SEMIPARAMETRIC REGRESSION ANALYSIS OF RIGHT- AND INTERVAL-CENSORED DATA

Fei Gao

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biostatistics in the Gillings School of Global Public Health.

Chapel Hill
2017

Approved by:

Donglin Zeng

Danyu Lin

David Couper

Gerardo Heiss

Michael G. Hudgens

ABSTRACT

Fei Gao: Semiparametric Regression Analysis of Right- and
Interval-Censored Data
(Under the direction of Donglin Zeng and Danyu Lin)

Right-censored data arise when the event time can only be observed up to the end of the follow-up, while interval-censored data arise when the event time is only known to lie within an interval. There is a large body of statistical literature on right-censored and interval-censored data, but the existing methods cannot properly handle certain complexities.

In the first project, we consider efficient semiparametric estimation of the accelerated failure time (AFT) model with partly interval-censored (PIC) data, which arise when the event time may be right-censored for some subjects and interval-censored for the others because of different observation schemes. We generalize the Buckley-James estimator to PIC data and develop a one-step estimator by deriving and estimating the efficient score for the regression parameters. We then establish the asymptotic properties of the estimators, conduct extensive simulation studies, and apply our methods to data derived from an AIDS study.

In the second project, we consider the setting when subjects may not complete the examination schedule for reasons related to the event of interest. To make a valid inference about the interval-censored event time of interest, we jointly model the event time of interest and the dropout time using transformation models with a shared random effect. We consider nonparametric maximum likelihood estimation (NPMLE) and develop a simple and stable Expectation-maximization (EM) algorithm. We then prove the asymptotic properties of the resulting estimators and show how to predict the event time of interest when dropout is an unavoidable terminal event. Finally, we provide an application to data on the incidence of diabetes from the Atherosclerosis Risk in Communities (ARIC) study.

In the third project, we formulate the effects of covariates on the joint distribution of multiple right- and interval-censored events through semiparametric proportional hazards models with random effects. We consider NPMLE, develop an EM algorithm, and establish the asymptotic properties of the resulting estimators. We leverage the joint modelling to provide dynamic prediction of disease incidence based on the evolving event history and provide an application to the ARIC study.

Keywords: Buckley-James estimator; Joint models; Nonparametric likelihood; Random effects; Semiparametric efficiency; Terminal event.

ACKNOWLEDGEMENTS

Over the past five years at the University of North Carolina at Chapel Hill, I have received enormous support and encouragement from a number of people. I would never have been able to finish my dissertation without their guidance and support.

First, I would like to express my deepest gratitude to my advisors, Dr. Donglin Zeng and Dr. Danyu Lin, for their guidance, support, and patience. They enlightened me by giving me my first glance of research, deepened my understanding of biostatistics, and led me to pursue research on topics for which I have a passion. They were always available to me despite their busy schedules. I have been very fortunate to have advisors like them.

I would also like to thank my committee members, Dr. David Couper, Dr. Gerardo Heiss, and Dr. Michael G. Hudgens, for their encouragement, insightful comments, and constructive questions at different stages of my research. Their recommendations and instructions have enabled me to assemble and finish the dissertation effectively. I am also very thankful to Dr. Joseph G. Ibrahim, who supported me through the years. The experience of collaboration under his guidance was invaluable.

Finally, this work would not have been possible without the unconditional support of my family and friends. I would like to thank my friends for all the great times we have shared, and I would like to thank my parents for their love, support, and sacrifices.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

In this chapter we introduce the concepts and ideas that will play a key role in the subsequent development of our thesis.

1.1 Interval-Censored Data

Interval-censored data arise when the timing of an event is not known precisely, but rather is known to lie within a time interval. Such data are frequently encountered in medical research, where the ascertainment of the disease of interest is made over a series of examination times. An example is the Atherosclerosis Risk in Communities (ARIC) study (The ARIC Investigators 1989), where subjects were examined for asymptomatic diseases, such as diabetes and hypertension, over five visits, each at least three years apart, such that the disease was only known to occur within a broad time interval.

There are several types of interval-censored data. The simplest type is called "Case-1" interval-censored data or current status data, which involves only one examination time for each subject. Case-1 interval-censored data is frequently encountered in cross-sectional studies and tumorigenicity experiments. A more general type is called "Case-2" or "Case-k" interval-censored data, when there are two or k examination times for each subject (Huang and Wellner, 1997). The most general and most common type is called "Mixed-case" interval-censored data, which allows for varying numbers of examination times among subjects (Schick and Yu, 2000).

A number of methods have been developed for regression analysis of interval-censored data. In particular, sieve estimation for the proportional odds model has been studied. Rossini and Tsiatis (1996) considered the Case-1 interval-censored data and approximated the baseline function by a uniformly spaced step function, where the number of jumps is predetermined by a Lipschitz-continuity assumption. They proposed an estimation procedure maximizing the sieve likelihood function and established the asymptotic properties for the estimators. Huang and Rossini (1997) studied the Case-2 interval-censored data, approximated the baseline log-odds function by linear functions, and discussed the conditions that allow for positive information for the regression parameter. Shen (1998) studied the same data and proposed sieve maximum likelihood methods approximating the baseline log-odds functions by monotone spline functions.

The nonparametric maximum likelihood estimation (NPMLE) for the regression analysis of interval-censored data has been studied. Huang (1995) and Huang (1996) studied the proportional odds and proportional hazards models with the Case-1 interval-censored data and proposed an iterative convex minorant algorithm for computation. They have shown that the estimators for the regression parameters are asymptotically normal with $n^{1/2}$ convergence rate and achieve the semiparametric efficiency bound and that the NPMLE for the baseline cumulative hazard function converges at $n^{1/3}$ rate. Zeng et al. (2016) considered the NPMLE for transformation models with the Mixed-case interval-censored data in the presence of time-dependent covariates. They established the asymptotic properties and devised an EM algorithm that converges stably.

Rank-based estimation methods for linear transformation models have also been studied. Sun and Sun (2005) considered the Case-1 interval-censored data and proposed the rank-based estimation, which can be solved by a standard root-finding method or the Newton-Raphson algorithm. Gu et al. (2005) considered the Case-2 interval-censored data and proposed a computational algorithm using Markov Chain Monte Carlo stochastic approximation. Zhang, Sun, Zhao and Sun (2005) considered the same data, proposed an estimating equation approach to estimate the regression parameters, and showed the asymptotic properties of the estimators. Zhang and Zhao (2013) proposed two empirical likelihood inference approaches for the rank-based regression parameters based on the generalized estimating equations.

1.2 Accelerated Failure Time Model

The accelerated failure time (AFT) model assumes that the logarithm of the failure time is linearly related to the covariates (Kalbfleisch and Prentice, 1980, pp. 32-34). Let T denote the failure time and $\boldsymbol{X}$ denote a set of covariates. The accelerated failure time model specifies that

$$\log(T) = \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X} + \epsilon,$$

where $\boldsymbol{\beta}$ is the regression coefficient, and ϵ is an error term with unknown distribution. Because of its direct physical interpretation, the AFT model is an appealing alternative to the proportional hazards model, especially when the response variable does not pertain to failure time and is a result of some mechanical process with a known sequence of intermediary stages. It may provide more accurate or more concise summarization of the data than the proportional hazards model in certain applications (Zeng and Lin, 2007).

A class of rank estimators have been proposed for the AFT model with right-censored data. Prentice (1978) first proposed the rank estimators based on the well-known weighted log-rank statistics. The asymptotic properties of the rank estimators were then rigorously studied by Tsiatis (1990) and Ying (1993) among others. Wei et al. (1990) developed an inference procedure based on the minimum-dispersion statistic, the calculation of which involves minimizations of discrete objective functions with potentially multiple local minima. Lin and Geyer (1992) suggested a computational method based on simulated annealing. Jin et al. (2003) proposed an iterative estimation procedure based on a class of monotone estimating functions and estimated the variance of the resulting parameter using a novel resampling procedure without involving high-dimensional, nonparametric density function estimates. Zhou (2005) proposed an empirical likelihood approach to derive a test and confidence interval for the rank-based estimator.

Buckley and James (1979) modified the least-squares estimator for the linear regression model to obtain an estimator for the AFT model with right-censored data. The asymptotic properties of the Buckley-James estimators were then rigorously studied by Ritov (1990) and Lai and Ying (1991) among others. Later, Jin et al. (2006) computed the Buckley-James estimator by iteratively applying Buckley-James estimating equation on an initial consistent estimator.

The rank-based estimators and Buckley-James estimators fail to achieve the semiparametric efficiency bound. To obtain an efficient estimator, Zeng and Lin (2007) constructed a smooth approximation to the profile likelihood function for the regression parameters of the AFT model using kernel smoothing and maximized the approximated profile likelihood function. Later, Lin and Chen (2013) proposed a one-step procedure based on a counting process martingale and kernel estimation of the hazard function.

1.3 Transformation Model

The class of linear transformation models relates an unknown transformation of the failure time T linearly to a vector of (time-independent) covariates $\boldsymbol{X}$:

$$H(T) = -\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X} + \epsilon,$$

where $H(\cdot)$ is an unspecified increasing function, $\boldsymbol{\beta}$ is a set of unknown regression parameters, and ϵ is a random error with a parametric distribution. The formulation can be extended to allow time-dependent covariates. In the time-dependent version, the cumulative hazard function for T given covariates $\boldsymbol{X}$ takes the form

$$\Lambda(t|\boldsymbol{X}) = G\left[\int_0^t \exp\left\{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}(s)\right\} d\Lambda(s)\right],$$

where G is a continuously differentiable and strictly increasing function, $\boldsymbol{\beta}$ is a set of unknown regression parameters, and $\Lambda(\cot)$ is an unspecified increasing function. The class of transformation models encompass the proportional hazards model and proportional odds model.

It is useful to consider the following class of frailty-induced transformations

$$G(x) = \int_0^\infty e^{-xt}\phi(t)dt,$$

where $\phi(t)$ is a density function of a frailty with support $[0, \infty)$. The choice of the gamma density with mean 1 and variance r yields the class of logarithmic transformations $G(x) = r^{-1}\log(1 + rx)$ $(r \geq 0)$.

CHAPTER 2: SEMIPARAMETRIC ESTIMATION OF THE ACCELERATED FAILURE TIME MODEL WITH PARTLY INTERVAL-CENSORED DATA

2.1 Introduction

Partly interval-censored (PIC) data consist of failure time observations, in which some of the failure times are exactly observed while others are only known to lie within certain intervals. Such data arise in clinical and epidemiological research when the occurrence of an asymptomatic event, such as diabetic nephropathy or HIV infection, is ascertained at clinic visits. If a subject takes frequent visits, then his or her failure time can be determined with sufficient accuracy. If the visits are infrequent, then the failure time is known to lie within an interval that may be too broad to be treated as exact.

Several statistical methods have been suggested to make inference with PIC data. Specifically, estimation of the survival function for PIC data was studied by Turnbull (1976) and Huang (1999), among others. Zhao et al. (2008) developed a generalized log-rank test for PIC data and established its asymptotic properties. Kim (2003) studied nonparametric maximum likelihood estimation (NPMLE) for the proportional hazards model.

In this chapter, we consider the accelerated failure time (AFT) model, which relates the logarithm of the failure time linearly to the covariates (Kalbfleisch and Prentice, 1980, pp. 32-34). Because of its direct physical interpretation, the AFT model is an appealing alternative to the proportional hazards model, especially when the response variable does not pertain to failure time. It may provide a more accurate or more concise summarization of the data than the proportional hazards model in certain applications (Zeng and Lin, 2007). However, semiparametric estimation of the AFT model is highly challenging, even in the case of right-censored data (Prentice, 1978; Buckley and James, 1979; Tsiatis, 1990; Lai and Ying, 1991; Zeng and Lin, 2007; Lin and Chen, 2013). For PIC data, we first propose an iterative algorithm similar to that of Buckley and James (1979). We show that the resulting estimator is consistent and asymptotically normal and its variance can be consistently estimated by bootstrap. We then propose an efficient estimator for the (vector-valued) regression parameter by the one-step Newton-Raphson update with the efficient score. We derive the efficient score and construct the one-step estimator using kernel estimation. The one-step estimator is shown to be consistent and asymptotically normal, with a limiting covariance matrix that attains the semiparametric efficiency bound and can be consistently estimated through bootstrap. We conduct extensive simulation studies to examine the performance of the Buckley-James and one-step estimators in realistic settings, and we use our methods to analyze data derived from an AIDS clinical trial.

2.2 Methods

2.2.1 Data and Model

Let T denote the failure time and $\boldsymbol{X}$ denote a d-vector of covariates. The AFT model specifies that

$$\log T = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta} + \epsilon,$$

where $\boldsymbol{\beta}$ is a d-vector of unknown regression parameters, and ϵ is an unobserved error independent of $\boldsymbol{X}$. The distribution of ϵ is arbitrary such that the model is semiparametric.

Let Δ indicate, by the values 1 versus 0, whether T is observed exactly or not. For $\Delta = 0$, there is a sequence of examination times $0 < U_1 < U_2 < \cdots < U_K < \infty$ that gives rise to the interval (L, R), where $L = \max\{U_k : U_k \leq T; k = 0, \ldots, K\}$, and $R = \min\{U_k : U_k \geq T; k = 1, \ldots, K+1\}$, with $U_0 = 0$ and $U_{K+1} = \infty$. We assume that the proportion of $\Delta = 1$ is not negligible, and the joint distribution of $(U_1, \ldots, U_K)$ is independent of T given $\boldsymbol{X}$ and $\Delta = 0$. Note that $L = 0$ represents a left-censored observation and $R = \infty$ represents a right-censored observation. For a random sample of n subjects, the PIC data consist of

$$\big\{\Delta_i, \Delta_i T_i, (1-\Delta_i)L_i, (1-\Delta_i)R_i, \boldsymbol{X}_i\big\} \ (i = 1, \ldots, n).$$

2.2.2 Generalized Buckley-James Estimation

If the failure time is observed for every subject, then the classical least-squares estimator for $\boldsymbol{\beta}$ is the solution to the estimating equation

$$\sum_{i=1}^{n}\left(\boldsymbol{X}_i-\overline{\boldsymbol{X}}\right)\left\{\left(Y_i-\overline{Y}\right)-\left(\boldsymbol{X}_i-\overline{\boldsymbol{X}}\right)^{\mathrm{T}}\boldsymbol{\beta}\right\}=\mathbf{0}, \tag{2.1}$$

where $Y_i=\log T_i$, $\overline{Y}=n^{-1}\sum_{i=1}^{n}Y_i$, and $\overline{\boldsymbol{X}}=n^{-1}\sum_{i=1}^{n}\boldsymbol{X}_i$. In the presence of censoring, some values of Y_i are not observed. Following the approach of Buckley and James (1979), we replace the unobserved Y_i by the conditional mean given the observed data. The conditional mean $\widehat{Y}_i(\boldsymbol{\beta},F)$ is given by

$$\begin{aligned}
&\Delta_iY_i+(1-\Delta_i)\,E\left(Y_i\middle|\max\{U_{ik}:U_{ik}<T_i\}=L_i,\min\{U_{ik}:U_{ik}\geq T_i\}=R_i,\boldsymbol{X}_i,\Delta_i=0,L_i,R_i\right)\\
&=\Delta_iY_i+(1-\Delta_i)\\
&\times\frac{E\left[E\left\{Y_iI\left(\max\{U_{ik}:U_{ik}<T_i\}=L_i,\min\{U_{ik}:U_{ik}\geq T_i\}=R_i\right)|U_1,\ldots,U_K,\boldsymbol{X}_i\right\}|\boldsymbol{X}_i\right]}{E\left\{\Pr\left(\max\{U_{ik}:U_{ik}<T_i\}=L_i,\min\{U_{ik}:U_{ik}\geq T_i\}=R_i|U_1,\ldots,U_K,\boldsymbol{X}_i\right)|\boldsymbol{X}_i\right\}}\\
&=\Delta_iY_i+(1-\Delta_i)\\
&\times\frac{E\left[\sum_{k=1}^{K}E\left\{Y_iI\left(U_{ik}=L_i,U_{i,k+1}=R_i,L_i<T_i\leq R_i\right)|U_1,\ldots,U_K,\boldsymbol{X}_i,\Delta_i=0\right\}|\boldsymbol{X}_i,\Delta_i=0\right]}{E\left\{\sum_{k=1}^{K}\Pr\left(U_{ik}=L_i,U_{i,k+1}=R_i,L_i<T_i\leq R_i|U_1,\ldots,U_K,\boldsymbol{X}_i,\Delta_i=0\right)|\boldsymbol{X}_i,\Delta_i=0\right\}}\\
&=\Delta_iY_i+(1-\Delta_i)\\
&\times\frac{E\left\{Y_iI(L_i<T_i\leq R_i)|\boldsymbol{X}_i,\Delta_i=0,L_i,R_i\right\}E\left[\sum_{k=1}^{K}I\left(U_{ik}=L_i,U_{i,k+1}=R_i\right)|\boldsymbol{X}_i,\Delta_i=0,L_i,R_i\right]}{\Pr(L_i<T_i\leq R_i|\boldsymbol{X}_i,\Delta_i=0,L_i,R_i)E\left[\sum_{k=1}^{K}I\left(U_{ik}=L_i,U_{i,k+1}=R_i\right)|\boldsymbol{X}_i,\Delta_i=0,L_i,R_i\right]}\\
&=\Delta_iY_{\boldsymbol{\beta},i}+(1-\Delta_i)\frac{\int_{L_{\boldsymbol{\beta},i}}^{R_{\boldsymbol{\beta},i}}udF(u)}{F(R_{\boldsymbol{\beta},i})-F(L_{\boldsymbol{\beta},i})}+\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta},
\end{aligned}$$

where $Y_{\boldsymbol{\beta},i}=Y_i-\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta}$, $L_{\boldsymbol{\beta},i}=\log L_i-\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta}$, $R_{\boldsymbol{\beta},i}=\log R_i-\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta}$, and F is the distribution function of ϵ. The third equality follows from the conditional independence of the failure time and the examination times. Replacement of Y_i in (2.1) by $\widehat{Y}_i(\boldsymbol{\beta},F)$ yields

$$\sum_{i=1}^{n}\left(\boldsymbol{X}_i-\overline{\boldsymbol{X}}\right)\left[\left\{\widehat{Y}_i(\boldsymbol{\beta},F)-\overline{Y}(\boldsymbol{\beta},F)\right\}-\left(\boldsymbol{X}_i-\overline{\boldsymbol{X}}\right)^{\mathrm{T}}\boldsymbol{\beta}\right]=\mathbf{0},$$

where $\overline{Y}(\boldsymbol{\beta}, F) = n^{-1}\sum_{i=1}^{n} \widehat{Y}_i(\boldsymbol{\beta}, F)$. Because F is unknown, we replace F by the self-consistency estimator $\widehat{F}_{\boldsymbol{\beta}}$ (Turnbull, 1976; Huang, 1999) based on the transformed PIC data $\{\Delta_i, \Delta_i Y_{\boldsymbol{\beta},i}, (1 - \Delta_i)L_{\boldsymbol{\beta},i}, (1-\Delta_i)R_{\boldsymbol{\beta},i}\}$ $(i = 1, \ldots, n)$. The estimator $\widehat{F}_{\boldsymbol{\beta}}$ solves the self-consistency equation

$$\widehat{F}_{\boldsymbol{\beta}}(t) \quad = \quad n^{-1}\sum_{i=1}^{n}\left\{\Delta_i I(Y_{\boldsymbol{\beta},i} \leq t) + (1-\Delta_i)\frac{\widehat{F}_{\boldsymbol{\beta}}(R_{\boldsymbol{\beta},i} \wedge t) - \widehat{F}_{\boldsymbol{\beta}}(L_{\boldsymbol{\beta},i} \wedge t)}{\widehat{F}_{\boldsymbol{\beta}}(R_{\boldsymbol{\beta},i}) - \widehat{F}_{\boldsymbol{\beta}}(L_{\boldsymbol{\beta},i})}\right\}, \qquad (2.2)$$

where $a \wedge b = \min(a, b)$. If all of the failure times are observed, the right-hand side of equation (2.2) is simply the empirical distribution function for $Y_{\boldsymbol{\beta}}$. When the failure times are subject to censoring, the right-hand side is the conditional probability of $Y_{\boldsymbol{\beta}} \leq t$ given the observed data under the probability measure induced by $\widehat{F}_{\boldsymbol{\beta}}$. The generalized Buckley-James estimator $\widehat{\boldsymbol{\beta}}$ is the root of $\boldsymbol{U}_n(\boldsymbol{\beta}, \boldsymbol{\beta}) = \mathbf{0}$, where

$$\boldsymbol{U}_n(\boldsymbol{\beta}, \boldsymbol{b}) = n^{-1}\sum_{i=1}^{n}\left(\boldsymbol{X}_i - \overline{\boldsymbol{X}}\right)\left[\left\{\widehat{Y}_i(\boldsymbol{b}, \widehat{F}_{\boldsymbol{b}}) - \overline{Y}(\boldsymbol{b}, \widehat{F}_{\boldsymbol{b}})\right\} - \left(\boldsymbol{X}_i - \overline{\boldsymbol{X}}\right)^{\mathrm{T}}\boldsymbol{\beta}\right].$$

The function $\boldsymbol{U}_n(\boldsymbol{\beta}, \boldsymbol{\beta})$ is not continuous in $\boldsymbol{\beta}$, so it is difficult to directly solve the estimating equation. We propose an iterative algorithm. With $(\boldsymbol{\beta}^{(0)}, F^{(0)})$ as the starting value, the algorithm proceeds as follows:

1. at step m, solve the self-consistency equation (2.2) with $\boldsymbol{\beta} = \boldsymbol{\beta}^{(m-1)}$ to obtain $F^{(m)} = \widehat{F}_{\boldsymbol{\beta}^{(m-1)}}$;
2. update $\boldsymbol{\beta}$ with the equation $\boldsymbol{\beta}^{(m)} = \boldsymbol{L}_n(\boldsymbol{\beta}^{(m-1)}, F^{(m)})$, where

$$\boldsymbol{L}_n(\boldsymbol{\beta}, F) = \left\{\sum_{i=1}^{n}\left(\boldsymbol{X}_i - \overline{\boldsymbol{X}}\right)^{\otimes 2}\right\}^{-1}\left[\sum_{i=1}^{n}\left(\boldsymbol{X}_i - \overline{\boldsymbol{X}}\right)\left\{\widehat{Y}_i(\boldsymbol{\beta}, F) - \overline{Y}(\boldsymbol{\beta}, F)\right\}\right]$$

   with $\boldsymbol{a}^{\otimes 2} = \boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}$; and
3. set $m = m + 1$, and repeat steps (a) and (b) until convergence.

Denote the resulting estimator of $(\boldsymbol{\beta}, F)$ as $(\widehat{\boldsymbol{\beta}}, \widehat{F})$, where $\widehat{F} = \widehat{F}_{\widehat{\boldsymbol{\beta}}}$. In Section 2.6.1, we show that $(\widehat{\boldsymbol{\beta}}, \widehat{F})$ is consistent for the true value $(\boldsymbol{\beta}_0, F_0)$ and asymptotically normal under mild regularity conditions. The covariance matrix for the limiting distribution is difficult to directly estimate due to the lack of an analytical form. Therefore, we approximate the asymptotic distribution by bootstrapping the observations $\{\Delta_i, \Delta_i T_i, (1-\Delta_i)L_i, (1-\Delta_i)R_i, \boldsymbol{X}_i\}$ $(i = 1, \ldots, n)$. Let $\widehat{\boldsymbol{\beta}}^*$ be the generalized Buckley-James estimator of a bootstrap sample. In Section 2.6.2, we show that the conditional distribution of $\sqrt{n}(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}})$ given the data converges weakly to the asymptotic distribution of $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$. The empirical distribution of $\widehat{\boldsymbol{\beta}}^*$ can then be used to approximate the distribution of $\widehat{\boldsymbol{\beta}}$. Confidence intervals for individual components of $\boldsymbol{\beta}_0$ can be constructed by the Wald method (with the variance of $\widehat{\boldsymbol{\beta}}^*$) or from the empirical percentiles of $\widehat{\boldsymbol{\beta}}^*$.

2.2.3 One-step Efficient Estimation

We wish to develop an estimator for $\boldsymbol{\beta}$ that attains the semiparametric efficiency bound for PIC data. Let $\widetilde{\boldsymbol{l}}_{\boldsymbol{\beta}}(\mathcal{O}, \boldsymbol{\beta}, F)$ be the efficient score for $\boldsymbol{\beta}$ under the AFT model with the observed data $\mathcal{O} \equiv \{\Delta, \Delta T, (1-\Delta)L, (1-\Delta)R, \boldsymbol{X}\}$. We can construct a semiparametric efficient estimator through the one-step Newton-Raphson update (Bickel et al., 1993, pp. 40-45) of the generalized Buckley-James estimator $(\widehat{\boldsymbol{\beta}}, \widehat{F})$,

$$\widetilde{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}} + \left\{\mathbb{P}_n \widetilde{\boldsymbol{l}}_{\beta}(\mathcal{O}; \widehat{\boldsymbol{\beta}}, \widehat{F})^{\otimes 2}\right\}^{-1} \left\{\mathbb{P}_n \widetilde{\boldsymbol{l}}_{\beta}(\mathcal{O}; \widehat{\boldsymbol{\beta}}, \widehat{F})\right\}, \tag{2.3}$$

where $\mathbb{P}_n$ is the empirical measure.

