 4

Additive-multiplicative rates model for recurrent event data with intermittently observed time-dependent covariates

4.1 Introduction

In various clinical and biomedical studies, the event of interest may happen multiple times to a subject, which is referred to as recurrent event. Examples include recurrent bleedings in patients with hematologic malignancies (Stanworth et al., 2015) and recurrent cardiovascular events in subjects with diabetes (Van Der Heijden et al., 2013). During the follow-up of recurrent events, it is common to have repeated measurements of time-dependent covariates and it is often of interest to investigate the effect of such covariates on the occurrence of recurrent events. Our motivation is from an observational study about pharyngitis among school children (Jose et al., 2018). Pharyngitis is often caused by viruses, but some bacteria including streptococci can cause pharyngitis as well. In this study, weekly visits were scheduled to monitor the recurrent occurrence of pharyngitis and the status of streptococci infection was determined for those diagnosed with pharyngitis. In the meantime, monthly visits were scheduled for each participant to monitor the streptococci infection status regularly. The goal of this study was to explore the effect of streptococci on the risk of pharyngitis.

For the analysis of recurrent events, Prentice et al. (1981b) and Andersen and Gill (1982) proposed the multiplicative model on the intensity function which is interpreted as the instantaneous risk of event conditioning on the event history. To achieve a desirable, marginal interpretation and to allow flexible dependence structure among the recurrent events, Pepe and Cai (1993); Lawless et al. (1997); Lin et al. (2001) discussed the regression models on mean/rate function of recurrent event process and assumed that the covariate effects were multiplicative. Alternatively, the covariate effects may be additive. Liu and Wu (2011) considered the additive intensity model, and Lin et al. (2001) proposed the additive rates model. The multiplicative and additive models have different assumptions on the relationship between the covariate effects and the event process, thus it is desirable to estimate both types of effects under a general model setting. For univariate survival data where the event of interest only happens once, Lin and Ying (1995) proposed the additive-multiplicative model for the hazard function and Scheike and Zhang (2002) considered the Cox-Aalen model to allow the covariate effects to be time-varying. Recently, Cai et al. (2017a) studied modeling additive and multiplicative effects simultaneously on the mean residual life function. For recurrent event data, Han et al. (2016) proposed the additive-multiplicative models focusing on the markers contingent on recurrent event with an informative terminal event. Liu et al. (2010) considered the additive-multiplicative rates models for recurrent events, which allows covariates to have both multiplicative and additive effects on the rate function of recurrent event process.

Although the additive-multiplicative rates model allows the covariates to be time-dependent, the history of the covariates for each individual is required to be observed completely. In practice, however, the time-dependent covariates are usually only measured intermittently during the follow-up. One way to deal with the infrequently updated covariates is to predict the missing covariate values by smoothing the observed values, as was discussed in Raboud et al. (1993); Tsiatis et al. (1995); Boscardin et al. (1998); Bycott and Taylor (1998); Dafni and Tsiatis (1998) and summarized in Andersen and Liestøl (2003). Another commonly used approach is to jointly model the longitudinal covariate process and the recurrent event process. The joint models of longitudinal data and time-to-event data have been studied extensively in literature. Many authors considered the joint models of the two processes through latent random effects, including Wulfsohn and Tsiatis (1997); Xu and Zeger (2001); Vonesh et al. (2006), among others. In the setting of recurrent event data, Henderson et al. (2000) proposed to model the relationship between the time-dependent covariates and recurrent events by a latent Gaussian process and Li (2016) considered the joint model of the recurrent event process and the binary covariate process. More complex models which considered the covariate process, recurrent event process and the terminal event simultaneously also have been studied, including Kim et al. (2012) and Cai et al. (2017b) among others. To our knowledge, little research has been done to explore the additive-multiplicative rates model with intermittently observed time-dependent covariates.

Li et al. (2016) proposed kernel smoothed estimators for the proportional rates model for recurrent event data with intermittently observed time-dependent covariates. In Chapter 3, we extend the kernel smoothing method to the additive rates model. In this chapter, we propose to extend the kernel smoothing method to the parameter estimation of the additive-multiplicative rates model for recurrent event data.

The rest of the chapter is organized as follows. The additive-multiplicative rates model and the proposed estimator are introduced in Section 4.2. Simulation studies evaluating the performance of the proposed estimator are presented in Section 4.3.

Section 4.4 includes the analysis of the Indian pharyngitis data as introduced in the motivating example. Finally, a conclusion remark is included in Section 4.5.

4.2 Model and the proposed estimator

Suppose that n subjects are recruited in a study. Let $i = 1, \ldots, n$ index the subjects. Let $N_i^*(t)$ denote the number of events that subject i has experienced at or prior to time t when there is no censoring. Let $\mathbf{W}_i(t) = (\mathbf{Z}_i(t)^\intercal, \mathbf{X}_i(t)^\intercal)^\intercal$ be a $p \times 1$ vector of possibly time-dependent covariates and let $\boldsymbol{\theta}_0 = (\boldsymbol{\gamma}_0^\intercal, \boldsymbol{\beta}_0^\intercal)^\intercal$ be the corresponding true regression parameters. We consider the model proposed by Liu et al. (2010) on the rate function of the counting process $N_i^*(t)$,

$$\lambda(t|\mathbf{W}_i(t)) = g\{\boldsymbol{\gamma}_0^\intercal \mathbf{Z}_i(t)\} + h\{\boldsymbol{\beta}_0^\intercal \mathbf{X}_i(t)\}\lambda_0(t), \tag{4.1}$$

where $\lambda_0(t)$ is an unspecified, baseline rate function and g and h are known link functions. Specifically, if we let $g(x) = x$ and $h(x) = \exp(x)$, then model (4.1) becomes

$$\lambda(t|\mathbf{W}_i(t)) = \boldsymbol{\gamma}_0^\intercal \mathbf{Z}_i(t) + \exp\{\boldsymbol{\beta}_0^\intercal \mathbf{X}_i(t)\}\lambda_0(t), \tag{4.2}$$

which can be regarded as a combination of the semiparametric additive rates model and proportional rates model for recurrent event process. When $\boldsymbol{\gamma}_0 = \mathbf{0}$, the model in (4.2) degenerates to a proportional rates model and when $\boldsymbol{\beta}_0 = \mathbf{0}$, it degenerates to an additive rates model.

Let C_i denote the censoring time of subject i and we assume that C_i is independent of the counting process $N_i^*(t)$. Let $Y_i(t) = I(C_i \geq t)$ and define $N_i(t)$ as the observed number of events which is equivalent to $N_i^*(t \wedge C_i)$ where $t \wedge C_i = \min(t, C_i)$. Let τ denote a pre-specified time point and the recurrent event process could potentially be observed beyond τ. For the model estimation, first we define the following process

$$M_i(t,\boldsymbol{\theta}) = N_i(t) - \int_0^t Y_i(u)\{g\{\boldsymbol{\gamma}^\intercal \mathbf{Z}_i(u)\}du + h\{\boldsymbol{\beta}^\intercal \mathbf{X}_i(u)\}d\mu_0(u)\},$$

where $\mu_0(t) = \int_0^t \lambda_0(u)du$ is the baseline mean function. Following Lin and Ying (1995), Liu et al. (2010) proposed the following estimating function for model (4.1):

$$U(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \{\mathbf{D}_i(\boldsymbol{\theta},u) - \bar{\mathbf{D}}(\boldsymbol{\theta},u)\}dM_i(u,\boldsymbol{\theta}), \tag{4.3}$$

where $\mathbf{D}_i(\boldsymbol{\theta},t)$ is a p-dimensional smooth process of $\mathbf{W}_i(t)$ and $\boldsymbol{\theta}$ but not involving $\lambda_0(t)$, and

$$\bar{\mathbf{D}}(\boldsymbol{\theta},t) = \frac{\frac{1}{n}\sum_{i=1}^n Y_i(t)\mathbf{D}_i(\boldsymbol{\theta},t)h\{\boldsymbol{\beta}^\intercal \mathbf{X}_i(t)\}}{\frac{1}{n}\sum_{i=1}^n Y_i(t)h\{\boldsymbol{\beta}^\intercal \mathbf{X}_i(t)\}}.$$

By solving $\sum_{i=1}^n \int_0^t dM_i(u,\boldsymbol{\theta}) = 0$, the baseline mean function $\mu_0(t)$ in (4.3) can be estimated by

$$\widehat{\mu}_0(t,\boldsymbol{\theta}) = \int_0^t \frac{\sum_{i=1}^n \{dN_i(u) - Y_i(u)g\{\boldsymbol{\gamma}^\intercal \mathbf{Z}_i(u)\}du\}}{\sum_{i=1}^n Y_i(u)h\{\boldsymbol{\beta}^\intercal \mathbf{X}_i(u)\}}, \tag{4.4}$$

thus the estimating function in (4.3) becomes

$$U(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \{\mathbf{D}_i(\boldsymbol{\theta},u) - \bar{\mathbf{D}}(\boldsymbol{\theta},u)\}\{dN_i(u) - Y_i(u)g\{\boldsymbol{\gamma}^\intercal \mathbf{Z}_i(u)\}du\}. \tag{4.5}$$

Since we are mostly interested in simultaneously modeling the relative and absolute difference in rate function due to the covariates, we focus on the model in (4.2) in the remaining of this chapter. Under model (4.2), the estimating function in (4.5) becomes

$$
\begin{aligned}
U(\boldsymbol{\theta}) =& \frac{1}{n}\sum_{i=1}^{n}\int_{0}^{\tau}\{\mathbf{D}_i(\boldsymbol{\theta},u)-\bar{\mathbf{D}}(\boldsymbol{\theta},u)\}dN_i(u) \\
&-\frac{1}{n}\sum_{i=1}^{n}\int_{0}^{\tau}\{\mathbf{D}_i(\boldsymbol{\theta},u)-\bar{\mathbf{D}}(\boldsymbol{\theta},u)\}Y_i(u)\boldsymbol{\gamma}^{\intercal}\mathbf{Z}_i(u)du \\
=& \frac{1}{n}\sum_{i=1}^{n}\int_{0}^{\tau}\{\mathbf{D}_i(\boldsymbol{\theta},u)-\bar{\mathbf{D}}(\boldsymbol{\theta},u)\}dN_i(u)-\Bigg[\int_{0}^{\tau}\frac{1}{n}\sum_{i=1}^{n}Y_i(u) \\
&\left\{\frac{\sum_{j=1}^{n}\mathbf{D}_j(\boldsymbol{\theta},u)Y_j(u)\mathbf{Z}_j^{\intercal}(u)}{\sum_{j=1}^{n}Y_j(u)}-\bar{\mathbf{D}}(\boldsymbol{\theta},u)\frac{\sum_{j=1}^{n}Y_j(u)\mathbf{Z}_j^{\intercal}(u)}{\sum_{j=1}^{n}Y_j(u)}\right\}du\Bigg]\boldsymbol{\gamma}.
\end{aligned}
\tag{4.6}
$$

Then the estimator $\widehat{\boldsymbol{\theta}}$ can be obtained by solving $U(\boldsymbol{\theta})=\mathbf{0}$. When time-dependent covariates are present, the evaluation of the estimating function requires the time-dependent covariates to be observed continuously throughout the entire follow-up time. As is illustrated in the motivating example, time-dependent covariates are usually measured at intermittent visits, and hence the values of such covariates between these measurement times remain unknown, which renders the estimating function in (4.6) not evaluable based on the observed data only. A simple approach to deal with the intermittently observed time-dependent covariates is to impute the missing values by carrying forward the last observed value. However, this approach imposes a strong assumption that the covariate processes are step functions and thus is expected to introduce bias in the model estimation, as shown in Cao et al. (2015) and Li et al. (2016). Another simple approach is to impute the unobserved values between two observation times by linear interpolation. The linear interpolation method assumes that the covariate value is a linear function of time between every two adjacent observation times. However, this assumption may not hold in practice, especially for binary covariates. In what follows, we propose a semiparametric estimator by applying the kernel smoothing method to deal with the problem caused by the intermittently observed time-dependent covariates. The proposed method kernel smooths functions involving time-dependent covariates across subjects instead of imputing the missing values of each individual, thus it is fundamentally different from the two simple imputation methods and is expected to be more accurate.

First, we show that the estimating function in (4.6) is a functional of empirical processes. According to Lin and Ying (1995) and Liu et al. (2010), a possible choice for $\mathbf{D}_i(\boldsymbol{\theta}, t)$ is

$$\mathbf{D}_i(\boldsymbol{\theta}, t) = \begin{pmatrix} g'\{\boldsymbol{\gamma}^\intercal \mathbf{Z}_i(t)\}\mathbf{Z}_i(t)/h\{\boldsymbol{\beta}^\intercal \mathbf{X}_i(t)\} \\ h'\{\boldsymbol{\beta}^\intercal \mathbf{X}_i(t)\}\mathbf{X}_i(t)/h\{\boldsymbol{\beta}^\intercal \mathbf{X}_i(t)\} \end{pmatrix}.$$

Under model (4.2), we have

$$\mathbf{D}_i(\boldsymbol{\theta}, t) = \begin{pmatrix} \mathbf{Z}_i(t)/\exp\{\boldsymbol{\beta}^\intercal \mathbf{X}_i(t)\} \\ \mathbf{X}_i(t) \end{pmatrix}$$

and thus

$$\bar{\mathbf{D}}(\boldsymbol{\theta}, t) = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n Y_i(t)\mathbf{Z}_i(t)/\frac{1}{n}\sum_{i=1}^n Y_i(t)\exp\{\boldsymbol{\beta}^\intercal \mathbf{X}_i(t)\} \\ \frac{1}{n}\sum_{i=1}^n Y_i(t)\mathbf{X}_i(t)\exp\{\boldsymbol{\beta}^\intercal \mathbf{X}_i(t)\}/\frac{1}{n}\sum_{i=1}^n Y_i(t)\exp\{\boldsymbol{\beta}^\intercal \mathbf{X}_i(t)\} \end{pmatrix}.$$

We further define $S_z^{(k)}(t) = n^{-1}\sum_{i=1}^n Y_i(t)\mathbf{Z}_i(t)^{\otimes k}$, $S_x^{(k)}(t, \boldsymbol{\beta}) = n^{-1}\sum_{i=1}^n Y_i(t)\mathbf{X}_i(t)^{\otimes k}$ $\exp\{\boldsymbol{\beta}^\intercal \mathbf{X}_i(t)\}$, $k = 0, 1, 2$, where $a^{\otimes 0} = 1$, $a^{\otimes 1} = a$, $a^{\otimes 2} = aa^\intercal$, and $S_{z2x}(t, \boldsymbol{\beta}) = n^{-1}\sum_{i=1}^n Y_i(t)\mathbf{Z}_i(t)^{\otimes 2}\exp\{-\boldsymbol{\beta}^\intercal \mathbf{X}_i(t)\}$, $S_{zx}(t) = n^{-1}\sum_{i=1}^n Y_i(t)\mathbf{X}_i(t)\mathbf{Z}_i^\intercal(t)$. Then the estimating function in (4.6) can be written as

$$U(\boldsymbol{\theta}) = \begin{pmatrix} U_1(\boldsymbol{\theta}) \\ U_2(\boldsymbol{\theta}) \end{pmatrix},$$

where

$$\begin{aligned} U_1(\boldsymbol{\theta}) = {} & \frac{1}{n}\sum_{i=1}^n \int_0^\tau \mathbf{Z}_i(u)\exp\{-\boldsymbol{\beta}^\intercal \mathbf{X}_i(t)\}dN_i(u) - \int_0^\tau \frac{S_z^{(1)}(u)}{S_x^{(0)}(u, \boldsymbol{\beta})}\left\{\frac{1}{n}\sum_{i=1}^n dN_i(u)\right\} \\ & - \left[\int_0^\tau \frac{1}{n}\sum_{i=1}^n Y_i(u)\left\{\frac{S_{z2x}(u, \boldsymbol{\beta})}{S_z^{(0)}(u)} - \frac{S_z^{(1)}(u)}{S_x^{(0)}(u, \boldsymbol{\beta})}\left(\frac{S_z^{(1)}(u)}{S_z^{(0)}(u)}\right)^\intercal\right\}du\right]\boldsymbol{\gamma}, \end{aligned}$$

$$
\begin{aligned}
U_2(\boldsymbol{\theta}) = & \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \mathbf{X}_i(u)dN_i(u) - \int_0^\tau \frac{S_x^{(1)}(u,\boldsymbol{\beta})}{S_x^{(0)}(u,\boldsymbol{\beta})}\left\{\frac{1}{n}\sum_{i=1}^{n} dN_i(u)\right\} \\
& - \left[\int_0^\tau \frac{1}{n}\sum_{i=1}^{n} Y_i(u)\left\{\frac{S_{zx}(u)}{S_z^{(0)}(u)} - \frac{S_x^{(1)}(u,\boldsymbol{\beta})}{S_x^{(0)}(u,\boldsymbol{\beta})}\left(\frac{S_z^{(1)}(u)}{S_z^{(0)}(u)}\right)^{\intercal}\right\} du\right]\boldsymbol{\gamma}. \qquad (4.7)
\end{aligned}
$$

Note that if $\mathbf{W} = \mathbf{X}$, U_2 degenerates to the estimating function of the proportional rates model; if $\mathbf{W} = \mathbf{Z}$, U_1 becomes the estimating function of the additive rates model.