According to the semiparametric efficiency theory (Bickel et al., 1993, chap. 3), the efficient score for $\boldsymbol{\beta}$ is the sum of the scores for $\boldsymbol{\beta}$ and F along the least favorable direction $\boldsymbol{g}$ that is orthogonal to the tangent set for F. After the derivations given in Section 2.6.3, we find that the least favorable direction $\boldsymbol{g}$ satisfies an integral equation. We replace the unknown quantities in the integral equation by appropriate sample estimators. The resulting function $\widehat{\boldsymbol{g}}$ satisfies the linear equation $\boldsymbol{A}(\widehat{\boldsymbol{g}}(t_1)^{\mathrm{T}}, \ldots, \widehat{\boldsymbol{g}}(t_m)^{\mathrm{T}})^{\mathrm{T}} = \boldsymbol{c}$, where $\boldsymbol{A} = (a_{jl}) \in \mathbb{R}^{m \times m}$, $\boldsymbol{c} = (\boldsymbol{c}_1^{\mathrm{T}}, \ldots, \boldsymbol{c}_m^{\mathrm{T}})^{\mathrm{T}}$,

$$a_{jl} = I(l=j)\widehat{P}\left(\Delta \middle| Y_{\boldsymbol{\beta}_0} = t_j\right) + \widehat{F}\{t_l\}\,\mathbb{P}_n\left[(1-\Delta)\frac{I(L_{\widehat{\boldsymbol{\beta}}} < t_j \le R_{\widehat{\boldsymbol{\beta}}}, L_{\widehat{\boldsymbol{\beta}}} < t_l \le R_{\widehat{\boldsymbol{\beta}}})}{\left\{\widehat{F}(R_{\widehat{\boldsymbol{\beta}}}) - \widehat{F}(L_{\widehat{\boldsymbol{\beta}}})\right\}^2}\right],$$

$$\boldsymbol{c}_j = \widehat{P}\left(\Delta \boldsymbol{X} \middle| Y_{\boldsymbol{\beta}_0} = t_j\right)\frac{\widehat{f}'(t_j)}{\widehat{f}(t_j)} + \mathbb{P}_n\left[(1-\Delta)\frac{\boldsymbol{X}\left\{\widehat{f}(R_{\widehat{\boldsymbol{\beta}}}) - \widehat{f}(L_{\widehat{\boldsymbol{\beta}}})\right\} I(L_{\widehat{\boldsymbol{\beta}}} < t_j \le R_{\widehat{\boldsymbol{\beta}}})}{\left\{\widehat{F}(R_{\widehat{\boldsymbol{\beta}}}) - \widehat{F}(L_{\widehat{\boldsymbol{\beta}}})\right\}^2}\right],$$

$Y_{\boldsymbol{\beta}} = Y - \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta}$, $L_{\boldsymbol{\beta}} = \log L - \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta}$, $R_{\boldsymbol{\beta}} = \log R - \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta}$, and $\widehat{F}\{t_l\}$ is the jump size of $\widehat{F}$ at t_l. Let f_0 and f_0' be the density function of ϵ and its derivative, respectively. The terms $\widehat{f}(t)$, $\widehat{f}'(t)$, $\widehat{P}\left(\Delta \middle| Y_{\boldsymbol{\beta}_0} = t\right)$, and $\widehat{P}\left(\Delta \boldsymbol{X} \middle| Y_{\boldsymbol{\beta}_0} = t\right)$ are kernel estimators of f_0, f_0', $E(\Delta | Y_{\boldsymbol{\beta}_0} = t)$, and

$E(\Delta \boldsymbol{X} | Y_{\boldsymbol{\beta}_0} = t)$, defined as

$$
\begin{aligned}
\widehat{f}(t) &= \frac{1}{a_n} \int_0^\infty K\left(\frac{s-t}{a_n}\right) d\widehat{F}(s), \\
\widehat{f}'(t) &= \frac{1}{b_n^3 \int u^2 K(u) du} \int_0^\infty (s-t) K\left(\frac{s-t}{b_n}\right) d\widehat{F}(s), \\
\widehat{P}\left(\Delta | Y_{\boldsymbol{\beta}_0} = t\right) &= \frac{1}{n a_n \widehat{f}(t)} \mathbb{P}_n \sum_{i=1}^n \Delta_i K\left(\frac{Y_{\widehat{\boldsymbol{\beta}},i} - t}{a_n}\right),
\end{aligned}
$$

and

$$
\widehat{P}\left(\Delta \boldsymbol{X} | Y_{\boldsymbol{\beta}_0} = t\right) = \frac{1}{n a_n \widehat{f}(t)} \sum_{i=1}^n \Delta_i \boldsymbol{X}_i K\left(\frac{Y_{\widehat{\boldsymbol{\beta}},i} - t}{a_n}\right),
$$

where $K(\cdot)$ is a smooth and symmetric kernel function, and a_n and b_n are bandwidths. The conditions for the choices of the kernel function and bandwidths can be found in Section 2.6.4.

The efficient score function can be estimated by

$$
\widehat{\boldsymbol{l}}(\widehat{\boldsymbol{\beta}}, \widehat{F}, \widehat{\boldsymbol{g}}) = -\left[\Delta\left\{\boldsymbol{X}\frac{\widehat{f}'(Y_{\widehat{\boldsymbol{\beta}}})}{\widehat{f}(Y_{\widehat{\boldsymbol{\beta}}})} + \widehat{\boldsymbol{g}}(Y_{\widehat{\boldsymbol{\beta}}})\right\} + (1-\Delta)\frac{\boldsymbol{X}\left\{\widehat{f}(R_{\widehat{\boldsymbol{\beta}}}) - \widehat{f}(L_{\widehat{\boldsymbol{\beta}}})\right\} + \int_{L_{\widehat{\boldsymbol{\beta}}}}^{R_{\widehat{\boldsymbol{\beta}}}} \widehat{\boldsymbol{g}}(u) d\widehat{F}(u)}{\widehat{F}(R_{\widehat{\boldsymbol{\beta}}}) - \widehat{F}(L_{\widehat{\boldsymbol{\beta}}})}\right].
$$

We replace the efficient score function $\widetilde{\boldsymbol{l}}_\beta(\mathcal{O}; \widehat{\boldsymbol{\beta}}, \widehat{F})$ in (2.3) by $\widehat{\boldsymbol{l}}(\widehat{\boldsymbol{\beta}}, \widehat{F}, \widehat{\boldsymbol{g}})$ to obtain the one-step estimator

$$
\widetilde{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}} + \left\{\mathbb{P}_n \widehat{\boldsymbol{l}}(\widehat{\boldsymbol{\beta}}, \widehat{F}, \widehat{\boldsymbol{g}})^{\otimes 2}\right\}^{-1}\left\{\mathbb{P}_n \widehat{\boldsymbol{l}}(\widehat{\boldsymbol{\beta}}, \widehat{F}, \widehat{\boldsymbol{g}})\right\}.
$$

In Section 2.6.4, we show that $\sqrt{n}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ converges in distribution to a mean-zero normal random vector with a covariance matrix that attains the semiparametric efficiency bound. We estimate the covariance matrix by bootstrapping the observations and applying the one-step procedure. The validity of the bootstrap is proved in Section 2.6.4. We also show that if the error ϵ is normally distributed, then the efficient score function is equivalent to the generalized Buckley-James estimating function. Thus, the generalized Buckley-James estimator is semiparametric efficient when the error is normally distributed.

2.3 Simulation Studies

We conducted extensive simulation studies to assess the performance of the proposed methods. We generated failure times from the AFT model: $\log T = -X_1 - X_2 - \epsilon$, where X_1 and X_2 are independent Bernoulli(0.5) and standard normal variables, respectively, and ϵ is independent of

(X_1, X_2). We considered four error distributions: standard normal distribution; standard extreme-value distribution; extreme-value distribution with location and scale parameters of -0.5 and 1.5, respectively; and logarithm of the gamma distribution with shape and scale parameters of 1 and 1, respectively. We simulated the time to loss to follow-up C from Uniform$[10, 15]$. For each subject, with probability p, we exactly observed the failure time T if $T \leq C$ and obtained a right-censored observation at C if $T > C$. With the remaining probability $1 - p$, we generated a sequence of examination times $U_k = U_{k-1}+$Uniform$[0.1, 1]$ $(k = 1, \ldots, K)$ such that $U_K < C$. We created the interval-censored observation $(L, R) \equiv (U_k, U_{k+1})$ if $U_k < T \leq U_{k+1}$ for $k = 0, \ldots, K$. The probability p depends on the covariates such that $p = p_0 - 0.1I(X_1 = 1)$, where p_0 was chosen to yield approximately 25% and 50% exact observations.

We considered the iterative algorithm convergent if both the norm of the difference for $\boldsymbol{\beta}$ and the integrated mean squared difference for F in two successive steps are less than 10^{-4} or the difference of the mean squared error $n^{-1} \sum_{i=1}^{n} \{\widehat{Y}(\boldsymbol{\beta}, \widehat{F}_{\boldsymbol{\beta}}) - \overline{Y}(\boldsymbol{\beta}, \widehat{F}_{\boldsymbol{\beta}}) - \left(\boldsymbol{X}_i - \overline{\boldsymbol{X}}\right)^{\mathrm{T}} \boldsymbol{\beta}\}^2$ between two successive steps is less than 10^{-2}. In all the scenarios we considered, the non-convergence rate was less than 1%. We estimated the standard error using the Wald method based on 200 bootstrap datasets.

Table 2.1 summarizes the results of the generalized Buckley-James estimation for sample sizes $n = 250$ and 500. The bias of the parameter estimator is small and tends to decrease as n increases. The standard error estimator accurately reflects the true variation, and the confidence intervals have proper coverage probabilities.

With the generalized Buckley-James estimator as the initial estimator, we carried out the one-step efficient estimation, and the results are shown in Table 2.2. We chose the Gaussian kernel for convenience. The optimal bandwidths for estimating the density and its derivative are $a_n = (4/3)^{1/5}\sigma n^{-1/5}$ and $b_n = (4/5)^{1/7}\sigma n^{-1/7}$(Swanepoel, 1988), where σ is the sample standard deviation of $\{Y_{\widehat{\boldsymbol{\beta}},i} : \Delta_i = 1\}$ $(i = 1, \ldots, n)$. We replaced σ by the minimum of the sample standard deviation and the interquartile range divided by 1.34, as suggested by Silverman (1986, p. 48).

The one-step estimator tends to be slightly positively biased, and the bias gets smaller as n increases. In the case of the normal error distribution, the one-step estimator has slightly larger standard error than the generalized Buckley-James estimator. This is not surprising because both estimators are asymptotically efficient when the error distribution is normal and the one-step estimator involves kernel approximation of the least favorable direction. For other error distributions, the

Table 2.1: Simulation results for the generalized Buckley-James estimator

Error Distribution	Exact Rate		$n=250$ Bias	SE	SEE	CP	$n=500$ Bias	SE	SEE	CP
$N(0,1)$	25%	β_1	0.001	0.144	0.144	0.944	0.001	0.103	0.101	0.948
		β_2	−0.001	0.076	0.077	0.948	0.001	0.054	0.054	0.949
	50%	β_1	0.000	0.136	0.135	0.948	0.001	0.094	0.095	0.946
		β_2	0.000	0.070	0.070	0.946	0.001	0.049	0.049	0.947
EV(0,1)	25%	β_1	−0.006	0.172	0.170	0.952	−0.003	0.120	0.119	0.950
		β_2	−0.008	0.092	0.094	0.946	−0.002	0.066	0.066	0.945
	50%	β_1	−0.002	0.167	0.166	0.949	−0.001	0.118	0.118	0.953
		β_2	−0.002	0.088	0.087	0.945	0.000	0.064	0.062	0.949
EV(−0.5,1.5)	25%	β_1	−0.015	0.251	0.256	0.953	−0.005	0.180	0.178	0.949
		β_2	−0.013	0.139	0.140	0.947	−0.005	0.098	0.097	0.948
	50%	β_1	0.003	0.250	0.249	0.950	0.000	0.176	0.176	0.948
		β_2	−0.003	0.130	0.130	0.945	−0.001	0.093	0.092	0.948
Gamma(1,1)	25%	β_1	−0.007	0.174	0.170	0.948	−0.002	0.119	0.120	0.953
		β_2	−0.007	0.095	0.094	0.949	−0.003	0.066	0.066	0.948
	50%	β_1	−0.002	0.165	0.166	0.950	0.001	0.116	0.118	0.945
		β_2	0.000	0.086	0.087	0.943	−0.001	0.062	0.062	0.945

Bias and SE are the bias and standard error, respectively, of the parameter estimator; SEE is the mean of the standard error estimator; and CP is the coverage probability of the 95% confidence interval. EV(a,b) denotes the extreme-value distribution with location parameter a and scale parameter b. Gamma(a,b) denotes the logarithm of the gamma distribution with shape parameter a and scale parameter b. Each entry is based on 10,000 replicates.

one-step estimator achieves up to 16% efficiency gain over the generalized Buckley-James estimator in terms of variance. The efficiency gain in terms of mean squared error of the estimators is similar. The standard error estimator becomes more accurate as n increases. The confidence intervals have satisfactory coverage probabilities.

PIC data often arise as an approximation to interval-censored data, where the observations with short intervals are treated as exactly observed failure times. We examined the performance of the proposed estimators in this practical setting. We simulated the failure time T and time to loss to follow-up C in the same manner as before. For each subject, we generated a sequence of examination times $U_k = U_{k-1}+\mathrm{Uniform}[a, b]$ $(k = 1, \ldots, K)$ such that $U_K < C$. We set $(a, b) = (0, 0.1)$ with probability p and $(a, b) = (0.1, 1)$ with probability $1-p$. We created the interval-censored observation $(L, R) \equiv (U_k, U_{k+1})$ if $U_k < T \leq U_{k+1}$ for $k = 0, \ldots, K$. If the interval length $R - L$ is smaller than 0.1, we treated the observation as exactly observed failure time at the geometric mid-point $\sqrt{LR}$. In this case, $\Delta = I(R - L < 0.1)$, and the exact observations are approximations to the true failure times.

We display the results for the proposed estimators with 50% exact observations in Table 2.3. The generalized Buckley-James estimator and one-step estimator have reasonably small bias. The standard error estimators accurately reflect the true variation, and the confidence intervals have satisfactory coverage probabilities. The one-step estimator achieves up to 13% efficiency gain for some of the considered error distributions.

Table 2.2: Simulation results for the one-step estimator

Error Distribution	Exact Rate		$n = 250$					$n = 500$				
			Bias	SE	SEE	CP	RE	Bias	SE	SEE	CP	RE
$N(0,1)$	25%	β_1	0.012	0.146	0.146	0.947	0.978	0.008	0.104	0.102	0.948	0.954
		β_2	0.009	0.076	0.077	0.943	0.991	0.008	0.054	0.054	0.945	1.005
	50%	β_1	0.009	0.138	0.137	0.944	0.962	0.007	0.095	0.096	0.946	1.000
		β_2	0.006	0.072	0.071	0.942	0.961	0.005	0.049	0.049	0.941	1.000
EV(0,1)	25%	β_1	0.001	0.169	0.169	0.953	1.031	0.001	0.118	0.118	0.951	1.026
		β_2	0.000	0.089	0.092	0.947	1.089	0.002	0.064	0.064	0.945	1.050
	50%	β_1	0.005	0.155	0.162	0.951	1.155	0.005	0.111	0.115	0.958	1.121
		β_2	0.003	0.084	0.085	0.949	1.090	0.004	0.060	0.061	0.954	1.068
EV(-0.5,1.5)	25%	β_1	-0.002	0.247	0.255	0.953	1.036	0.005	0.178	0.176	0.953	1.001
		β_2	-0.001	0.136	0.137	0.948	1.045	0.003	0.095	0.095	0.947	1.037
	50%	β_1	0.016	0.232	0.239	0.951	1.166	0.011	0.168	0.169	0.950	1.105
		β_2	0.007	0.123	0.125	0.944	1.129	0.007	0.088	0.089	0.949	1.094
Gamma(1,1)	25%	β_1	0.000	0.169	0.169	0.951	1.053	0.001	0.118	0.118	0.951	1.022
		β_2	0.000	0.092	0.092	0.952	1.061	0.000	0.063	0.064	0.947	1.079
	50%	β_1	0.005	0.158	0.161	0.956	1.087	0.006	0.112	0.116	0.953	1.110
		β_2	0.004	0.082	0.085	0.949	1.097	0.003	0.059	0.061	0.950	1.110

See the Note to Table 2.1. RE is the relative efficiency, defined as the ratio of the variance of the generalized Buckley-James estimator to that of the one-step estimator.

Table 2.3: Simulation results for the PIC approximation

Error Distribution			Generalized Buckley-James				One-Step				
			Bias	SE	SEE	CP	Bias	SE	SEE	CP	RE
$N(0,1)$	$n=250$	β_1	−0.007	0.135	0.138	0.947	0.006	0.137	0.139	0.947	0.960
		β_2	−0.008	0.072	0.072	0.947	0.003	0.072	0.072	0.946	1.002
	$n=500$	β_1	−0.007	0.098	0.097	0.952	0.004	0.099	0.098	0.952	0.978
		β_2	−0.007	0.051	0.051	0.946	0.001	0.051	0.051	0.949	0.994
EV(0,1)	$n=250$	β_1	−0.006	0.163	0.161	0.949	0.008	0.154	0.155	0.953	1.126
		β_2	−0.005	0.085	0.084	0.949	0.006	0.081	0.081	0.949	1.118
	$n=500$	β_1	−0.007	0.116	0.115	0.949	0.003	0.112	0.111	0.952	1.073
		β_2	−0.007	0.061	0.060	0.949	0.001	0.058	0.058	0.952	1.118
EV(−0.5,1.5)	$n=250$	β_1	−0.006	0.233	0.231	0.949	0.014	0.223	0.224	0.949	1.089
		β_2	−0.005	0.122	0.121	0.950	0.014	0.117	0.117	0.948	1.086
	$n=500$	β_1	0.005	0.164	0.166	0.948	0.018	0.158	0.161	0.948	1.078
		β_2	−0.004	0.085	0.087	0.944	0.009	0.083	0.084	0.948	1.071
Gamma(1,1)	$n=250$	β_1	−0.004	0.163	0.161	0.947	0.005	0.153	0.155	0.948	1.136
		β_2	−0.006	0.084	0.085	0.944	0.006	0.079	0.081	0.948	1.122
	$n=500$	β_1	−0.005	0.115	0.115	0.949	0.004	0.109	0.112	0.950	1.108
		β_2	−0.007	0.060	0.060	0.946	0.001	0.057	0.058	0.951	1.096

See the Note to Table 2.2.

Table 2.4: Simulation results for the original Buckley-James estimator

Error Distribution			Right End Bias	Right End SE	Mid-point Bias	Mid-point SE
$N(0,1)$	$n=250$	β_1	0.291	0.115	0.131	0.124
		β_2	0.243	0.062	0.119	0.065
	$n=500$	β_1	0.291	0.084	0.133	0.088
		β_2	0.243	0.044	0.119	0.046
EV(0,1)	$n=250$	β_1	0.388	0.125	0.233	0.139
		β_2	0.319	0.069	0.195	0.074
	$n=500$	β_1	0.390	0.090	0.233	0.099
		β_2	0.318	0.049	0.196	0.053
EV($-$0.5,1.5)	$n=250$	β_1	0.536	0.154	0.396	0.177
		β_2	0.432	0.084	0.322	0.093
	$n=500$	β_1	0.537	0.111	0.401	0.124
		β_2	0.432	0.059	0.321	0.064
Gamma(1,1)	$n=250$	β_1	0.390	0.127	0.233	0.141
		β_2	0.319	0.069	0.196	0.074
	$n=500$	β_1	0.389	0.088	0.232	0.097
		β_2	0.319	0.048	0.196	0.052

See the Note to Table 2.1.

A naive approach to analyzing interval-censored data is to approximate all interval-censored observations by single values and then apply the methodology for potentially right-censored data. We examined this approach in the second simulation setting by treating each interval-censored observation as exact failure time at the right end or the mid-point of the interval and applying the original Buckley-James estimator. As shown in Table 2.4, both approximations yield estimators with smaller standard error than the generalized Buckley-James and one-step estimators but induce severe bias in the parameter estimation.

2.4 An AIDS Example

We considered an AIDS Clinical Trial Group (ACTG) study (Goggins and Finkelstein, 2000). In this clinical trial, blood and urine samples were collected at clinical visits to test for the presence of opportunistic infection cytomegalovirus (CMV), which is also known as shedding of the virus. The blood and urine samples were originally scheduled to be collected about every 12 and 4 weeks, respectively. The CMV shedding times in both blood and urine are interval-censored in that the events are only known to occur between the last negative and first positive tests.

The data set consists of 204 HIV-infected patients with at least one blood and urine samples taken during the study. For CMV shedding time in blood, 7 patients have left-censored observations, 174 patients have right-censored observations, and 23 patients have interval-censored observations. For CMV shedding time in urine, the corresponding numbers are 49, 88 and 67. The data set also includes the patient's baseline CD4 cell count as an indicator of less than versus greater than 75 (cells/μl). It is of interest to determine whether the baseline CD4 cell count is predictive of CMV shedding time.

This data set was previously analyzed by Goggins and Finkelstein (2000) using the proportional hazards model for bivariate interval-censored data. To illustrate the proposed methods, we generated a PIC version of the data. Specifically, we defined the failure time as the minimum of the shedding times in blood and in urine. If the shedding times in blood and in urine are $(L_b, R_b]$ and $(L_u, R_u]$, then the failure time is known to lie within $(L_b \wedge L_u, R_b \wedge R_u]$. The numbers of left-, interval-, and right-censored observations are 51, 65, and 88, respectively. The interval lengths for the interval-censored observations range from 1 month to 9 months. We treated interval-censored observations with interval lengths less than 2 months as exact observations at the geometric mid-point of the interval to obtain 46 exact observations.

We fit the AFT model to the generated PIC data. We estimated the standard error of the generalized Buckley-James estimator using the Wald method based on 1,000 bootstrap datasets. We used the optimal bandwidths described in the previous section for the one-step estimation. For comparisons, we also fit the proportion hazards model using the NPMLE method described in Kim (2003). The results are summarized in Table 2.5.

The estimates of the regression parameter in the AFT model are negative and thus indicate that patients with higher CD4 cell counts tend to have longer time to CMV shedding. The one-step estimator yields a larger estimate of the effect size than the generalized Buckley-James estimator, with a slightly larger standard error estimate, resulting in a slightly smaller p-value. Not surprisingly, the estimate of the regression parameter under the proportional hazards model has an opposite sign.

2.5 Discussion

It is much more challenging, both computationally and theoretically, to deal with PIC data under the AFT model than under the proportional hazards model. We developed a generalization of the Buckley-James estimator and a one-step efficient estimator, both of which perform well in realistic settings. We tackled the theoretical challenges through careful use of modern empirical

Table 2.5: Regression analysis for the ACTG study

Model	Est	Std error	Z-statistic	p-value	95% CI
Proportional hazards model	0.814	0.205	3.974	<0.0001	(0.412, 1.215)
AFT model					
Generalized Buckley-James	-1.212	0.335	-3.616	<0.0001	(-1.835, -0.560)
One-step	-1.256	0.343	-3.664	<0.0001	(-1.802, -0.563)

95% CI is the 95% confidence interval based on the Wald method (for proportional hazards model) or the empirical percentiles of the bootstrap samples (for AFT model).

process theory and semiparametric efficiency theory.

A non-negligible proportion of exact observations is a crucial assumption for the proposed methods. It plays an important role in establishing the asymptotic properties. With this assumption, there are some subjects with exactly observed failure times, so the estimator for the survival function of ϵ can be estimated accurately at those points. This leads to the $\sqrt{n}$ convergence rate, a faster rate than with only interval-censored observations. Computationally, we let the survival function be a step function with jumps at the exact failure times. Without exact observations, a natural estimator for the survival function would be a step function with potential jumps at all interval endpoints, such that the likelihood becomes non-concave and the estimation becomes unstable.

In practice, certain bootstrap samples may contain too few or no exact observations. We suggest to delete those samples provided that they account for a small proportion of all bootstrap samples. An alternative strategy is to perform parametric bootstrap, which requires modeling of the censoring distribution (Efron and Tibshirani, 1993, pp. 90-92).

We used kernel estimation for density and its derivative in constructing the one-step estimator. The estimation for this one-dimensional distribution is relatively stable and accurate. If the density or its derivative is estimated with bias, the resulting function will depart from the efficient score function. However, the function is still a valid score function, such that the one-step estimator remains consistent.

For the accelerated failure time model with right-censored data, the rank-based estimator (Gehan, 1965), which solves the gradient of a weighted probability for the observed rank, can be easily calculated via the linear programming technique. Lin and Chen (2013) proposed a one-step efficient estimation procedure using the rank-based estimator as the initial estimator. For PIC data, due to the existence of interval-censored observations, we cannot recover the rank structure to obtain rank-based estimating equations.

In most medical studies, the events of interest are asymptomatic such that the failure times are intrinsically interval-censored. A common practice is to apply the methodology for right-censored data by treating the time of the first detection or the mid-point of the interval as the exact failure time. However, this strategy can induce severe bias in the estimation, as shown in our simulation studies. The PIC methodology as presented in this chapter provides a better approximation to interval-censored data by treating only the small intervals as exact observations.

2.6 Technical Details

2.6.1 Asymptotic Properties of the Buckley-James Estimator

To study the asymptotic properties of $(\widehat{\boldsymbol{\beta}}, \widehat{F})$, we impose the following regularity conditions:

Condition 1. The true value of regression parameters $\boldsymbol{\beta}_0$ belongs to a known compact set $\mathcal{B}$ in $\mathbb{R}^d$. The covariates belong to a bounded set $\mathcal{X}$ in $\mathbb{R}^d$.

Condition 2. The true distribution function F_0 is positive, continuous, and strictly increasing with a positive and twice-continuously differentiable density f_0.

Condition 3. The distribution of Δ depends only on the observed data $\{\Delta T, (1-\Delta)L, (1-\Delta)R, \boldsymbol{X}\}$. There exists a constant $c_0 > 0$ such that $\Pr(\Delta = 1|\boldsymbol{X}) > c_0$ with probability 1.

Condition 4. Conditional on $\Delta = 0$, the joint density of the examination times $(U_1, \ldots, U_K)$ is continuous and differentiable in their support with respect to some dominating measure. There exists a constant $\epsilon_0 > 0$ such that $\Pr\{\min_{0 \le k \le K-1} (U_{k+1} - U_k) > \epsilon_0 | \boldsymbol{X}, K, \Delta = 0\} = 1$.

Remark 2.1. *Condition 1 states the compactness of the Euclidean parameter space and the boundedness of covariates, which are standard assumptions in regression analysis. Conditions 2 and 4 are the smoothness conditions imposed on the underlying density functions. In practice, the examination may be scheduled at a fixed time, but patients may come randomly such that the joint density for the examination times is smooth. Condition 3 assumes coarsening at random, allowing* Δ *to depend on* T *and* (L, R)*, which is similar to the missing at random assumption. If* $\Delta = \xi I(T \le \tau)$*, where* $0 < \tau < \infty$ *is the study duration, and* ξ *is a Bernoulli random variable with success probability* p*, then* T *is exactly observed with probability* $p\Pr(T \le \tau)$*. If* $\Delta = I(R - L \le \eta_0)$*, where* η_0 *is some small positive number, then* T *is exactly observed with probability* $\Pr(R - L \le \eta_0)$*. Condition 3 also ensures* *that the proportion of exact observations is non-negligible, which is crucial to the* $\sqrt{n}$*-convergence rate for* $\widehat{F}$ *and invertibility of the information matrix.*

The consistency and asymptotic normality of the generalized Buckley-James estimator are stated below.

Theorem 2.1. *Suppose that Conditions 1-4 hold. Then, there is a root to the generalized Buckley-James estimating equation* $\boldsymbol{U}_n(\boldsymbol{\beta}, \boldsymbol{\beta}) = \mathbf{0}$ *such that* $(\widehat{\boldsymbol{\beta}}, \widehat{F})$ *is strongly consistent for* $(\boldsymbol{\beta}_0, F_0)$*, and* $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0, \widehat{F} - F_0)$ *converges weakly to a zero-mean Gaussian process in the metric space* $\mathbb{R}^d \times \overline{lin}(BV_1(\mathbb{R}))$*, where* $BV_1(\mathbb{R})$ *denotes the set of functions with total variation bounded by 1 on* $\mathbb{R}$*,* $\overline{lin}(BV_1(\mathbb{R}))$ *denotes the closed linear span for linear functionals of* $BV_1(\mathbb{R})$*, and* $(\widehat{F} - F_0)(h) = \int h(t) d(\widehat{F} - F_0)(t)$ *for* $h \in BV_1(\mathbb{R})$*.*

Proof. Let P denote the true probability measure. The generalized Buckley-James estimator $(\widehat{\boldsymbol{\beta}}, \widehat{F})$ is a Z-estimator solving the estimating equation $\mathbb{P}_n \begin{pmatrix} \boldsymbol{\Phi}^{(1)}(\boldsymbol{\beta}, F) \\ \Phi^{(2)}(\boldsymbol{\beta}, F)(t) \end{pmatrix} = \mathbf{0}$ for all t, where

$$\boldsymbol{\Phi}^{(1)}(\boldsymbol{\beta}, F) = (\boldsymbol{X} - \overline{\boldsymbol{X}}_n) \left\{ \Delta Y_{\boldsymbol{\beta}} + (1 - \Delta) \frac{\int_{L_{\boldsymbol{\beta}}}^{R_{\boldsymbol{\beta}}} u dF(u)}{F(R_{\boldsymbol{\beta}}) - F(L_{\boldsymbol{\beta}})} \right\},$$

and

$$\Phi^{(2)}(\boldsymbol{\beta}, F)(t) = \Delta I(Y_{\boldsymbol{\beta}} \leq t) + (1 - \Delta) \frac{F(R_{\boldsymbol{\beta}} \wedge t) - F(L_{\boldsymbol{\beta}} \wedge t)}{F(R_{\boldsymbol{\beta}}) - F(L_{\boldsymbol{\beta}})} - F(t).$$

We first replace the function $\Phi^{(2)}(\boldsymbol{\beta}, F)(t)$ by a function of the bounded variation function $h \in BV_1(\mathbb{R})$. Specifically, we define

$$\Phi^{(2)}(\boldsymbol{\beta}, F)(h) = \Delta h(Y_{\boldsymbol{\beta}}) + (1 - \Delta) \frac{\int_{L_{\boldsymbol{\beta}}}^{R_{\boldsymbol{\beta}}} h(t) dF(t)}{F(R_{\boldsymbol{\beta}}) - F(L_{\boldsymbol{\beta}})} - \int h(t) dF(t).$$

Write $\widetilde{h}(t) = \sum_{j=1}^{m'} h(t_j^*) I(t_{j-1}^* < t \leq t_j^*)$, where $t_j^* \in \{Y_{\boldsymbol{\beta},i} : \Delta_i = 1\}$ $(i = 1, \ldots, n)$. The step function $\widetilde{h}$ can be written as a finite sum of simple functions, denoted as $\widetilde{h}(t) = \sum_{j=1}^{m'} \alpha_j I(t \leq t_j^*)$. Then,

$$\mathbb{P}_n \Phi^{(2)}(\widehat{\boldsymbol{\beta}}, \widehat{F})(h) = \mathbb{P}_n \Phi^{(2)}(\widehat{\boldsymbol{\beta}}, \widehat{F})(\widetilde{h}) = \sum_{j=1}^{m'} \alpha_j \mathbb{P}_n \Phi^{(2)}(\widehat{\boldsymbol{\beta}}, \widehat{F})(t_j^*) = 0.$$

Thus, $(\widehat{\boldsymbol\beta}, \widehat{F})$ is the root of $\mathbb{P}_n \boldsymbol{\Phi}(\boldsymbol\beta, F)(h) \equiv \mathbb{P}_n \begin{pmatrix} \boldsymbol{\Phi}^{(1)}(\boldsymbol\beta, F) \\ \Phi^{(2)}(\boldsymbol\beta, F)(h) \end{pmatrix} = \mathbf{0}$ for all $h \in BV_1(\mathbb{R})$.

To prove the local consistency of $(\widehat{\boldsymbol\beta}, \widehat{F})$, we appeal to the implicit function theorem (Schwartz, 1969, p. 15). The distribution function F is contained in the Banach space $\overline{\text{lin}}(BV_1(\mathbb{R}))$, where $F(h) = \int h(t)dF(t)$ for $h \in BV_1(\mathbb{R})$. The corresponding norm is defined as $\|F\|_\rho = \sup_{\|h\|_{BV} \leq 1} |\int h(t)dF(t)|$, where $\|\cdot\|_{BV}$ is the bounded variation norm. The function $\mathbb{P}_n \boldsymbol{\Phi}(\boldsymbol\beta, F)$ is then a map from $\mathbb{R}^d \times \overline{\text{lin}}(BV_1(\mathbb{R}))$ to $\mathbb{R}^d \times \overline{\text{lin}}(BV_1(\mathbb{R}))$. For any $(\boldsymbol\beta, F)$ in $B_\delta(\boldsymbol\beta_0, F_0) \equiv \{(\boldsymbol\beta, F) : |\boldsymbol\beta - \boldsymbol\beta_0| + \|F - F_0\|_\rho < \delta\}$ and $(\boldsymbol\beta^*, F^*)$ in $\mathbb{R}^d \times \overline{\text{lin}}(BV_1(\mathbb{R}))$, the path-wise derivative of $\mathbb{P}_n \boldsymbol{\Phi}(\boldsymbol\beta, F)(h)$ along the direction $(\boldsymbol\beta + \eta\boldsymbol\beta^*, F + \eta F^*)$ is

$$\boldsymbol{C}_{\boldsymbol\beta,F,n}(\boldsymbol\beta^*, F^*)(h) \equiv \begin{pmatrix} \dfrac{\partial \mathbb{P}_n \boldsymbol{\Phi}^{(1)}(\boldsymbol\beta, F)}{\partial \boldsymbol\beta}(\boldsymbol\beta^*) & \dfrac{\partial \mathbb{P}_n \Phi^{(2)}(\boldsymbol\beta, F)(h)}{\partial \boldsymbol\beta}(\boldsymbol\beta^*) \\ \dfrac{\partial \mathbb{P}_n \boldsymbol{\Phi}^{(1)}(\boldsymbol\beta, F)}{\partial F}(F^*) & \dfrac{\partial \mathbb{P}_n \Phi^{(2)}(\boldsymbol\beta, F)(h)}{\partial F}(F^*) \end{pmatrix}.$$

By Lemma 2.1, $\boldsymbol{C}_{\boldsymbol\beta,F,n}(\boldsymbol\beta^*, F^*)(h) - \boldsymbol{C}_{\boldsymbol\beta,F}(\boldsymbol\beta^*, F^*)(h) = o_p(1)$ uniformly in $(\boldsymbol\beta, F, \boldsymbol\beta^*, F^*, h)$ in $B_\delta(\boldsymbol\beta_0, F_0) \times \mathbb{R}^d \times \overline{\text{lin}}(BV_1(\mathbb{R})) \times BV_1(\mathbb{R})$ for some $\delta > 0$, where

$$\begin{aligned} \boldsymbol{C}_{\boldsymbol\beta,F}(\boldsymbol\beta^*, F^*)(h) &= \begin{pmatrix} \dfrac{\partial P \boldsymbol{\Phi}^{(1)}(\boldsymbol\beta, F)}{\partial \boldsymbol\beta}(\boldsymbol\beta^*) & \dfrac{\partial P \Phi^{(2)}(\boldsymbol\beta, F)(h)}{\partial \boldsymbol\beta}(\boldsymbol\beta^*) \\ \dfrac{\partial P \boldsymbol{\Phi}^{(1)}(\boldsymbol\beta, F)}{\partial F}(F^*) & \dfrac{\partial P \Phi^{(2)}(\boldsymbol\beta, F)(h)}{\partial F}(F^*) \end{pmatrix} \\ &= \begin{pmatrix} \dfrac{\partial P \boldsymbol{\Phi}^{(1)}(\boldsymbol\beta, F)}{\partial \boldsymbol\beta}(\boldsymbol\beta^*) & \dfrac{\partial P \Phi^{(2)}(\boldsymbol\beta, F)(h)}{\partial \boldsymbol\beta}(\boldsymbol\beta^*) \\ \int \boldsymbol{w}_{\boldsymbol\beta,F}(t)dF^*(t) & \int \mathcal{H}_{\boldsymbol\beta,F}(h)(t)dF^*(t) \end{pmatrix}, \end{aligned} \tag{2.4}$$

with

$$\boldsymbol{w}_{\boldsymbol\beta,F}(t) = E\left((\boldsymbol{X} - E\boldsymbol{X})(1-\Delta) I\left(L_{\boldsymbol\beta} < t \leq R_{\boldsymbol\beta}\right) \left[\frac{\{F(R_{\boldsymbol\beta}) - F(L_{\boldsymbol\beta})\} t - \int_{L_{\boldsymbol\beta}}^{R_{\boldsymbol\beta}} u dF(u)}{\{F(R_{\boldsymbol\beta}) - F(L_{\boldsymbol\beta})\}^2} \right] \right),$$

and

$$\begin{aligned} \mathcal{H}_{\boldsymbol\beta,F}(h)(t) &= E\left((1-\Delta) I\left(L_{\boldsymbol\beta} < t \leq R_{\boldsymbol\beta}\right) \left[\frac{\{F(R_{\boldsymbol\beta}) - F(L_{\boldsymbol\beta})\} h(t) - \int_{L_{\boldsymbol\beta}}^{R_{\boldsymbol\beta}} h(u) dF(u)}{\{F(R_{\boldsymbol\beta}) - F(L_{\boldsymbol\beta})\}^2} \right] \right) \\ &\quad -h(t). \end{aligned} \tag{2.5}$$

By Lemma 2.2, $\boldsymbol{C}_{\boldsymbol{\beta},F}$ is invertible at $(\boldsymbol{\beta}_0, F_0)$ and continuous in $B_\delta(\boldsymbol{\beta}_0, F_0)$. Thus, the derivative operator $\boldsymbol{C}_{\boldsymbol{\beta},F,n}$ is invertible at $(\boldsymbol{\beta}_0, F_0)$ and continuous in $B_\delta(\boldsymbol{\beta}_0, F_0)$. The implicit function theorem yields that $\mathbb{P}_n\boldsymbol{\Phi}(\boldsymbol{\beta}, F)$ is one-to-one in $B_\delta(\boldsymbol{\beta}_0, F_0)$. Simple algebraic manipulation yields

$$P\begin{pmatrix}\boldsymbol{\Phi}^{(1)}(\boldsymbol{\beta}_0, F_0)\\ \Phi^{(2)}(\boldsymbol{\beta}_0, F_0)(h)\end{pmatrix} = \begin{pmatrix}E\left\{(\boldsymbol{X} - E\boldsymbol{X})E\left(\epsilon|\boldsymbol{X}\right)\right\}\\ E\left\{\Delta h(\epsilon) + (1-\Delta)h(\epsilon)\right\} - E\left\{h(\epsilon)\right\}\end{pmatrix} = \mathbf{0}.$$

Thus, $\mathbb{P}_n\boldsymbol{\Phi}(\boldsymbol{\beta}_0, F_0) = o_p(1)$. For an arbitrary small $\delta > 0$, there exists a large enough n such that there exists $(\widehat{\boldsymbol{\beta}}, \widehat{F})$ with $(\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| + \|\widehat{F} - F_0\|_\rho) < \delta$ and $\mathbb{P}_n\boldsymbol{\Phi}(\widehat{\boldsymbol{\beta}}, \widehat{F}) = \mathbf{0}$.

Next, we prove the asymptotic normality of $(\widehat{\boldsymbol{\beta}}, \widehat{F})$. Write $\Psi(\boldsymbol{\beta}, F)(\boldsymbol{a}, h) = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{\Phi}^{(1)}(\boldsymbol{\beta}, F) + \Phi^{(2)}(\boldsymbol{\beta}, F)(h)$ for $\boldsymbol{a} \in \mathbb{R}^d$ and $h \in BV_1(\mathbb{R})$. By the Taylor series expansion,

$$\begin{aligned}
0 &= \left\{\mathbb{P}_n\Psi(\widehat{\boldsymbol{\beta}}, \widehat{F})(\boldsymbol{a}, h) - P\Psi(\widehat{\boldsymbol{\beta}}, \widehat{F})(\boldsymbol{a}, h)\right\} + \left\{P\Psi(\widehat{\boldsymbol{\beta}}, \widehat{F})(\boldsymbol{a}, h) - P\Psi(\boldsymbol{\beta}_0, F_0)(\boldsymbol{a}, h)\right\}\\
&= \left(\mathbb{P}_n - P\right)\Psi(\widehat{\boldsymbol{\beta}}, \widehat{F})(\boldsymbol{a}, h) + \boldsymbol{a}^{\mathrm{T}}\frac{\partial P\boldsymbol{\Phi}^{(1)}(\boldsymbol{\beta}_0, F_0)}{\partial\boldsymbol{\beta}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + \frac{\partial P\Phi^{(2)}(\boldsymbol{\beta}_0, F_0)(h)}{\partial\boldsymbol{\beta}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\\
&\quad + \boldsymbol{a}^{\mathrm{T}}\int \boldsymbol{w}_{\boldsymbol{\beta}_0,F_0}(t)d(\widehat{F} - F_0)(t) + \int \mathcal{H}_{\boldsymbol{\beta}_0,F_0}(h)(t)d(\widehat{F} - F_0)(t) + o_p\left(\left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right\|\right)\\
&\quad + o_p\left(\left\|\widehat{F} - F_0\right\|_\rho\right).
\end{aligned} \tag{2.6}$$

By Lemma 2.2, the operator $\mathcal{H}_{\boldsymbol{\beta}_0,F_0}$ is continuously invertible from $BV_1(\mathbb{R})$ to $BV_1(\mathbb{R})$. For any $\boldsymbol{\mu} \in \mathbb{R}^d$, the Taylor series expansion in (2.6) with $\boldsymbol{a} = \boldsymbol{\mu}$ and $h = -\mathcal{H}^{-1}_{\boldsymbol{\beta}_0,F_0}(\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{w}_{\boldsymbol{\beta}_0,F_0})$ yields

$$\begin{aligned}
&\sqrt{n}\left[\boldsymbol{\mu}^{\mathrm{T}}\frac{\partial P\boldsymbol{\Phi}^{(1)}(\boldsymbol{\beta}_0, F_0)}{\partial\boldsymbol{\beta}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + \frac{\partial P\Phi^{(2)}(\boldsymbol{\beta}_0, F_0)\{-\mathcal{H}^{-1}_{\boldsymbol{\beta}_0,F_0}(\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{w}_{\boldsymbol{\beta}_0,F_0})\}}{\partial\boldsymbol{\beta}}\right](\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\\
= \;& -\mathbb{G}_n\Psi(\widehat{\boldsymbol{\beta}}, \widehat{F})\left\{\boldsymbol{\mu}, -\mathcal{H}^{-1}_{\boldsymbol{\beta}_0,F_0}(\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{w}_{\boldsymbol{\beta}_0,F_0})\right\} + o_p\left(\sqrt{n}\left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right\| + \sqrt{n}\left\|\widehat{F} - F_0\right\|_\rho\right),
\end{aligned}$$

where $\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - P)$, which is the empirical process. Write

$$\boldsymbol{D} = \frac{\partial P\boldsymbol{\Phi}^{(1)}(\boldsymbol{\beta}_0, F_0)}{\partial\boldsymbol{\beta}} - \frac{\partial P\Phi^{(2)}(\boldsymbol{\beta}_0, F_0)\left\{\mathcal{H}^{-1}_{\boldsymbol{\beta}_0,F_0}(\boldsymbol{w}_{\boldsymbol{\beta}_0,F_0})\right\}}{\partial\boldsymbol{\beta}}, \tag{2.7}$$

and

$$\boldsymbol{\psi}\left(\boldsymbol{\beta}, F\right) = \boldsymbol{\Phi}^{(1)}(\boldsymbol{\beta}, F) - \Phi^{(2)}(\boldsymbol{\beta}, F)\left\{\mathcal{H}^{-1}_{\boldsymbol{\beta}_0,F_0}(\boldsymbol{w}_{\boldsymbol{\beta}_0,F_0})\right\}.$$

The matrix $\boldsymbol{D}$ is invertible by Lemma 2.2. It then follows from the arbitrariness of $\boldsymbol{\mu}$ that

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = -\boldsymbol{D}^{-1}\mathbb{G}_n\boldsymbol{\psi}\left(\widehat{\boldsymbol{\beta}}, \widehat{F}\right) + o_p\left(\sqrt{n}\left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right\| + \sqrt{n}\left\|\widehat{F} - F_0\right\|_\rho\right)$$

For any function $\nu(\cdot) \in BV_1(\mathbb{R})$, equation (2.6) with $\boldsymbol{a} = \boldsymbol{0}$ and $h = \mathcal{H}^{-1}_{\boldsymbol{\beta}_0,F_0}(\nu)$ yields

$$\begin{aligned}
& \sqrt{n}\int \nu(t)d(\widehat{F} - F_0)(t) \\
= \quad & -\mathbb{G}_n\Phi^{(2)}(\widehat{\boldsymbol{\beta}}, \widehat{F})\left\{\mathcal{H}^{-1}_{\boldsymbol{\beta}_0,F_0}(\nu)\right\} + \frac{\partial P\Phi^{(2)}(\boldsymbol{\beta}_0, F_0)\left\{\mathcal{H}^{-1}_{\boldsymbol{\beta}_0,F_0}(\nu)\right\}}{\partial \boldsymbol{\beta}}\boldsymbol{D}^{-1}\mathbb{G}_n\boldsymbol{\psi}\left(\widehat{\boldsymbol{\beta}}, \widehat{F}\right) \\
& + o_p\left(\sqrt{n}\left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right\| + \sqrt{n}\left\|\widehat{F} - F_0\right\|_\rho\right).
\end{aligned}$$

Then, for any $(\boldsymbol{\mu}, \nu) \in \mathbb{R}^d \times BV_1(\mathbb{R})$,

$$\begin{aligned}
& \sqrt{n}\left\{\boldsymbol{\mu}^{\mathrm{T}}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) + \int \nu(t)d(\widehat{F} - F_0)(t)\right\} \\
= \quad & -\mathbb{G}_n\left(\left[\boldsymbol{\mu}^{\mathrm{T}} - \frac{\partial P\Phi^{(2)}(\boldsymbol{\beta}_0, F_0)\left\{\mathcal{H}^{-1}_{\boldsymbol{\beta}_0,F_0}(\nu)\right\}}{\partial \boldsymbol{\beta}}\right]\boldsymbol{D}^{-1}\boldsymbol{\psi}\left(\boldsymbol{\beta}_0, F_0\right) + \Phi^{(2)}(\boldsymbol{\beta}_0, F_0)\left\{\mathcal{H}^{-1}_{\boldsymbol{\beta}_0,F_0}(\nu)\right\}\right) \\
& + o_p(1).
\end{aligned}$$

By Lemma 2.1, $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0, \widehat{F} - F_0)$ converges weakly to a zero-mean Gaussian process in the metric space $\mathbb{R}^d \times \overline{\mathrm{lin}}(BV_1(\mathbb{R}))$. □

2.6.2 Asymptotic Properties of the Bootstrap Variance Estimator

The following theorem states the asymptotic properties of the bootstrap estimator, thereby validating the bootstrap procedure.

Theorem 2.2. *Suppose that Conditions 1-4 hold. Then, the conditional distribution of $\sqrt{n}(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}})$ given the data converges weakly to the asymptotic distribution of $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$.*

Proof. Let $\widehat{F}^*$ be the bootstrap estimator for the distribution function of the error term. The estimator $(\widehat{\boldsymbol{\beta}}^*, \widehat{F}^*)$ solves the bootstrap version of the estimating equation $\widehat{\mathbb{P}}_n\boldsymbol{\Phi}(\widehat{\boldsymbol{\beta}}^*, \widehat{F}^*)(h) = \boldsymbol{0}$, where $\widehat{\mathbb{P}}_n$ denote the bootstrap empirical distribution. The consistency and convergence rate of $\widehat{\boldsymbol{\beta}}^*$

and $\widehat{F}^*$ follow from the arguments in Section 2.6.1. By the Taylor series expansion,

$$\begin{aligned}
0 &= \widehat{\mathbb{P}}_n\left\{\Psi(\widehat{\boldsymbol{\beta}}^*, \widehat{F}^*)(\boldsymbol{a}, h) - \Psi(\widehat{\boldsymbol{\beta}}, \widehat{F})(\boldsymbol{a}, h)\right\} + \left(\widehat{\mathbb{P}}_n - \mathbb{P}_n\right)\Psi(\widehat{\boldsymbol{\beta}}, \widehat{F})(\boldsymbol{a}, h) \\
&= \boldsymbol{a}^{\mathrm{T}}\frac{\partial\widehat{\mathbb{P}}_n\boldsymbol{\Phi}^{(1)}(\widehat{\boldsymbol{\beta}}, \widehat{F})}{\partial\boldsymbol{\beta}}(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}) + \frac{\partial\widehat{\mathbb{P}}_n\Phi^{(2)}(\widehat{\boldsymbol{\beta}}, \widehat{F})(h)}{\partial\boldsymbol{\beta}}(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}) + \boldsymbol{a}^{\mathrm{T}}\frac{\partial\widehat{\mathbb{P}}_n\boldsymbol{\Phi}^{(1)}(\widehat{\boldsymbol{\beta}}, \widehat{F})}{\partial F}(\widehat{F}^* - \widehat{F}_0) \\
&\quad + \frac{\partial\widehat{\mathbb{P}}_n\Phi^{(2)}(\widehat{\boldsymbol{\beta}}, \widehat{F})(h)}{\partial F}(\widehat{F}^* - \widehat{F}) + o_p\left(\left\|\widehat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}_0\right\| + \left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right\|\right) \\
&\quad + o_p\left(\left\|\widehat{F}^* - F_0\right\|_\rho + \left\|\widehat{F} - F_0\right\|_\rho\right) + \left(\widehat{\mathbb{P}}_n - \mathbb{P}_n\right)\Psi(\widehat{\boldsymbol{\beta}}, \widehat{F})(\boldsymbol{a}, h).
\end{aligned}$$

By Theorem 3.6.1 of van der Vaart and Wellner (1996), the conditional distribution of $(\widehat{\mathbb{P}}_n - \mathbb{P}_n)\Psi(\widehat{\boldsymbol{\beta}}, \widehat{F})(\boldsymbol{a}, h)$ given the data is asymptotically equivalent to the distribution of $(\mathbb{P}_n - P)\Psi(\widehat{\boldsymbol{\beta}}, \widehat{F})(\boldsymbol{a}, h)$. For any $\boldsymbol{\mu} \in \mathbb{R}^d$, we let $\boldsymbol{a} = \boldsymbol{\mu}$ and $h = -\mathcal{H}^{-1}_{\boldsymbol{\beta}_0, F_0}(\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{w}_{\boldsymbol{\beta}_0, F_0})$. Then,

$$\begin{aligned}
\sqrt{n}\boldsymbol{\mu}^{\mathrm{T}}\left(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}\right) &= -\boldsymbol{D}^{-1}\left[\sqrt{n}\left(\widehat{\mathbb{P}}_n - \mathbb{P}_n\right)\Psi(\widehat{\boldsymbol{\beta}}, \widehat{F})\left\{\boldsymbol{\mu}, -\mathcal{H}^{-1}_{\boldsymbol{\beta}_0, F_0}(\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{w}_{\boldsymbol{\beta}_0, F_0})\right\} + o_p(1)\right] \\
&= -\boldsymbol{D}^{-1}\boldsymbol{\mu}^{\mathrm{T}}\mathbb{G}_n\boldsymbol{\psi}(\boldsymbol{\beta}_0, F_0) + o_p(1).
\end{aligned}$$

It then follows from the arbitrariness of $\boldsymbol{\mu}$ that

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}\right) = -\boldsymbol{D}^{-1}\mathbb{G}_n\boldsymbol{\psi}(\boldsymbol{\beta}_0, F_0) + o_p(1).$$

Therefore, $\sqrt{n}(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}})$ converges weakly to a zero-mean Gaussian variable, and $\sqrt{n}(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}})$ and $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ have the same asymptotic distribution.