As can be seen from Equations (4.7), the estimating function consists of the processes $n^{-1}\sum_{i=1}^{n}\int_0^\tau \mathbf{Z}_i(t)\exp\{-\boldsymbol{\beta}^\intercal\mathbf{X}_i(t)\}dN_i(u)$, $n^{-1}\sum_{i=1}^{n}\int_0^\tau \mathbf{X}_i(u)dN_i(u)$, $n^{-1}\sum_{i=1}^{n} dN_i(u)$, $S_z^{(k)}(t)$, $S_x^{(k)}(t,\boldsymbol{\beta})$, $k = 0, 1$, $S_{z2x}(t,\boldsymbol{\beta})$ and $S_{zx}(t)$. Among them, the processes $n^{-1}\sum_{i=1}^{n}\int_0^\tau \mathbf{Z}_i(t)\exp\{-\boldsymbol{\beta}^\intercal\mathbf{X}_i(t)\}dN_i(u)$, $n^{-1}\sum_{i=1}^{n}\int_0^\tau \mathbf{X}_i(u)dN_i(u)$ and $n^{-1}\sum_{i=1}^{n} dN_i(u)$ can be calculated based on the observed data, while the ratios $S_z^{(1)}(t)/S_x^{(0)}(t,\boldsymbol{\beta})$, $S_{z2x}(t,\boldsymbol{\beta})/S_z^{(0)}(t)$, $S_z^{(1)}(t)/S_z^{(0)}(t)$, $S_{zx}(t)/S_z^{(0)}(t)$ and $S_x^{(1)}(t,\boldsymbol{\beta})/S_x^{(0)}(t,\boldsymbol{\beta})$ are required to be known throughout the follow-up period, which is not satisfied when $\mathbf{Z}_i(t)$ and $\mathbf{X}_i(t)$ are not continuously observed. Next we present how to approximate these ratios based on the observed data by applying the kernel smoothing method.

We define the limiting processes $s_z^{(k)}(t) = \mathrm{E}\{Y_i(t)\mathbf{Z}_i(t)^{\otimes k}\}$, $s_x^{(k)}(t,\boldsymbol{\beta}) = \mathrm{E}\left[Y_i(t)\mathbf{X}_i(t)^{\otimes k}\exp\{\boldsymbol{\beta}^\intercal\mathbf{X}_i(t)\}\right]$, $k = 0, 1$, $s_{z2x}(t,\boldsymbol{\beta}) = \mathrm{E}\left[\mathbf{Z}_i(t)^{\otimes 2}\exp\{-\boldsymbol{\beta}^\intercal\mathbf{X}_i(t)\}\right]$, $s_{zx}(t) = \mathrm{E}\{\mathbf{X}_i(t)\mathbf{Z}_i^\intercal(t)\}$. We approximate the ratios involving unobserved data by estimating the ratios of the limiting processes using the kernel smoothing method. Let $O_i(t)$ be the number of measurements of the covariates at pre-scheduled visits at or prior to time t. $O_i(0) = 0$ and $O_i(t)$ increases by 1 at the time of each pre-scheduled regular visit of subject i. We assume that the observation process $O_i(\cdot)$ is independent of $\mathbf{Z}_i(\cdot)$ and the censoring time C_i. The proposed kernel smoothed processes are as follows.

$$
\widehat{S}_{z,h}^{(k)}(t) = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau K_h(t-u)Y_i(u)\mathbf{Z}_i(u)^{\otimes k}dO_i(u),
$$

$$
\begin{aligned}
\widehat{S}^{(k)}_{x,h}(t,\boldsymbol{\beta}) &= \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} K_h(t-u)Y_i(u)\mathbf{X}_i(t)^{\otimes k}\exp\{\boldsymbol{\beta}^{\intercal}\mathbf{X}_i(t)\}dO_i(u),\\
\widehat{S}_{z2x,h}(t,\boldsymbol{\beta}) &= \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} K_h(t-u)Y_i(u)\mathbf{Z}_i(t)^{\otimes 2}\exp\{-\boldsymbol{\beta}^{\intercal}\mathbf{X}_i(t)\}dO_i(u),\\
\widehat{S}_{zx,h}(t) &= \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} K_h(t-u)Y_i(u)\mathbf{X}_i(t)\mathbf{Z}_i(t)^{\intercal}dO_i(u),
\end{aligned}
\tag{4.8}
$$

for $t \in [h, \tau-h]$, where $K_h(t) = K(t/h)/h$ with h as the bandwidth, $0 < h < \tau/2$, and the kernel function $K(t)$ has a bounded support and satisfies $\int_{-1}^{1} K(t)dt = 1$ and $\int_{-1}^{1} tK(t)dt = 0$. For boundary correction, we let $\widehat{S}^{(k)}_{z,h}(t) = \widehat{S}^{(k)}_{z,h}(h)$, $\widehat{S}^{(k)}_{x,h}(t,\boldsymbol{\beta}) = \widehat{S}^{(k)}_{x,h}(h,\boldsymbol{\beta})$, $\widehat{S}_{z2x,h}(t,\boldsymbol{\beta}) = \widehat{S}_{z2x,h}(h,\boldsymbol{\beta})$, $\widehat{S}_{zx,h}(t) = \widehat{S}_{zx,h}(h)$ for $t \in [0,h)$ and $\widehat{S}^{(k)}_{z,h}(t) = \widehat{S}^{(k)}_{z,h}(\tau-h)$, $\widehat{S}^{(k)}_{x,h}(t,\boldsymbol{\beta}) = \widehat{S}^{(k)}_{x,h}(\tau-h,\boldsymbol{\beta})$, $\widehat{S}^{(k)}_{z2x,h}(t,\boldsymbol{\beta}) = \widehat{S}_{z2x,h}(\tau-h,\boldsymbol{\beta})$, $\widehat{S}_{zx,h}(t) = \widehat{S}_{zx,h}(\tau-h)$ for $t \in (\tau-h,\tau]$. We define $m(t) = \mathrm{E}\{O_i(t)\}$, then we can show that the kernel smoothed processes in (4.8) converge to $s^{(k)}_z(t)m(t)$, $s^{(k)}_x(t,\boldsymbol{\beta})m(t)$, $s_{z2x}(t,\boldsymbol{\beta})m(t)$, $s_{zx}(t)m(t)$, respectively. Since $m(t)$ cancels out in the ratios, we can show that the ratios of the kernel smoothed processes converge to the ratios of the corresponding limiting processes. Thus, the ratios $S^{(1)}_z(t)/S^{(0)}_x(t,\boldsymbol{\beta})$, $S_{z2x}(t,\boldsymbol{\beta})/S^{(0)}_z(t)$, $S^{(1)}_z(t)/S^{(0)}_z(t)$, $S_{zx}(t)/S^{(0)}_z(t)$ and $S^{(1)}_x(t,\boldsymbol{\beta})/S^{(0)}_x(t,\boldsymbol{\beta})$ in (4.7) can be replaced by the kernel smoothed counterparts to lead to the proposed estimating function $\widehat{U}(\boldsymbol{\theta}) = \begin{pmatrix}\widehat{U}_1(\boldsymbol{\theta})\\ \widehat{U}_2(\boldsymbol{\theta})\end{pmatrix}$, where

$$
\begin{aligned}
\widehat{U}_1(\boldsymbol{\theta}) &= \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\mathbf{Z}_i(u)\exp\{-\boldsymbol{\beta}^{\intercal}\mathbf{X}_i(u)\}dN_i(u) - \int_0^{\tau}\frac{\widehat{S}^{(1)}_{z,h}(u)}{\widehat{S}^{(0)}_{x,h}(u,\boldsymbol{\beta})}\left\{\frac{1}{n}\sum_{i=1}^{n}dN_i(u)\right\}\\
&\quad - \left[\int_0^{\tau}\frac{1}{n}\sum_{i=1}^{n}Y_i(u)\left\{\frac{\widehat{S}_{z2x,h}(u,\boldsymbol{\beta})}{\widehat{S}^{(0)}_{z,h}(u)} - \frac{\widehat{S}^{(1)}_{z,h}(u)}{\widehat{S}^{(0)}_{x,h}(u,\boldsymbol{\beta})}\left(\frac{\widehat{S}^{(1)}_{z,h}(u)}{\widehat{S}^{(0)}_{z,h}(u)}\right)^{\intercal}\right\}du\right]\boldsymbol{\gamma},\\
\widehat{U}_2(\boldsymbol{\theta}) &= \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\mathbf{X}_i(u)dN_i(u) - \int_0^{\tau}\frac{\widehat{S}^{(1)}_{x,h}(u,\boldsymbol{\beta})}{\widehat{S}^{(0)}_{x,h}(u,\boldsymbol{\beta})}\left\{\frac{1}{n}\sum_{i=1}^{n}dN_i(u)\right\}
\end{aligned}
$$

$$-\left[\int_0^\tau \frac{1}{n}\sum_{i=1}^n Y_i(u)\left\{\frac{\widehat{S}_{zx,h}(u)}{\widehat{S}^{(0)}_{z,h}(u)} - \frac{\widehat{S}^{(1)}_{x,h}(u,\boldsymbol{\beta})}{\widehat{S}^{(0)}_{x,h}(u,\boldsymbol{\beta})}\left(\frac{\widehat{S}^{(1)}_{z,h}(u)}{\widehat{S}^{(0)}_{z,h}(u)}\right)^\intercal\right\}du\right]\boldsymbol{\gamma}. \tag{4.9}$$

The proposed estimator $\widehat{\boldsymbol{\theta}}_h$ can be obtained by solving $\widehat{U}(\boldsymbol{\theta}) = \mathbf{0}$. The large sample properties of $\widehat{\boldsymbol{\theta}}_h$ is summarized in Theorem 3. A detailed proof is included in Appendix C.

Theorem 3. Under conditions 1-10 in Appendix C, $\widehat{\boldsymbol{\theta}}_h$ is a consistent estimator of $\boldsymbol{\theta}_0$, and as $n \to \infty$, $\sqrt{n}(\widehat{\boldsymbol{\theta}}_h - \boldsymbol{\theta}_0)$ converges to a normal distribution with zero mean and variance $A(\boldsymbol{\theta}_0)^{-1}V(\boldsymbol{\theta}_0)\{A(\boldsymbol{\theta}_0)^{-1}\}^\intercal$ with $A(\boldsymbol{\theta}_0)$ and $V(\boldsymbol{\theta}_0)$ defined in Appendix C.

For the estimation of the baseline mean function, we show that the estimator in (4.4) can be written as

$$\begin{aligned}
\widehat{\mu}_0(t,\boldsymbol{\theta}) &= \int_0^t \frac{\sum_{i=1}^n\{dN_i(u) - Y_i(u)\boldsymbol{\gamma}^\intercal \mathbf{Z}_i(u)du\}}{\sum_{i=1}^n Y_i(u)\exp\{\boldsymbol{\beta}^\intercal \mathbf{X}_i(u)\}} \\
&= \sum_{i=1}^n \int_0^t \frac{1}{S_x^{(0)}(u,\boldsymbol{\beta})}\frac{1}{Y.(u)}dN_i(u) - \int_0^t \frac{S_z^{(1)}(u)}{S_x^{(0)}(u,\boldsymbol{\beta})}du\boldsymbol{\gamma},
\end{aligned}$$

where $Y.(t) = \sum_{i=1}^n Y_i(t)$. Thus, the baseline mean function can be estimated by

$$\widehat{\mu}_{0,h}(t,\widehat{\boldsymbol{\theta}}_h) = \sum_{i=1}^n \int_0^t \frac{1}{\widehat{S}^{(0)}_{x,h}(u,\widehat{\boldsymbol{\beta}}_h)}\frac{1}{Y.(u)}dN_i(u) - \int_0^t \frac{\widehat{S}^{(1)}_{z,h}(u)}{\widehat{S}^{(0)}_{x,h}(u,\widehat{\boldsymbol{\beta}}_h)}du\widehat{\boldsymbol{\gamma}}_h.$$

To ensure that the estimated baseline mean function is monotone, we can use

$$\widetilde{\mu}_{0,h}(t,\widehat{\boldsymbol{\theta}}_h) = \max_{0\le u\le t}\widehat{\mu}_{0,h}(u,\widehat{\boldsymbol{\theta}}_h).$$

4.3 Simulation

Simulation studies were conducted to evaluate the finite-sample performance of the proposed method. A total of 1000 data replicates with sample size 100 and 200 were generated for each simulation scenario. The resampling size in the bootstrap method for variance estimation was set to be 50. We simulated the recurrent events based on the following intensity model

$$\lambda\{t|\mathbf{W}_i(t), \gamma_i\} = u_i\left[\gamma_0 Z_i(t) + \exp\{\beta_0 X_i(t)\}\lambda_0(t)\right], \tag{4.10}$$

where u_i is a frailty variable with mean 1 and variance σ^2. The frailty variable induces the within-subject correlations. Note that the intensity model in (4.10) implies a marginal rate model $\lambda\{t|\mathbf{W}_i(t)\} = \gamma_0 Z_i(t) + \exp\{\beta_0 X_i(t)\}\lambda_0(t)$. We explored two distributions of the frailty variable, gamma distribution and log-normal distribution, and two values of the variance $\sigma^2 = 0.2, 0.4$, for different levels of correlations. The baseline intensity function $\lambda_0(t) = 0.3\mathrm{I}(t \leq 10) + 0.5\mathrm{I}(10 < t \leq 20)$. To evaluate the proposed method on different types of covariates, we let $Z_i(t)$ be a continuous covariate and $X_i(t)$ be a binary covariate. The continuous $Z_i(t)$ was simulated by a linear function of time t as $Z_i(t) = b_{0i} + b_{1i}t$, where the random intercept b_{0i} was generated from a normal distribution with mean 0.5 and variance 0.05 and the random slope from a normal distribution with mean -0.05 and variance $5 \cdot 10^{-4}$. With a negative mean of the random slope, the covariate $Z(\cdot)$ has a decreasing time trend at the population level. For the binary covariate $X_i(t)$, we first generated the baseline $X_i(0)$ from a Bernoulli distribution with probability 0.2. Then the binary covariate process was assumed to alternate between states 0 and 1. We assumed that the duration of state 0 of subject i followed an exponential distribution with rate function $1/\{\xi_i g(t)\}$ and the duration of state 1 followed an exponential distribution with rate $1/\xi_i$, where ξ_i was a subject-specific random effect which followed a gamma distribution with mean 1 and variance 0.25. The value of $g(t)$ was 4 for $t \in [0, 10]$ and changed to 6 afterwards, which indicates a decreasing time trend at the population level. For the values of regression coefficients, we considered three scenarios: (1) the true model included both an additive part and a multiplicative part: $\beta_0 = 0.5, \gamma_0 = 0.2$; (2) the additive-multiplicative model degenerated to the additive rates model: $\beta_0 = 0, \gamma_0 = 0.2$; (3) the additive-multiplicative degenerated to the proportional rates model: $\beta_0 = 0.5, \gamma_0 = 0$.

In all scenarios, we assumed that the covariates of a subject were always observed at the event times of the same subject. For the regular visits, we assumed that the covariates were measured at the baseline visit (time 0) and at each of 20 pre-scheduled regular visits with each visit time being simulated from a uniform distribution from $j-1$ to j, $j = 1, \ldots, 20$. The censoring time was randomly generated from a uniform distribution from 0 to 20.