2.6.3 Derivation of the Efficient Score

Let $P_{\boldsymbol{\beta}, F}$ be the likelihood under the AFT model with PIC data. Let $F_\eta(h)(\cdot) = \int^{(\cdot)}[1 + \eta\{h(t) - \int h(u)dF_0(u)\}]dF_0(t)$ for $h \in L_2(P)$, and $\eta \in [-\delta, \delta]$ for some small $\delta > 0$. The score function for $\boldsymbol{\beta}$ is

$$\dot{\boldsymbol{l}}_{\boldsymbol{\beta}}(\mathcal{O}; \boldsymbol{\beta}_0, F_0) = -\boldsymbol{X}\left\{\Delta\frac{f_0'(Y_{\boldsymbol{\beta}_0})}{f_0(Y_{\boldsymbol{\beta}_0})} + (1 - \Delta)\frac{f_0(R_{\boldsymbol{\beta}_0}) - f_0(L_{\boldsymbol{\beta}_0})}{F_0(R_{\boldsymbol{\beta}_0}) - F_0(L_{\boldsymbol{\beta}_0})}\right\}.$$

The score function for F along the one-dimensional submodel $\{P_{\boldsymbol{\beta}, F_\eta(h)} : h \in L_2(P), \eta \in [-\delta, \delta]\}$ is

$$\dot{l}_F(\mathcal{O}; \boldsymbol{\beta}_0, F_0)(h) = \Delta h(Y_{\boldsymbol{\beta}_0}) + (1 - \Delta)\frac{\int_{L_{\boldsymbol{\beta}_0}}^{R_{\boldsymbol{\beta}_0}} h(t)dF_0(t)}{F_0(R_{\boldsymbol{\beta}_0}) - F_0(L_{\boldsymbol{\beta}_0})} - \int h(t)dF_0(t).$$

The efficient score $\widetilde{\boldsymbol{l}}_{\boldsymbol{\beta}}(\mathcal{O};\boldsymbol{\beta}_0,F_0)$ is defined as the linear combination of the scores that is orthogonal to the tangent set for F (Bickel et al., 1993, chap. 3). Thus,

$$\widetilde{\boldsymbol{l}}_{\boldsymbol{\beta}}(\mathcal{O};\boldsymbol{\beta}_0,F_0)=\dot{\boldsymbol{l}}_{\boldsymbol{\beta}}(\mathcal{O};\boldsymbol{\beta}_0,F_0)+\dot{l}_F(\mathcal{O};\boldsymbol{\beta}_0,F_0)(\boldsymbol{h}^*),$$

with

$$E\left[\left\{\dot{\boldsymbol{l}}_{\boldsymbol{\beta}}(\mathcal{O};\boldsymbol{\beta}_0,F_0)+\dot{l}_F(\mathcal{O};\boldsymbol{\beta}_0,F_0)(\boldsymbol{h}^*)\right\}\dot{l}_F(\mathcal{O};\boldsymbol{\beta}_0,F_0)(h)\right]=\mathbf{0} \tag{2.8}$$

for all $h\in L_2(P)$, where $\boldsymbol{h}^*$ is a d-vector of functions in $L_2(P)$. Therefore, for any $h\in L_2(P)$,

$$\begin{aligned}
&E\left[\Delta h(Y_{\boldsymbol{\beta}_0})\left\{\boldsymbol{h}^*(Y_{\boldsymbol{\beta}_0})-\int\boldsymbol{h}^*dF_0-\boldsymbol{X}\frac{f'(Y_{\boldsymbol{\beta}_0})}{f(Y_{\boldsymbol{\beta}_0})}\right\}\right.\\
&\left.+(1-\Delta)\frac{\int_{L_{\boldsymbol{\beta}_0}}^{R_{\boldsymbol{\beta}_0}}h(t)dF_0(t)}{F_0(R_{\boldsymbol{\beta}_0})-F_0(L_{\boldsymbol{\beta}_0})}\left\{\frac{\int_{L_{\boldsymbol{\beta}_0}}^{R_{\boldsymbol{\beta}_0}}\boldsymbol{h}^*(t)dF_0(t)}{F_0(R_{\boldsymbol{\beta}_0})-F_0(L_{\boldsymbol{\beta}_0})}-\int\boldsymbol{h}^*dF_0-\boldsymbol{X}\frac{f_0(R_{\boldsymbol{\beta}_0})-f_0(L_{\boldsymbol{\beta}_0})}{F_0(R_{\boldsymbol{\beta}_0})-F_0(L_{\boldsymbol{\beta}_0})}\right\}\right]\\
&=\int\left\{E\left(\Delta|Y_{\boldsymbol{\beta}_0}=t\right)\left\{\boldsymbol{h}^*(t)-\int\boldsymbol{h}^*dF_0\right\}-E\left(\Delta\boldsymbol{X}|Y_{\boldsymbol{\beta}_0}=t\right)\frac{f_0'(t)}{f_0(t)}\right.\\
&\left.+E\left((1-\Delta)I(L_{\boldsymbol{\beta}_0}<t\le R_{\boldsymbol{\beta}_0})\left[\frac{\int_{L_{\boldsymbol{\beta}_0}}^{R_{\boldsymbol{\beta}_0}}\left\{\boldsymbol{h}^*(u)-\int\boldsymbol{h}^*dF_0\right\}dF_0(u)-\boldsymbol{X}\{f_0(R_{\boldsymbol{\beta}_0})-f_0(L_{\boldsymbol{\beta}_0})\}}{\left\{F_0(R_{\boldsymbol{\beta}_0})-F_0(L_{\boldsymbol{\beta}_0})\right\}^2}\right]\right)\right\}\\
&\times h(t)dF_0(t)=\mathbf{0},
\end{aligned}$$

such that

$$\begin{aligned}
&E\left(\Delta|Y_{\boldsymbol{\beta}_0}=t\right)\left\{\boldsymbol{h}^*(t)-\int\boldsymbol{h}^*dF_0\right\}-E\left(\Delta\boldsymbol{X}|Y_{\boldsymbol{\beta}_0}=t\right)\frac{f_0'(t)}{f_0(t)}\\
&+E\left((1-\Delta)I(L_{\boldsymbol{\beta}_0}<t\le R_{\boldsymbol{\beta}_0})\left[\frac{\int_{L_{\boldsymbol{\beta}_0}}^{R_{\boldsymbol{\beta}_0}}\left\{\boldsymbol{h}^*(u)-\int\boldsymbol{h}^*dF_0\right\}dF_0(u)-\boldsymbol{X}\{f_0(R_{\boldsymbol{\beta}_0})-f_0(L_{\boldsymbol{\beta}_0})\}}{\left\{F_0(R_{\boldsymbol{\beta}_0})-F_0(L_{\boldsymbol{\beta}_0})\right\}^2}\right]\right)\\
&=\mathbf{0}
\end{aligned}$$

for any t. Note that if $\boldsymbol{h}^*$ is a solution to (2.8), then $\boldsymbol{h}^*+\boldsymbol{c}$ is also a solution for arbitrary d-vector of constant functions $\boldsymbol{c}$. We denote $\boldsymbol{g}(\cdot)=\boldsymbol{h}^*(\cdot)-\int\boldsymbol{h}^*dF_0$, which is the unique solution of (2.8) with $\int\boldsymbol{g}(t)dF_0(t)=\mathbf{0}$. Then, the function $\boldsymbol{g}$ satisfies

$$E\left(\Delta|Y_{\boldsymbol{\beta}_0}=t\right)\boldsymbol{g}(t)-E\left(\Delta\boldsymbol{X}|Y_{\boldsymbol{\beta}_0}=t\right)\frac{f_0'(t)}{f_0(t)}$$

$$+ E\left((1-\Delta)I(L_{\boldsymbol{\beta}_0} < t \leq R_{\boldsymbol{\beta}_0})\left[\frac{\int_{L_{\boldsymbol{\beta}_0}}^{R_{\boldsymbol{\beta}_0}} \boldsymbol{g}(u)dF_0(u) - \boldsymbol{X}\left\{f_0(R_{\boldsymbol{\beta}_0}) - f_0(L_{\boldsymbol{\beta}_0})\right\}}{\left\{F_0(R_{\boldsymbol{\beta}_0}) - F_0(L_{\boldsymbol{\beta}_0})\right\}^2}\right]\right) = \mathbf{0}. \qquad (2.9)$$

If the error ϵ is normally distributed, then $\boldsymbol{g}(t) = tE(\boldsymbol{X})$, such that the efficient score function is equivalent to the generalized Buckley-James estimating function. Thus, the generalized Buckley-James estimator is semiparametric efficient when the error is normally distributed.

To implement $\widetilde{\boldsymbol{l}}_{\boldsymbol{\beta}}(\mathcal{O};\boldsymbol{\beta}_0, F_0)$, we need an appropriate estimator for $\boldsymbol{g}$. Because $\widehat{F}$ only takes jumps at t_l $(l = 1,\ldots,m)$, we approximate $\boldsymbol{g}$ by the step function $\widehat{\boldsymbol{g}}$ with jumps at $\{t_1,\ldots,t_m\}$. Solving equation (2.9) at $t = t_j$ and replacing f_0, f_0', $E\left(\Delta|Y_{\boldsymbol{\beta}_0} = t\right)$ and $E\left(\Delta\boldsymbol{X}|Y_{\boldsymbol{\beta}_0} = t\right)$ by the kernel estimators, we obtain

$$\sum_{l=1}^{m} a_{jl}\widehat{\boldsymbol{g}}(t_l) = \boldsymbol{c}_j$$

for $j = 1,\ldots,m$. Replacing $\boldsymbol{h}^*$ by $\widehat{\boldsymbol{g}}$ in the efficient score, we obtain the terms $\widehat{\boldsymbol{l}}(\widehat{\boldsymbol{\beta}},\widehat{F},\widehat{\boldsymbol{g}})$ in the Newton-Raphson update. $\square$

Remark 2.2. *The derivation of the efficient score is based on the projection on the tangent space of the nuisance parameter, which is similar to Example 25.28 in van der Vaart (1998). This derivation is different from the work of Ritov and Wellner (1988) for the AFT model with right-censored data. The latter is based on the martingale structure in right-censored data such that the right-censored counterpart of the left hand side of (2.8) is computed as the expectation of the predictable covariation process. If only right-censored data are observed, the least favorable direction has the closed form*

$$\begin{aligned}\boldsymbol{g}(t) \quad &= \quad E(\boldsymbol{X}|C - \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta}_0 \geq t)\left\{\frac{f_0'(t)}{f_0(t)} + \frac{f_0(t)}{1-F_0(t)}\right\} \\ &\quad - \int^t E(\boldsymbol{X}|C - \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta}_0 \geq s)\left[\frac{f_0'(s)}{1-F_0(s)} + \frac{f_0^2(s)}{\{1-F_0(s)\}^2}\right]\lambda'(s)ds,\end{aligned}$$

where C denote the censoring time.

2.6.4 Asymptotic Properties of the One-step Estimator

To describe the asymptotic properties of the one-step estimator $\widetilde{\boldsymbol{\beta}}$, we impose some conditions on the kernel function and bandwidths:

Condition 5. The kernel function K is twice-continuously differentiable, and $K^{(r)}(r = 1, 2)$ has bounded total variation in $(-\infty,\infty)$. The bandwidths $a_n = n^{-\nu_1}$ with $\nu_1 \in (0, 1/3)$ and $b_n = n^{-\nu_2}$

with $\nu_2 \in (0, 1/6)$.

Remark 2.3. *Condition 5 ensures that the kernel smoothed estimators for the efficient score functions are consistent approximations. This condition is satisfied by many kernel functions, including Gaussian kernels and smooth kernels with bounded support.*

The consistency and asymptotic efficiency for the one-step estimator are stated below.

Theorem 2.3. *Suppose that Conditions 1-5 hold. Then, $\sqrt{n}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ converges in distribution to a mean-zero normal random vector with a covariance matrix that attains the semiparametric efficiency bound and can be consistently estimated by $\left\{\mathbb{P}_n \widehat{\boldsymbol{l}}(\widehat{\boldsymbol{\beta}}, \widehat{F}, \widehat{\boldsymbol{g}})^{\otimes 2}\right\}^{-1}$.*

Proof. We first prove that the estimator $\widehat{\boldsymbol{g}}(\cdot)$ is consistent for $\boldsymbol{g}(\cdot)$, which satisfies the equation $M\left\{\boldsymbol{g}(t)\right\} = \mathbf{0}$ for all t, where

$$\begin{aligned} M\left\{\boldsymbol{g}(t)\right\} =& E\left(\Delta | Y_{\boldsymbol{\beta}_0} = t\right) \boldsymbol{g}(t) - E\left(\Delta \boldsymbol{X} | Y_{\boldsymbol{\beta}_0} = t\right) \frac{f_0'(t)}{f_0(t)} \\ &+ E\left((1-\Delta) I(L_{\boldsymbol{\beta}_0} < t \le R_{\boldsymbol{\beta}_0}) \left[\frac{\int_{L_{\boldsymbol{\beta}_0}}^{R_{\boldsymbol{\beta}_0}} \boldsymbol{g}(u) dF_0(u) - \boldsymbol{X}\left\{f_0(R_{\boldsymbol{\beta}_0}) - f_0(L_{\boldsymbol{\beta}_0})\right\}}{\left\{F_0(R_{\boldsymbol{\beta}_0}) - F_0(L_{\boldsymbol{\beta}_0})\right\}^2}\right]\right). \end{aligned}$$

The estimator $\widehat{\boldsymbol{g}}(t)$ is a d-vector of step functions such that $M_n\left\{\widehat{\boldsymbol{g}}(t)\right\} = \mathbf{0}$ at $t_1, \ldots, t_m$, where

$$\begin{aligned} M_n\left\{\widehat{\boldsymbol{g}}(t_l)\right\} =& \widehat{P}\left(\Delta | Y_{\widehat{\boldsymbol{\beta}}} = t_l\right) \widehat{\boldsymbol{g}}(t_l) - \widehat{P}\left(\Delta \boldsymbol{X} | Y_{\widehat{\boldsymbol{\beta}}} = t_l\right) \frac{\widehat{f}'(t_l)}{\widehat{f}(t_l)} \\ &+ \mathbb{P}_n\left((1-\Delta) I(L_{\widehat{\boldsymbol{\beta}}} < t_l \le R_{\widehat{\boldsymbol{\beta}}}) \left[\frac{\int_{L_{\boldsymbol{\beta}}}^{R_{\boldsymbol{\beta}}} \widehat{\boldsymbol{g}}(u) d\widehat{F}(u) - \boldsymbol{X}\left\{\widehat{f}(R_{\widehat{\boldsymbol{\beta}}}) - \widehat{f}(L_{\widehat{\boldsymbol{\beta}}})\right\}}{\left\{\widehat{F}(R_{\widehat{\boldsymbol{\beta}}}) - \widehat{F}(L_{\widehat{\boldsymbol{\beta}}})\right\}^2}\right]\right). \end{aligned}$$

Define

$$\begin{aligned} \widetilde{f}(t) &= \frac{1}{a_n} \int_0^\infty K\left(\frac{s-t}{a_n}\right) dF_0(s), \\ \widetilde{f}'(t) &= \frac{1}{b_n^3 \int u^2 K(u) du} \int_0^\infty (s-t) K\left(\frac{s-t}{b_n}\right) dF_0(s), \\ \widetilde{P}\left(\Delta | Y_{\boldsymbol{\beta}_0} = t\right) &= \frac{1}{n a_n f_0(t)} \sum_{i=1}^n \Delta_i K\left(\frac{Y_{\boldsymbol{\beta}_0, i} - t}{a_n}\right), \end{aligned}$$

and

$$\widetilde{P}\left(\Delta \boldsymbol{X} | Y_{\boldsymbol{\beta}_0} = t\right) = \frac{1}{n a_n f_0(t)} \sum_{i=1}^n \Delta_i \boldsymbol{X}_i K\left(\frac{Y_{\boldsymbol{\beta}_0, i} - t}{a_n}\right).$$

Under Condition 5, $\widetilde{P}(\Delta | Y_{\boldsymbol{\beta}_0} = t)$, $\widetilde{P}(\Delta \boldsymbol{X} | Y_{\boldsymbol{\beta}_0} = t)$, $\widetilde{f}'(t)$, and $\widetilde{f}(t)$ are consistent for $\Pr(\Delta | Y_{\boldsymbol{\beta}_0} = t)$, $\Pr(\Delta \boldsymbol{X} | Y_{\boldsymbol{\beta}_0} = t)$, $f_0'(t)$, and $f_0(t)$, respectively, by Theorem 2.5 of Schuster (1969). It then follows from the consistency of $(\widehat{\boldsymbol{\beta}}, \widehat{F})$ that the kernel estimators $\widehat{P}(\Delta | Y_{\boldsymbol{\beta}_0} = t)$, $\widehat{P}(\Delta \boldsymbol{X} | Y_{\boldsymbol{\beta}_0} = t)$, $\widehat{f}'(t)$, and $\widehat{f}(t)$ are strongly consistent. Then, for any d-vector of functions $\boldsymbol{g}(\cdot)$ with bounded variation on $\mathbb{R}$,

$$\sup_t \left| M_n \left\{ \boldsymbol{g}(t) \right\} - M \left\{ \boldsymbol{g}(t) \right\} \right| \to_p 0.$$

By Theorem 2.5 of Schuster (1969), the derivative of $\widehat{f}'(t)/\widehat{f}(t)$ converges to the derivative of $f_0'(t)/f_0(t)$. The last term of $M_n(\cdot)$ is the empirical measure of the product of a step function and a bounded function, so it has bounded total variation. Thus, the function $\widehat{\boldsymbol{g}}(\cdot)$ is a bounded variation function. By the Taylor series expansion,

$$\begin{aligned}
0 &= \int M_n \left\{ \widehat{\boldsymbol{g}}(t) \right\}^{\mathrm{T}} \left(\widehat{\boldsymbol{g}} - \boldsymbol{g} \right)(t) d\widehat{F}(t) \\
&= \int M_n \left\{ \widehat{\boldsymbol{g}}(t) \right\}^{\mathrm{T}} \left(\widehat{\boldsymbol{g}} - \boldsymbol{g} \right)(t) dF_0(t) + o_p(1) \\
&= \int \left(\left[M_n \left\{ \widehat{\boldsymbol{g}}(t) \right\} - M \left\{ \widehat{\boldsymbol{g}}(t) \right\} \right] + \left[M \left\{ \widehat{\boldsymbol{g}}(t) \right\} - M \left\{ \boldsymbol{g}(t) \right\} \right] \right)^{\mathrm{T}} \left(\widehat{\boldsymbol{g}} - \boldsymbol{g} \right)(t) dF_0(t) + o_p(1) \\
&= \int \left[M \left\{ \widehat{\boldsymbol{g}}(t) \right\} - M \left\{ \boldsymbol{g}(t) \right\} \right]^{\mathrm{T}} \left(\widehat{\boldsymbol{g}} - \boldsymbol{g} \right)(t) dF_0(t) + o_p(1) \\
&= -E \left\{ \Delta \left\| \widehat{\boldsymbol{g}}(Y_{\boldsymbol{\beta}_0}) - \boldsymbol{g}(Y_{\boldsymbol{\beta}_0}) \right\|^2 + (1 - \Delta) \left\| \frac{\int_{L_{\boldsymbol{\beta}_0}}^{R_{\boldsymbol{\beta}_0}} \left(\widehat{\boldsymbol{g}} - \boldsymbol{g} \right)(u) dF_0(u)}{F_0(R_{\boldsymbol{\beta}_0}) - F_0(L_{\boldsymbol{\beta}_0})} \right\|^2 \right\} + o_p(1).
\end{aligned}$$

Therefore, the difference $(\widehat{\boldsymbol{g}} - \boldsymbol{g})$ converges to zero in probability in the normed space $L_2(P)$. This implies the consistency of $\widehat{\boldsymbol{g}}$.

By Lemma 2.1, the class $\{\widetilde{\boldsymbol{l}}_{\boldsymbol{\beta}}(\boldsymbol{\beta}, F) \equiv \widetilde{\boldsymbol{l}}_{\boldsymbol{\beta}}(\mathcal{O}; \boldsymbol{\beta}, F) : (\boldsymbol{\beta}, F) \in B_\delta(\boldsymbol{\beta}_0, F_0)\}$ is Donsker for some fixed $\delta > 0$. It then follows from the consistency of $\widehat{\boldsymbol{\beta}}$, $\widehat{F}$, and $\widehat{\boldsymbol{g}}$ that $\mathbb{P}_n \widehat{\boldsymbol{l}}(\widehat{\boldsymbol{\beta}}, \widehat{F}, \widehat{\boldsymbol{g}}) = P \widetilde{\boldsymbol{l}}_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0, F_0) + o_p(1)$ and $\mathbb{P}_n \widehat{\boldsymbol{l}}(\widehat{\boldsymbol{\beta}}, \widehat{F}, \widehat{\boldsymbol{g}})^{\otimes 2} = P \widetilde{\boldsymbol{l}}_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0, F_0)^{\otimes 2} + o_p(1)$. The consistency of the one-step estimator $\tilde{\boldsymbol{\beta}}$ thus follows.

To derive the asymptotic distribution of $\widetilde{\boldsymbol{\beta}}$, we first consider the limit of $\left\{ \mathbb{P}_n \widehat{\boldsymbol{l}}(\widehat{\boldsymbol{\beta}}, \widehat{F}, \widehat{\boldsymbol{g}})^{\otimes 2} \right\}^{-1}$. Suppose that $P \widetilde{\boldsymbol{l}}_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0, F_0)^{\otimes 2}$ is singular. Then, $\widetilde{\boldsymbol{l}}_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0, F_0) = \boldsymbol{0}$ almost surely. We choose $\Delta = 1$ and $Y_{\boldsymbol{\beta}} = t$ such that $\boldsymbol{g}(t) + \boldsymbol{X} f_0'(t)/f_0(t) = \boldsymbol{0}$. No deterministic $\boldsymbol{g}$ can be found for arbitrary $\boldsymbol{X}$ to satisfy this equation. Thus, the information matrix $P \widetilde{\boldsymbol{l}}_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0, F_0)^{\otimes 2}$ is nonsingular, and

$\left\{\mathbb{P}_n\widehat{\boldsymbol{l}}(\widehat{\boldsymbol{\beta}},\widehat{F},\widehat{\boldsymbol{g}})^{\otimes 2}\right\}^{-1} \to_p \{P\widetilde{\boldsymbol{l}}_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0,F_0)^{\otimes 2}\}^{-1}$. By the Taylor series expansion,

$$
\begin{aligned}
& \sqrt{n}\left(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right) \\
= & \sqrt{n}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right)+\sqrt{n}\left\{\mathbb{P}_n\widehat{\boldsymbol{l}}\left(\widehat{\boldsymbol{\beta}},\widehat{F},\widehat{\boldsymbol{g}}\right)^{\otimes 2}\right\}^{-1} \\
& \times\left[\mathbb{P}_n\widehat{\boldsymbol{l}}\left(\boldsymbol{\beta}_0,\widehat{F},\widehat{\boldsymbol{g}}\right)-\left\{\mathbb{P}_n\widehat{\boldsymbol{l}}\left(\boldsymbol{\beta}_0,\widehat{F},\widehat{\boldsymbol{g}}\right)^{\otimes 2}\right\}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right)+o_p\left(\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right\|\right)\right] \\
= & \left[1-\left\{\mathbb{P}_n\widehat{\boldsymbol{l}}(\widehat{\boldsymbol{\beta}},\widehat{F},\widehat{\boldsymbol{g}})^{\otimes 2}\right\}^{-1}\left\{\mathbb{P}_n\widehat{\boldsymbol{l}}(\boldsymbol{\beta}_0,\widehat{F},\widehat{\boldsymbol{g}})^{\otimes 2}\right\}\right]\sqrt{n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)+o_p(1) \\
& +\sqrt{n}\left\{\mathbb{P}_n\widehat{\boldsymbol{l}}(\widehat{\boldsymbol{\beta}},\widehat{F},\widehat{\boldsymbol{g}})^{\otimes 2}\right\}^{-1}\left[\mathbb{P}_n\widehat{\boldsymbol{l}}(\boldsymbol{\beta}_0,\widehat{F},\widehat{\boldsymbol{g}})+\frac{\partial\mathbb{P}_n\widehat{\boldsymbol{l}}(\boldsymbol{\beta}_0,\widehat{F},\boldsymbol{g})}{\partial\boldsymbol{g}}(\widehat{\boldsymbol{g}}-\boldsymbol{g})\left\{1+O_p\left(\|\widehat{\boldsymbol{g}}-\boldsymbol{g}\|_{L_2(P)}\right)\right\}\right]
\end{aligned}
$$

Note that

$$
\begin{aligned}
\frac{\partial\mathbb{P}_n\widehat{\boldsymbol{l}}(\boldsymbol{\beta}_0,\widehat{F},\boldsymbol{g})}{\partial\boldsymbol{g}}(\widehat{\boldsymbol{g}}-\boldsymbol{g}) = & -\left(\mathbb{P}_n-P\right)\Phi^{(2)}(\boldsymbol{\beta}_0,F_0)(\widehat{\boldsymbol{g}}-\boldsymbol{g}) \\
& -\mathbb{P}_n\left((1-\Delta)\left[\frac{\int_{L_{\boldsymbol{\beta}_0}}^{R_{\boldsymbol{\beta}_0}}(\widehat{\boldsymbol{g}}-\boldsymbol{g})(u)d\left(\widehat{F}-F_0\right)(u)}{\widehat{F}(R_{\boldsymbol{\beta}_0})-\widehat{F}(L_{\boldsymbol{\beta}_0})}\right.\right. \\
& \left.\left.-\frac{\left\{\left(\widehat{F}-F_0\right)(R_{\boldsymbol{\beta}_0})-\left(\widehat{F}-F_0\right)(L_{\boldsymbol{\beta}_0})\right\}\int_{L_{\boldsymbol{\beta}_0}}^{R_{\boldsymbol{\beta}_0}}(\widehat{\boldsymbol{g}}-\boldsymbol{g})(u)dF_0(u)}{\left\{F_0(R_{\boldsymbol{\beta}_0})-F_0(L_{\boldsymbol{\beta}_0})\right\}\left\{\widehat{F}(R_{\boldsymbol{\beta}_0})-\widehat{F}(L_{\boldsymbol{\beta}_0})\right\}}\right]\right).
\end{aligned}
$$

Because $(\widehat{\boldsymbol{g}}-\boldsymbol{g})$ has bounded total variation, $\Phi^{(2)}(\boldsymbol{\beta}_0,F_0)(\widehat{\boldsymbol{g}}-\boldsymbol{g})$ belongs to a Donsker class according to Lemma 2.1. It then follows from the consistency of $\widehat{\boldsymbol{g}}$ that the first term on the right side of the above equation is $o_p(n^{-1/2})$. The second term is $o_p(n^{-1/2})$ by the $\sqrt{n}$-consistency of $\widehat{F}$ and the consistency of $\widehat{\boldsymbol{g}}$. Thus,

$$
\frac{\partial\mathbb{P}_n\widehat{\boldsymbol{l}}(\boldsymbol{\beta}_0,\widehat{F},\boldsymbol{g})}{\partial\boldsymbol{g}}(\widehat{\boldsymbol{g}}-\boldsymbol{g})=o_p(n^{-1/2}).
$$

Therefore,

$$
\begin{aligned}
\sqrt{n}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) &= \sqrt{n}\left\{\mathbb{P}_n\widehat{\boldsymbol{l}}(\widehat{\boldsymbol{\beta}},\widehat{F},\widehat{\boldsymbol{g}})^{\otimes 2}\right\}^{-1}\mathbb{P}_n\widehat{\boldsymbol{l}}(\boldsymbol{\beta}_0,\widehat{F},\boldsymbol{g})+o_p(1) \\
&= \left\{P\widetilde{\boldsymbol{l}}_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0,F_0)^{\otimes 2}\right\}^{-1}\mathbb{G}_n\widetilde{\boldsymbol{l}}_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0,F_0)+o_p(1).
\end{aligned}
$$

Hence, $\widetilde{\boldsymbol{\beta}}$ is asymptotically efficient for $\boldsymbol{\beta}_0$. □

Theorem 2.4. *Suppose that Conditions 1-5 hold. Let $\widetilde{\boldsymbol{\beta}}^*$ be the one-step estimator of a bootstrap* *sample. Then, the conditional distribution of* $\sqrt{n}(\widetilde{\boldsymbol{\beta}}^* - \widetilde{\boldsymbol{\beta}})$ *given the data converges weakly to the asymptotic distribution of* $\sqrt{n}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$.