The results for simulated datasets with gamma and lognormal frailty are presented in Table 4.1 and Table 4.2, respectively. We provide the relative bias (Bias) and Monte Carlo standard deviation (SD) of the point estimations for the proposed method and two simple imputation methods, the last covariate carried forward (LCCF) and the linear interpolation method. For the proposed method, we also report the average of the estimated standard errors by the bootstrap method (ASE) and the coverage percentage (CP). When the true model is the additive-multiplicative model ($\beta_0 = 0.5, \gamma_0 = 0.2$), the LCCF method gives biased estimations for both β and γ. The linear interpolation method has small bias for the estimation of γ, which is likely due to the linear feature of covariate $Z(\cdot)$, but gives biased estimations for β. The proposed method provides virtually unbiased estimations for both regression coefficients. The ASEs are close to the empirical SDs and the coverage percentages are around 95%. As the variance of the frailty increases from 0.2 to 0.4, the Monte Carlo SDs of the estimations increases. When the true model degenerates to the additive model or the multiplicative model with one significant covariate ($\beta = 0, \gamma = 0.2$; $\beta = 0.5, \gamma = 0$), the proposed method provides unbiased estimations for both coefficients, which indicates that it is valid to fit the additive-multiplicative rates model under such

Table 4.1: Simulation results for the gamma frailty model: Bias is the relative bias computed by dividing the difference of the mean of the 1000 estimated parameters and the true value by the true value (if the true value is 0, Bias is the mean of the 1000 estimated parameters); SD is the standard deviation of the 1000 estimated values; ASE is the mean of the 1000 estimated standard errors by bootstrap method; CP is the proportion of 95% confidence intervals covering the true value.

				LCCF		Linear		Proposed			
n	σ^2	$\bar{m}$		Bias	SD	Bias	SD	Bias	SD	ASE	CP
						$\beta = 0.5, \gamma = 0.2$					
100	0.2	4.23	β	-0.136	0.126	0.145	0.158	0.012	0.144	0.148	0.933
			γ	-0.575	0.109	0.009	0.104	0.034	0.105	0.103	0.937
	0.4	4.24	β	-0.131	0.139	0.150	0.176	0.011	0.157	0.155	0.948
			γ	-0.621	0.128	-0.033	0.123	-0.013	0.123	0.123	0.938
200	0.2	4.25	β	-0.133	0.092	0.153	0.114	0.009	0.103	0.102	0.929
			γ	-0.570	0.082	0.006	0.078	0.021	0.078	0.073	0.931
	0.4	4.24	β	-0.139	0.099	0.143	0.124	0.000	0.111	0.108	0.946
			γ	-0.602	0.096	-0.023	0.092	-0.009	0.092	0.088	0.935
						$\beta = 0, \gamma = 0.2$					
100	0.2	3.82	β	-0.012	0.148	-0.005	0.195	-0.003	0.169	0.172	0.953
			γ	-0.540	0.105	-0.008	0.100	0.033	0.101	0.100	0.934
	0.4	3.83	β	-0.016	0.163	-0.011	0.216	-0.006	0.181	0.183	0.946
			γ	-0.526	0.128	0.008	0.123	0.051	0.123	0.119	0.938
200	0.2	3.84	β	-0.010	0.101	-0.002	0.131	-0.002	0.113	0.120	0.952
			γ	-0.546	0.075	-0.021	0.073	0.016	0.073	0.070	0.943
	0.4	3.82	β	-0.017	0.106	-0.009	0.140	-0.006	0.119	0.125	0.968
			γ	-0.550	0.092	-0.024	0.089	0.012	0.089	0.086	0.937
						$\beta = 0.5, \gamma = 0$					
100	0.2	3.92	β	-0.148	0.122	0.143	0.152	0.007	0.136	0.143	0.955
			γ	-0.112	0.108	0.001	0.103	0.006	0.103	0.099	0.931
	0.4	3.90	β	-0.163	0.136	0.121	0.172	-0.021	0.153	0.150	0.938
			γ	-0.124	0.131	-0.011	0.124	-0.006	0.125	0.120	0.930
200	0.2	3.91	β	-0.138	0.086	0.157	0.107	0.013	0.097	0.098	0.952
			γ	-0.113	0.074	-0.002	0.070	0.002	0.071	0.071	0.940
	0.4	3.91	β	-0.136	0.093	0.161	0.117	0.015	0.103	0.104	0.949
			γ	-0.113	0.096	-0.002	0.090	0.002	0.090	0.086	0.929

Table 4.2: Simulation results for the lognormal frailty model: Bias is the relative bias computed by dividing the difference of the mean of the 1000 estimated parameters and the true value by the true value (if the true value is 0, Bias is the mean of the 1000 estimated parameters); SD is the standard deviation of the 1000 estimated values; ASE is the mean of the 1000 estimated standard errors by bootstrap method; CP is the proportion of 95% confidence intervals covering the true value.

				LCCF		Linear		Proposed			
n	σ^2	$\bar{m}$		Bias	SD	Bias	SD	Bias	SD	ASE	CP
						$\beta = 0.5, \gamma = 0.2$					
100	0.2	4.24	β	-0.141	0.130	0.141	0.161	0.004	0.145	0.147	0.950
			γ	-0.605	0.110	-0.021	0.106	0.005	0.106	0.102	0.930
	0.4	4.22	β	-0.141	0.135	0.140	0.167	0.004	0.150	0.157	0.954
			γ	-0.575	0.134	0.010	0.129	0.032	0.130	0.123	0.938
200	0.2	4.25	β	-0.141	0.088	0.141	0.110	0.006	0.100	0.102	0.943
			γ	-0.560	0.078	0.015	0.076	0.030	0.076	0.074	0.940
	0.4	4.24	β	-0.130	0.093	0.152	0.114	0.007	0.102	0.107	0.959
			γ	-0.584	0.093	-0.008	0.090	0.010	0.090	0.088	0.943
						$\beta = 0, \gamma = 0.2$					
100	0.2	3.83	β	-0.029	0.160	-0.027	0.210	-0.019	0.177	0.173	0.934
			γ	-0.554	0.105	-0.020	0.101	0.026	0.102	0.099	0.940
	0.4	3.82	β	-0.016	0.164	-0.012	0.215	-0.008	0.181	0.180	0.947
			γ	-0.568	0.130	-0.034	0.126	0.010	0.127	0.117	0.932
200	0.2	3.82	β	-0.019	0.107	-0.013	0.140	-0.010	0.120	0.120	0.944
			γ	-0.524	0.073	-0.001	0.071	0.037	0.071	0.070	0.939
	0.4	3.82	β	-0.014	0.110	-0.007	0.144	-0.003	0.122	0.125	0.941
			γ	-0.559	0.092	-0.033	0.089	0.001	0.089	0.085	0.940
						$\beta = 0.5, \gamma = 0$					
100	0.2	3.90	β	-0.149	0.118	0.142	0.147	0.007	0.134	0.140	0.954
			γ	-0.112	0.108	0.001	0.102	0.006	0.103	0.099	0.931
	0.4	3.91	β	-0.155	0.129	0.133	0.161	-0.009	0.146	0.150	0.951
			γ	-0.111	0.127	0.002	0.120	0.007	0.121	0.119	0.947
200	0.2	3.91	β	-0.146	0.086	0.148	0.108	0.005	0.098	0.098	0.957
			γ	-0.110	0.075	0.001	0.071	0.004	0.071	0.070	0.944
	0.4	3.91	β	-0.147	0.093	0.146	0.116	0.000	0.103	0.104	0.948
			γ	-0.113	0.092	-0.002	0.087	0.003	0.087	0.086	0.945

degenerated scenarios.

4.4 Real data analysis

We analyzed the Indian pharyngitis data using the proposed method. Pharyngitis is the infection of the back of the throat and it can be caused by viruses or bacteria. When the cause is group A streptococcus (GAS), pharyngitis is also known as strep throat. The symptoms of GAS pharyngitis include sore throat, fever, nausea and it may cause some rare but serious diseases including rheumatic heart disease if left untreated. GAS pharyngitis is common in children from age 5 to age 15 and can be transmitted through saliva or nasal secretions. At the same time, other bacteria including group C and G streptococcus (GCS, GGS) can cause pharyngitis with similar clinical symptoms as well. In the motivating example, 307 school children aged 7 to 11 years old in a rural area in Velore, India were recruited to investigate the relationship between streptococcal infections and the risk of pharyngitis. Each child was examined weekly for the symptoms of pharyngitis. For those who were diagnosed with pharyngitis, throat cultures were obtained to test if GAS, GCS and GGS were positive. In the meantime, to monitor the streptococci status regularly, monthly regular visits were scheduled for each child. Since the regular visits were pre-scheduled, it is reasonable to assume that they are independent of the covariate processes and the censoring time, as is required by the proposed method.

In order to avoid the potential multicollinearity problem, we first examined the association between GAS, GCS and GGS by conducting McNemar's test for pairwise comparison using the first observed value of regular visits of each child. The results show that GCS was significantly correlated with both GAS and GGS, and GAS and GGS were not significantly correlated, so we included only GAS and GGS in the model. We considered the following four candidate models: (1) both covariates have multiplicative effects (MM); (2) both covariates have additive effects (AA); (3) GAS has an additive effect and GGS has a multiplicative effect (AM); (4) GAS

Table 4.3: Analysis of Indian pharyngitis data: MM is the proportional rates model; AA is the additive rates model; AM is the additive-multiplicative rates model which includes GAS in the additive part and GGS in the multiplicative part; MA is the additive-multiplicative rates model which includes GAS in the multiplicative part and GGS in the additive part; Est is the estimated regression coefficient; SE is the standard error estimated by bootstrap with resampling size 100; CI is confidence interval.

Model	GAS			GGS		
	Est	SE	95% CI	Est	SE	95% CI
MM	0.418	0.117	(0.189, 0.647)	0.146	0.118	(-0.085, 0.377)
AA	0.067	0.020	(0.028, 0.106)	0.020	0.017	(-0.013, 0.053)
AM	0.067	0.019	(0.030, 0.104)	0.148	0.121	(-0.089, 0.385)
MA	0.437	0.120	(0.202, 0.672)	0.021	0.023	(-0.024, 0.066)

has a multiplicative effect and GGS has an additive effect (MA). Table 4.3 shows the analysis results for the four models. All four models indicate that GAS has a significant effect on the risk of pharyngitis, while GGS is not significant in any model.

4.5 Discussion

In this chapter, we propose a semiparametric estimator for the regression coefficients of the additive-multiplicative rates model to deal with the intermittently observed time-dependent covariates. The additive-multiplicative rates model combines the proportional rates model and the additive rates model, and hence allows some covariates to have multiplicative effects on the risk of recurrent events and others to have additive effects. The proposed method applies the nonparametric kernel smoothing approach to estimate the mean processes of the time-dependent covariates, thus it does not rely on any assumption of the covariate distribution or any specification of the covariate process and is expected to be more robust.

In practice, a practical problem is to determine the covariates included in $\mathbf{Z}(t)$ and $\mathbf{X}(t)$. If a covariate is expected to greatly influence the risk difference or the researchers are more interested in the absolute risks, then it should be included in $\mathbf{Z}(t)$. Otherwise, if a covariate is expected to strongly influence the risk ratios or the researchers are more interested in the relative risks, then it should be included in $\mathbf{X}(t)$. If the underlying biological process is not clear and the number of covariates is small, one can consider all the possible candidate models and examine the estimated effects in different models. Research on model selection procedure is warranted.

An R package, **rectime**, has been developed to implement the proposed estimator for the additive-multiplicative rates model with intermittently observed time-dependent covariates, as well as the estimators for the proportional rates model and the additive rates model with such covariates. The Indian pharyngitis data is included in the package as a data example.

Chapter 5

Discussion

Recurrent event data are commonly encountered in biomedical and epidemiological studies. The focus of analyzing recurrent event data is either on the gap times between consecutive events or on the total time to events. In Chapter 2, we focus on the analysis of the recurrent gap time data and propose a smooth and monotone estimating function for the AFT model, which facilitates more computationally stable and efficient point and variance estimations compared to the existing method. In Chapter 3 and Chapter 4, we focus on analyzing total time to events data by the rates models. First, we propose a novel semiparametric estimator by kernel smoothing the mean covariate processes for the additive rates model to deal with the intermittently observed time-dependent covariates in Chapter 3, and then we extend the method to the more general additive-multiplicative rates model in Chapter 4. The proposed method does not rely on any assumption of the underlying covariate process and thus is expected to be more robust compared to the simple approaches.

In this dissertation, the proposed estimators for the additive rates model and the additive-multiplicative rates model enrich the toolbox, in addition to the Cox-type model, for analyzing recurrent event data with intermittently observed time-dependent covariates. As a future research direction, it is of interest to explore model checking techniques to test if the multiplicative or the additive effects assumption holds for the covariates, which can provide guidance in model selection as well. Lin et al. (2000) proposed a standardized score-type process to check the multiplicative assumption of the proportional rates model for recurrent event data, but it requires that time-dependent covariates (if there are any) are observed continuously. Indeed, the performance of the method in Lin et al. (2000) has not been examined for the situations when (either continuously or intermittently observed) time-dependent covariates are present. For additive models, Yin (2007) proposed a score-type process to check the additive hazards assumption for the additive hazards model with multivariate survival data (i.e., clustered survival data). To our knowledge, no model checking methods for the additive rates model or the additive-multiplicative rates model have been studied for recurrent event data.

The additive rates model and the additive-multiplicative rates model provide alternative ways in analyzing recurrent event data when the multiplicative effect assumption imposed by the Cox-type model is invalid. It is worthwhile to investigate other models, such as the general transformation model (Lin et al., 2001; Zeng and Lin, 2006) which includes the proportional odds model as a special case, and extend the proposed method to accommodate the intermittently observed time-dependent covariates.

References

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

Andersen, P. K. and Liestøl, K. (2003). Attenuation caused by infrequently updated covariates in survival analysis. *Biostatistics*, 4(4):633–649.

Arcones, M. A. (1998). Asymptotic theory for m-estimators over a convex kernel. *Econometric Theory*, 14(4):387–422.

Barker, J. N., Hough, R. E., van Burik, J.-A., DeFor, T. E., MacMillan, M. L., OBrien, M. R., and Wagner, J. E. (2005). Serious infections after unrelated donor transplantation in 136 children: Impact of stem cell source. *Biology of Blood and Marrow Transplantation*, 11(5):362–370.

Boscardin, W. J., Taylor, J. M., and Law, N. (1998). Longitudinal models for AIDS marker data. *Statistical Methods in Medical Research*, 7(1):13–27.

Brown, B. M. and Wang, Y.-G. (2007). Induced smoothing for rank regression with censored survival times. *Statistics in Medicine*, 26(4):828–836.

Bycott, P. and Taylor, J. (1998). A comparison of smoothing techniques for CD4 data measured with error in a time-dependent Cox proportional hazards model. *Statistics in Medicine*, 17(18):2061–2077.

Cai, J., He, H., Song, X., and Sun, L. (2017a). An additive-multiplicative mean residual life model for right-censored data. *Biometrical Journal*, 59(3):579–592.

Cai, Q., Wang, M.-C., and Chan, K. C. G. (2017b). Joint modeling of longitudinal, recurrent events and failure time data for survivor's population. *Biometrics*, 73(4):1150–1160.

Cao, H., Churpek, M. M., Zeng, D., and Fine, J. P. (2015). Analysis of the proportional hazards model with sparse longitudinal covariates. *Journal of the American Statistical Association*, 110(511):1187–1196.

Chang, S.-H. (2004). Estimating marginal effects in accelerated failure time models for serial sojourn times among repeated events. *Lifetime Data Analysis*, 10(2):175–190.

Chiang, C.-T., Wang, M.-C., and Huang, C.-Y. (2005). Kernel estimation of rate function for recurrent event data. *Scandinavian Journal of Statistics*, 32(1):77–91.

Chiou, S., Kang, S., and Yan, J. (2015a). Rank-based estimating equations with general weight for accelerated failure time models: an induced smoothing approach. *Statistics in Medicine*, 34(9):1495–1510.

Chiou, S., Kang, S., and Yan, J. (2015b). Semiparametric accelerated failure time modeling for clustered failure times from stratified sampling. *Journal of the American Statistical Association*, 110(510):621–629.

Cook, R. J. and Lawless, J. F. (2007). *The Statistical Analysis of Recurrent Events*. Springer Science & Business Media.

Dafni, U. G. and Tsiatis, A. A. (1998). Evaluating surrogate markers of clinical outcome when measured with error. *Biometrics*, 54(4):1445–1462.

Du, P. (2009). Nonparametric modeling of the gap time in recurrent event data. *Lifetime Data Analysis*, 15(2):256–277.

Du, P., Jiang, Y., and Wang, Y. (2011). Smoothing spline anova frailty model for recurrent event data. *Biometrics*, 67(4):1330–1339.

Eaton, W. W., Mortensen, P. B., Herrman, H., Freeman, H., Bilker, W., Burgess, P., and Wooff, K. (1992). Long-term course of hospitalization for schizophrenia: Part I. Risk for rehospitalization. *Schizophrenia Bulletin*, 18(2):217–228.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. New York: Chapman & Hall.

Fygenson, M. and Ritov, Y. (1994). Monotone estimating equations for censored data. *The Annals of Statistics*, 22(2):732–746.

Gehan, E. A. (1965). A generalized wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika*, 52(1-2):203–223.

Han, M., Song, X., Sun, L., and Liu, L. (2016). An additive-multiplicative mean model for marker data contingent on recurrent event with an informative terminal event. *Statistica Sinica*, 26(3):1197–1218.

Henderson, R., Diggle, P., and Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics*, 1(4):465–480.

Heslin, K. C. and Weiss, A. J. (2015). Hospital readmissions involving psychiatric disorders, 2012. *Statistical Brief*, 189.

Huang, Y. and Chen, Y. Q. (2003). Marginal regression of gaps between recurrent events. *Lifetime Data Analysis*, 9(3):293–303.

Jin, Z., Lin, D. Y., Wei, L. J., and Ying, Z. (2003). Rank-based inference for the accelerated failure time model. *Biometrika*, 90(2):341–353.

Jin, Z., Lin, D. Y., and Ying, Z. (2006). Rank regression analysis of multivariate failure time data based on marginal linear models. *Scandinavian Journal of Statistics*, 33(1):1–23.

Johnson, L. M. and Strawderman, R. L. (2009). Induced smoothing for the semiparametric accelerated failure time model: asymptotics and extensions to clustered data. *Biometrika*, 96(3):577–590.

Jose, J. J. M., Brahmadathan, K. N., Abraham, V. J., Huang, C.-Y., Morens, D., Hoe, N. P., Follmann, D. A., and Krause, R. M. (2018). Streptococcal group a, c, and g pharyngitis in school children: A prospective cohort study in southern india. *Epidemiology and Infection*. doi:10.1017/S095026881800064X.

Kalbfleisch, J. D. and Prentice, R. L. (2011). *The Statistical Analysis of Failure Time Data*. John Wiley & Sons.

Kang, F., Sun, L., and Zhao, X. (2015). A class of transformed hazards models for recurrent gap times. *Computational Statistics and Data Analysis*, 83:151–167.

Kim, S., Zeng, D., Chambless, L., and Li, Y. (2012). Joint models of longitudinal data and recurrent events with informative terminal event. *Statistics in Biosciences*, 4(2):262–281.

Kohli, P., Wallentin, L., Reyes, E., Horrow, J., Husted, S., Angiolillo, D. J., Ardissino, D., Maurer, G., Morais, J., Nicolau, J. C., et al. (2013). Reduction in first and recurrent cardiovascular events with ticagrelor compared with clopidogrel in the plato study. *Circulation*, 127(6):673–680.

Lawless, J. F. and Nadeau, C. (1995). Some simple robust methods for the analysis of recurrent events. *Technometrics*, 37(2):158–168.

Lawless, J. F., Nadeau, C., and Cook, R. J. (1997). Analysis of mean and rate functions for recurrent events. In *Proceedings of the First Seattle Symposium in Biostatistics*, pages 37–49. New York: Springer.

Li, H. and Yin, G. (2009). Generalized method of moments estimation for linear regression with clustered failure time data. *Biometrika*, 96(2):293–306.

Li, S. (2016). Joint modeling of recurrent event processes and intermittently observed time-varying binary covariate processes. *Lifetime Data Analysis*, 22(1):145–160.

Li, S., Sun, Y., Huang, C.-Y., Follmann, D. A., and Krause, R. (2016). Recurrent event data analysis with intermittently observed time-varying covariates. *Statistics in Medicine*, 35(18):3049–3065.

Lin, D., Wei, L., and Ying, Z. (2001). Semiparametric transformation models for point processes. *Journal of the American Statistical Association*, 96(454):620–628.

Lin, D. and Ying, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika*, 81(1):61–71.

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*, 62(4):711–730.

Lin, D. Y. and Ying, Z. (1995). Semiparametric analysis of general additive-multiplicative hazard models for counting processes. *The Annals of Statistics*, 23(5):1712–1734.

Liu, L. and Huang, X. (2009). Joint analysis of correlated repeated measures and recurrent events processes in the presence of death, with application to a study on acquired immune deficiency syndrome. *Journal of the Royal Statistical Society: Series C*, 58(1):65–81.

Liu, Y., Wu, Y., Cai, J., and Zhou, H. (2010). Additive–multiplicative rates model for recurrent events. *Lifetime Data Analysis*, 16(3):353–373.

Liu, Y. Y. and Wu, Y. S. (2011). Semiparametric additive intensity model with frailty for recurrent events. *Acta Mathematica Sinica, English Series*, 27(9):1831–1842.

Lu, W. (2005). Marginal regression of multivariate event times based on linear transformation models. *Lifetime Data Analysis*, 11(3):389–404.

Luo, X. and Huang, C.-Y. (2011). Analysis of recurrent gap time data using the weighted risk-set method and the modified within-cluster resampling method. *Statistics in Medicine*, 30(4):301–311.

Luo, X., Huang, C.-Y., and Wang, L. (2013). Quantile regression for recurrent gap time data. *Biometrics*, 69(2):375–385.

Lyu, T., Luo, X., Xu, G., and Huang, C.-Y. (2018). Induced smoothing for rank-based regression with recurrent gap time data. *Statistics in Medicine*, 37(7):1086–1100.

Munk-Jørgensen, P. and Mortensen, P. B. (1997). The danish psychiatric central register. *Danish Medical Bulletin*, 44(1):82–84.

Nelson, W. (1995). Confidence limits for recurrence dataapplied to cost or number of product repairs. *Technometrics*, 37(2):147–157.

Newey, W. K., Hsieh, F., and Robins, J. M. (2004). Twicing kernels and a small bias property of semiparametric estimators. *Econometrica*, 72(3):947–962.

Parzen, M. I., Wei, L. J., and Ying, Z. (1994). A resampling method based on pivotal estimating functions. *Biometrika*, 81(2):341–350.

Peña, E. A., Strawderman, R. L., and Hollander, M. (2001). Nonparametric estimation with recurrent event data. *Journal of the American Statistical Association*, 96(456):1299–1315.

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(423):811–820.

Prentice, R. L. (1978). Linear rank tests with right censored data. *Biometrika*, 65(1):167–179.

Prentice, R. L., Williams, B. J., and Peterson, A. V. (1981a). On the regression analysis of multivariate failure time data. *Biometrika*, 68(2):373–379.

Prentice, R. L., Williams, B. J., and Peterson, A. V. (1981b). On the regression analysis of multivariate failure time data. *Biometrika*, 68(2):373–379.

Raboud, J., Reid, N., Coates, R. A., and Farewell, V. T. (1993). Estimating risks of progressing to AIDS when covariates are measured with error. *Journal of the Royal Statistical Society. Series A*, 156(3):393–406.

Rizopoulos, D. (2012). *Joint Models for Longitudinal and Time-to-Event Data: With Applications in R*. CRC Press. Taylor & Francis Group, LLC.

Schaubel, D. E., Zeng, D., and Cai, J. (2006). A semiparametric additive rates model for recurrent event data. *Lifetime Data Analysis*, 12(4):389–406.

Scheike, T. H. and Zhang, M.-J. (2002). An additive–multiplicative cox–aalen regression model. *Scandinavian Journal of Statistics*, 29(1):75–88.

Stanworth, S. J., Hudson, C. L., Estcourt, L. J., Johnson, R. J., and Wood, E. M. (2015). Risk of bleeding and use of platelet transfusions in patients with hematologic malignancies: recurrent event analysis. *Haematologica*, 100(6):740–747.

Strawderman, R. L. (2005). The accelerated gap times model. *Biometrika*, 92(3):647–666.

Sun, L. Q., Park, D. H., and Sun, J. G. (2006). The additive hazards model for recurrent gap times. *Statistica Sinica*, 16(3):919–932.

Tsiatis, A. A. (1990). Estimating regression parameters using linear rank tests for censored data. *The Annals of Statistics*, 18(1):354–372.

Tsiatis, A. A. and Davidian, M. (2004). Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica*, 14(3):809–834.

Tsiatis, A. A., Degruttola, V., and Wulfsohn, M. S. (1995). Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and CD4 counts in patients with AIDS. *Journal of the American Statistical Association*, 90(429):27–37.

Van Der Heijden, A. A., vant Riet, E., Bot, S. D., Cannegieter, S. C., Stehouwer, C. D., Baan, C. A., Dekker, J. M., and Nijpels, G. (2013). Risk of a recurrent cardiovascular event in individuals with type 2 diabetes or intermediate hyperglycemia. *Diabetes Care*, 36(11):3498–3502.

Van Der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. New York: Springer-Verlag.

Vonesh, E. F., Greene, T., and Schluchter, M. D. (2006). Shared parameter models for the joint analysis of longitudinal data and event times. *Statistics in Medicine*, 25(1):143–163.

Wang, M.-C. and Chang, S.-H. (1999). Nonparametric estimation of a recurrent survival function. *Journal of the American Statistical Association*, 94(445):146–153.

Wei, L., Ying, Z., and Lin, D. (1990). Linear regression analysis of censored survival data based on rank tests. *Biometrika*, 77(4):845–851.

Wulfsohn, M. S. and Tsiatis, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics*, 53(1):330–339.

Xu, J. and Zeger, S. L. (2001). Joint analysis of longitudinal data comprising repeated measures and times to events. *Journal of the Royal Statistical Society: Series C*, 50(3):375–387.

Yin, G. (2007). Model checking for additive hazards model with multivariate survival data. *Journal of Multivariate Analysis*, 98(5):1018–1032.

Ying, Z. (1993). A large sample study of rank estimation for censored regression data. *The Annals of Statistics*, 21(1):76–99.

Zeng, D. and Lin, D. (2006). Efficient estimation of semiparametric transformation models for counting processes. *Biometrika*, 93(3):627–640.

Zeng, D. and Lin, D. Y. (2008). Efficient resampling methods for nonsmooth estimating functions. *Biostatistics*, 9(2):355–363.

Appendix A

Proof of Theorem 1

We provide a brief proof of consistency and asymptotic normality of $\hat{\boldsymbol{\beta}}_{\mathrm{G}}^{(s)}$ by following the proofs for Theorem 1 and 2 in Johnson and Strawderman (2009). We assume the following regularity conditions:

Condition A1. The parameter space $\mathbb{B}$ containing $\boldsymbol{\beta}_0$ is a compact subset of $\mathbb{R}^p$.

Condition A2. $\|Z_i\| + m_i^*$ is uniformly bounded almost surely by a nonrandom constant $(i = 1, \ldots, n)$.

Condition A3. $\mathrm{Var}\,(\epsilon_{11}) < \infty$.

Condition A4. The matrix A and V defined in Theorem 1 exist and A is not singular.

Condition A5. Let $f_0(\cdot)$ denote the marginal density associated with model error term ϵ_{11}. Assume $f_0(\cdot)$ and $f_0'(\cdot)$ are bounded functions on $\mathbb{R}$ with

$$\int_{\mathbb{R}} \left\{ \frac{f_0'(t)}{f_0(t)} \right\}^2 f_0(t)dt < \infty.$$

Condition A6. The marginal distribution of C_i is absolutely continuous and has a uniformly bounded density $g_i(\cdot)$ on $\mathbb{R}$ for $i = 1, \ldots, n$.

Among the above conditions, A1, A2, A4, A5, A6 are standard conditions to ensure consistency and the asymptotic normality of the estimator from Equation (2.6) according to Johnson and Strawderman (2009). Since $|\mathrm{Cov}(\epsilon_{ij}, \epsilon_{ik})| \leq \mathrm{Var}(\epsilon_{11}), i =$

$1, \ldots, n, j, k = 1, \ldots, m_i^*$, Condition A3 ensures that the covariances between the error terms of recurrent events of the same person are bounded.

A.1 Proof of consistency

We know that the estimating function in Equation (2.6) is the gradient of convex objective function $L_{\mathrm{G}}(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i^*}\sum_{l=1}^{n}\sum_{k=1}^{m_l^*}(m_i^* m_l^*)^{-1}\delta_{ij}\{e_{lk}(\boldsymbol{\beta}) - e_{ij}(\boldsymbol{\beta})\}I\{e_{lk}(\boldsymbol{\beta}) \geq e_{ij}(\boldsymbol{\beta})\}$ which is continuous almost everywhere. Using a similar approach as the proofs for Lemmas 1 and 2 in Johnson and Strawderman (2009), we can prove that $\sup_{\boldsymbol{\beta}\in\mathbb{B}}|\frac{1}{n}L_{\mathrm{G}}(\boldsymbol{\beta}) - L_0(\boldsymbol{\beta})| \to 0$ almost surely where $L_0(\boldsymbol{\beta})$ is convex for $\boldsymbol{\beta} \in \mathbb{B}$ and $\sup_{\boldsymbol{\beta}\in\mathbb{B}}|\frac{1}{n}L_{\mathrm{G}}^{(s)}(\boldsymbol{\beta}) - L_0(\boldsymbol{\beta})| \to 0$ almost surely. Condition A4 implies that $L_0(\boldsymbol{\beta})$ is strictly convex at $\boldsymbol{\beta}_0$ and thus $\boldsymbol{\beta}_0$ is a unique minimizer of $L_0(\boldsymbol{\beta})$. Let $\hat{\boldsymbol{\beta}}_{\mathrm{G}}$ be the minimizer of $L_{\mathrm{G}}(\boldsymbol{\beta})$ and $\hat{\boldsymbol{\beta}}_{\mathrm{G}}^{(s)}$ be the minimizer of $L_{\mathrm{G}}^{(s)}(\boldsymbol{\beta})$. According to Theorem II.1 and Corollary II.2 in Andersen and Gill (1982), we can conclude that both $\hat{\boldsymbol{\beta}}_{\mathrm{G}}$ and $\hat{\boldsymbol{\beta}}_{\mathrm{G}}^{(s)}$ converge almost surely to $\boldsymbol{\beta}_0$.

A.2 Proof of asymptotic normality

First we prove the asymptotic normality of $n^{1/2}(\hat{\boldsymbol{\beta}}_{\mathrm{G}} - \boldsymbol{\beta}_0)$. Using similar arguments as in Theorem 2 in Ying (1993), we can show that

$$n^{1/2}(\hat{\boldsymbol{\beta}}_{\mathrm{G}} - \boldsymbol{\beta}_0) = -A^{-1}n^{-1/2}U_{\mathrm{G}}(\boldsymbol{\beta}_0) + o_p(1 + \sqrt{n}\|\hat{\boldsymbol{\beta}}_{\mathrm{G}} - \boldsymbol{\beta}_0\|).$$

We define

$$\begin{aligned} M_i^*(\boldsymbol{\beta}, t) &= \frac{1}{m_i^*}\sum_{j=1}^{m_i^*} M_{ij}(\boldsymbol{\beta}, t), \\ M_{ij}(\boldsymbol{\beta}, t) &= N_{ij}(\boldsymbol{\beta}, t) - \int_{-\infty}^{t} R_{ij}(\boldsymbol{\beta}, u)\lambda_0(u)du, \text{ and} \end{aligned}$$

$$\bar{z}(\boldsymbol{\beta}, t) = \frac{\mathrm{E}\{S_1^*(\boldsymbol{\beta}, t)\}}{\mathrm{E}\{S_0^*(\boldsymbol{\beta}, t)\}},$$

where $\lambda_0(\cdot)$ is the common hazard function of $\epsilon_{ij}, i = 1, \ldots, n, j = 1, \ldots, m_i^*$. Let $s_0(\boldsymbol{\beta}, x) = \mathrm{E}\{S_0(\boldsymbol{\beta}, x)\}$ and $s_1(\boldsymbol{\beta}, x) = \mathrm{E}\{S_1(\boldsymbol{\beta}, x)\}$. Following similar argument as in Luo and Huang (2011), we can show that $\mathrm{E}\left\{S_0^*(\boldsymbol{\beta}, t)\right\} = \mathrm{E}\left[n^{-1}\sum_{l=1}^n \mathrm{I}\{e_{l1}(\boldsymbol{\beta}) \geq t\}\right] = s_0(\boldsymbol{\beta}, t)$ and $\mathrm{E}\{S_1^*(\boldsymbol{\beta}, t)\} = \mathrm{E}\left[n^{-1}\sum_{l=1}^n Z_l \mathrm{I}\{e_{l1}(\boldsymbol{\beta}) \geq t\}\right] = s_1(\boldsymbol{\beta}, t)$, and hence we prove that $\bar{z}(\boldsymbol{\beta}, t) = s_1(\boldsymbol{\beta}, t)/s_0(\boldsymbol{\beta}, t)$. Then, following Jin et al. (2006), we have

$$\frac{1}{n}U_{\mathrm{G}}(\boldsymbol{\beta}_0) = \frac{1}{n}\sum_{i=1}^n u_i + o_p(n^{-1/2}),$$

where

$$u_i = \int_{-\infty}^{\tau} s_0(\boldsymbol{\beta}_0, t)\{Z_i - \bar{z}(\boldsymbol{\beta}, t)\} dM_i^*(\boldsymbol{\beta}, t).$$

According to central limit theorem, we have $\sqrt{n}\{n^{-1}U_{\mathrm{G}}(\boldsymbol{\beta}_0)\}$ converge in distribution to $N(0, V)$, thus, $\sqrt{n}(\hat{\boldsymbol{\beta}}_{\mathrm{G}} - \boldsymbol{\beta}_0)$ converges in distribution to $N(0, A^{-1}V(A^{-1})^\intercal)$.