Proof. Let $\widehat{\boldsymbol{g}}^*$ be the bootstrap version of $\widehat{\boldsymbol{g}}$. By the arguments in the proof of Theorem 2.3, $\widehat{\boldsymbol{g}}^* - \widehat{\boldsymbol{g}}$ converges to zero in probability in the normed space $L_2(P)$, and $\widehat{\mathbb{P}}_n \widehat{\boldsymbol{l}}(\widehat{\boldsymbol{\beta}}^*, \widehat{F}^*, \widehat{\boldsymbol{g}}^*) = P\widetilde{\boldsymbol{l}}_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0, F_0)^{\otimes 2} + o_P(1)$. It follows from the Taylor series expansion that

$$
\begin{aligned}
& \sqrt{n}\left(\widetilde{\boldsymbol{\beta}}^* - \widetilde{\boldsymbol{\beta}}\right) \\
= \;& \sqrt{n}\left(\widetilde{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}^*\right) + \sqrt{n}\left(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}\right) + \sqrt{n}\left(\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}}\right) \\
= \;& \sqrt{n}\left\{\widehat{\mathbb{P}}_n \widehat{\boldsymbol{l}}\left(\widehat{\boldsymbol{\beta}}^*, \widehat{F}^*, \widehat{\boldsymbol{g}}^*\right)^{\otimes 2}\right\}^{-1} \left[\widehat{\mathbb{P}}_n \widehat{\boldsymbol{l}}\left(\widehat{\boldsymbol{\beta}}, \widehat{F}^*, \widehat{\boldsymbol{g}}^*\right) - \left\{\widehat{\mathbb{P}}_n \widehat{\boldsymbol{l}}\left(\widehat{\boldsymbol{\beta}}, \widehat{F}^*, \widehat{\boldsymbol{g}}^*\right)^{\otimes 2}\right\}\left(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}\right)\right. \\
& \left. + o_p\left(\left\|\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}\right\|\right)\right] + \sqrt{n}(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}) - \sqrt{n}\left\{\mathbb{P}_n \widehat{\boldsymbol{l}}(\widehat{\boldsymbol{\beta}}, \widehat{F}, \widehat{\boldsymbol{g}})^{\otimes 2}\right\}^{-1} \mathbb{P}_n \widehat{\boldsymbol{l}}(\widehat{\boldsymbol{\beta}}, \widehat{F}, \widehat{\boldsymbol{g}}) \\
= \;& \left[1 - \left\{\widehat{\mathbb{P}}_n \widehat{\boldsymbol{l}}(\widehat{\boldsymbol{\beta}}^*, \widehat{F}^*, \widehat{\boldsymbol{g}}^*)^{\otimes 2}\right\}^{-1}\left\{\widehat{\mathbb{P}}_n \widehat{\boldsymbol{l}}(\widehat{\boldsymbol{\beta}}, \widehat{F}^*, \widehat{\boldsymbol{g}}^*)^{\otimes 2}\right\}\right]\sqrt{n}(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}) + o_p(1) \\
& + \sqrt{n}\left\{\widehat{\mathbb{P}}_n \widehat{\boldsymbol{l}}(\widehat{\boldsymbol{\beta}}^*, \widehat{F}^*, \widehat{\boldsymbol{g}}^*)^{\otimes 2}\right\}^{-1} \widehat{\mathbb{P}}_n \widehat{\boldsymbol{l}}(\widehat{\boldsymbol{\beta}}, \widehat{F}^*, \widehat{\boldsymbol{g}}^*) - \sqrt{n}\left\{\mathbb{P}_n \widehat{\boldsymbol{l}}(\widehat{\boldsymbol{\beta}}, \widehat{F}, \widehat{\boldsymbol{g}})^{\otimes 2}\right\}^{-1} \mathbb{P}_n \widehat{\boldsymbol{l}}(\widehat{\boldsymbol{\beta}}, \widehat{F}, \widehat{\boldsymbol{g}}) \\
= \;& \left\{P\widetilde{\boldsymbol{l}}_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0, F_0)^{\otimes 2}\right\}^{-1} \sqrt{n}\left(\widehat{\mathbb{P}}_n - \mathbb{P}_n\right)\widehat{\boldsymbol{l}}(\widehat{\boldsymbol{\beta}}, \widehat{F}, \widehat{\boldsymbol{g}}) + o_P(1).
\end{aligned}
$$

By Theorem 3.6.1 of van der Vaart and Wellner (1996), the conditional distribution of $\sqrt{n}(\widehat{\mathbb{P}}_n - \mathbb{P}_n)\widehat{\boldsymbol{l}}(\widehat{\boldsymbol{\beta}}, \widehat{F}, \widehat{\boldsymbol{g}})$ given the data is asymptotically equivalent to the distribution of $\sqrt{n}(\mathbb{P}_n - P)\widehat{\boldsymbol{l}}(\widehat{\boldsymbol{\beta}}, \widehat{F}, \widehat{\boldsymbol{g}})$. Therefore, $\sqrt{n}(\widetilde{\boldsymbol{\beta}}^* - \widetilde{\boldsymbol{\beta}})$ converges weakly to a zero-mean Gaussian variable, and $\sqrt{n}(\widetilde{\boldsymbol{\beta}}^* - \widetilde{\boldsymbol{\beta}})$ and $\sqrt{n}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ have the same asymptotic distribution. □

Remark 2.4. *The asymptotic theory presented in Theorems 2.1-2.4 was established under the condition that the true failure times are observed for a subset of study subjects. However, the conclusions of Theorems 2.1-2.3 are expected to hold when the "exact" observations are not the true failure times but rather the mid-points of small time intervals, provided that the lengths of the intervals are of the order* $o(1/n)$. *The corresponding proofs are substantially more difficult.*

2.6.5 Some Useful Lemmas

Lemma 2.1. *The class of functions* $\{\boldsymbol{\Phi}(\boldsymbol{\beta}, F)(h) : (\boldsymbol{\beta}, F) \in B_\delta(\boldsymbol{\beta}_0, F_0), h \in BV_1(\mathbb{R})\}$ *is Donsker for some fixed* $\delta > 0$.

Proof. By Theorem 9.3 and Lemma 9.11 of Kosorok (2007) and Lemma 2.6.13 of van der Vaart and Wellner (1996), the components of $\boldsymbol{\Phi}(\boldsymbol{\beta}, F)(h)$ are Donsker. It suffices to show that the denominator

$F(R_{\boldsymbol{\beta}}) - F(L_{\boldsymbol{\beta}})$ is bounded above zero. Note that

$$\begin{aligned} F(R_{\boldsymbol{\beta}}) - F(L_{\boldsymbol{\beta}}) &= F_0(R_{\boldsymbol{\beta}}) - F_0(L_{\boldsymbol{\beta}}) + \int I(L_{\boldsymbol{\beta}} < t \leq R_{\boldsymbol{\beta}}) d(F - F_0)(t) \\ &\geq F_0(R_{\boldsymbol{\beta}}) - F_0(L_{\boldsymbol{\beta}}) - 2\delta. \end{aligned}$$

If $L = U_0 = 0$ and $R = U_1$, which is strictly greater than ϵ_0 by Condition 4, then

$$F_0(R_{\boldsymbol{\beta}}) - F_0(L_{\boldsymbol{\beta}}) = F_0(R_{\boldsymbol{\beta}}) \geq F_0(R_{\boldsymbol{\beta}}) - F_0(R_{\boldsymbol{\beta}} - c_1) = c_1 f_0(t_1^*)$$

for a finite t_1^* and a finite positive constant c_1. Otherwise, $L = U_k$ and $R = U_{k+1}$ for some $k \in \{1, \ldots, K\}$. Then,

$$F_0(R_{\boldsymbol{\beta}}) - F_0(L_{\boldsymbol{\beta}}) = f_0(t_k^*) \left\{ \log \left(\frac{U_k + \epsilon_0}{U_k} \right) \right\} \geq c_2 f_0(t_k^*)$$

for a finite t_k^* and a finite positive constant c_2. By Condition 2, for any positive constant $M < \infty$, there exists some $\epsilon_M > 0$ such that $f_0(t) \geq \epsilon_M$ for any $t \in [-M, M]$. Thus, there exists $\epsilon^* > 0$ such that

$$F(R_{\boldsymbol{\beta}}) - F(L_{\boldsymbol{\beta}}) \geq \epsilon^* - 2\delta.$$

Therefore, $F(R_{\boldsymbol{\beta}}) - F(L_{\boldsymbol{\beta}})$ is strictly positive for $0 < \delta < \epsilon^*/2$. □

Lemma 2.2. *The operator $\boldsymbol{C}_{\boldsymbol{\beta},F}(\boldsymbol{\beta}^*, F^*)(h)$, as defined in (2.4), is continuous in the neighborhood of $(\boldsymbol{\beta}_0, F_0)$ and invertible at $(\boldsymbol{\beta}_0, F_0)$ from $\mathbb{R}^d \times \overline{lin}(BV_1(\mathbb{R}))$ to $\mathbb{R}^d \times \overline{lin}(BV_1(\mathbb{R}))$. The operator $\mathcal{H}_{\boldsymbol{\beta}_0,F_0}(h)$, as defined in (2.5), is continuously invertible from $BV_1(\mathbb{R})$ to $BV_1(\mathbb{R})$. The matrix $\boldsymbol{D}$, as defined in (2.7), is invertible.*

Proof. First, we consider the invertibility of the operator $\mathcal{H}_{\boldsymbol{\beta}_0,F_0}(h)$. Denote

$$J(t) = E\left\{ \frac{(1-\Delta) I\left(L_{\boldsymbol{\beta}_0} < t \leq R_{\boldsymbol{\beta}_0}\right)}{F_0(R_{\boldsymbol{\beta}_0}) - F_0(L_{\boldsymbol{\beta}_0})} \right\} - 1$$

and

$$\mathcal{G}\left\{h(t)\right\} = \mathcal{H}_{\boldsymbol{\beta}_0,F_0}\left\{h(t)\right\} - J(t) = -E\left[\frac{(1-\Delta)I\left(L_{\boldsymbol{\beta}_0} < t \le R_{\boldsymbol{\beta}_0}\right)\int_{L_{\boldsymbol{\beta}_0}}^{R_{\boldsymbol{\beta}_0}} h(u)dF_0(u)}{\left\{F_0(R_{\boldsymbol{\beta}_0}) - F_0(L_{\boldsymbol{\beta}_0})\right\}^2}\right].$$

The function $J(t)$ is strictly negative because

$$\begin{aligned} J(t) &= E\left[E\{(1-\Delta)I(L_{\boldsymbol{\beta}_0} < \epsilon \le R_{\boldsymbol{\beta}_0})|L_{\boldsymbol{\beta}_0}, R_{\boldsymbol{\beta}_0}\}\frac{I\left(L_{\boldsymbol{\beta}_0} < t \le R_{\boldsymbol{\beta}_0}\right)}{F_0(R_{\boldsymbol{\beta}_0}) - F_0(L_{\boldsymbol{\beta}_0})}\right] - 1 \\ &= E\left[E\left\{(1-\Delta)|L_{\boldsymbol{\beta}_0}, R_{\boldsymbol{\beta}_0}\right\} I\left(L_{\boldsymbol{\beta}_0} < t \le R_{\boldsymbol{\beta}_0}\right)\right] - 1 \\ &\le E(1-\Delta) - 1 < -c_0, \end{aligned}$$

where the last inequality follows from Condition 3. By Condition 4, the operator $\mathcal{G}$ projects $h \in BV_1(\mathbb{R})$ to a continuously differentiable function, so it is a compact operator. The operator $\mathcal{H}_{\boldsymbol{\beta}_0,F_0} = J + \mathcal{G}$ is then a Fredholm operator. By the theory of Fredholm operator (Rudin, 1973, pp. 99–103), the invertibility of $\mathcal{H}_{\boldsymbol{\beta}_0,F_0}$ holds if $\mathcal{H}_{\boldsymbol{\beta}_0,F_0}$ is one-to-one. Suppose that $\mathcal{H}_{\boldsymbol{\beta}_0,F_0}(h) = 0$ for some $h \in BV_1(\mathbb{R})$. Then,

$$0 = \int \mathcal{H}_{\boldsymbol{\beta}_0,F_0}(h)dh = P\left\{\frac{\partial \Phi^{(2)}(\boldsymbol{\beta}_0, F_0)(h)}{\partial F}(h)\right\} = P\left\{\left(\Phi^{(2)}(\boldsymbol{\beta}_0, F_0)(h)\right\}^2\right],$$

where the last equality follows from the fact that $\Phi^{(2)}(\boldsymbol{\beta}, F)(h)$ is the score for F along the direction indexed by h. Thus,

$$\Phi^{(2)}(\boldsymbol{\beta}_0, F_0)(h) = \Delta h(Y_{\boldsymbol{\beta}_0}) + (1-\Delta)\frac{\int_{L_{\boldsymbol{\beta}_0}}^{R_{\boldsymbol{\beta}_0}} h(t)dF_0(t)}{F_0(R_{\boldsymbol{\beta}_0}) - F_0(L_{\boldsymbol{\beta}_0})} - \int h(t)dF_0(t) = 0$$

almost surely. We choose $\Delta = 1$ and $Y_{\boldsymbol{\beta}_0} = t$ such that h is a constant function. In addition, $\mathcal{H}_{\boldsymbol{\beta}_0,F_0}(h) = 0$, so $h = 0$. The operator $\mathcal{H}_{\boldsymbol{\beta}_0,F_0}$ is one-to-one and thus is continuously invertible.

Clearly, $\boldsymbol{C}_{\boldsymbol{\beta},F}(\boldsymbol{\beta}^*, F^*)$ is continuous in the neighborhood of $(\boldsymbol{\beta}_0, F_0)$. Write

$$\boldsymbol{C}_{\boldsymbol{\beta}_0,F_0}(\boldsymbol{\beta}^*, F^*) = \begin{pmatrix} \boldsymbol{C}_{11}(\boldsymbol{\beta}^*) & \boldsymbol{C}_{12}(\boldsymbol{\beta}^*) \\ \boldsymbol{C}_{21}(F^*) & C_{22}(F^*) \end{pmatrix}.$$

Let $\mathcal{H}^*_{\boldsymbol{\beta}_0,F_0}$ denote the dual operator of $\mathcal{H}_{\boldsymbol{\beta}_0,F_0}$, which is also continuously invertible. Then, the operator

$$C_{22}\left\{F^*(h)\right\} = \int \mathcal{H}_{\boldsymbol{\beta}_0,F_0}\left\{h(t)\right\} dF^*(t) = \int h(t) d\mathcal{H}^*_{\boldsymbol{\beta}_0,F_0}\left\{F^*(t)\right\}$$

is a continuously invertible map from $\overline{\text{lin}}BV_1(\mathbb{R})$ to $\overline{\text{lin}}BV_1(\mathbb{R})$. If $\boldsymbol{C}_{\boldsymbol{\beta}_0,F_0}$ is invertible, then its inverse operator is

$$\boldsymbol{C}^{-1}_{\boldsymbol{\beta}_0,F_0} = \begin{pmatrix} \left(\boldsymbol{C}_{11} - \boldsymbol{C}_{12}C_{22}^{-1}\boldsymbol{C}_{21}\right)^{-1} & -\left(\boldsymbol{C}_{11} - \boldsymbol{C}_{12}C_{22}^{-1}\boldsymbol{C}_{21}\right)^{-1}\boldsymbol{C}_{12}C_{22}^{-1} \\ -C_{22}^{-1}\boldsymbol{C}_{21}\left(\boldsymbol{C}_{11} - \boldsymbol{C}_{12}C_{22}^{-1}\boldsymbol{C}_{21}\right)^{-1} & C_{22}^{-1} + C_{22}^{-1}\boldsymbol{C}_{21}\left(\boldsymbol{C}_{11} - \boldsymbol{C}_{12}C_{22}^{-1}\boldsymbol{C}_{21}\right)^{-1}\boldsymbol{C}_{12}C_{22}^{-1} \end{pmatrix}.$$

The invertibility of $\boldsymbol{C}_{\boldsymbol{\beta}_0,F_0}$ holds if the matrix

$$\begin{aligned} & \boldsymbol{C}_{11} - \boldsymbol{C}_{12}C_{22}^{-1}\boldsymbol{C}_{21} \\ = \; & \frac{\partial P\boldsymbol{\Phi}^{(1)}(\boldsymbol{\beta}_0, F_0)}{\partial \boldsymbol{\beta}} - \frac{\partial P\Phi^{(2)}(\boldsymbol{\beta}_0, F_0)}{\partial \boldsymbol{\beta}}\left[\left\{\frac{\partial P\Phi^{(2)}(\boldsymbol{\beta}_0, F_0)(h)}{\partial F}\right\}^{-1}\left\{\frac{\partial P\boldsymbol{\Phi}^{(1)}(\boldsymbol{\beta}_0, F_0)}{\partial F}\right\}\right] \\ = \; & \frac{\partial P\boldsymbol{\Phi}^{(1)}(\boldsymbol{\beta}_0, F_0)}{\partial \boldsymbol{\beta}} - \frac{\partial P\Phi^{(2)}(\boldsymbol{\beta}_0, F_0)\left\{\mathcal{H}^{-1}_{\boldsymbol{\beta}_0,F_0}(\boldsymbol{w}_{\boldsymbol{\beta}_0,F_0})\right\}}{\partial \boldsymbol{\beta}} = \boldsymbol{D} \end{aligned}$$

is invertible. Denote $\boldsymbol{Q}_{\boldsymbol{X}}(t) = \left\{\mathcal{H}^{-1}_{\boldsymbol{\beta}_0,F_0}(\boldsymbol{w}_{\boldsymbol{\beta}_0,F_0})\right\}(t) - \boldsymbol{X}t$. The matrix $\boldsymbol{D}$ can be written as

$$\begin{aligned} \boldsymbol{D} \; = \; & E\Bigg\{\boldsymbol{X}\Bigg(\Delta\boldsymbol{Q}'_{\boldsymbol{X}}(Y_{\boldsymbol{\beta}_0}) + (1-\Delta)\Bigg[\frac{\boldsymbol{Q}_{\boldsymbol{X}}(R_{\boldsymbol{\beta}_0})f_0(R_{\boldsymbol{\beta}_0}) - \boldsymbol{Q}_{\boldsymbol{X}}(L_{\boldsymbol{\beta}_0})f_0(L_{\boldsymbol{\beta}_0})}{F_0(R_{\boldsymbol{\beta}_0}) - F_0(L_{\boldsymbol{\beta}_0})} \\ & - \frac{\left\{f_0(R_{\boldsymbol{\beta}_0}) - f_0(L_{\boldsymbol{\beta}_0})\right\}\int_{L_{\boldsymbol{\beta}_0}}^{R_{\boldsymbol{\beta}_0}}\boldsymbol{Q}_{\boldsymbol{X}}(u)dF_0(u)}{\left\{F_0(R_{\boldsymbol{\beta}_0}) - F_0(L_{\boldsymbol{\beta}_0})\right\}^2}\Bigg]\Bigg)^{\mathrm{T}}\Bigg\}, \end{aligned}$$

and $\boldsymbol{Q}_{\boldsymbol{X}}(t)$ satisfies

$$\begin{aligned} & E\left((1-\Delta)I\left(L_{\boldsymbol{\beta}_0} < t \le R_{\boldsymbol{\beta}_0}\right)\left[\frac{\left\{F_0(R_{\boldsymbol{\beta}_0}) - F_0(L_{\boldsymbol{\beta}_0})\right\}\boldsymbol{Q}_{\boldsymbol{X}}(t) - \int_{L_{\boldsymbol{\beta}_0}}^{R_{\boldsymbol{\beta}_0}}\boldsymbol{Q}_{\boldsymbol{X}}(u)dF_0(u)}{\left\{F_0(R_{\boldsymbol{\beta}_0}) - F_0(L_{\boldsymbol{\beta}_0})\right\}^2}\right]\right) \\ & -E\left\{\boldsymbol{Q}_{\boldsymbol{X}}(t)\right\} = \boldsymbol{0}. \end{aligned} \tag{2.10}$$

We can treat the AFT model with PIC data as a submodel of a larger model, where the error ϵ can potentially depend on $\boldsymbol{X}$. The log-likelihood for the larger model is

$$\Delta \log f_{\epsilon|\boldsymbol{X}}(Y_{\boldsymbol{\beta}}) + (1-\Delta)\log\left\{F_{\epsilon|\boldsymbol{X}}(R_{\boldsymbol{\beta}}) - F_{\epsilon|\boldsymbol{X}}(L_{\boldsymbol{\beta}})\right\},$$

where $F_{\epsilon|\boldsymbol{X}}$ and $f_{\epsilon|\boldsymbol{X}}$ are, respectively, the distribution function and density function of ϵ given $\boldsymbol{X}$. The score of this larger model with respect to the direction

$$
\begin{aligned}
&\{\boldsymbol{\beta}_\eta(\boldsymbol{a}), F_{\eta,\boldsymbol{X}}(\cdot)\} \\
&\equiv \left[\boldsymbol{\beta}_0 + \eta\boldsymbol{a}, F_{\epsilon|\boldsymbol{X}}(\cdot) + \eta\int^{(\cdot)}\left\{\frac{f'_{\epsilon|\boldsymbol{X}}}{f_{\epsilon|\boldsymbol{X}}}(t)\boldsymbol{X} - \boldsymbol{Q}_{\boldsymbol{X}}(t) + \int \boldsymbol{Q}_{\boldsymbol{X}}(t)dF_{\epsilon|\boldsymbol{X}}(t)\right\}dF_{\epsilon|\boldsymbol{X}}(t)\right],
\end{aligned}
$$

where $\boldsymbol{a} \in \mathbb{R}^d$, is given by

$$
\boldsymbol{s}\left(\boldsymbol{\beta}, F_{\epsilon|\boldsymbol{X}}\right) = -\Delta\boldsymbol{Q}_{\boldsymbol{X}}(Y_{\boldsymbol{\beta}}) - (1-\Delta)\frac{\int_{L_{\boldsymbol{\beta}}}^{R_{\boldsymbol{\beta}}}\boldsymbol{Q}_{\boldsymbol{X}}(t)dF_{\epsilon|\boldsymbol{X}}(t)}{F_{\epsilon|\boldsymbol{X}}(R_{\boldsymbol{\beta}}) - F_{\epsilon|\boldsymbol{X}}(L_{\boldsymbol{\beta}})} + \int \boldsymbol{Q}_{\boldsymbol{X}}(t)dF_{\epsilon|\boldsymbol{X}}(t).
$$

The corresponding information matrix along that direction at the true value $(\boldsymbol{\beta}_0, F_0)$ is given by

$$
\begin{aligned}
&\boldsymbol{D} + E\Bigg\{(1-\Delta)\Bigg(\frac{\int_{L_{\boldsymbol{\beta}_0}}^{R_{\boldsymbol{\beta}_0}}\boldsymbol{Q}_{\boldsymbol{X}}(t)\left\{\boldsymbol{X}f'_0(t) - \boldsymbol{Q}_{\boldsymbol{X}}(t)f_0(t)\right\}^{\mathrm{T}}dt}{F_0(R_{\boldsymbol{\beta}_0}) - F_0(L_{\boldsymbol{\beta}_0})} \\
&-\frac{\int_{L_{\boldsymbol{\beta}_0}}^{R_{\boldsymbol{\beta}_0}}\boldsymbol{Q}_{\boldsymbol{X}}(t)dF_0(t)\left[\boldsymbol{X}\left\{f_0(R_{\boldsymbol{\beta}_0}) - f_0(L_{\boldsymbol{\beta}_0})\right\} - \int_{L_{\boldsymbol{\beta}_0}}^{R_{\boldsymbol{\beta}_0}}\boldsymbol{Q}_{\boldsymbol{X}}(t)dF_0(t)\right]^{\mathrm{T}}}{\left\{F_0(R_{\boldsymbol{\beta}_0}) - F_0(L_{\boldsymbol{\beta}_0})\right\}^2} \\
&-\int \boldsymbol{Q}_{\boldsymbol{X}}(t)\left\{\boldsymbol{X}f'_0(t) - \boldsymbol{Q}_{\boldsymbol{X}}(t)f_0(t)\right\}^{\mathrm{T}}dt\Bigg)\Bigg\}.
\end{aligned}
$$

The information matrix is equal to $\boldsymbol{D}$ because the second term is exact zero, which can be shown to be true if we multiply (2.10) by $\{\boldsymbol{X}f'_0(t) - \boldsymbol{Q}_{\boldsymbol{X}}(t)f_0(t)\}^{\mathrm{T}}$ and integrate t out. Thus, the matrix $\boldsymbol{D}$ is the information matrix of the larger model. It is singular only if the score $\boldsymbol{s}(\boldsymbol{\beta}_0, F_0)$ is zero almost surely. We can choose $\Delta = 1$ and $Y_{\boldsymbol{\beta}_0} = t$ to find the contradiction. Thus, the matrix $\boldsymbol{D}$ is nonsingular, and $\boldsymbol{C}_{\boldsymbol{\beta}_0,F_0}(\boldsymbol{\beta}^*, F^*)$ is invertible from $\mathbb{R}^d \times \overline{\mathrm{lin}}(BV_1(\mathbb{R}))$ to $\mathbb{R}^d \times \overline{\mathrm{lin}}(BV_1(\mathbb{R}))$. □

CHAPTER 3: SEMIPARAMETRIC REGRESSION ANALYSIS OF INTERVAL-CENSORED DATA WITH INFORMATIVE DROPOUT

3.1 Introduction

Interval-censored data arise when the timing of an event is not known precisely but rather is known to lie within a time interval. Such data are frequently encountered in medical research, where the ascertainment of the disease of interest is made over a series of examination times. An example is the Atherosclerosis Risk in Communities (ARIC) study (The ARIC Investigators, 1989), where subjects were examined for asymptomatic diseases, such as diabetes and hypertension, over five visits, with the first four each approximately three years apart and then a gap of about 15 years before the fifth visit, such that the disease was only known to occur within a broad time interval.

A number of methods have been developed for regression analysis of interval-censored data. In particular, nonparametric maximum likelihood estimation for the proportional odds, proportional hazards, and transformation models have been studied by Huang (1995), Huang (1996), and Zeng et al. (2016), respectively. Sieve estimation for the proportional odds and proportional hazards models has been suggested by Rossini and Tsiatis (1996), Huang and Rossini (1997), Shen (1998), and Cai and Betensky (2003). Rank-based estimation methods for linear transformation models have been proposed by Gu et al. (2005), Sun and Sun (2005), Zhang, Sun, Zhao and Sun (2005), and Zhang and Zhao (2013).

All aforementioned work assumes that the examination process is independent of the event of interest, possibly conditional on covariates. This assumption is often violated in chronic disease research because subjects may drop out of the study prematurely for health-related reasons. For example, in the ARIC study, a large number of subjects died before their last scheduled visit. When dropout is correlated with the event of interest, the existing methods may yield invalid inference. In the situation where dropout is caused by a terminal event, such as death, the existing methods, which fail to account for the fact that the event of interest cannot occur after the terminal event, will provide incorrect estimation of disease incidence even if dropout is independent of the event of interest.

In this chapter, we adjust for informative dropout through the use of a random effect. Specifically, we consider a broad class of joint models, under which the event time of interest follows a semiparametric transformation model with a random effect and the dropout time follows a different semiparametric transformation model but with the same random effect. The transformation models encompass the proportional hazards and proportional odds models. We study nonparametric maximum likelihood estimation for the joint models and develop a stable EM algorithm for its implementation. We establish the asymptotic properties of the resulting estimators, with different rates of convergence for the cumulative hazard functions of the event time of interest and the dropout time. In addition, we show how to predict the incidence for the event of interest when its occurrence is precluded by the development of a terminal event. Furthermore, we demonstrate the advantages of the proposed methods over the existing ones through realistic simulation studies and provide a detailed illustration with data derived from the ARIC study. Finally, we show technical details and additional figures.

3.2 Methods

3.2.1 Models and Likelihood

We consider a random sample of n subjects. For $i = 1, \ldots, n$, let T_i denote the event time or failure time of interest, D_i the dropout time, and $\boldsymbol{X}_i(\cdot)$ a p-vector of possibly time-dependent external covariates for the ith subject. We characterize the dependence between T_i and D_i through a random effect b_i, which is assumed to be normal with mean zero and variance σ^2. Let $\boldsymbol{X}_i$ denote the entire history of the covariates. Conditional on b_i and $\boldsymbol{X}_i$, the cumulative hazard functions for T_i and D_i follow the transformation models

$$\Lambda(t|b_i, \boldsymbol{X}_i) = G\left\{\int_0^t e^{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_i(s)+b_i} d\Lambda(s)\right\} \tag{3.1}$$

and

$$A(t|b_i, \boldsymbol{X}_i) = H\left\{\int_0^t e^{\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{X}_i(s)+b_i} dA(s)\right\}, \tag{3.2}$$

respectively, where $G(\cdot)$ and $H(\cdot)$ are specific transformation functions, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are unknown regression parameters, and $\Lambda(\cdot)$ and $A(\cdot)$ are arbitrary cumulative baseline hazard functions. For notational simplicity, we use the same $\boldsymbol{X}_i$ in models (3.1) and (3.2), although it is straightforward to use different sets of covariates.