Next we prove the asymptotic normality of $\sqrt{n}(\hat{\boldsymbol{\beta}}_{\mathrm{G}}^{(s)} - \boldsymbol{\beta}_0)$ and show that $\sqrt{n}(\hat{\boldsymbol{\beta}}_{\mathrm{G}}^{(s)} - \boldsymbol{\beta}_0)$ and $\sqrt{n}(\hat{\boldsymbol{\beta}}_{\mathrm{G}} - \boldsymbol{\beta}_0)$ converge to the same limiting distribution. First, following a similar approach as in Johnson and Strawderman (2009), Lemma 3, we can prove that $\|\dot{U}_{\mathrm{G}}^{(s)}(\boldsymbol{\beta}_0) - A\| \to 0$. Second, since we know that $A^{-1}\left\{n^{-1/2}U_{\mathrm{G}}(\boldsymbol{\beta}_0)\right\}$ is asymptotically normal with mean zero and variance $A^{-1}VA^{-1}$, then if we can prove

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{\mathrm{G}}^{(s)} - \boldsymbol{\beta}_0) + A^{-1}n^{-1/2}U_{\mathrm{G}}(\boldsymbol{\beta}_0) \to 0 \tag{A.1}$$

in probability, it will imply that $\sqrt{n}(\hat{\boldsymbol{\beta}}_{\mathrm{G}}^{(s)} - \boldsymbol{\beta}_0)$ converge in distribution to $N(0, A^{-1}VA^{-1})$.

Following Arcones (1998), let $G_n(\boldsymbol{\beta}) = L_{\mathrm{G}}^{(s)}(\boldsymbol{\beta})$, $\eta_n = n^{-1/2}U_{\mathrm{G}}(\boldsymbol{\beta}_0)$, $M_n = n^{1/2}I_p$,

$V_n = (1/2)A$. Then (A.1) can be written as

$$M_n \left(\hat{\boldsymbol{\beta}}_{\mathrm{G}}^{(s)} - \boldsymbol{\beta}_0 \right) + \frac{1}{2} V_n^{-1} \eta_n \to 0 \tag{A.2}$$

in probability. According to Theorem 3 in Arcones (1998), (A.2) holds if the following conditions are met:

Condition B1. $G_n(\boldsymbol{\beta})$ is convex and $\hat{\boldsymbol{\beta}}_{\mathrm{G}}^{(s)}$ is a sequence satisfying $G_n(\hat{\boldsymbol{\beta}}_{\mathrm{G}}^{(s)}) \leq \inf_{\boldsymbol{\beta} \in \mathbb{B}} G_n(\boldsymbol{\beta}) + o_p(1)$.

Condition B2. $\eta_n = O_p(1)$, $\liminf_{n\to\infty} \inf_{|\boldsymbol{\beta}|=1} \boldsymbol{\beta}' V_n \boldsymbol{\beta} > 0$ and $\limsup_{n\to\infty} \sup_{|\boldsymbol{\beta}|=1} \boldsymbol{\beta}' V_n \boldsymbol{\beta} < \infty$.

Condition B3. For each $\boldsymbol{\beta} \in \mathbb{R}^P$, $G_n(\boldsymbol{\beta}_0 + M_n^{-1}\boldsymbol{\beta}) - G_n(\boldsymbol{\beta}_0) - \boldsymbol{\beta}'\eta_n - \boldsymbol{\beta}' V_n \boldsymbol{\beta} = o_p(1)$.

It is easy to show that Conditions B1 and B2 hold when Conditions A1-A6 hold. We need to prove that Condition B3 holds. By Taylor expansion, we have

$$\begin{aligned} G_n(\boldsymbol{\beta}_0 + M_n^{-1}\boldsymbol{\beta}) =& G_n(\boldsymbol{\beta}_0) + (M_n^{-1}\boldsymbol{\beta})^{\intercal} \left\{ \frac{\partial}{\partial \boldsymbol{\beta}} G_n(\boldsymbol{\beta}_0) \right\} + \frac{1}{2} (M_n^{-1}\boldsymbol{\beta})^{\intercal} \left\{ \frac{\partial^2}{\partial \boldsymbol{\beta}^2} G_n(\boldsymbol{\beta}_n^*) \right\} \\ & (M_n^{-1}\boldsymbol{\beta}) + o_p(1), \end{aligned}$$

then

$$G_n \left(\boldsymbol{\beta}_0 + M_n^{-1}\boldsymbol{\beta} \right) - G_n(\boldsymbol{\beta}_0) - \boldsymbol{\beta}' \left\{ n^{-1/2} U_{\mathrm{G}}^{(s)}(\boldsymbol{\beta}_0) \right\} - \frac{1}{2} \boldsymbol{\beta}' \left\{ \dot{U}_{\mathrm{G}}^{(s)}(\boldsymbol{\beta}_n^*) \right\} \boldsymbol{\beta} = o_p(1), \tag{A.3}$$

where $\|\boldsymbol{\beta}_n^* - \boldsymbol{\beta}_0\| \leq \|M_n^{-1}\boldsymbol{\beta}\|$. Since $\{n^{-1} U_{\mathrm{G}}^{(s)}(\boldsymbol{\beta})\}$ is a sequence of bounded, continuously differentiable functions and $\|\dot{U}_{\mathrm{G}}^{(s)}(\boldsymbol{\beta}_0) - A\| \to 0$, $\dot{U}_{G}^{(s)}(\boldsymbol{\beta}_n^*)$ in (A.3) can be replaced by A. Thus we have

$$G_n(\boldsymbol{\beta}_0 + M_n^{-1}\boldsymbol{\beta}) - G_n(\boldsymbol{\beta}_0) - \boldsymbol{\beta}' \left\{ n^{-1/2} U_{\mathrm{G}}^{(s)}(\boldsymbol{\beta}_0) \right\} - \boldsymbol{\beta}' V_n \boldsymbol{\beta} = o_p(1). \tag{A.4}$$

Then Condition B3 holds if we can prove

$$n^{-1/2}\|U_{\mathrm{G}}^{(s)}(\boldsymbol{\beta}_0) - U_{\mathrm{G}}(\boldsymbol{\beta}_0)\| \to 0 \tag{A.5}$$

in probability.

By the definition of $U_{\mathrm{G}}^{(s)}(\boldsymbol{\beta}_0)$, we have

$$U_{\mathrm{G}}^{(s)}(\boldsymbol{\beta}_0) - U_{\mathrm{G}}(\boldsymbol{\beta}_0) = \int_{\mathbb{R}^P} \{U_{\mathrm{G}}(\boldsymbol{\beta}_0 + n^{-1/2}u) - U_{\mathrm{G}}(\boldsymbol{\beta}_0)\}\phi(u)du, \tag{A.6}$$

where $\phi(u)$ is the pdf of W. Define $K_n(u;\boldsymbol{\beta}_0,\Theta) = \|\frac{1}{\sqrt{n}}\{U_{\mathrm{G}}(\boldsymbol{\beta}_0 + n^{-1/2}u) - U_{\mathrm{G}}(\boldsymbol{\beta}_0)\} - \Theta u\|$ where Θ is a fixed matrix that satisfies $\|\Theta\| \leq M$ and $M < \infty$. We know that $\mathrm{E}(W) = \int_{\mathbb{R}^P} u\phi(u) = 0$, so we can derive

$$\begin{aligned}
n^{-1/2}\|U_{\mathrm{G}}^{(s)}(\boldsymbol{\beta}_0) - U_{\mathrm{G}}(\boldsymbol{\beta}_0)\| =& \Bigg\|\int_{\mathbb{R}^P} \left[\frac{1}{\sqrt{n}}\left\{U_{\mathrm{G}}(\boldsymbol{\beta}_0 + n^{-1/2}u) - U_{\mathrm{G}}(\boldsymbol{\beta}_0)\right\} - \Theta u\right] \\
&\phi(u)du + \int_{\mathbb{R}^P} \Theta u\phi(u)du\Bigg\| \\
\leq& \Bigg\|\int_{\mathbb{R}^P} \left[\frac{1}{\sqrt{n}}\left\{U_{\mathrm{G}}(\boldsymbol{\beta}_0 + n^{-1/2}u) - U_{\mathrm{G}}(\boldsymbol{\beta}_0)\right\} - \Theta u\right] \\
&\phi(u)du\Bigg\| + \Bigg\|\int_{\mathbb{R}^P} \Theta u\phi(u)du\Bigg\| \\
=& \int_{\mathbb{R}^P} K_n(u;\boldsymbol{\beta}_0,\Theta)\phi(u)du \\
=& I_1 + I_2,
\end{aligned}$$

where $I_1 = \int_{\|u\|\leq\epsilon_n} K_n(u;\boldsymbol{\beta}_0,\Theta)\phi(u)du$ and $I_2 = \int_{\|u\|>\epsilon_n} K_n(u;\boldsymbol{\beta}_0,\Theta)\phi(u)du$ for any $\epsilon_n > 0$. Following a similar approach as in Theorem 2 in Ying (1993), we have

$$\sup_{\|b-\boldsymbol{\beta}_0\|\leq d_n} \frac{\left\|\frac{1}{\sqrt{n}}\left\{U_{\mathrm{G}}(b) - U_{\mathrm{G}}(\boldsymbol{\beta}_0)\right\} - A\sqrt{n}(b-\boldsymbol{\beta}_0)\right\|}{1 + n^{1/2}\|b-\boldsymbol{\beta}_0\|} = o_p(1) \tag{A.7}$$

for any positive sequence $d_n \to 0$. Let $b = \boldsymbol{\beta}_0 + n^{-1/2}u$, $d_n = n^{-1/2}\epsilon_n$, $\Theta = A$ and suppose that $\epsilon_n = o(\sqrt{n})$, then it follows Equation (A.7) that

$$\sup_{\|u\| \le \epsilon_n} \frac{K_n(u; \boldsymbol{\beta}_0, \Theta)}{1 + \|u\|} = o_p(1),$$

which implies $I_1 \to 0$ in probability.

Let $\Theta = A$. Because of the triangle inequality, we have

$$\begin{aligned}
I_2 =& \int_{\|u\|>\epsilon_n} \left\| \frac{1}{\sqrt{n}} \left\{ U_{\mathrm{G}} \left(\boldsymbol{\beta}_0 + n^{-1/2}u \right) - U_{\mathrm{G}}(\boldsymbol{\beta}_0) \right\} - Au \right\| \phi(u) du \\
\le& \sqrt{n} \int_{\|u\|>\epsilon_n} \left\| \frac{1}{n} \left\{ U_{\mathrm{G}} \left(\boldsymbol{\beta}_0 + n^{-1/2}u \right) - U_{\mathrm{G}}(\boldsymbol{\beta}_0) \right\} \right\| \phi(u) du \\
& + \int_{\|u\|>\epsilon_n} \|Au\| \phi(u) du \\
\le& \sup_{\|u\|>\epsilon_n} \left\| \frac{1}{n} \left\{ U_{\mathrm{G}}(\boldsymbol{\beta}_0 + n^{-1/2}u) - U_{\mathrm{G}}(\boldsymbol{\beta}_0) \right\} \right\| \sqrt{n} \int_{\|u\|>\epsilon_n} \phi(u) du \\
& + \|A\| \int_{\|u\|>\epsilon_n} \|u\| \phi(u) du. \qquad \text{(A.8)}
\end{aligned}$$

Since there is a constant $Q < \infty$ such that $n^{-1}U_{\mathrm{G}}(\boldsymbol{\beta}) < Q$ based on Condition A2, we can derive that the first component in (A.8) is $\le 2Q\sqrt{n}P(\|W\| > \epsilon_n)$. It is easy to show that a sequence of ϵ_n can be selected so that $\epsilon_n = o(\sqrt{n})$, $\epsilon_n \to \infty$ as $n \to \infty$ and $2Q\sqrt{n}P(\|W\| > \epsilon_n) \to 0$, $\int_{\|u\|>\epsilon_n} \|u\|\phi(u)du \to 0$ as $n \to \infty$. Thus, we have shown that (A.8) $\to 0$ in probability, which implies $I_2 \to 0$ in probability, then (A.5) holds. Therefore, the asymptotic normality of $\hat{\boldsymbol{\beta}}_{\mathrm{G}}^{(s)}$ is proved.

Appendix B

Proof of Theorem 2

Similar to the proofs in Li et al. (2016), we impose the following assumptions:

1. $\{N_i(\cdot), O_i(\cdot), Y_i(\cdot), \mathbf{Z}_i(\cdot)\}, i = 1, \ldots, n$, are independent and identically distributed.

2. $N_i(\tau)$ is bounded. Define $\lambda^c(\cdot)$ as the rate function of $N_i(\cdot)$ and $\lambda^c(\cdot)$ is of bounded variation.

3. The true parameter $\boldsymbol{\beta}_0$ is in a compact set $\mathcal{B}$ in $\mathcal{R}^p$ and the baseline rate function $\lambda_0(t)$ is absolutely continuous.

4. For each element in the covariates $\mathbf{Z}_i(t)$, the covariate process $Z_{ij}(t)$ has uniformly bounded total variation, namely, $\int_0^\tau |dZ_{ij}(t)| + |Z_{ij}(0)| \leq c$ for some $c > 0$ for all i and j . Without loss of generality, we assume $Z_{ij}(t) \geq 0$.

5. The censoring time C_i is independent of $N_i^*(\cdot)$ conditional on $\mathbf{Z}_i(\cdot)$ with $G(\tau) = P(C_i \geq \tau) > 0$.

6. The function $s^{(k)}(t) = \mathrm{E}\{Y_i(t)\mathbf{Z}_i(t)^{\otimes k}\}, k = 0, 1, 2$ have bounded second derivatives for $t \in [0, \tau]$.

7. The observation time process $O_i(t)$ is independent of $\{N_i^*(\cdot), Y_i(\cdot), \mathbf{Z}_i(\cdot)\}$ and is bounded. Moreover, the covariate collection rate function $m(t)$, defined by $m(t)dt = \mathrm{E}\{dO_i(t)\}$, is positive and has bounded second derivative for $t \in [0, \tau]$.

8. The kernel function $K(\cdot)$ is a symmetric density function with bounded support which satisfies: $\int_{-1}^{1} K(t)dt = 1$, $\int_{-1}^{1} tK(t)dt = 0$ and $\int_{-1}^{1} t^2K(t)dt$ is a positive constant.

9. The bandwidth $h = O(n^{-v})$, where $1/4 < v < 1/2$.

B.1 Proof of consistency

Define $\Psi(u) = \mathrm{E}\{Y(u)\mathbf{Z}(u)^{\otimes k}\}m(u)$, then $s^{(k)}(t)m(t) = \Psi(t)$. The expectation of the kernel smoothed processes $\mathrm{E}\{\widehat{S}_h^{(k)}(t)\} = \int_0^\tau K_h(t-u)\mathrm{E}\{Y(u)\mathbf{Z}(u)^{\otimes k}\}m(u)du$, then we have

$$\begin{aligned}\mathrm{E}\{\widehat{S}_h^{(k)}(t)\} &= \int_0^\tau K_h(t-u)\Psi(u)du = \int_{\frac{t-\tau}{h}}^{\frac{t}{h}} K(\bar{u})\Psi(t-h\bar{u})d\bar{u} \\ &= \Psi(t)\int_{\frac{t-\tau}{h}}^{\frac{t}{h}} K(\bar{u})d\bar{u} - \int_{\frac{t-\tau}{h}}^{\frac{t}{h}} h\bar{u}K(\bar{u})\Psi'(t)d\bar{u} + \int_{\frac{t-\tau}{h}}^{\frac{t}{h}} h^2\bar{u}^2K(\bar{u})d\bar{u}\Psi''(t^*).\end{aligned}$$

It is easy to see that $\sup_{t\in[h,\tau-h]}|\mathrm{E}\{\widehat{S}_h^{(k)}(t)\} - s^{(k)}(t)m(t)| = O(h^2)$ under the assumption (8). Also, it is straightforward to show that $\sup_{t\in[0,h)}|s^{(k)}(t)m(t) - s^{(k)}(h)m(h)|$ and $\sup_{t\in(\tau-h,\tau]}|s^{(k)}(t)m(t) - s^{(k)}(\tau-h)m(\tau-h)| = O(h)$.

Next, we show the convergence of $\widehat{S}_h^{(k)}(t) - \mathrm{E}\{\widehat{S}_h^{(k)}(t)\}$. We define $\widehat{R}^{(k)}(t) = n^{-1}\sum_{i=1}^n \int_0^\tau Y_i(u)\,\mathbf{Z}_i(u)^{\otimes k}dO_i(u)$ and $r^{(k)}(t) = \mathrm{E}\{\int_0^\tau Y(u)\mathbf{Z}(u)^{\otimes k}dO(u)\}$, so $\widehat{S}_h^{(k)}(t) = \int_0^\tau K_h(t-u)d\widehat{R}^{(k)}(u)$ and $\mathrm{E}\{\widehat{S}_h^{(k)}(t)\} = \int_0^\tau K_h(t-u)dr^{(k)}(u)$. Then we have

$$\sup_{t\in[h,\tau-h]}|\widehat{S}_h^{(k)}(t) - \mathrm{E}\{\widehat{S}_h^{(k)}(t)\}| \le h^{-1}\sup_{t\in[0,\tau]}|\widehat{R}^{(k)}(t) - r^{(k)}(t)|V(K), \tag{B.1}$$

where $V(K)$ is the variation of the kernel function. Also, since the function classes $\mathcal{F}_k = \{\int_0^t Y(u)\mathbf{Z}(u)^{\otimes k}dO(u) : t \in [0,\tau]\}$ are monotone, by Theorem 2.14.9 in Van Der Vaart and Wellner (1996), $P\left(\sup_{t\in[0,\tau]}\sqrt{n}|\widehat{R}^{(k)}(t) - r^{(k)}(t)| > x\right) \le c_k x^{v_k}e^{-b_kx^2}$, where c_k, v_k, b_k are constants. Therefore for any ϵ, we have

$$P\left(\sup_{t\in[0,\tau]} h^{-1}|\widehat{R}^{(k)}(t) - r^{(k)}(t)| > \epsilon\right) = P\left(\sup_{t\in[0,\tau]} \sqrt{n}|\widehat{R}^{(k)}(t) - r^{(k)}(t)| > \sqrt{n}h\epsilon\right) \\ \leq c_k(\sqrt{n}h\epsilon)^{v_k} e^{-b_k(\sqrt{n}h\epsilon)^2}. \tag{B.2}$$

According to (B.1) and (B.2), we have $\sup_{t\in[h,\tau-h]} |\widehat{S}_h^{(k)}(t) - \mathrm{E}\{\widehat{S}_h^{(k)}(t)\}|$ converges to 0 when $nh^2 \to \infty$. Previously we have shown that $\sup_{t\in[h,\tau-h]} |\mathrm{E}\{\widehat{S}_h^{(k)}(t)\} - s^{(k)}(t)m(t)| = O(h^2)$, so the consistency of $\widehat{S}_h^{(k)}(t)$ has been proved. By the law of large numbers, we know that $n^{-1}\sum_{i=1}^n N_i(t)$ converges to $\mathrm{E}\{N_i(t)\}$ and $n^{-1}\sum_{i=1}^n \int_0^\tau \mathbf{Z}_i(t)\, dN_i(t)$ converges to $\int_0^\tau \mathrm{E}\{\mathbf{Z}_i(t)dN_i(t)\}$. Thus, we have that $\widehat{\boldsymbol{\beta}}$ converges in probability to $\boldsymbol{\beta}_0$.

B.2 Proof of asymptotic normality

We first prove that $\sqrt{n}\widehat{U}_h(\boldsymbol{\beta}_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \phi_i(\boldsymbol{\beta}_0) + o_p(1)$. From Equation (3.5), we have

$$\begin{aligned}\sqrt{n}\widehat{U}_h(\boldsymbol{\beta}_0) =& \frac{1}{\sqrt{n}}\sum_{i=1}^n \int_0^\tau \mathbf{Z}_i(t)dN_i(t) - \frac{1}{\sqrt{n}}\sum_{i=1}^n \int_0^\tau \widehat{\xi}_h^{(1)}(t)dN_i(t) \\ & - \frac{1}{\sqrt{n}}\sum_{i=1}^n \int_0^\tau Y_i(t)\widehat{\xi}_h^{(2)}(t)dt\boldsymbol{\beta}_0 + \frac{1}{\sqrt{n}}\sum_{i=1}^n \int_0^\tau Y_i(t)\widehat{\xi}_h^{(1)}(t)^{\otimes 2}dt\boldsymbol{\beta}_0 \\ & \overset{def}{=} I_1 + I_2 + I_3 + I_4.\end{aligned}$$

We show that

$$I_2 = -\frac{1}{\sqrt{n}}\sum_{i=1}^n \int_0^\tau \widehat{\xi}_h^{(1)}(t)dN_i(t)$$

$$
\begin{aligned}
&= -\sqrt{n}\int_0^\tau \widehat{\xi}_h^{(1)}(t) d\left\{\frac{1}{n}\sum_{i=1}^n N_i(t) - \mathrm{E}\{N_i(t)\}\right\} - \sqrt{n}\int_0^\tau \widehat{\xi}_h^{(1)}(t) d\mathrm{E}\{N_i(t)\} \\
&= -\sqrt{n}\int_0^\tau \frac{s^{(1)}(t)}{s^{(0)}(t)} d\left\{\frac{1}{n}\sum_{i=1}^n N_i(t) - \mathrm{E}\{N_i(t)\}\right\} - \sqrt{n}\int_0^\tau \widehat{\xi}_h^{(1)}(t) d\mathrm{E}\{N_i(t)\} \\
&\quad + o_p(1).
\end{aligned}
$$

Since $\lambda^c(t)$ is the rate function of $N_i(t)$, we have $\lambda^c(t)dt = d\mathrm{E}\{N_i(t)\}$. Also,

$$
\begin{aligned}
&\sqrt{n}\int_0^\tau \widehat{\xi}_h^{(1)}(t)\lambda^c(t)dt - \sqrt{n}\int_0^\tau \frac{s^{(1)}(t)}{s^{(0)}(t)}\lambda^c(t)dt \\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^n \left\{\int_0^\tau \frac{\lambda^c(t)}{s^{(0)}(t)m(t)} Y_i(t)\mathbf{Z}_i(t)dO_i(t) - \int_0^\tau \frac{s^{(1)}(t)\lambda^c(t)}{s^{(0)}(t)^2 m(t)} Y_i(t)dO_i(t)\right\} \\
&+ o_p(1),
\end{aligned}
$$

thus we have $I_2 = \frac{1}{\sqrt{n}}\sum_{i=1}^n \phi_{2i} + o_p(1)$, where

$$
\phi_{2i} = -\int_0^\tau \frac{s^{(1)}(t)}{s^{(0)}(t)} dN_i(t) - \int_0^\tau \frac{\lambda^c(t)}{s^{(0)}(t)m(t)} Y_i(t)\mathbf{Z}_i(t)dO_i(t) + \int_0^\tau \frac{s^{(1)}(t)\lambda^c(t)}{s^{(0)}(t)^2 m(t)} Y_i(t)dO_i(t).
$$

Define $R(t) = \mathrm{E}\{Y_i(t)\}$, we have

$$
\begin{aligned}
I_3 &= -\frac{1}{\sqrt{n}}\sum_{i=1}^n \int_0^\tau Y_i(t)\widehat{\xi}_h^{(2)}(t)dt\boldsymbol{\beta}_0 \\
&= -\frac{1}{\sqrt{n}}\sum_{i=1}^n \int_0^\tau Y_i(t)\frac{s^{(2)}(t)}{s^{(0)}(t)}dt\boldsymbol{\beta}_0 + \sqrt{n}\int_0^\tau R(t)\frac{s^{(2)}(t)}{s^{(0)}(t)}dt\boldsymbol{\beta}_0 \\
&\quad - \sqrt{n}\int_0^\tau \widehat{\xi}_h^{(2)}(t)R(t)dt\boldsymbol{\beta}_0 + o_p(1) \\
&= -\frac{1}{\sqrt{n}}\sum_{i=1}^n \left\{\int_0^\tau Y_i(t)\frac{s^{(2)}(t)}{s^{(0)}(t)}dt - \int_0^\tau \frac{R(t)}{s^{(0)}(t)m(t)} Y_i(t)\mathbf{Z}_i(t)^{\otimes 2}dO_i(t) + \right. \\
&\qquad\qquad \left. \int_0^\tau \frac{s^{(2)}(t)R(t)}{s^{(0)}(t)^2 m(t)} Y_i(t)dO_i(t)\right\}\boldsymbol{\beta}_0 + o_p(1)
\end{aligned}
$$

$$\stackrel{def}{=} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_{3i}(\boldsymbol{\beta}_0) + o_p(1),$$

and

$$\begin{aligned}
I_4 =& \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^\tau Y_i(t) \widehat{\xi}_h^{(1)}(t)^{\otimes 2} dt \boldsymbol{\beta}_0 \\
=& \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^\tau Y_i(t) \frac{s^{(1)}(t)^{\otimes 2}}{s^{(0)}(t)^2} dt \boldsymbol{\beta}_0 - \sqrt{n} \int_0^\tau R(t) \frac{s^{(1)}(t)^{\otimes 2}}{s^{(0)}(t)^2} dt \boldsymbol{\beta}_0 \\
&+ \sqrt{n} \int_0^\tau R(t) \widehat{\xi}_h^{(1)}(t)^{\otimes 2} dt \boldsymbol{\beta}_0 + o_p(1) \\
=& \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \int_0^\tau Y_i(t) \frac{s^{(1)}(t)^{\otimes 2}}{s^{(0)}(t)^2} dt + \int_0^\tau \frac{2 s^{(1)}(t) R(t)}{s^{(0)}(t)^2 m(t)} Y_i(t) \mathbf{Z}_i(t) dO_i(t) \right. \\
&\left. - \int_0^\tau \frac{2 s^{(1)}(t)^2 R(t)}{s^{(0)}(t)^3 m(t)} Y_i(t) dO_i(t) \right\} \boldsymbol{\beta}_0 + o_p(1) \\
\stackrel{def}{=}& \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_{4i}(\boldsymbol{\beta}_0) + o_p(1).
\end{aligned}$$

Thus we have $\sqrt{n}\widehat{U}_h(\boldsymbol{\beta}_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_i(\boldsymbol{\beta}_0) + o_p(1)$, where $\phi_i(\boldsymbol{\beta}_0) = \int_0^\tau \mathbf{Z}_i(t) dN_i(t) + \phi_{2i}(\boldsymbol{\beta}_0) + \phi_{3i}(\boldsymbol{\beta}_0) + \phi_{4i}(\boldsymbol{\beta}_0)$. Therefore, $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ converges in distribution to a normal random variable with mean zero and variance $\Sigma = A(\boldsymbol{\beta}_0)^{-1} V(\boldsymbol{\beta}_0) \{A(\boldsymbol{\beta}_0)^{-1}\}^\intercal$, where $A(\boldsymbol{\beta}_0) = \int_0^\tau R(t) \left[\frac{s^{(2)}(t)}{s^{(0)}(t)} - \left\{ \frac{s^{(1)}(t)}{s^{(0)}(t)} \right\}^{\otimes 2} \right] dt$ and $V(\boldsymbol{\beta}_0) = E\{\phi_1(\boldsymbol{\beta}_0) \phi_1(\boldsymbol{\beta}_0)^\intercal\}$.

Appendix C

Proof of Theorem 3

Assumptions:

1. $\{N_i(\cdot), O_i(\cdot), Y_i(\cdot), \mathbf{W}_i(\cdot)\}$ are independent and identically distributed.

2. $N_i(\tau)$ is bounded. Define $\lambda^c(\cdot)$ as the rate function of $N_i(\cdot)$ and $\lambda^c(\cdot)$ is of bounded variation.

3. The true parameter $\boldsymbol{\theta}_0$ is in a compact set Θ in $\mathcal{R}^p$ and the baseline rate function $\lambda_0(t)$ is absolutely continuous.

4. For each element in the covariate vector $\mathbf{W}_i(t)$, the covariate process $W_{ij}(t)$ has uniformly bounded total variation, namely, $\int_0^\tau |dW_{ij}(t)| + |W_{ij}(0)| \leq c$ for some $c > 0$ for all i and j. Without loss of generality, we assume $W_{ij}(t) \geq 0$.

5. The censoring time C_i is independent of $N_i^*(\cdot)$ conditional on $\mathbf{W}_i(\cdot)$ with $G(\tau) = P(C_i \geq \tau) > 0$.

6. The functions $s_z^{(k)}(t) = \mathrm{E}[Y_i(t)\mathbf{Z}_i(t)^k]$, $s_x^{(k)}(t, \boldsymbol{\beta}) = \mathrm{E}[Y_i(t)\mathbf{X}_i(t)^k \exp\{\boldsymbol{\beta}^\intercal \mathbf{X}_i(t)\}]$, $k = 0, 1$ and $s_{z2x}(t, \boldsymbol{\beta}) = \mathrm{E}[Y_i(t)\mathbf{Z}_i(t)^{\otimes 2} \exp\{-\boldsymbol{\beta}^\intercal \mathbf{X}_i(t)\}]$, $s_{zx}(t) = \mathrm{E}[Y_i(t)\mathbf{X}_i(t)\mathbf{Z}_i(t)^\intercal]$ have bounded second derivatives for $t \in [0, \tau]$.

7. The observation time process $O_i(t)$ is independent of $\{N_i^*(\cdot), Y_i(\cdot), \mathbf{W}_i(\cdot)\}$ and is

bounded. Moreover, the covariate collection function $m(t)$, defined by $m(t)dt = \mathrm{E}[dO_i(t)]$, is positive and has bounded second derivative for $t \in [0, \tau]$.

8. The matrix $A = \mathrm{E}\{-\frac{\partial U(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\}$ is nonsingular.

9. The kernel function $K(\cdot)$ is a symmetric density function with bounded support which satisfies: $\int_{-1}^{1} K(t)dt = 1$, $\int_{-1}^{1} tK(t)dt = 0$ and $\int_{-1}^{1} t^2 K(t)dt$ is a positive constant.