Remark 3.1. *We allow different transformation functions for the event of interest and dropout and let the data determine the best choices. Since there are only two events per subject, one shared random effect b_i is sufficient to capture the dependence and additional parameters would not be identifiable.*

The transformation functions $G(\cdot)$ and $H(\cdot)$ include completely monotonic functions

$$G(x) = -\log \int_0^\infty e^{-xt} f_G(t) dt \tag{3.3}$$

and

$$H(x) = -\log \int_0^\infty e^{-xt} f_H(t) dt, \tag{3.4}$$

where $f_G(\cdot)$ and $f_H(\cdot)$ are density functions with support on $[0, \infty)$. Particularly, the class of logarithmic transformations $r^{-1} \log(1 + rx)$ $(r \geq 0)$ is generated by the gamma density function with mean 1 and variance r. The choice of $r = 0$ or 1 yields the proportional hazards or proportional odds model, respectively.

Suppose that the event of interest, such as diabetes, is asymptomatic, such that its occurrence can only be detected through periodic examinations. By contrast, dropout (e.g., death), can be observed exactly. There is a sequence of potential examination times for each subject. Obviously, no examination can occur after dropout. There exists non-informative censoring (e.g., end of the study), after which examination cannot occur either.

Specifically, let $0 < U_{i1} < \cdots < U_{i,M_i} < \infty$ denote the ith subject's potential examination times, which have finite support $\mathcal{U}$ with least upper bound τ. Let C_i denote the noninformative censoring time on D_i, such that we observe $Y_i \equiv \min(D_i, C_i)$ and $\Delta_i \equiv I(D_i \leq C_i)$, where $I(\cdot)$ is the indicator function. The examination for the ith subject does not occur after Y_i. Because examination typically does not occur at the time of dropout or the end of the study, Y_i is not assumed equal to U_{ij} for some j. Thus, the failure time T_i is known to lie in the interval (L_i, R_i), where $L_i = \max\{U_{im} : U_{im} < T_i, U_{im} \leq Y_i, m = 0, \ldots, M_i\}$, $R_i = \min\{U_{im} : U_{im} \geq T_i, U_{im} \leq Y_i, m = 1, \ldots, M_i\}$, and $U_{i0} = 0$. We let $R_i = \infty$ if the latter set is empty. If $Y_i < U_{i1}$, then no examination is performed and

$(L_i, R_i) = (0, \infty)$. The observed data consist of $\mathcal{O}_i$ $(i = 1, \ldots, n)$, where $\mathcal{O}_i = \{L_i, R_i, Y_i, \Delta_i, \boldsymbol{X}_i(\cdot)\}$.

Assume that M_i, $\{U_{im} : m = 1, \ldots, M_i\}$, and C_i are independent of (T_i, D_i, b_i) conditional on $\boldsymbol{X}_i$. The observed-data likelihood under models (3.1) and (3.2) is

$$\prod_{i=1}^{n} \int_{b_i} \left(\exp\left[-G\left\{ \int_0^{L_i} e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_i(s) + b_i} d\Lambda(s) \right\} \right] - \exp\left[-G\left\{ \int_0^{R_i} e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_i(s) + b_i} d\Lambda(s) \right\} \right] \right)$$
$$\times \left[e^{\boldsymbol{\gamma}^{\mathrm{T}} \boldsymbol{X}_i(Y_i) + b_i} A'(Y_i) H' \left\{ \int_0^{Y_i} e^{\boldsymbol{\gamma}^{\mathrm{T}} \boldsymbol{X}_i(s) + b_i} dA(s) \right\} \right]^{\Delta_i}$$
$$\times \exp\left[-H\left\{ \int_0^{Y_i} e^{\boldsymbol{\gamma}^{\mathrm{T}} \boldsymbol{X}_i(s) + b_i} dA(s) \right\} \right] \phi(b_i; \sigma^2) db_i,$$

where $g'(\cdot)$ denotes the derivative of the function $g(\cdot)$, $\phi(b; \sigma^2) = (2\pi\sigma^2)^{-1/2} \exp(-b^2/2\sigma^2)$, and we define $\exp[-G\{\int_0^{\infty} e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_i(s) + b_i} d\Lambda(s)\}] = 0$.

3.2.2 Estimation Procedure

We adopt the nonparametric maximum likelihood approach, under which the estimators for the cumulative baseline hazard functions Λ and A are step functions with jumps at the unique endpoints of the intervals, $0 < t_1 < \cdots < t_{m_1} < \infty$, and at the uncensored dropout times, $0 < s_1 < \cdots < s_{m_2} < \infty$, where m_1 and m_2 are the total numbers of potential jump points. We denote the step sizes for Λ as $\lambda_1, \ldots, \lambda_{m_1}$ and the step sizes for A as $\alpha_1, \ldots, \alpha_{m_2}$. Write $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma^2)$ and $\mathcal{A} = (\Lambda, A)$. We maximize the objective function

$$L_n(\boldsymbol{\theta}, \mathcal{A}) \equiv \prod_{i=1}^{n} \int_{b_i} L_i^{(1)}(b_i; \boldsymbol{\beta}, \Lambda) L_i^{(2)}(b_i; \boldsymbol{\gamma}, A) \phi(b_i; \sigma^2) db_i$$

over $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, σ^2, $(\lambda_1, \ldots, \lambda_{m_1})$, and $(\alpha_1, \ldots, \alpha_{m_2})$, where

$$L_i^{(1)}(b_i; \boldsymbol{\beta}, \Lambda) = \exp\left\{ -G\left(\sum_{t_l \le L_i} e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_{il} + b_i} \lambda_l \right) \right\} - \exp\left\{ -G\left(\sum_{t_l \le R_i} e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_{il} + b_i} \lambda_l \right) \right\},$$

and

$$L_i^{(2)}(b_i; \boldsymbol{\gamma}, A) = \left[A\{Y_i\} e^{\boldsymbol{\gamma}^{\mathrm{T}} \boldsymbol{X}_i(Y_i) + b_i} H'\left(\sum_{s_l \le Y_i} e^{\boldsymbol{\gamma}^{\mathrm{T}} \boldsymbol{X}_{il}^* + b_i} \alpha_l \right) \right]^{\Delta_i}$$
$$\times \exp\left\{ -H\left(\sum_{s_l \le Y_i} e^{\boldsymbol{\gamma}^{\mathrm{T}} \boldsymbol{X}_{il}^* + b_i} \alpha_l \right) \right\},$$

with $\boldsymbol{X}_{il} = \boldsymbol{X}_i(t_l)$ for $l = 1, \ldots, m_1$, $\boldsymbol{X}^*_{il} = \boldsymbol{X}_i(s_l)$ for $l = 1, \ldots, m_2$, and $A\{Y_i\}$ being the jump size of A at Y_i.

Direct maximization of $L_n(\boldsymbol{\theta}, \mathcal{A})$ is difficult due to the lack of analytical expressions for the parameters $\lambda_1, \ldots, \lambda_{m_1}$ and $\alpha_1, \ldots, \alpha_{m_2}$. We introduce some latent random variables to form a likelihood function equivalent to $L_n(\boldsymbol{\theta}, \mathcal{A})$ such that the maximization can be carried out by a simple EM algorithm. First, we introduce two latent frailties ξ_i and ψ_i with density functions $f_G(\cdot)$ and $f_H(\cdot)$ given in equations (3.3) and (3.4), respectively. We then introduce independent Poisson random variables W_{il} $(l = 1, \ldots, m_1, t_l \leq R^*_i)$ with means $\lambda_l \xi_i \exp(\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_{il} + b_i)$, where $R^*_i = L_i I(R_i = \infty) + R_i I(R_i < \infty)$. Conditional on (ξ_i, b_i), the likelihood function of $\{W_{il}; l = 1, \ldots, m_1, t_l \leq R^*_i\}$ is

$$\prod_{l=1, t_l \leq R^*_i}^{m_1} \left\{ \frac{1}{W_{il}!} \left(\lambda_l \xi_i e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_{il} + b_i} \right)^{W_{il}} \exp\left(-\lambda_l \xi_i e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_{il} + b_i} \right) \right\}.$$

Let $N_{1i} = \sum_{t_l \leq L_i} W_{il}$ and $N_{2i} = I(R_i < \infty) \sum_{L_i < t_l \leq R_i} W_{il}$. Suppose that we observe $N_{1i} = 0$ and $N_{2i} > 0$. The observed-data likelihood for $N_{1i} = 0$ and $N_{2i} > 0$ given ξ_i and b_i is equal to

$$g_{i1}(\xi_i, b_i) \equiv \exp\left(-\xi_i \sum_{t_l \leq L_i} e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_{il} + b_i} \lambda_l \right) - I(R_i < \infty) \exp\left(-\xi_i \sum_{t_l \leq R_i} e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_{il} + b_i} \lambda_l \right).$$

In addition, the observed-data likelihood for (Y_i, Δ_i) given ψ_i and b_i is

$$g_{i2}(\psi_i, b_i) \equiv \left\{ \psi_i A'(Y_i) e^{\boldsymbol{\gamma}^{\mathrm{T}} \boldsymbol{X}_i(Y_i) + b_i} \right\}^{\Delta_i} \exp\left\{ -\psi_i \int_0^{Y_i} e^{\boldsymbol{\gamma}^{\mathrm{T}} \boldsymbol{X}_i(s) + b_i} dA(s) \right\}.$$

Therefore, $L^{(1)}_i(b_i; \boldsymbol{\beta}, \Lambda) = \int_{\xi_i} g_{i1}(\xi_i, b_i) f_G(\xi_i) d\xi_i$, and $L^{(2)}_i(b_i; \boldsymbol{\gamma}, A) = \int_{\psi_i} g_{i2}(\psi_i, b_i) f_H(\psi_i) d\psi_i$. In other words, $L_n(\boldsymbol{\theta}, \mathcal{A})$ can be viewed as the observed-data likelihood for $\{N_{1i} = 0, N_{2i} > 0, Y_i, \Delta_i\}$ with $(W_{il}, \xi_i, \psi_i, b_i)$ $(l = 1, \ldots, m_1, t_l \leq R^*_i)$ as latent variables. Based on the foregoing results, we propose an EM algorithm treating $(W_{il}, \xi_i, \psi_i, b_i)$ $(i = 1, \ldots, n; l = 1, \ldots, m_1, t_l \leq R^*_i)$ as complete data.

In the M-step, we maximize the conditional expectation of the complete-data log-likelihood given the observed data so as to update the parameters. Specifically, we update $\boldsymbol{\beta}$ by solving the equation

$$\sum_{i=1}^{n} \sum_{l=1}^{m_1} \widehat{E}\left(W_{il}\right) I(t_l \leq R^*_i) \left[\boldsymbol{X}_{il} - \frac{\sum_{j=1}^{n} \boldsymbol{X}_{jl} I(t_l \leq R^*_j) \widehat{E}\left\{ \xi_j \exp\left(\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_{jl} + b_j \right) \right\}}{\sum_{j=1}^{n} I(t_l \leq R^*_j) \widehat{E}\left\{ \xi_j \exp\left(\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_{jl} + b_j \right) \right\}} \right] = \boldsymbol{0},$$

and we update Λ by

$$\lambda_l = \frac{\sum_{i=1}^n I(t_l \leq R_i^*)\widehat{E}\left(W_{il}\right)}{\sum_{i=1}^n I(t_l \leq R_i^*)\widehat{E}\left\{\xi_i \exp\left(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_{il} + b_i\right)\right\}},\ l = 1, \ldots, m_1,$$

where $\widehat{E}(\cdot)$ denotes the conditional expectation given the observed data $\widetilde{\mathcal{O}}_i \equiv \{N_{1i} = 0, N_{2i} > 0, Y_i, \Delta_i, \boldsymbol{X}_i(\cdot)\}$ $(i = 1, \ldots, n)$. In addition, we update $\boldsymbol{\gamma}$ by solving the equation

$$\sum_{i=1}^n \Delta_i \left(\boldsymbol{X}_i(Y_i) - \frac{\sum_{j=1}^n I(Y_j \geq Y_i)\boldsymbol{X}_j(Y_i)\widehat{E}\left[\psi_j \exp\left\{\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{X}_j(Y_i) + b_j\right\}\right]}{\sum_{j=1}^n I(Y_j \geq Y_i)\widehat{E}\left[\psi_j \exp\left\{\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{X}_j(Y_i) + b_j\right\}\right]}\right) = \mathbf{0},$$

and we update A by

$$\alpha_l = \frac{\sum_{i=1}^n \Delta_i I(Y_i = s_l)}{\sum_{i=1}^n I(Y_i \geq s_l)\widehat{E}\left\{\psi_i \exp\left(\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{X}_{il}^* + b_i\right)\right\}},\ l = 1, \ldots, m_2.$$

Finally, we update σ^2 by $\sigma^2 = n^{-1}\sum_{i=1}^n \widehat{E}(b_i^2)$.

In the E-step, we evaluate the conditional expectation of W_{il} $(l = 1, \ldots, m_1, t_l \leq R_i^*)$ and the other terms of ξ_i, ψ_i, and b_i given the observed data $\widetilde{\mathcal{O}}_i$ for $i = 1, \ldots, n$. Specifically, the conditional expectation of W_{il} $(l = 1, \ldots, m, t_l \leq R_i^*)$ given $\widetilde{\mathcal{O}}_i$, ξ_i, and b_i is

$$I(L_i < t_l \leq R_i < \infty)\frac{\lambda_l \xi_i \exp\left(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_{il} + b_i\right)}{1 - \exp\left(-\sum_{L_i < t_{l'} \leq R_i} \lambda_{l'}\xi_i e^{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_{il'} + b_i}\right)}.$$

Note that the joint density of (ξ_i, ψ_i, b_i) given $\widetilde{\mathcal{O}}_i$ is proportional to $g_{i1}(\xi_i, b_i)f_G(\xi_i)g_{i2}(\psi_i, b_i)f_H(\psi_i)$ $\times\phi(b_i; \sigma^2)$. We evaluate the conditional expectation of W_{il} and the other terms through numerical integration over ξ_i, ψ_i, and b_i with Gaussian quadratures.

We iterate between the E-step and the M-step until convergence. We denote the final estimators for $\boldsymbol{\theta}$ and $\mathcal{A}$ as $\widehat{\boldsymbol{\theta}} \equiv (\widehat{\boldsymbol{\beta}}, \widehat{\gamma}, \widehat{\sigma}^2)$ and $\widehat{\mathcal{A}} \equiv (\widehat{\Lambda}, \widehat{A})$. The survival function for the failure time of interest, $P(T \geq t|\boldsymbol{X})$, can be estimated by

$$\int_b \exp\left[-G\left\{\sum_{t_l \leq t} e^{\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{X}(t_l) + b}\widehat{\lambda}_l\right\}\right]\phi\left(b; \widehat{\sigma}^2\right) db.$$

Remark 3.2. *The proposed EM algorithm has several desirable features. First, large-scale op-*

timization is avoided as jump sizes are updated explicitly in the M-step. Second, the regression parameters are updated by solving estimating equations similar to the partial likelihood score equations via one-step Newton-Raphson. Finally, the E-step involves only 2-dimensional numerical integration.

If dropout is a terminal event, which cannot be avoided, then we have a semi-competing risks set-up (Fine et al., 2001) in that the occurrence of the terminal event precludes the development of the event of interest but not vice versa. It is more meaningful to consider the cumulative incidence function for the failure time of interest

$$\begin{aligned}
& P(T \le t, T \le D|\boldsymbol{X}) \\
= \; & P(T \le t \le D|\boldsymbol{X}) + P(T \le D < t|\boldsymbol{X}) \\
= \; & \int_b \left(1 - \exp\left[-G\left\{\int_0^t e^{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}(u)+b} d\Lambda(u)\right\}\right]\right) \exp\left[-H\left\{\int_0^t e^{\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{X}(u)+b} dA(u)\right\}\right] \phi(b;\sigma^2)db \\
& + \int_b \left[\int_0^t \left\{e^{\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{X}(s)+b} H'\left\{\int_0^s e^{\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{X}(u)+b} dA(u)\right\} \exp\left[-H\left\{\int_0^s e^{\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{X}(u)+b} dA(u)\right\}\right]\right.\right. \\
& \left.\left. \times \left(1 - \exp\left[-G\left\{\int_0^s e^{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}(u)+b} d\Lambda(u)\right\}\right]\right)\right\} dA(s)\right] \phi(b;\sigma^2)db.
\end{aligned}$$

We estimate this quantity by

$$\begin{aligned}
& \int_b \left(1 - \exp\left[-G\left\{\sum_{t_l \le t} e^{\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{X}(t_l)+b} \widehat{\lambda}_l\right\}\right]\right) \exp\left[-H\left\{\sum_{s_l \le t} e^{\widehat{\boldsymbol{\gamma}}^{\mathrm{T}}\boldsymbol{X}(s_l)+b} \widehat{\alpha}_l\right\}\right] \phi(b;\sigma^2)db \\
& + \int_b \left[\sum_{s_l \le t} \left\{\widehat{\alpha}_l e^{\widehat{\boldsymbol{\gamma}}^{\mathrm{T}}\boldsymbol{X}(s_l)+b} H'\left\{\sum_{l' \le l} e^{\widehat{\boldsymbol{\gamma}}^{\mathrm{T}}\boldsymbol{X}(s_{l'})+b} \widehat{\alpha}_{l'}\right\} \exp\left[-H\left\{\sum_{l' \le l} e^{\widehat{\boldsymbol{\gamma}}^{\mathrm{T}}\boldsymbol{X}(s_{l'})+b} \widehat{\alpha}_{l'}\right\}\right]\right.\right. \\
& \left.\left. \times \left(1 - \exp\left[-G\left\{\sum_{t_{l'} \le s_l} e^{\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{X}(t_{l'})+b} \widehat{\lambda}_{l'}\right\}\right]\right)\right\}\right] \phi(b;\widehat{\sigma}^2)db,
\end{aligned}$$

where the integral is evaluated by numerical integration with Gaussian quadratures, and $\widehat{\lambda}_l$ and $\widehat{\alpha}_l$ are the estimators of λ_l and α_l, respectively.

We have implicitly assumed that the transformation functions are known. In practice, we consider a variety of transformation models and choose the one that best fits the data according to, say, the Akaike information criterion.

3.2.3 Asymptotic Theory

We establish the asymptotic properties of $(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}})$ under the following regularity conditions.

Condition 1. The true value of $\boldsymbol{\theta}$, denoted by $\boldsymbol{\theta}_0 \equiv (\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, \sigma_0^2)$, belongs to the interior of a known compact set $\Theta \equiv \mathcal{B}_1 \times \mathcal{B}_2 \times \mathcal{S}$, where $\mathcal{B}_1, \mathcal{B}_2 \subset \mathbb{R}^p$, and $\mathcal{S} \subset (0, \infty)$.

Condition 2. The true value $\Lambda_0(\cdot)$ of $\Lambda(\cdot)$ is strictly increasing and continuously differentiable on $\mathcal{U}$ with $\Lambda_0(0) = 0$. The true value $A_0(\cdot)$ of $A(\cdot)$ is strictly increasing and continuously differentiable on $[0, \tau]$ with $A_0(0) = 0$.

Condition 3. There exists some positive constant δ^* such that $\Pr(C \geq \tau | \boldsymbol{X}) = \Pr(C = \tau | \boldsymbol{X}) \geq \delta^*$ almost surely.

Condition 4. The number of potential examination times M is positive with $E(M) < \infty$. There exists a positive constant η such that $\Pr\{\min_{0 \leq m < M}(U_{m+1} - U_m) \geq \eta | M, \boldsymbol{X}\} = 1$. In addition, there exists a probability measure μ in $\mathcal{U}$ such that the bivariate distribution function of (U_m, U_{m+1}) conditional on $(M, \boldsymbol{X})$ is dominated by $\mu \times \mu$ and its Randon-Nikodym derivative, denoted by $\widetilde{f}_m(u, v; M, \boldsymbol{X})$, can be expanded to a positive and twice-continuously differentiable function in the set $\{(u, v) : 0 \leq u \leq \tau, 0 \leq v \leq \tau, v - u \geq \eta\}$.

Condition 5. With probability 1, $\boldsymbol{X}(\cdot)$ has bounded total variation in $[0, \tau]$. If there exists a deterministic function $a_1(t)$ and a constant vector $\boldsymbol{a}_2$ such that $a_1(t) + \boldsymbol{a}_2^{\mathrm{T}} \boldsymbol{X}(t) = 0$ for any $t \in \mathcal{U}$ with probability 1, then $a_1(t) = 0$ for any $t \in \mathcal{U}$ and $\boldsymbol{a}_2 = \mathbf{0}$.

Condition 6. The function $G(\cdot)$ is twice differentiable with $G(0) = 0$ and $G'(x) > 0$. The function $H(\cdot)$ is three-times differentiable with $H(0) = 0$ and $H'(x) > 0$. The lth derivative of $\exp\{-G(\cdot)\}$ is bounded for $l = 1, 2$, and the lth derivative of $\exp\{-H(\cdot)\}$ is bounded for $l = 1, 2, 3$. There exists a positive constant ρ such that

$$\limsup_{x \to \infty} (1 + x)^{\rho} \exp\{-G(x)\} < \infty,$$

$$\limsup_{x \to \infty} (1 + x)^{\rho} \exp\{-H(x)\} < \infty,$$

and

$$\limsup_{x \to \infty} (1 + x)^{1+\rho} H'(x) \exp\{-H(x)\} < \infty.$$

Condition 7. For any pair of parameters $(\boldsymbol{\beta}_{(1)}, \boldsymbol{\gamma}_{(1)}, \sigma_{(1)}^2, \Lambda_{(1)}, A_{(1)})$ and $(\boldsymbol{\beta}_{(2)}, \boldsymbol{\gamma}_{(2)}, \sigma_{(2)}^2, \Lambda_{(2)}, A_{(2)})$, where $\Lambda_{(1)}$, $\Lambda_{(2)}$, $A_{(1)}$, and $A_{(2)}$ are strictly increasing, if with probability 1,

$$\int_b \left(\exp\left[-G\left\{ \int_0^{u_1} e^{\boldsymbol{\beta}_{(1)}^{\mathrm{T}} \boldsymbol{X}(s)+b} d\Lambda_{(1)}(s) \right\} - H\left\{ \int_0^{u_2} e^{\boldsymbol{\gamma}_{(1)}^{\mathrm{T}} \boldsymbol{X}(s)+b} dA_{(1)}(s) \right\} \right] \right) \phi(b;\sigma_{(1)}^2) db$$

$$= \int_b \left(\exp\left[-G\left\{ \int_0^{u_1} e^{\boldsymbol{\beta}_{(2)}^{\mathrm{T}} \boldsymbol{X}(s)+b} d\Lambda_{(2)}(s) \right\} - H\left\{ \int_0^{u_2} e^{\boldsymbol{\gamma}_{(2)}^{\mathrm{T}} \boldsymbol{X}(s)+b} dA_{(2)}(s) \right\} \right] \right) \phi(b;\sigma_{(2)}^2) db$$

for any $u_1 \in \mathcal{U}$ and $u_2 \in [0,\tau]$, then $\boldsymbol{\beta}_{(1)} = \boldsymbol{\beta}_{(2)}$, $\boldsymbol{\gamma}_{(1)} = \boldsymbol{\gamma}_{(2)}$, $\sigma_{(1)}^2 = \sigma_{(2)}^2$, $\Lambda_{(1)}(u_1) = \Lambda_{(2)}(u_1)$ for any $u_1 \in \mathcal{U}$, and $A_{(1)}(u_2) = A_{(2)}(u_2)$ for any $u_2 \in [0,\tau]$.

Condition 8. If there exist functions $c_1(t)$ and $c_2(t)$ and a constant c_3 such that

$$\begin{aligned}
&\int_b \Bigg[G'\left\{ \int_0^{u_1} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}(s)+b} d\Lambda_0(s) \right\} \int_0^{u_1} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}(s)+b} c_1(s) d\Lambda_0(s) \\
&+ H'\left\{ \int_0^{u_2} e^{\boldsymbol{\gamma}_0^{\mathrm{T}} \boldsymbol{X}(s)+b} dA_0(s) \right\} \int_0^{u_2} e^{\boldsymbol{\gamma}_0^{\mathrm{T}} \boldsymbol{X}(s)+b} c_2(s) dA_0(s) - c_3 \frac{\phi'(b,\sigma_0^2)}{\phi(b;\sigma_0^2)} \Bigg] \\
&\times \exp\left[-G\left\{ \int_0^{u_1} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}(s)+b} d\Lambda_0(s) \right\} - H\left\{ \int_0^{u_2} e^{\boldsymbol{\gamma}_0^{\mathrm{T}} \boldsymbol{X}(s)+b} dA_0(s) \right\} \right] \phi(b,\sigma_0^2) db = 0
\end{aligned}$$

with probability 1 for any $u_1 \in \mathcal{U}$ and $u_2 \in [0,\tau]$, where ϕ' is the derivative of ϕ with respect to σ^2, then $c_1(u_1) = 0$, $c_2(u_2) = 0$, and $c_3 = 0$ for any $u_1 \in \mathcal{U}$ and $u_2 \in [0,\tau]$.

Remark 3.3. *Conditions 1, 2, and 5 are standard conditions for failure time regression with time-dependent covariates. Condition 3 implies that there is a positive probability for dropout to be observed in the time interval* $[0,\tau]$*. Condition 4 pertains to the joint distribution of examination times; it requires that two adjacent examination times be separated by at least* η*. The dominating measure* μ *is chosen as the Lebesgue measure if the examination times are continuous random variables and as the counting measure if examinations occur only at a finite number of time points. The number of potential examination times* M *can be fixed or random, is possibly different among study subjects, and is allowed to depend on covariates. Condition 6 pertains to the transformation functions and holds for the logarithmic transformations. Conditions 7 and 8 are identifiability conditions for the consistency of the estimators and nonsingularity of the information matrix, respectively.*

We state the strong consistency of $(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}})$ and weak convergence of $\widehat{\boldsymbol{\theta}}$ in two theorems.