10. The bandwidth $h = O(n^{-v})$, where $1/4 < v < 1/2$.

C.1 Proof of consistency

To show the consistency of $\widehat{\boldsymbol{\theta}}_h$, it is sufficient to prove that the processes that constitute the estimating function $\widehat{U}(\boldsymbol{\theta})$, including $n^{-1}\sum_{i=1}^{n}\int_0^\tau \mathbf{Z}_i(t)\exp\{-\boldsymbol{\beta}^\intercal \mathbf{X}_i(t)\}dN_i(u)$, $n^{-1}\sum_{i=1}^{n}\int_0^\tau \mathbf{X}_i(u)dN_i(u)$, $n^{-1}\sum_{i=1}^{n} dN_i(u)$, $\widehat{S}_{z,h}^{(k)}(t)$, $\widehat{S}_{x,h}^{(k)}(t,\boldsymbol{\beta})$, $k = 0, 1$, $\widehat{S}_{z2x,h}(t,\boldsymbol{\beta})$ and $\widehat{S}_{zx,h}(t)$, converge to their limits uniformly. We know $\boldsymbol{\theta}_0 = (\boldsymbol{\gamma}_0^\intercal, \boldsymbol{\beta}_0^\intercal)^\intercal$ where $\boldsymbol{\gamma}_0$ is a $m\times 1$ vector, $\boldsymbol{\beta}_0$ is a $q\times 1$ vector and $m+q = p$. Since $\boldsymbol{\theta}_0$ is in a compact set Θ in $\mathcal{R}^p$ by assumption 3, $\boldsymbol{\beta}_0$ is contained in a compact set $\mathcal{B}$ in $\mathcal{R}^q$. The function classes $\mathcal{F}_{z,k} = \{\int_0^t Y(u)\mathbf{Z}(u)^{\otimes k}dO(u) : t \in [0,\tau]\}$, $\mathcal{F}_{zx} = \{\int_0^t Y(u)\mathbf{X}(u)\mathbf{Z}(u)^\intercal dO(u) : t \in [0,\tau]\}$ are monotone and $\mathcal{F}_{x,k} = \{\int_0^t Y(u)\mathbf{X}(u)^{\otimes k}\exp\{\boldsymbol{\beta}\mathbf{X}(u)\}dO(u) : \boldsymbol{\beta} \in \mathcal{B}, t \in [0,\tau]\}$, $\mathcal{F}_{z2x} = \{\int_0^t Y(u)\mathbf{Z}(u)^{\otimes 2}\exp\{-\boldsymbol{\beta}\mathbf{X}(u)\}dO(u) : \boldsymbol{\beta} \in \mathcal{B}, t \in [0,\tau]\}$ have bracketing number of polynomial order $1/\epsilon^4$.

According to Theorem 2.14.9 in Van Der Vaart and Wellner (1996) and following similar steps in the Appendix of Li et al. (2016), we can show that $\sup_{t\in[h,\tau-h]}|\widehat{S}_{z,h}^{(k)}(t) - \mathrm{E}\{\widehat{S}_{z,h}^{(k)}(t)\}|$, $\sup_{\boldsymbol{\beta}\in\mathcal{B},t\in[h,\tau-h]}|\widehat{S}_{x,h}^{(k)}(t,\boldsymbol{\beta}) - \mathrm{E}\{\widehat{S}_{x,h}^{(k)}(t,\boldsymbol{\beta})\}|$, $\sup_{\boldsymbol{\beta}\in\mathcal{B},t\in[h,\tau-h]}|\widehat{S}_{z2x,h}^{(k)}(t,\boldsymbol{\beta}) - \mathrm{E}\{\widehat{S}_{z2x,h}^{(k)}(t,\boldsymbol{\beta})\}|$, $\sup_{t\in[h,\tau-h]}|\widehat{S}_{zx,h}^{(k)}(t) - \mathrm{E}\{\widehat{S}_{zx,h}^{(k)}(t)\}|$ converge to 0 in probability when $nh^2 \to \infty$. Since $\sup_{t\in[h,\tau-h]}|\mathrm{E}\{\widehat{S}_{z,h}^{(k)}(t)\} - s_z^{(k)}(t)m(t)| = O(h^2)$, $\sup_{\boldsymbol{\beta}\in\mathcal{B},t\in[h,\tau-h]}|\mathrm{E}\{\widehat{S}_{x,h}^{(k)}(t,\boldsymbol{\beta})\} - s_x^{(k)}(t,\boldsymbol{\beta})m(t)| = O(h^2)$, $\sup_{\boldsymbol{\beta}\in\mathcal{B},t\in[h,\tau-h]}|\mathrm{E}\{\widehat{S}_{z2x,h}^{(k)}(t,\boldsymbol{\beta})\} -$

$s_{z2x}^{(k)}(t,\boldsymbol{\beta})m(t)| = O(h^2)$, $\sup_{t\in[h,\tau-h]} |\mathrm{E}\{\widehat{S}_{zx,h}^{(k)}(t)\} - s_{zx}^{(k)}(t)m(t)| = O(h^2)$ and $\sup_{t\in[0,h)} |s_z^{(k)}(t)m(t) - s_z^{(k)}(h)m(h)|$, $\sup_{t\in(\tau-h,\tau]} |s_z^{(k)}(t)m(t) - s_z^{(k)}(\tau-h)m(\tau-h)| = O(h)$, $\sup_{\boldsymbol{\beta}\in\mathcal{B},t\in[0,h)} |s_x^{(k)}(t,\boldsymbol{\beta})m(t) - s_x^{(k)}(h,\boldsymbol{\beta})m(h)|$, $\sup_{\boldsymbol{\beta}\in\mathcal{B},t\in(\tau-h,\tau]} |s_x^{(k)}(t,\boldsymbol{\beta})m(t) - s_x^{(k)}(\tau-h,\boldsymbol{\beta})m(\tau-h)| = O(h)$, $\sup_{\boldsymbol{\beta}\in\mathcal{B},t\in[0,h)} |s_{z2x}^{(k)}(t,\boldsymbol{\beta})m(t) - s_{z2x}^{(k)}(h,\boldsymbol{\beta})m(h)|$, $\sup_{\boldsymbol{\beta}\in\mathcal{B},t\in(\tau-h,\tau]} |s_{z2x}^{(k)}(t,\boldsymbol{\beta})m(t) - s_{z2x}^{(k)}(\tau-h,\boldsymbol{\beta})m(\tau-h)| = O(h)$, $\sup_{t\in[0,h)} |s_{zx}^{(k)}(t)m(t) - s_{zx}^{(k)}(h)m(h)|$, $\sup_{t\in(\tau-h,\tau]} |s_{zx}^{(k)}(t)m(t) - s_{zx}^{(k)}(\tau-h)m(\tau-h)| = O(h)$, the uniform consistency of $\widehat{S}_{z,h}^{(k)}(t)$, $\widehat{S}_{x,h}^{(k)}(t,\boldsymbol{\beta})$, $k=0,1$, $\widehat{S}_{z2x,h}(t,\boldsymbol{\beta})$ and $\widehat{S}_{zx,h}(t)$ have been proved. By the law of large numbers, we can show that $n^{-1}\sum_{i=1}^n N_i(t)$ converges to $\mathrm{E}\{N_i(t)\}$, $n^{-1}\sum_{i=1}^n \int_0^\tau \mathbf{Z}_i(t)\exp\{-\boldsymbol{\beta}^\intercal\mathbf{X}_i(t)\}dN_i(t)$ converges to $\int_0^\tau \mathrm{E}\{\mathbf{Z}_i(t)\exp\{-\boldsymbol{\beta}^\intercal\mathbf{X}_i(t)\} dN_i(t)\}$ and $n^{-1}\sum_{i=1}^n \int_0^\tau \mathbf{X}_i(t)dN_i(t)$ converges to $\int_0^\tau \mathrm{E}\{\mathbf{X}_i(t)dN_i(t)\}$. Therefore, the proposed estimator $\widehat{\boldsymbol{\theta}}_h$ converges to the true parameter $\boldsymbol{\theta}_0$ in probability.

C.2 Proof of asymptotic normality

We prove the asymptotic normality of $\sqrt{n}\widehat{U}(\boldsymbol{\theta}_0) = (\sqrt{n}\widehat{U}_1(\boldsymbol{\theta}_0)^\intercal, \sqrt{n}\widehat{U}_2(\boldsymbol{\theta}_0)^\intercal)^\intercal$. From Equations (4.9), we show that $\sqrt{n}\widehat{U}_1(\boldsymbol{\theta}_0)$ has the form

$$\begin{aligned}
\sqrt{n}\widehat{U}_1(\boldsymbol{\theta}_0) =& \frac{1}{\sqrt{n}}\sum_{i=1}^n \int_0^\tau \mathbf{Z}_i(u)\exp\{-\boldsymbol{\beta}_0^\intercal\mathbf{X}_i(u)\}dN_i(u) - \frac{1}{\sqrt{n}}\int_0^\tau \frac{\widehat{S}_{z,h}^{(1)}(u)}{\widehat{S}_{x,h}^{(0)}(u,\boldsymbol{\beta}_0)} \\
&\left\{\sum_{i=1}^n dN_i(u)\right\} - \frac{1}{\sqrt{n}}\int_0^\tau \sum_{i=1}^n Y_i(u)\left\{\frac{\widehat{S}_{z2x,h}(u,\boldsymbol{\beta}_0)}{\widehat{S}_{z,h}^{(0)}(u)}\right\}du\boldsymbol{\gamma} \\
&+ \frac{1}{\sqrt{n}}\int_0^\tau \sum_{i=1}^n Y_i(u)\left\{\frac{\widehat{S}_{z,h}^{(1)}(u)}{\widehat{S}_{x,h}^{(0)}(u,\boldsymbol{\beta}_0)}\left(\frac{\widehat{S}_{z,h}^{(1)}(u)}{\widehat{S}_{z,h}^{(0)}(u)}\right)^\intercal\right\}du\boldsymbol{\gamma} \\
&\stackrel{def}{=} I_1 + I_2 + I_3 + I_4.
\end{aligned}$$

$\lambda^c(t)$ is the rate function of $N(\cdot)$, so we have $\lambda^c(t)dt = d\mathrm{E}\{N_i(t)\}$. Then we have

$$
\begin{aligned}
I_2 =& -\frac{1}{\sqrt{n}}\int_0^\tau \frac{\widehat{S}^{(1)}_{z,h}(u)}{\widehat{S}^{(0)}_{x,h}(u,\boldsymbol{\beta}_0)}\left\{\sum_{i=1}^n dN_i(u)\right\} \\
=& -\sqrt{n}\int_0^\tau \frac{s^{(1)}_{z}(u)}{s^{(0)}_{x}(u,\boldsymbol{\beta}_0)} d\left\{\frac{1}{n}\sum_{i=1}^n N_i(u) - \mathrm{E}\{N_i(u)\}\right\} \\
& -\sqrt{n}\int_0^\tau \frac{\widehat{S}^{(1)}_{z,h}(u)}{\widehat{S}^{(0)}_{x,h}(u,\boldsymbol{\beta}_0)} d\mathrm{E}\{N_i(u)\} + o_p(1) \\
=& -\frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^\tau Y_i(u)\left(\mathbf{Z}_i(u) - \frac{s^{(1)}_{z}(u)}{s^{(0)}_{x}(u,\boldsymbol{\beta}_0)}\exp\{\boldsymbol{\beta}_0^\intercal\mathbf{X}_i(u)\}\right)\frac{\lambda^c(u)}{s^{(0)}_{x}(u,\boldsymbol{\beta}_0)m(u)} \\
& dO_i(u) - \frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^\tau \frac{s^{(1)}_{z}(u)}{s^{(0)}_{x}(u,\boldsymbol{\beta}_0)} dN_i(u) + o_p(1).
\end{aligned}
$$

Define $R(t) = \mathrm{E}\{Y_i(t)\}$, then

$$
\begin{aligned}
I_3 =& -\frac{1}{\sqrt{n}}\int_0^\tau \sum_{i=1}^n Y_i(u)\left\{\frac{\widehat{S}_{z2x,h}(u,\boldsymbol{\beta}_0)}{\widehat{S}^{(0)}_{z,h}(u)}\right\} du\boldsymbol{\gamma} \\
=& -\sqrt{n}\int_0^\tau \frac{s_{z2x,h}(u,\boldsymbol{\beta}_0)}{s^{(0)}_{z}(u)}\left\{\frac{1}{n}\sum_{i=1}^n Y_i(u) - R(u)\right\} du\boldsymbol{\gamma} \\
& -\sqrt{n}\int_0^\tau \left\{\frac{\widehat{S}_{z2x,h}(u,\boldsymbol{\beta}_0)}{\widehat{S}^{(0)}_{z,h}(u)}\right\} R(u) du\boldsymbol{\gamma} + o_p(1) \\
=& -\frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^\tau Y_i(u)\left(\mathbf{Z}_i(u)^{\otimes 2}\exp\{-\boldsymbol{\beta}^\intercal\mathbf{X}_i(u)\} - \frac{s_{z2x,h}(u,\boldsymbol{\beta}_0)}{s^{(0)}_{z}(u)}\right) \\
& \frac{R(u)}{s^{(0)}_{z}(u)m(u)} dO_i(u)\boldsymbol{\gamma} - \frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^\tau Y_i(u)\frac{s_{z2x}(u,\boldsymbol{\beta}_0)}{s^{(0)}_{z}(u)} du\boldsymbol{\gamma} + o_p(1),
\end{aligned}
$$

and

$$
I_4 = \frac{1}{\sqrt{n}}\int_0^\tau \sum_{i=1}^n Y_i(u)\left\{\frac{\widehat{S}^{(1)}_{z,h}(u)}{\widehat{S}^{(0)}_{x,h}(u,\boldsymbol{\beta}_0)}\left(\frac{\widehat{S}^{(1)}_{z,h}(u)}{\widehat{S}^{(0)}_{z,h}(u)}\right)^\intercal\right\} du\boldsymbol{\gamma}
$$

$$=\sqrt{n}\int_0^\tau \left\{\frac{1}{n}\sum_{i=1}^n Y_i(u) - R(u)\right\}\left\{\frac{s_{z,h}^{(1)}(u)}{s_{x,h}^{(0)}(u,\boldsymbol{\beta}_0)}\left(\frac{s_{z,h}^{(1)}(u)}{s_{z,h}^{(0)}(u)}\right)^\intercal\right\}du\boldsymbol{\gamma}$$

$$+\sqrt{n}\int_0^\tau R(t)\left\{\frac{\widehat{S}_{z,h}^{(1)}(u)}{\widehat{S}_{x,h}^{(0)}(u,\boldsymbol{\beta}_0)}\left(\frac{\widehat{S}_{z,h}^{(1)}(u)}{\widehat{S}_{z,h}^{(0)}(u)}\right)^\intercal\right\}du\boldsymbol{\gamma}$$

$$=\frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^\tau \frac{s_z^{(1)}(u)}{s_x^{(0)}(u,\boldsymbol{\beta}_0)}Y_i(u)\left\{\mathbf{Z}_i(u)^\intercal-\left(\frac{s_z^{(1)}(u)}{s_z^{(0)}(u)}\right)^\intercal\right\}\frac{R(u)}{s_z^{(0)}(u)m(u)}dO_i(u)\boldsymbol{\gamma}$$

$$+\frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^\tau Y_i(u)\left(\mathbf{Z}_i(u)-\frac{s_z^{(1)}(u)}{s_x^{(0)}(u,\boldsymbol{\beta}_0)}\exp\{\boldsymbol{\beta}_0^\intercal\mathbf{X}_i(u)\}\right)\left(\frac{s_z^{(1)}(u)}{s_z^{(0)}(u)}\right)^\intercal$$

$$\frac{R(u)}{s_x^{(0)}(u,\boldsymbol{\beta}_0)m(u)}dO_i(u)\boldsymbol{\gamma}+\frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^\tau Y_i(u)\left\{\frac{s_z^{(1)}(u)}{s_x^{(0)}(u,\boldsymbol{\beta}_0)}\left(\frac{s_z^{(1)}(u)}{s_z^{(0)}(u)}\right)^\intercal\right\}du\boldsymbol{\gamma}$$

$$+o_p(1).$$

Thus we have $\sqrt{n}\widehat{U}_1(\boldsymbol{\theta}_0)=\frac{1}{\sqrt{n}}\sum_{i=1}^n\phi_i^{(1)}(\boldsymbol{\theta}_0)+o_p(1)$, where $\phi_i^{(1)}(\boldsymbol{\theta}_0)=\int_0^\tau\mathbf{Z}_i(u)$ $\exp\{-\boldsymbol{\beta}_0^\intercal\mathbf{X}_i(u)\}dN_i(u)+\phi_{2i}^{(1)}(\boldsymbol{\theta}_0)+\phi_{3i}^{(1)}(\boldsymbol{\theta}_0)+\phi_{4i}^{(1)}(\boldsymbol{\theta}_0)+o_p(1)$ and

$$\phi_{2i}^{(1)}(\boldsymbol{\theta}_0)=-\int_0^\tau Y_i(u)\left(\mathbf{Z}_i(u)-\frac{s_z^{(1)}(u)}{s_x^{(0)}(u,\boldsymbol{\beta}_0)}\exp\{\boldsymbol{\beta}_0^\intercal\mathbf{X}_i(u)\}\right)\frac{\lambda^c(u)}{s_x^{(0)}(u,\boldsymbol{\beta}_0)m(u)}$$