Theorem 3.1. *Under Conditions 1–7,* $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| \rightarrow_{a.s.} 0$*,* $\|\widehat{\Lambda} - \Lambda_0\|_{l^\infty(\mathcal{U})} \rightarrow_{a.s.} 0$*, and* $\|\widehat{A} - A_0\|_{l^\infty[0,\tau]} \rightarrow_{a.s.} 0$*, where* $\|\cdot\|_{l^\infty(\mathcal{B})}$ *denotes the supremum norm on* $\mathcal{B}$*.*

Theorem 3.2. *Under Conditions 1–8, $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ converges weakly to a $(2p+1)$-variate zero-mean normal random vector with a covariance matrix that attains the semiparametric efficiency bound.*

The proofs of Theorems 3.1 and 3.2 are given in Section 3.6. To estimate the covariance matrix of $\widehat{\boldsymbol{\theta}}$, we use the profile likelihood (Murphy and Van der Vaart, 2000). Specifically, we define the profile log-likelihood function

$$pl_n(\boldsymbol{\theta}) = \max_{\mathcal{A} \in \mathcal{C}_1 \times \mathcal{C}_2} l_n(\boldsymbol{\theta}, \mathcal{A}),$$

where $l_n(\boldsymbol{\theta}, \mathcal{A}) = \log L_n(\boldsymbol{\theta}, \mathcal{A})$, $\mathcal{C}_1$ is the set of step functions with non-negative jumps at t_l ($l = 1, \ldots, m_1$), and $\mathcal{C}_2$ is the set of step functions with non-negative jumps at s_l ($l = 1, \ldots, m_2$). We estimate the covariance matrix of $\widehat{\boldsymbol{\theta}}$ by the inverse of

$$\sum_{i=1}^{n} \begin{pmatrix} \dfrac{pl_i(\widehat{\boldsymbol{\theta}} + h_n e_1) - pl_i(\widehat{\boldsymbol{\theta}})}{h_n} \\ \vdots \\ \dfrac{pl_i(\widehat{\boldsymbol{\theta}} + h_n e_{2p+1}) - pl_i(\widehat{\boldsymbol{\theta}})}{h_n} \end{pmatrix}^{\otimes 2},$$

where pl_i is the ith subject's contribution to pl_n, e_j is the jth canonical vector in $\mathbb{R}^{2p+1}$, $\boldsymbol{a}^{\otimes 2} = \boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}$, and h_n is a constant of order $n^{-1/2}$. To evaluate the profile likelihood, we use the EM algorithm of Section 2.3 but only update Λ and A in the M-step.

3.3 Simulation Studies

We conducted simulation studies to assess the performance of the proposed methods. We considered one time-independent covariate $X_1 \sim \mathrm{Unif}(0, 1)$ and one time-dependent covariate $X_2(t) = \widetilde{B}_1 I(t \leq V) + \widetilde{B}_2 I(t > V)$, where $\widetilde{B}_1$ and $\widetilde{B}_2$ are independent Bernoulli(0.5), $V \sim \mathrm{Unif}(0, \tau)$, and $\tau = 4$. We considered logarithmic transformation functions $G(x) = r_G^{-1} \log(1 + r_G x)$ and $H(x) = r_H^{-1} \log(1 + r_H x)$. We set $\boldsymbol{\beta} \equiv (\beta_1, \beta_2)^{\mathrm{T}} = (0.5, 0.4)^{\mathrm{T}}$, $\boldsymbol{\gamma} \equiv (\gamma_1, \gamma_2)^{\mathrm{T}} = (0.5, 0.2)^{\mathrm{T}}$, $\sigma^2 = 0.5$, $\Lambda(t) = 0.5t$, and $A(t) = \log(1 + 0.5t)$. We generated the potential examination times $U_m \sim U_{m-1} + 0.1 + \mathrm{Unif}(0, \tau/5)$ with $U_0 = 0$ and the censoring time C from $\mathrm{Unif}(2\tau/3, \tau)$. The number of actual examinations is approximately 2.4 per subject. The event time of interest is left-censored for 19% subjects, interval-censored for 28% subjects, and right-censored for 53% subjects. We set $n = 100, 200$, or 400 and used 10,000 replicates. The variance estimators were obtained with $h_n = 5/\sqrt{n}$.

Table 3.1: Summary statistics for the proposed estimators

		$r_G = r_H = 0$				$r_G = r_H = 1$			
		Bias	SE	SEE	CP	Bias	SE	SEE	CP
$n = 100$	β_1	0.027	0.728	0.680	0.95	0.030	0.921	0.880	0.95
	β_2	0.028	0.422	0.382	0.94	0.037	0.533	0.481	0.94
	γ_1	0.002	0.544	0.527	0.96	−0.012	0.723	0.727	0.97
	γ_2	0.009	0.288	0.289	0.96	0.000	0.381	0.383	0.96
	σ^2	−0.019	0.555	0.627	0.98	−0.112	0.781	1.035	0.97
$n = 200$	β_1	0.006	0.482	0.463	0.95	0.009	0.623	0.602	0.95
	β_2	0.016	0.277	0.261	0.94	0.021	0.356	0.329	0.94
	γ_1	0.002	0.368	0.364	0.96	0.010	0.498	0.502	0.96
	γ_2	0.000	0.198	0.200	0.96	0.004	0.260	0.262	0.96
	σ^2	−0.002	0.367	0.409	0.97	−0.044	0.556	0.698	0.96
$n = 400$	β_1	0.000	0.329	0.321	0.95	0.011	0.429	0.419	0.94
	β_2	0.007	0.187	0.182	0.95	0.012	0.244	0.229	0.94
	γ_1	0.005	0.255	0.254	0.95	0.002	0.354	0.351	0.95
	γ_2	0.003	0.140	0.139	0.95	0.003	0.182	0.182	0.95
	σ^2	0.002	0.250	0.272	0.97	−0.015	0.401	0.476	0.96

SE, SEE, and CP stand, respectively, for the empirical standard error, mean standard error estimator, and empirical coverage percentage of the 95% confidence interval. For σ^2, bias and SEE are based on the median instead of the mean, and the confidence interval is based on the log transformation. Each entry is based on 10,000 replicates.

Table 3.1 summarizes the results on the estimation of $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and σ^2 for different values of n, r_G, and r_H. The biases for all parameter estimators are small and decrease as n increases. The variance estimators for $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\gamma}}$ are accurate, especially for large n. The variance estimator for $\widehat{\sigma}^2$ tends to overestimate the actual variability. The 95% confidence intervals for $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and σ^2 have reasonable coverage probabilities.

We also evaluated the method of Zeng et al. (2016), which does not account for informative dropout. The results for this naive method are shown in Table 3.2. The estimator for $\boldsymbol{\beta}$ is biased, and the coverage probability of the corresponding confidence interval is poor.

Figure 3.1(a) shows the estimation of the baseline survival function for the event of interest when dropout is regarded as voluntary patient withdrawal. The proposed estimator is virtually unbiased, whereas the naive method (Zeng et al. 2016) overestimates the survival function. Figure 3.1(b) shows the estimation of the baseline cumulative incidence function for the event of interest when dropout is treated as a terminal event. The proposed estimator is again virtually unbiased; the naive estimator has severe positive bias since it does not acknowledge the fact that the event of interest

Table 3.2: Summary statistics for the naive method

		$r_G = r_H = 0$				$r_G = r_H = 1$			
		Bias	SE	SEE	CP	Bias	SE	SEE	CP
$n = 100$	β_1	-0.125	0.600	0.558	0.93	-0.054	0.829	0.780	0.93
	β_2	-0.058	0.355	0.323	0.93	-0.015	0.483	0.433	0.92
$n = 200$	β_1	-0.143	0.400	0.382	0.93	-0.074	0.559	0.541	0.94
	β_2	-0.067	0.236	0.223	0.93	-0.025	0.324	0.299	0.93
$n = 400$	β_1	-0.148	0.274	0.266	0.91	-0.069	0.387	0.379	0.94
	β_2	-0.075	0.160	0.156	0.92	-0.033	0.222	0.210	0.93

See the Note to Table 3.1.

cannot occur after the terminal event.

3.4 ARIC Study

ARIC is a prospective epidemiological study conducted in four U.S. communities: Forsyth County, NC; Jackson, MS; suburbs of Minneapolis, MN; and Washington County, MD (The ARIC investigators, 1989). One important objective is to investigate risk factors for diabetes. A total of 14,751 Caucasian and African-American participants underwent a baseline examination between 1987 and 1989 and were scheduled for four subsequent examinations to take place in 1990−1992, 1993−1995, 1996−1998, and 2011−2013. Diabetes status (defined as fasting glucose $\geq$ 126 mg/dL, non-fasting glucose $\geq$ 200 mg/dL, self-reported physician diagnosis of diabetes, or use of diabetic medication) was determined at each examination.

We related the incidence of diabetes and death to race, gender, community, and five baseline risk factors: age, body mass index, glucose level, systolic blood pressure, and diastolic blood pressure. We excluded 1,933 subjects with prevalent diabetes or unknown diabetes status at baseline and 13 subjects with missing baseline covariate values to obtain a total of 12,805 subjects. Among those subjects, 11,686 (91.3%), 10,557 (82.4%), 9,533 (74.4%), and 5,035 (39.3%) completed the second, third, fourth, and fifth visits, respectively. As shown in Figure 3.3 in Section 3.6, there are sufficient overlaps of the visit times for us to study diabetes onset from year 2 to year 12 and from year 22 to year 27. A total of 2,492 (19.5%) subjects developed diabetes during the study, and 4,363 (34.1%) subjects died before the end of the study.

We fit models (3.1) and (3.2) with logarithmic transformation functions indexed by parameters r_G and r_H for diabetes and death, respectively. The likelihood is maximized at $r_G = 2.3$ and $r_H = 0$,

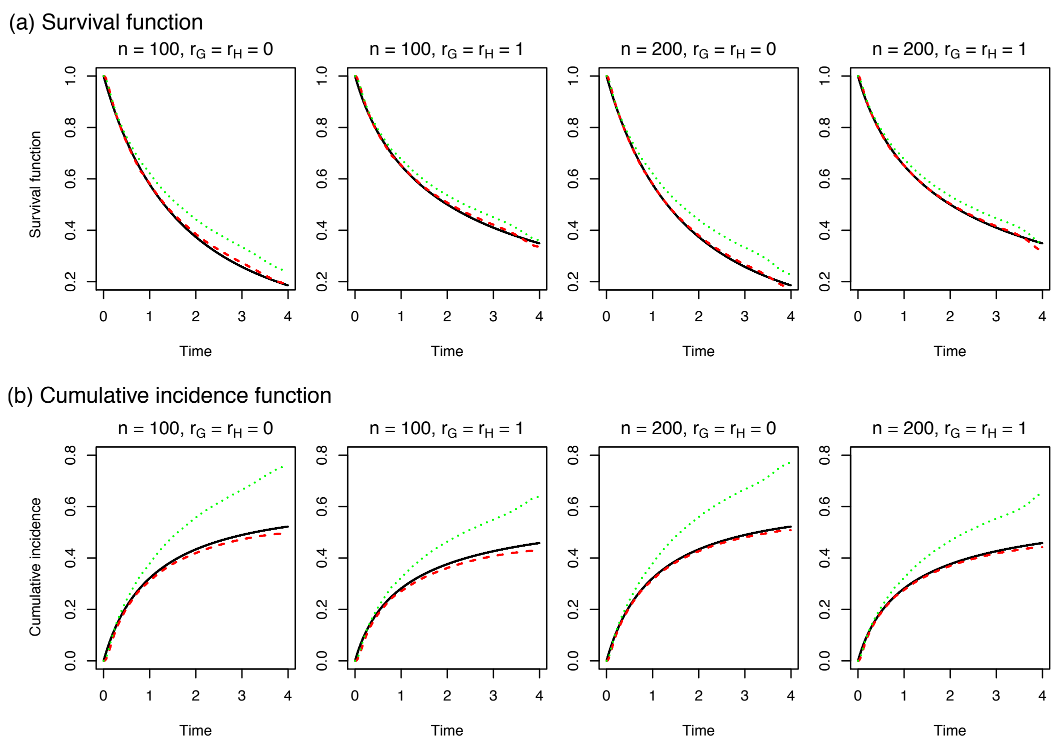

Figure 3.1: Estimation of (a) the baseline survival function and (b) the baseline cumulative incidence function. The solid black curve, dashed red curve, and dotted green curve pertain, respectively, to the true value, mean estimate from the proposed method, and mean estimate from the naive method.

which is the combination that would be selected by the Akaike information criterion. For easy interpretation, we set $r_G = 1$ and $r_H = 0$.

Table 3.3 shows the estimation results for the proportional hazards models for both events ($r_G = r_H = 0$), the proportional odds models for both events ($r_G = r_H = 1$), and the combination of the proportional odds model for diabetes and the proportional hazards model for death ($r_G = 1, r_H = 0$). The log-likelihood values are approximately -48724.5, -48707.3, and -48681.1 for the three combinations of transformation parameters. The variance component σ^2 was estimated to be 0.561, 0.681, and 0.530 with standard errors 0.063, 0.096, and 0.070, respectively, for the three combinations of transformation parameters, indicating strong dependence between diabetes and death. Under all considered models, an African-American individual has a higher risk of diabetes than a Caucasian individual. In addition, higher baseline body mass index, glucose level, and systolic blood pressure are associated with increased risk for diabetes.

The results from the naive method, which are shown in Table 3.4, are considerably different from ours. In particular, the naive method identifies a negative association between age and risk of diabetes, which contradicts the established positive association in the literature. The proposed method adjusting for death finds no significant negative association. The relationship between age and risk of diabetes identified by the naive method is likely a spurious finding that reflects the strong correlation between age and death.

Figure 3.2 compares the estimated cumulative incidence functions for an African-American male versus a Caucasian male with the same values of other risk factors. The risk of diabetes is considerably higher for the African-American individual than the Caucasian individual under all considered models, with appreciably different estimates between the proportional hazards and proportional odds models. The estimated probabilities from the proposed method are lower in the tail than their naive counterparts, especially under the proportional odds model, highlighting the importance of adjusting for death. The estimated cumulative baseline hazard functions for diabetes and death are shown in Figure 3.4 in Section 3.6.

3.5 Discussion

In this chapter, we study efficient nonparametric maximum likelihood estimation of joint models for interval-censored data with informative dropout. We establish the asymptotic properties for the estimators through innovative use of modern empirical process theory. In the proofs, separate

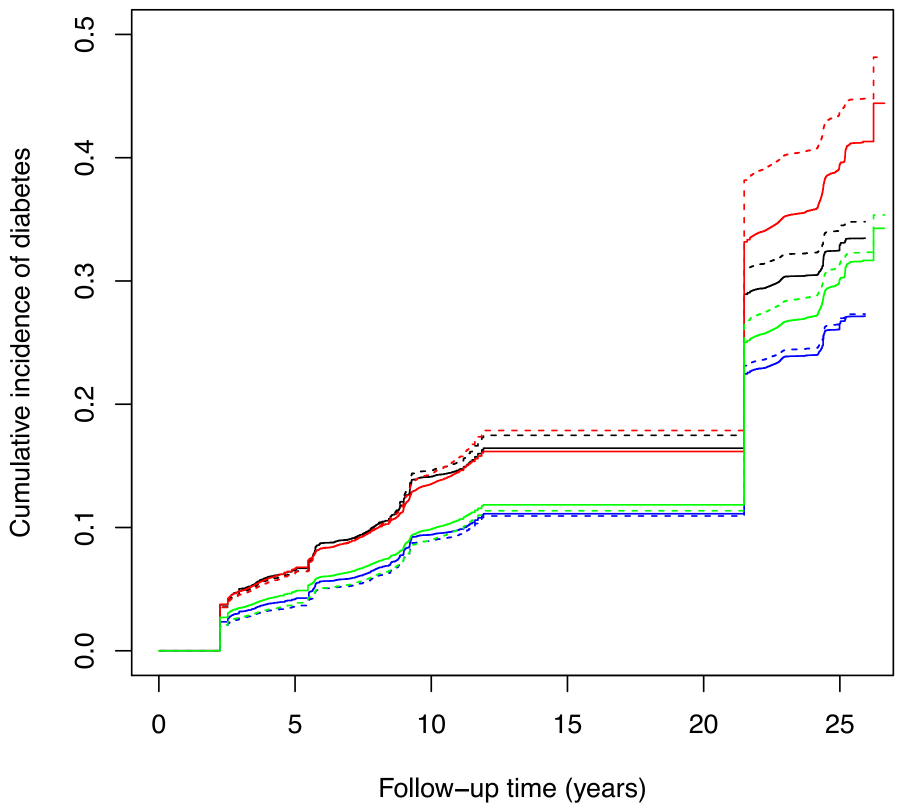

Figure 3.2: Estimation of cumulative incidence functions for an African-American male versus a Caucasian male residing in Forsyth County, NC, aged 54 years, body mass index 27 kg/m^2, glucose value 98 mg/dl, systolic blood pressure 118 mmHg, and diastolic blood pressure 73 mmHg. The red solid and dashed curves pertain to the African-American individual with the proportional hazards and proportional odds models, respectively, from the naive method. The green solid and dashed curves pertain to the Caucasian individual with the proportional hazards and proportional odds models, respectively, from the naive method. The black solid and dashed curves pertain to the African-American individual with the proportional hazards and proportional odds models, respectively, from the proposed method, where the dropout time is modeled by the proportional hazards model. The blue solid and dashed curves pertain to the Caucasian individual with the proportional hazards and proportional odds models, respectively, from the proposed method, where the dropout time is modeled by the proportional hazards model.

Table 3.3: Regression analysis for diabetes in the ARIC study with adjustments for death

r_G	r_H	Covariate	Diabetes			Death		
			Est	Std. Err.	p-value	Est	Std. Err.	p-value
0	0	Jackson	−0.185	0.126	0.142	0.039	0.106	0.714
		Minneapolis Suburbs	−0.424	0.069	$< 10^{-4}$	−0.032	0.052	0.535
		Washington County	0.101	0.066	0.124	0.088	0.050	0.081
		Age	−0.004	0.004	0.359	0.104	0.004	$< 10^{-4}$
		Male	−0.010	0.047	0.833	0.565	0.037	$< 10^{-4}$
		White	−0.485	0.130	0.0002	−0.503	0.108	$< 10^{-4}$
		Body mass index	0.075	0.004	$< 10^{-4}$	−0.003	0.004	0.363
		Glucose	0.102	0.003	$< 10^{-4}$	0.006	0.002	0.002
		Systolic blood pressure	0.006	0.002	0.001	0.017	0.001	$< 10^{-4}$
		Diastolic blood pressure	−0.001	0.003	0.793	−0.014	0.002	$< 10^{-4}$
1	1	Jackson	−0.242	0.143	0.090	0.096	0.129	0.459
		Minneapolis Suburbs	−0.510	0.083	$< 10^{-4}$	−0.034	0.063	0.588
		Washington County	0.118	0.079	0.136	0.121	0.062	0.051
		Age	−0.008	0.005	0.160	0.124	0.004	$< 10^{-4}$
		Male	−0.033	0.057	0.561	0.680	0.045	$< 10^{-4}$
		White	−0.648	0.151	$< 10^{-4}$	−0.604	0.132	$< 10^{-4}$
		Body mass index	0.096	0.005	$< 10^{-4}$	−0.005	0.004	0.266
		Glucose	0.124	0.003	$< 10^{-4}$	0.007	0.002	0.003
		Systolic blood pressure	0.007	0.002	0.001	0.020	0.002	$< 10^{-4}$
		Diastolic blood pressure	−0.001	0.004	0.891	−0.017	0.003	$< 10^{-4}$
1	0	Jackson	−0.232	0.145	0.109	0.038	0.106	0.716
		Minneapolis Suburbs	−0.502	0.081	$< 10^{-4}$	−0.033	0.052	0.528
		Washington County	0.114	0.078	0.141	0.086	0.050	0.086
		Age	−0.007	0.005	0.196	0.103	0.004	$< 10^{-4}$
		Male	−0.030	0.056	0.592	0.562	0.037	$< 10^{-4}$
		White	−0.629	0.151	$< 10^{-4}$	−0.500	0.107	$< 10^{-4}$
		Body mass index	0.094	0.005	$< 10^{-4}$	−0.003	0.004	0.411
		Glucose	0.122	0.003	$< 10^{-4}$	0.006	0.002	0.001
		Systolic blood pressure	0.007	0.002	0.001	0.017	0.001	$< 10^{-4}$
		Diastolic blood pressure	0.000	0.004	0.894	−0.014	0.002	$< 10^{-4}$

Forsyth County, NC, is the reference group for the field center variables.

Table 3.4: Regression analysis for diabetes in the ARIC study without adjustments for death

r_G	Covariate	Est	Std. Err.	p-value
0	Jackson	-0.141	0.102	0.169
	Minneapolis Suburbs	-0.373	0.062	$< 10^{-4}$
	Washington County	0.095	0.058	0.099
	Age	-0.010	0.004	0.009
	Male	-0.042	0.041	0.303
	White	-0.336	0.106	0.001
	Body mass index	0.067	0.003	$< 10^{-4}$
	Glucose	0.090	0.002	$< 10^{-4}$
	Systolic blood pressure	0.005	0.002	0.004
	Diastolic blood pressure	0.000	0.003	0.992
1	Jackson	-0.210	0.134	0.116
	Minneapolis Suburbs	-0.464	0.076	$< 10^{-4}$
	Washington County	0.112	0.072	0.121
	Age	-0.013	0.005	0.009
	Male	-0.062	0.052	0.230
	White	-0.530	0.139	$< 10^{-4}$
	Body mass index	0.089	0.005	$< 10^{-4}$
	Glucose	0.114	0.003	$< 10^{-4}$
	Systolic blood pressure	0.006	0.002	0.007
	Diastolic blood pressure	0.001	0.003	0.813

See the Note to Table 3.3.

treatments are given to the estimators of the cumulative baseline hazard functions for the event of interest and dropout. We avoid the assumption of Zeng et al. (2016) that a subset of study subjects are examined at the end of the study by carefully evaluating the bracket covering number for a class of functions that involves unbounded Λ.

We applied our methods to data derived from the ARIC study, where diabetes is the event of interest and death is the terminal event. In the ARIC study, there are other outcomes of interest that are either interval-censored (e.g. hypertension, peripheral artery disease) or right-censored (e.g. myocardial infarction, stroke). The proposed framework can be extended to incorporate multiple interval-censored events and multiple right-censored events and thereby analyze an enriched version of data from the ARIC study.

The class of transformation models is very broad and thus allows accurate prediction in a variety of situations. In practice, one would need to determine which model best fits the data. One strategy is to use the Akaike information criterion to select the best transformations, as we did for the ARIC study. It would be worthwhile to develop additional methods for model selection and model checking.

3.6 Technical Details

Let $\mathbb{P}_n$ denote the empirical measure for n independent subjects, $\mathbb{P}$ denote the true probability measure, and $\mathbb{G}_n \equiv \sqrt{n}(\mathbb{P}_n - \mathbb{P})$ denote the empirical process. The proofs of Theorems 3.1 and 3.2 make use of three lemmas, which are stated and proved in Section 3.6.3.

3.6.1 Proof of Theorem 3.1

We first show the existence of the estimator $(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}})$. Let $\widetilde{M} = \sup_{t\in\mathcal{U}} \sup_{\boldsymbol{X}(t),\boldsymbol{\beta}} |\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}(t)| + \sup_{t\in[0,\tau]} \sup_{\boldsymbol{X}(t),\boldsymbol{\gamma}} |\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{X}(t)|$. For any $(\boldsymbol{\theta}, \mathcal{A})$ in the parameter space, the integrand in the ith term of $l_n(\boldsymbol{\theta}, \mathcal{A})$ is bounded by

$$O(1)\left(A\{Y_i\}e^{\widetilde{M}+|b|}\right)^{\Delta_i}\left\{1+\int_0^{Y_i} e^{\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{X}_i(s)+b}dA(s)\right\}^{-(\Delta_i+\rho)}\phi(b,\sigma^2)$$

under Condition 6. Thus, $l_n(\boldsymbol{\theta}, \mathcal{A})$ attains the maximum for finite A values, so the estimator $(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}})$ exists by allowing $\widehat{\Lambda}(\tau) = \infty$.

We prove that $\limsup_n \widehat{\Lambda}(\tau-\epsilon) < \infty$ with probability 1 for any $\epsilon > 0$ and that $\limsup_n \widehat{A}(\tau) < \infty$ with probability 1. By definition, $l_n(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}}) - l_n(\boldsymbol{\theta}, \mathcal{A}) \geq 0$ for any $(\boldsymbol{\theta}, \mathcal{A})$ in the parameter space. We wish to show that if $\widehat{A}(\tau)$ diverges, then this difference must be negative, which is a contradiction.

The key is to construct a suitable function in the parameter space that converges uniformly to $\mathcal{A}_0$.

For Λ, we define a step function $\widetilde{\Lambda}$ satisfying $\widetilde{\Lambda}(t) = \Lambda_0(t)$ for $t = t_1, \ldots, t_{m_1}$ such that it converges uniformly to Λ_0. For A, we construct a function $\widetilde{A}$ by imitating $\widehat{A}$. By differentiating $l_n(\boldsymbol{\theta}, \mathcal{A})$ with respect to $A\{Y_i\}$ and setting the derivative to 0, we find that $\widehat{A}$ satisfies the equation

$$\frac{\Delta_i}{\widehat{A}\{Y_i\}} = \sum_{j=1}^{n} \frac{\int_b K_1\left(b, \mathcal{O}_j; \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}}, \widehat{\mathcal{A}}\right) K_2\left(Y_i, b, \mathcal{O}_j; \widehat{\boldsymbol{\gamma}}, \widehat{A}\right) \phi\left(b; \widehat{\sigma}^2\right) db}{\int_b K_1\left(b, \mathcal{O}_j; \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}}, \widehat{\mathcal{A}}\right) \phi\left(b; \widehat{\sigma}^2\right) db}, \tag{3.5}$$

where

$$K_1(b, \mathcal{O}; \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathcal{A}) = \left(\exp\left[-G\left\{\int_0^L e^{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}(s)+b} d\Lambda(s)\right\}\right] - \exp\left[-G\left\{\int_0^R e^{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}(s)+b} d\Lambda(s)\right\}\right]\right)$$

$$\times \left[e^{\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{X}(Y)+b} H'\left\{\int_0^Y e^{\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{X}(s)+b} dA(s)\right\}\right]^{\Delta} \exp\left[-H\left\{\int_0^Y e^{\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{X}(s)+b} dA(s)\right\}\right],$$

$$\begin{aligned} K_2(t, b, \mathcal{O}; \boldsymbol{\gamma}, A) \quad &= \quad I(Y \geq t) e^{\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{X}(t)+b} \left[\Delta \frac{H''\left\{\int_0^Y e^{\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{X}(s)+b} dA(s)\right\}}{H'\left\{\int_0^Y e^{\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{X}(s)+b} dA(s)\right\}}\right. \\ &\qquad \left. - H'\left\{\int_0^Y e^{\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{X}(s)+b} dA(s)\right\}\right], \end{aligned}$$

and $H''(\cdot)$ is the second derivative of $H(\cdot)$. We replace $\widehat{\boldsymbol{\theta}}$ and $\widehat{\mathcal{A}}$ on the right side of equation (4.8) by $\boldsymbol{\theta}_0$ and $\mathcal{A}_0$, respectively, to obtain a similar function. We denote the solution as $\widetilde{A}$. By the Glivenko-Cantelli result in Lemma 3.1, $\widetilde{A}$ converges uniformly to A_0 in $[0, \tau]$.