$$dO_i(u)-\int_0^\tau\frac{s_z^{(1)}(u)}{s_x^{(0)}(u,\boldsymbol{\beta}_0)}dN_i(u),$$

$$\phi_{3i}^{(1)}(\boldsymbol{\theta}_0)=-\int_0^\tau Y_i(u)\left(\mathbf{Z}_i(u)^{\otimes 2}\exp\{-\boldsymbol{\beta}^\intercal\mathbf{X}_i(u)\}-\frac{s_{z2x,h}(u,\boldsymbol{\beta}_0)}{s_z^{(0)}(u)}\right)\frac{R(u)}{s_z^{(0)}(u)m(u)}$$

$$dO_i(u)\boldsymbol{\gamma}-\int_0^\tau Y_i(u)\frac{s_{z2x}(u,\boldsymbol{\beta}_0)}{s_z^{(0)}(u)}du\boldsymbol{\gamma},$$

$$\phi_{4i}^{(1)}(\boldsymbol{\theta}_0)=\int_0^\tau\frac{s_z^{(1)}(u)}{s_x^{(0)}(u,\boldsymbol{\beta}_0)}Y_i(u)\left\{\mathbf{Z}_i(u)^\intercal-\left(\frac{s_z^{(1)}(u)}{s_z^{(0)}(u)}\right)^\intercal\right\}\frac{R(u)}{s_z^{(0)}(u)m(u)}dO_i(u)\boldsymbol{\gamma}$$

$$+\int_0^\tau Y_i(u)\left(\mathbf{Z}_i(u)-\frac{s_z^{(1)}(u)}{s_x^{(0)}(u,\boldsymbol{\beta}_0)}\exp\{\boldsymbol{\beta}_0^\intercal\mathbf{X}_i(u)\}\right)\left(\frac{s_z^{(1)}(u)}{s_z^{(0)}(u)}\right)^\intercal$$

$$\frac{R(u)}{s_x^{(0)}(u,\boldsymbol{\beta}_0)m(u)}dO_i(u)\boldsymbol{\gamma}+\int_0^\tau Y_i(u)\left\{\frac{s_z^{(1)}(u)}{s_x^{(0)}(u,\boldsymbol{\beta}_0)}\left(\frac{s_z^{(1)}(u)}{s_z^{(0)}(u)}\right)^\intercal\right\}du\boldsymbol{\gamma}.$$

Similarly, it can be shown that $\sqrt{n}\widehat{U}_2(\boldsymbol{\theta}_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \phi_i^{(2)}(\boldsymbol{\theta}_0) + o_p(1)$, where $\phi_i^{(2)}(\boldsymbol{\theta}_0) = \int_0^\tau \mathbf{X}_i(u)dN_i(u) + \phi_{2i}^{(2)}(\boldsymbol{\theta}_0) + \phi_{3i}^{(2)}(\boldsymbol{\theta}_0) + \phi_{4i}^{(2)}(\boldsymbol{\theta}_0) + o_p(1)$, and

$$\begin{aligned}
\phi_{2i}^{(2)}(\boldsymbol{\theta}_0) =& -\int_0^\tau Y_i(u)\left(\mathbf{X}_i(u) - \frac{s_x^{(1)}(u,\boldsymbol{\beta}_0)}{s_x^{(0)}(u,\boldsymbol{\beta}_0)}\right)\exp\{\boldsymbol{\beta}_0^\intercal\mathbf{X}_i(u)\}\frac{\lambda^c(u)}{s_x^{(0)}(u,\boldsymbol{\beta}_0)m(u)} \\
& dO_i(u) - \int_0^\tau \frac{s_x^{(1)}(u,\boldsymbol{\beta}_0)}{s_x^{(0)}(u,\boldsymbol{\beta}_0)}dN_i(u), \\
\phi_{3i}^{(2)}(\boldsymbol{\theta}_0) =& -\int_0^\tau Y_i(u)\left(\mathbf{X}_i(u)\mathbf{Z}_i(u)^\intercal - \frac{s_{zx}(u)}{s_z^{(0)}(u)}\right)\frac{R(u)}{s_z^{(0)}(u)m(u)}dO_i(u)\boldsymbol{\gamma} \\
& -\int_0^\tau Y_i(u)\frac{s_{zx}(u)}{s_z^{(0)}(u)}du\boldsymbol{\gamma}, \\
\phi_{4i}^{(2)}(\boldsymbol{\theta}_0) =& \int_0^\tau \frac{s_x^{(1)}(u,\boldsymbol{\beta}_0)}{s_x^{(0)}(u,\boldsymbol{\beta}_0)}Y_i(u)\left\{\mathbf{Z}_i(u)^\intercal - \left(\frac{s_z^{(1)}(u)}{s_z^{(0)}(u)}\right)^\intercal\right\}\frac{R(u)}{s_z^{(0)}(u)m(u)}dO_i(u)\boldsymbol{\gamma} \\
& + \int_0^\tau Y_i(u)\left(\mathbf{X}_i(u) - \frac{s_x^{(1)}(u,\boldsymbol{\beta}_0)}{s_x^{(0)}(u,\boldsymbol{\beta}_0)}\right)\left(\frac{s_z^{(1)}(u)}{s_z^{(0)}(u)}\right)^\intercal\frac{R(u)\exp\{\boldsymbol{\beta}_0^\intercal\mathbf{X}_i(u)\}}{s_x^{(0)}(u,\boldsymbol{\beta}_0)m(u)} \\
& dO_i(u)\boldsymbol{\gamma} + \int_0^\tau Y_i(u)\left\{\frac{s_x^{(1)}(u,\boldsymbol{\beta}_0)}{s_x^{(0)}(u,\boldsymbol{\beta}_0)}\left(\frac{s_z^{(1)}(u)}{s_z^{(0)}(u)}\right)^\intercal\right\}du\boldsymbol{\gamma}.
\end{aligned}$$

Define $\phi_i(\boldsymbol{\theta}_0) = (\phi_i^{(1)}(\boldsymbol{\theta}_0)^\intercal, \phi_i^{(2)}(\boldsymbol{\theta}_0)^\intercal)^\intercal$, then we have $\sqrt{n}\widehat{U}(\boldsymbol{\theta}_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \phi_i(\boldsymbol{\theta}_0) + o_p(1)$.

Define $A = \mathrm{E}\{-\frac{\partial U(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\} = \begin{pmatrix} \mathrm{E}\{-\frac{\partial U_1(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}}\} & \mathrm{E}\{-\frac{\partial U_1(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma}}\} \\ \mathrm{E}\{-\frac{\partial U_2(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}}\} & \mathrm{E}\{-\frac{\partial U_2(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma}}\} \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$. We define $s_{z2x}^{(2)}(t,\boldsymbol{\theta}) = \mathrm{E}[Y_i(u)\mathbf{Z}_i(u)\mathbf{X}_i(u)^\intercal\boldsymbol{\gamma}^\intercal\mathbf{Z}_i(u)\exp\{-\boldsymbol{\beta}^\intercal\mathbf{X}_i(u)\}]$, then we have

$$\begin{aligned}
A_{11} =& \mathrm{E}\left[\int_0^\tau \mathbf{Z}_i(u)\mathbf{X}_i(u)^\intercal\exp\{-\boldsymbol{\beta}^\intercal\mathbf{X}_i(u)\}dN_i(u)\right] - \int_0^\tau \frac{s_z^{(1)}(u)}{s_x^{(0)}(u,\boldsymbol{\beta})}\left(\frac{s_x^{(1)}(u,\boldsymbol{\beta})}{s_x^{(0)}(u,\boldsymbol{\beta})}\right)^\intercal \\
& \lambda^c(u)du - \int_0^\tau R(u)\left\{\frac{s_{z2x}^{(2)}(u,\boldsymbol{\theta})}{s_z^{(0)}(u)} - \frac{s_z^{(1)}(u)}{s_x^{(0)}(u,\boldsymbol{\beta})}\left(\frac{s_x^{(1)}(u,\boldsymbol{\beta})}{s_x^{(0)}(u,\boldsymbol{\beta})}\right)^\intercal\boldsymbol{\gamma}^\intercal\frac{s_z^{(1)}(u)}{s_z^{(0)}(u)}\right\}du, \\
A_{12} =& \int_0^\tau R(u)\left\{\frac{s_{z2x}(u,\boldsymbol{\beta})}{s_z^{(0)}(u)} - \frac{s_z^{(1)}(u)}{s_x^{(0)}(u,\boldsymbol{\beta})}\left(\frac{s_z^{(1)}(u)}{s_z^{(0)}(u)}\right)^\intercal du\right\},
\end{aligned}$$

$$
\begin{aligned}
A_{21} =& \int_0^\tau \left\{ \frac{s_x^{(2)}(u,\boldsymbol{\beta})}{s_x^{(0)}(u,\boldsymbol{\beta})} - \left(\frac{s_x^{(1)}(u,\boldsymbol{\beta})}{s_x^{(0)}(u,\boldsymbol{\beta})} \right)^{\otimes 2} \right\} \lambda^c(u) du \\
& - \int_0^\tau R(u) \left\{ \frac{s_x^{(2)}(u,\boldsymbol{\beta})}{s_x^{(0)}(u,\boldsymbol{\beta})} - \left(\frac{s_x^{(1)}(u,\boldsymbol{\beta})}{s_x^{(0)}(u,\boldsymbol{\beta})} \right)^{\otimes 2} \right\} \boldsymbol{\gamma}^\intercal \frac{s_z^{(1)}(u)}{s_z^{(0)}(u)} du, \\
A_{22} =& \int_0^\tau R(u) \left\{ \frac{s_{zx}(u)}{s_z^{(0)}(u)} - \frac{s_x^{(1)}(u,\boldsymbol{\beta})}{s_x^{(0)}(u,\boldsymbol{\beta})} \left(\frac{s_z^{(1)}(u)}{s_z^{(0)}(u)} \right)^\intercal \right\} du.
\end{aligned}
$$

Also, we have $\widehat{A} = -\frac{\partial \widehat{U}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{pmatrix} -\frac{\partial \widehat{U}_1(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} & -\frac{\partial \widehat{U}_1(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma}} \\ -\frac{\partial \widehat{U}_2(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} & -\frac{\partial \widehat{U}_2(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma}} \end{pmatrix} = \begin{pmatrix} \widehat{A}_{11} & \widehat{A}_{12} \\ \widehat{A}_{21} & \widehat{A}_{22} \end{pmatrix}$. Define $\widehat{S}_{z2x}^{(2)}(t,\boldsymbol{\theta})$ $= \frac{1}{n}\sum_{i=1}^n \int_0^\tau K_h(t-u) Y_i(u) \mathbf{Z}_i(u) \mathbf{X}_i(u)^\intercal \boldsymbol{\gamma}^\intercal \mathbf{Z}_i(u) \exp\{-\boldsymbol{\beta}^\intercal \mathbf{X}_i(u)\} dO_i(u)$, then we have

$$
\begin{aligned}
\widehat{A}_{11} =& \frac{1}{n} \sum_{i=1}^n \int_0^\tau \left\{ \mathbf{Z}_i(u) \mathbf{X}_i(u)^\intercal \exp\{-\boldsymbol{\beta}^\intercal \mathbf{X}_i(u)\} - \int_0^\tau \frac{\widehat{S}_{z,h}^{(1)}(u)}{\widehat{S}_{x,h}^{(0)}(u,\boldsymbol{\beta})} \left(\frac{\widehat{S}_{x,h}^{(1)}(u,\boldsymbol{\beta})}{\widehat{S}_{x,h}^{(0)}(u,\boldsymbol{\beta})} \right)^\intercal \right\} \\
& dN_i(u) - \frac{1}{n} \sum_{i=1}^n \int_0^\tau Y_i(u) \left\{ \frac{\widehat{S}_{z2x,h}^{(2)}(u,\boldsymbol{\theta})}{\widehat{S}_{z,h}^{(0)}(u)} - \frac{\widehat{S}_{z,h}^{(1)}(u)}{\widehat{S}_{x,h}^{(0)}(u,\boldsymbol{\beta})} \left(\frac{\widehat{S}_{x,h}^{(1)}(u,\boldsymbol{\beta})}{\widehat{S}_{x,h}^{(0)}(u,\boldsymbol{\beta})} \right)^\intercal \right. \\
& \left. \boldsymbol{\gamma}^\intercal \left(\frac{\widehat{S}_{z,h}^{(1)}(u)}{\widehat{S}_{z,h}^{(0)}(u)} \right) \right\} du, \\
\widehat{A}_{12} =& \frac{1}{n} \sum_{i=1}^n \int_0^\tau Y_i(u) \left\{ \frac{\widehat{S}_{z2x,h}(u,\boldsymbol{\beta})}{\widehat{S}_{z,h}^{(0)}(u)} - \frac{\widehat{S}_{z,h}^{(1)}(u)}{\widehat{S}_{x,h}^{(0)}(u,\boldsymbol{\beta})} \left(\frac{\widehat{S}_{z,h}^{(1)}(u)}{\widehat{S}_{z,h}^{(0)}(u)} \right)^\intercal \right\} du, \\
\widehat{A}_{21} =& \frac{1}{n} \sum_{i=1}^n \int_0^\tau \left(\frac{\widehat{S}_{x,h}^{(2)}(u,\boldsymbol{\beta})}{\widehat{S}_{x,h}^{(0)}(u,\boldsymbol{\beta})} - \left\{ \frac{\widehat{S}_{x,h}^{(1)}(u,\boldsymbol{\beta})}{\widehat{S}_{x,h}^{(0)}(u,\boldsymbol{\beta})} \right\}^{\otimes 2} \right) dN_i(u) \\
& - \frac{1}{n} \sum_{i=1}^n \int_0^\tau Y_i(u) \left\{ \frac{\widehat{S}_{x,h}^{(2)}(u,\boldsymbol{\beta})}{\widehat{S}_{x,h}^{(0)}(u,\boldsymbol{\beta})} - \left(\frac{\widehat{S}_{x,h}^{(1)}(u,\boldsymbol{\beta})}{\widehat{S}_{x,h}^{(0)}(u,\boldsymbol{\beta})} \right)^{\otimes 2} \right\} \boldsymbol{\gamma}^\intercal \frac{\widehat{S}_{z,h}^{(1)}(u)}{\widehat{S}_{z,h}^{(0)}(u)} du, \\
\widehat{A}_{22} =& \frac{1}{n} \sum_{i=1}^n \int_0^\tau Y_i(u) \left\{ \frac{\widehat{S}_{zx,h}(u)}{\widehat{S}_{z,h}^{(0)}(u)} - \frac{\widehat{S}_{x,h}^{(1)}(u,\boldsymbol{\beta})}{\widehat{S}_{x,h}^{(0)}(u,\boldsymbol{\beta})} \left(\frac{\widehat{S}_{z,h}^{(1)}(u)}{\widehat{S}_{z,h}^{(0)}(u)} \right)^\intercal \right\} du.
\end{aligned}
$$

By Taylor expansion, we have $\widehat{U}(\widehat{\boldsymbol{\theta}}_h) - \widehat{U}(\boldsymbol{\theta}_0) = \widehat{A}(\boldsymbol{\theta}^*)(\widehat{\boldsymbol{\theta}}_h - \boldsymbol{\theta}_0)$ where $\boldsymbol{\theta}^*$ satisfies that for each $j = 1, \dots, p$, $\boldsymbol{\theta}_j^*$ is on the line segment between $\boldsymbol{\theta}_{0,j}$ and $\widehat{\boldsymbol{\theta}}_{h,j}$. Since we can show that $\widehat{A}(\boldsymbol{\theta}^*)$ converges to $A(\boldsymbol{\theta})$, we have $\sqrt{n}(\widehat{\boldsymbol{\theta}}_h - \boldsymbol{\theta}_0)$ converges to a normal distribution with mean 0 and variance $A(\boldsymbol{\theta}_0)^{-1}V(\boldsymbol{\theta}_0)\{A(\boldsymbol{\theta}_0)^{-1}\}^\intercal$, where $V(\boldsymbol{\theta}_0) = \mathrm{E}\{\phi_1(\boldsymbol{\theta}_0)\phi_1(\boldsymbol{\theta}_0)^\intercal\}$.