Clearly, $n^{-1}\left\{l_n(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}}) - l_n(\boldsymbol{\theta}_0, \widetilde{\mathcal{A}})\right\} \geq 0$. Let $\delta_{im} = I(U_{im} < T_i \leq U_{i,m+1})$ for $i = 1, \ldots, n$ and $m = 0, \ldots, M_i$, where $U_{i,M_i+1} = \infty$. By Condition 6 and the fact that $e^{-|x|}(1+y) \leq 1 + e^x y \leq e^{|x|}(1+y)$, we obtain

$$\begin{aligned} 0 \quad &\leq \quad n^{-1} l_n(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}}) - n^{-1} l_n(\boldsymbol{\theta}_0, \widetilde{\mathcal{A}}) \\ &\leq \quad O(1) + n^{-1} \sum_{i=1}^{n} \log\left(n\widehat{A}\{Y_i\}\right) \\ &\quad + n^{-1} \sum_{i=1}^{n} \left[\log \int_b \left\{e^{\widehat{\boldsymbol{\gamma}}^{\mathrm{T}}\boldsymbol{X}_i(Y_i)+b}\right\}^{\Delta_i} \left\{1 + \int_0^{Y_i} e^{\widehat{\boldsymbol{\gamma}}^{\mathrm{T}}\boldsymbol{X}_i(t)+b} d\widehat{A}(t)\right\}^{-\Delta_i - \rho} \phi\left(b; \widehat{\sigma}^2\right) db\right] \end{aligned}$$

$$
\begin{aligned}
&\le O(1)+n^{-1}\sum_{i=1}^{n}\log\left(n\widehat{A}\{Y_i\}\right)\\
&\quad +n^{-1}\sum_{i=1}^{n}\left(\log\int_b\left(e^{\widetilde{M}+b}\right)^{\Delta_i}\left[e^{-\widetilde{M}-|b|}\left\{1+\widehat{A}(Y_i)\right\}\right]^{-\Delta_i-\rho}\phi\left(b;\widehat{\sigma}^2\right)db\right)\\
&\le O(1)+n^{-1}\sum_{i=1}^{n}\log\left(n\widehat{A}\{Y_i\}\right)-n^{-1}\sum_{i=1}^{n}\left[(\Delta_i+\rho)\log\left\{1+\widehat{A}(Y_i)\right\}\right].
\end{aligned}
$$

We first show that $\limsup_n \widehat{A}(\tau)<\infty$ using the partitioning idea of Murphy (1994). Specifically, we construct a sequence $u_0=\tau>u_1>\cdots>u_Q=0$. Then,

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^{n}\log\left(n\widehat{A}\{Y_i\}\right)-n^{-1}\sum_{i=1}^{n}\left[(\Delta_i+\rho)\log\left\{1+\widehat{A}(Y_i)\right\}\right]\\
\le\ & O(1)+\sum_{q=0}^{Q}n^{-1}\sum_{i=1}^{n}I(Y_i\in[u_{q+1},u_q))\log\left(n\widehat{A}\{Y_i\}\right)-n^{-1}\sum_{i=1}^{n}I(Y_i=\tau)\rho\log\left\{1+\widehat{A}(\tau)\right\}\\
& -\sum_{q=0}^{Q}n^{-1}\sum_{i=1}^{n}I\left(Y_i\in[u_{q+1},u_q)\right)\log\left\{1+\widehat{A}(u_{q+1})\right\},
\end{aligned}
$$

which is further bounded by

$$
\begin{aligned}
& -(2n)^{-1}\sum_{i=1}^{n}(\rho+\Delta_i)I(Y_i=\tau)\log\left\{1+\widehat{A}(\tau)\right\}\\
& -\left\{(2n)^{-1}\sum_{i=1}^{n}(\rho+\Delta_i)I(Y_i=\tau)-n^{-1}\sum_{i=1}^{n}\Delta_i I\left(Y_i\in[u_1,u_0)\right)\right\}\log\left\{1+\widehat{A}(\tau)\right\}\\
& -\sum_{q=1}^{Q}\left\{n^{-1}\sum_{i=1}^{n}(\rho+\Delta_i)I\left(Y_i\in[u_q,u_{q-1})\right)-n^{-1}\sum_{i=1}^{n}\Delta_i I\left(Y_i\in[u_{q+1},u_q)\right)\right\}\log\left\{1+\widehat{A}(u_q)\right\}.
\end{aligned}
$$

Note that u_q is chosen such that the coefficients in front of $\log\{1+\widehat{A}(u_q)\}$ are all negative when n is large enough. Thus, the corresponding terms cannot diverge to ∞. However, if $\widehat{A}(\tau)$ diverges to ∞, then the first term diverges to $-\infty$. We conclude that there exists some $M_A<\infty$ such that $\limsup_n \widehat{A}(\tau)\le M_A$. Therefore,

$$
\begin{aligned}
0 &\le n^{-1}l_n(\widehat{\boldsymbol{\theta}},\widehat{\mathcal{A}})-n^{-1}l_n(\boldsymbol{\theta}_0,\widetilde{\mathcal{A}})\\
&\le O(1)+n^{-1}\sum_{i=1}^{n}\left\{\log\int_b\left(\exp\left[-G\left\{e^{\widetilde{M}+|b|}\widehat{\Lambda}(U_{i,M_i})\right\}\right]\right)^{\delta_{i,M_i}}\phi\left(b;\widehat{\sigma}^2\right)db\right\}
\end{aligned}
$$

$$
\begin{aligned}
&\leq \quad O(1)+n^{-1}\sum_{i=1}^{n}\left\{\log\int_{|b|\leq 1}\left(\exp\left[-G\left\{e^{\widetilde{M}+|b|}\widehat{\Lambda}(U_{i,M_i})\right\}\right]\right)^{\delta_{i,M_i}}\phi\left(b;\widehat{\sigma}^2\right)db\right\}\\
&\qquad +n^{-1}\sum_{i=1}^{n}\left\{\log\int_{|b|>1}\phi\left(b;\widehat{\sigma}^2\right)db\right\}\\
&\leq \quad O(1)-n^{-1}\sum_{i=1}^{n}\delta_{i,M_i}G\left\{e^{\widetilde{M}+1}\widehat{\Lambda}(U_{i,M_i})\right\}.
\end{aligned}
$$

If $\limsup_n \widehat{\Lambda}(\tau-\epsilon)=\infty$, then $G\{e^{\widetilde{M}+1}\widehat{\Lambda}(\tau-\epsilon)\}=\infty$ with probability 1 under Condition 6. This is a contradiction. Therefore, $\limsup_n \widehat{\Lambda}(\tau-\epsilon)<\infty$ with probability 1 for any $\epsilon>0$. By choosing a sequence of ϵ decreasing to 0, it then follows from Helly's selection lemma that along a subsequence, $\widehat{\Lambda}\to\Lambda_*$ pointwise on any interior set of $\mathcal{U}$, $\widehat{A}\to A_*$ weakly on $[0,\tau]$, and $\widehat{\boldsymbol{\theta}}\to\boldsymbol{\theta}_*\equiv(\boldsymbol{\beta}_*,\boldsymbol{\gamma}_*,\sigma_*^2)$. We denote $\mathcal{A}_*=(\Lambda_*,A_*)$.

We now show that $\boldsymbol{\theta}_*=\boldsymbol{\theta}_0$ and $\mathcal{A}_*=\mathcal{A}_0$. We define

$$
m(\boldsymbol{\theta},\mathcal{A})=\log\left\{\frac{L(\boldsymbol{\theta},\mathcal{A})+L(\boldsymbol{\theta}_0,\widetilde{\mathcal{A}})}{2}\right\}
$$

and

$$
\mathcal{M}=\left\{m(\boldsymbol{\theta},\mathcal{A}):\boldsymbol{\theta}\in\Theta,\Lambda\in\mathcal{D}_\infty,A\in\mathcal{D}_{M_A}\right\}, \tag{3.6}
$$

where $L(\boldsymbol{\theta},\mathcal{A})$ is the objective function for a single subject, and $\mathcal{D}_c=\{\Lambda:\Lambda$ is increasing with $\Lambda(0)=0,\Lambda(\tau)\leq c\}$. By the concavity of the log function,

$$
\mathbb{P}_n m(\widehat{\boldsymbol{\theta}},\widehat{\mathcal{A}})\geq\frac{1}{2}\left\{\mathbb{P}_n\log L(\widehat{\boldsymbol{\theta}},\widehat{\mathcal{A}})+\mathbb{P}_n\log L(\boldsymbol{\theta}_0,\widetilde{\mathcal{A}})\right\}\geq\mathbb{P}_n l(\boldsymbol{\theta}_0,\widetilde{\mathcal{A}})=\mathbb{P}_n m(\boldsymbol{\theta}_0,\widetilde{\mathcal{A}}).
$$

Note that

$$
\begin{aligned}
\mathbb{P}m\left(\widehat{\boldsymbol{\theta}},\widehat{\mathcal{A}}\right)-\mathbb{P}m\left(\boldsymbol{\theta}_0,\widetilde{\mathcal{A}}\right) &= \mathbb{P}\log\left[\frac{L(\widehat{\boldsymbol{\theta}},\widehat{\mathcal{A}})+L(\boldsymbol{\theta}_0,\widetilde{\mathcal{A}})}{2L(\boldsymbol{\theta}_0,\widetilde{\mathcal{A}})}\right]\\
&= \mathbb{P}\log\left\{\frac{1}{2}+\frac{\widehat{A}\{Y\}^{\Delta}\int_b K_1\left(b,\mathcal{O};\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\gamma}},\widehat{\mathcal{A}}\right)\phi(b,\widehat{\sigma}^2)db}{2\widetilde{A}\{Y\}^{\Delta}\int_b K_1\left(b,\mathcal{O};\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0,\widetilde{\mathcal{A}}\right)\phi(b,\sigma_0^2)db}\right\}.
\end{aligned}
$$

By the definition of $\widetilde{A}$, $\widehat{A}(t)$ is absolutely continuous with respect to $\widetilde{A}(t)$, and

$$\widehat{A}(t) = \int_0^t \frac{\mathbb{P}_n \nu\left(s, \mathcal{O}; \boldsymbol{\theta}_0, \mathcal{A}_0\right)}{\left|\mathbb{P}_n \nu\left(s, \mathcal{O}; \widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}}\right)\right|} d\widetilde{A}(s), \tag{3.7}$$

where

$$\nu(t, \mathcal{O}; \boldsymbol{\theta}, \mathcal{A}) = \frac{\int_b K_1\left(b, \mathcal{O}; \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathcal{A}\right) K_2(t, b, \mathcal{O}; \boldsymbol{\gamma}, A)\phi(b; \sigma^2) db}{\int_b K_1\left(b, \mathcal{O}; \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathcal{A}\right) \phi(b; \sigma^2) db}.$$

To take limits on both sides of equation (4.4), we first show that the denominator of the integrand is uniformly bounded away from zero. It follows from the Glivenko-Cantelli property in Lemma 3.1 that

$$\sup_{t \in [0,\tau]} \left|\mathbb{P}_n \nu\left(t, \mathcal{O}; \boldsymbol{\theta}_0, \mathcal{A}_0\right) - \mathbb{P} \nu\left(t, \mathcal{O}; \boldsymbol{\theta}_0, \mathcal{A}_0\right)\right| \to_{a.s.} 0$$

and

$$\sup_{t \in [0,\tau]} \left|\mathbb{P}_n \nu\left(t, \mathcal{O}; \widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}}\right) - \mathbb{P} \nu\left(t, \mathcal{O}; \boldsymbol{\theta}_*, \mathcal{A}_*\right)\right| \to_{a.s.} 0.$$

Note that for any $\epsilon > 0$,

$$\limsup_n \widehat{A}(\tau) \geq \int_0^\tau \frac{\nu\left(t, \mathcal{O}; \boldsymbol{\theta}_0, \mathcal{A}_0\right)}{\epsilon + \left|\mathbb{P} \nu\left(t, \mathcal{O}; \boldsymbol{\theta}_*, \mathcal{A}_*\right)\right|} dA_0(s).$$

Let $\epsilon \to 0$. It follows from the Monotone Convergence Theorem that

$$\int_0^\tau \frac{\mathbb{P} \nu\left(t, \mathcal{O}; \boldsymbol{\theta}_0, \mathcal{A}_0\right)}{\left|\mathbb{P} \nu\left(t, \mathcal{O}; \boldsymbol{\theta}_*, \mathcal{A}_*\right)\right|} dA_0(t) < \infty.$$

We claim that $\min_{t \in [0,\tau]} |\mathbb{P}\nu(t, \mathcal{O}; \boldsymbol{\theta}_*, \mathcal{A}_*)| > 0$. If this inequality does not hold, then there exists some $t_* \in [0, \tau]$ such that $\mathbb{P}\nu(t_*, \mathcal{O}; \boldsymbol{\theta}_*, \mathcal{A}_*) = 0$. The function $\mathbb{P}\nu(t_*, \mathcal{O}; \boldsymbol{\theta}_*, \mathcal{A}_*)$ is right-differentiable almost everywhere. Thus, there exists $\delta > 0$ such that for $t \in (t_*, t_* + \delta)$,

$$\left|\mathbb{P} \nu\left(t, \mathcal{O}; \boldsymbol{\theta}_*, \mathcal{A}_*\right)\right| = \left|\mathbb{P} \nu\left(t, \mathcal{O}; \boldsymbol{\theta}_*, \mathcal{A}_*\right) - \mathbb{P} \nu\left(t_*, \mathcal{O}; \boldsymbol{\theta}_*, \mathcal{A}_*\right)\right| \leq O(1)|t - t_*|$$

almost everywhere. Hence,

$$\int_{t_*}^{t_*+\delta} \frac{1}{|t - t_*|} dA_0(t) < \infty,$$

which is a contradiction. By taking the limits on both sides of (4.4), we conclude that $A_*(t)$ is absolutely continuous with respect to $A_0(t)$, so that $A_*(t)$ is differentiable with respect to t. In addition, $d\widehat{A}(t)/d\widetilde{A}(t)$ converges to $dA_*(t)/dA_0(t)$ uniformly in t. It then follows from Lemma 3.1 that the class $\mathcal{M}$, as defined in (3.6), is Glivenko-Cantelli. Thus,

$$
\begin{aligned}
0 &\leq \mathbb{P}_n m\left(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}}\right) - \mathbb{P}_n m\left(\boldsymbol{\theta}_0, \widetilde{\mathcal{A}}\right) \\
&= \mathbb{P}\left\{m\left(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}}\right) - m\left(\boldsymbol{\theta}_0, \widetilde{\mathcal{A}}\right)\right\} + o_P(1) \\
&\to \mathbb{P}\left[\log\left\{\frac{1}{2} + \frac{A_*'(Y)^{\Delta}\int_b K_1\left(b, \mathcal{O}; \boldsymbol{\beta}_*, \boldsymbol{\gamma}_*, \mathcal{A}_*\right)\phi(b, \sigma_*^2)db}{2A_0'(Y)^{\Delta}\int_b K_1\left(b, \mathcal{O}; \boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, \mathcal{A}_0\right)\phi(b, \sigma_0^2)db}\right\}\right]
\end{aligned}
$$

such that the negative Kullback-Leibler information is positive. Therefore,

$$
\begin{aligned}
&\int_b \left\{\sum_{m=0}^{M} \delta_m \left(\exp\left[-G\left\{\int_0^{U_m} e^{\boldsymbol{\beta}_*^{\mathrm{T}}\boldsymbol{X}(s)+b} d\Lambda_*(s)\right\}\right] - \exp\left[-G\left\{\int_0^{U_{m+1}} e^{\boldsymbol{\beta}_*^{\mathrm{T}}\boldsymbol{X}(s)+b} d\Lambda_*(s)\right\}\right]\right)\right\} \\
&\times \left[e^{\boldsymbol{\gamma}_*^{\mathrm{T}}\boldsymbol{X}(Y)+b} A_*(Y) H'\left\{\int_0^{Y} e^{\boldsymbol{\gamma}_*^{\mathrm{T}}\boldsymbol{X}(s)+b} dA_*(s)\right\}\right]^{\Delta} \exp\left[-H\left\{\int_0^{Y} e^{\boldsymbol{\gamma}_*^{\mathrm{T}}\boldsymbol{X}(s)+b} dA_*(s)\right\}\right] \phi(b; \sigma_*^2)db \\
&= \int_b \left\{\sum_{m=0}^{M} \delta_m \left(\exp\left[-G\left\{\int_0^{U_m} e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}(s)+b} d\Lambda_0(s)\right\}\right] - \exp\left[-G\left\{\int_0^{U_{m+1}} e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}(s)+b} d\Lambda_0(s)\right]\right)\right\}\right\} \\
&\times \left[e^{\boldsymbol{\gamma}_0^{\mathrm{T}}\boldsymbol{X}(Y)+b} A_0(Y) H'\left\{\int_0^{Y} e^{\boldsymbol{\gamma}_0^{\mathrm{T}}\boldsymbol{X}(s)+b} dA_0(s)\right\}\right]^{\Delta} \exp\left[-H\left\{\int_0^{Y} e^{\boldsymbol{\gamma}_0^{\mathrm{T}}\boldsymbol{X}(s)+b} dA_0(s)\right\}\right] \phi(b; \sigma_0^2)db
\end{aligned}
$$

with probability 1. For any $m \in \{0, \ldots, M\}$, we set $\delta_{m'} = 1$ in the above equation for $m' = m, \ldots, M$ and take the sum of the resulting equations to obtain

$$
\begin{aligned}
&\int_b \exp\left[-G\left\{\int_0^{U_m} e^{\boldsymbol{\beta}_*^{\mathrm{T}}\boldsymbol{X}(s)+b} d\Lambda_*(s)\right\}\right] \left[e^{\boldsymbol{\gamma}_*^{\mathrm{T}}\boldsymbol{X}(Y)+b} A_*(Y) H'\left\{\int_0^{Y} e^{\boldsymbol{\gamma}_*^{\mathrm{T}}\boldsymbol{X}(s)+b} dA_*(s)\right\}\right]^{\Delta} \\
&\exp\left[-H\left\{\int_0^{Y} e^{\boldsymbol{\gamma}_*^{\mathrm{T}}\boldsymbol{X}(s)+b} dA_*(s)\right\}\right] \phi(b; \sigma_*^2)db \\
= &\int_b \exp\left[-G\left\{\int_0^{U_m} e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}(s)+b} d\Lambda_0(s)\right\}\right] \left[e^{\boldsymbol{\gamma}_0^{\mathrm{T}}\boldsymbol{X}(Y)+b} A_0(Y) H'\left\{\int_0^{Y} e^{\boldsymbol{\gamma}_0^{\mathrm{T}}\boldsymbol{X}(s)+b} dA_0(s)\right\}\right]^{\Delta} \\
&\exp\left[-H\left\{\int_0^{Y} e^{\boldsymbol{\gamma}_0^{\mathrm{T}}\boldsymbol{X}(s)+b} dA_0(s)\right\}\right] \phi(b; \sigma_0^2)db.
\end{aligned}
$$

Because m is arbitrary, we can replace U_m in the above equation by any $t_1 \in \mathcal{U}$. We set $\Delta = 1$ and integrate Y from 0 to $t_2 \in [0,\tau]$ to obtain

$$
\begin{aligned}
& \int_b \exp\left[-G\left\{\int_0^{t_1} e^{\boldsymbol{\beta}_*^{\mathrm{T}}\boldsymbol{X}(s)+b} d\Lambda_*(s)\right\} - H\left\{\int_0^{t_2} e^{\boldsymbol{\gamma}_*^{\mathrm{T}}\boldsymbol{X}(s)+b} dA_*(s)\right\}\right] \phi(b;\sigma_*^2)db \\
= & \int_b \exp\left[-G\left\{\int_0^{t_1} e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}(s)+b} d\Lambda_0(s)\right\} - H\left\{\int_0^{t_2} e^{\boldsymbol{\gamma}_0^{\mathrm{T}}\boldsymbol{X}(s)+b} dA_0(s)\right\}\right] \phi(b;\sigma_0^2)db.
\end{aligned}
$$

By Condition 7, we have $\boldsymbol{\theta}_* = \boldsymbol{\theta}_0$ and $\mathcal{A}_* = \mathcal{A}_0$. We conclude that $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| \to 0$, $|\widehat{\Lambda}(t_1) - \Lambda_0(t_1)| \to 0$, and $|\widehat{A}(t_2) - A_0(t_2)| \to 0$ for any $t_1 \in \mathcal{U}$ and $t_2 \in [0,\tau]$. Because $\mathcal{A}_0$ is continuous, $\widehat{\mathcal{A}}$ converges uniformly to $\mathcal{A}_0$ on $\mathcal{U} \times [0,\tau]$.

3.6.2 Proof of Theorem 3.2

Let

$$
H_1(t,u,v,\mathcal{O};\boldsymbol{\theta},\mathcal{A}) = \frac{\int K_1(b,\mathcal{O};\boldsymbol{\beta},\boldsymbol{\gamma},\mathcal{A}) Q_1(t,u,v,b,\mathcal{O};\boldsymbol{\beta},\Lambda)\phi(b;\sigma^2)db}{\int K_1(b,\mathcal{O};\boldsymbol{\beta},\boldsymbol{\gamma},\mathcal{A})\phi(b;\sigma^2)db}
$$

and

$$
H_2(t,\mathcal{O};\boldsymbol{\theta},\mathcal{A}) = \frac{\int K_1(b,\mathcal{O};\boldsymbol{\beta},\boldsymbol{\gamma},\mathcal{A}) Q_2(t,Y,b,\mathcal{O};\boldsymbol{\gamma},A)\phi(b;\sigma^2)db}{\int K_1(b,\mathcal{O};\boldsymbol{\beta},\boldsymbol{\gamma},\mathcal{A})\phi(b;\sigma^2)db},
$$

where

$$
\begin{aligned}
& Q_1(t,u,v,b,\mathcal{O};\boldsymbol{\beta},\Lambda) \\
= \; & e^{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}(t)+b} \left(\frac{I(v \ge t)\exp\left[-G\left\{\int_0^v e^{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}(s)+b}d\Lambda(s)\right\}\right] G'\left\{\int_0^v e^{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}(s)+b}d\Lambda(s)\right\}}{\exp\left[-G\left\{\int_0^u e^{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}(s)+b}d\Lambda(s)\right\}\right] - \exp\left[-G\left\{\int_0^v e^{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}(s)+b}d\Lambda(s)\right\}\right]} \right. \\
& \left. - \frac{I(u \ge t)\exp\left[-G\left\{\int_0^u e^{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}(s)+b}d\Lambda(s)\right\}\right] G'\left\{\int_0^u e^{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}(s)+b}d\Lambda(s)\right\}}{\exp\left[-G\left\{\int_0^u e^{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}(s)+b}d\Lambda(s)\right\}\right] - \exp\left[-G\left\{\int_0^v e^{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}(s)+b}d\Lambda(s)\right\}\right]} \right),
\end{aligned}
$$

and

$$
\begin{aligned}
Q_2(t,u,b,\mathcal{O};\boldsymbol{\gamma},A) \;=\; & I(u \ge t) e^{\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{X}(t)+b} \\
& \times \left[\Delta \frac{H''\left\{\int_0^u e^{\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{X}(s)+b}dA(s)\right\}}{H'\left\{\int_0^u e^{\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{X}(s)+b}dA(s)\right\}} - H'\left\{\int_0^u e^{\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{X}(s)+b}dA(s)\right\}\right].
\end{aligned}
$$

Then, the score function for $\boldsymbol{\theta}$ is $\boldsymbol{l}_{\boldsymbol{\theta}}(\boldsymbol{\theta},\mathcal{A}) = (\boldsymbol{l}_{\boldsymbol{\beta}}(\boldsymbol{\theta},\mathcal{A})^{\mathrm{T}}, \boldsymbol{l}_{\boldsymbol{\gamma}}(\boldsymbol{\theta},\mathcal{A})^{\mathrm{T}}, l_{\sigma^2}(\boldsymbol{\theta},\mathcal{A}))^{\mathrm{T}}$, where

$$
\boldsymbol{l}_{\boldsymbol{\beta}}(\boldsymbol{\theta},\mathcal{A}) \;=\; \sum_{m=0}^{M} \delta_m \int_0^\tau H_1(t,U_m,U_{m+1},\mathcal{O};\boldsymbol{\theta},\mathcal{A})\boldsymbol{X}(t)d\Lambda(t),
$$

$$\boldsymbol{l}_{\gamma}(\boldsymbol{\theta}, \mathcal{A}) = \Delta \boldsymbol{X}(Y) + \int_0^\tau H_2(t, \mathcal{O}; \boldsymbol{\theta}, \mathcal{A}) \boldsymbol{X}(t) dA(t),$$

and

$$l_{\sigma^2}(\boldsymbol{\theta}, \mathcal{A}) = \frac{\int K_1(b, \mathcal{O}; \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathcal{A}) \phi'(b; \sigma^2) db}{\int K_1(b, \mathcal{O}; \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathcal{A}) \phi(b; \sigma^2) db}.$$

The score operator for $\mathcal{A}$ along the submodel $d\mathcal{A}_{\epsilon,\boldsymbol{h}} = ((1+\epsilon h_1)d\Lambda, (1+\epsilon h_2)dA)^{\mathrm{T}}$ for $\boldsymbol{h} = (h_1, h_2)$ with $h_1 \in L_2(\mu)$ and $h_2 \in BV_1[0,\tau]$ is

$$\begin{aligned} l_{\mathcal{A}}(\boldsymbol{\theta}, \mathcal{A})(\boldsymbol{h}) &= \sum_{m=0}^{M} \delta_m \int_0^\tau H_1(t, U_m, U_{m+1}, \mathcal{O}; \boldsymbol{\theta}, \mathcal{A}) h_1(t) d\Lambda(t) \\ &\quad + \Delta h_2(Y) A'(Y) + \int_0^\tau H_2(t, \mathcal{O}; \boldsymbol{\theta}, \mathcal{A}) h_2(t) dA(t), \end{aligned}$$

where $BV_1(\mathcal{B})$ denotes the set of functions on $\mathcal{B}$ with total variation bounded by 1.

Clearly,

$$\mathbb{G}_n \left\{ \boldsymbol{l}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}}) \right\} = -\sqrt{n} \mathbb{P} \left\{ \boldsymbol{l}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}}) - \boldsymbol{l}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, \mathcal{A}_0) \right\},$$

and

$$\mathbb{G}_n \left\{ l_{\mathcal{A}}(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}})(\boldsymbol{h}) \right\} = -\sqrt{n} \mathbb{P} \left\{ l_{\mathcal{A}}(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}})(\boldsymbol{h}) - l_{\mathcal{A}}(\boldsymbol{\theta}_0, \mathcal{A}_0)(\boldsymbol{h}) \right\}.$$

We apply the Taylor series expansions at $(\boldsymbol{\theta}_0, \mathcal{A}_0)$ to the right sides of the above two equations. In light of Lemma 3.3, the second-order terms are bounded by

$$\begin{aligned} & O_P(1)\sqrt{n} E \left[\sum_{m=0}^{M} \left\{ \widehat{\Lambda}(U_m) - \Lambda_0(U_m) \right\}^2 + \left\{ \widehat{A}(Y) - A_0(Y) \right\}^2 + \left\| \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \right\|^2 \right. \\ & \left. + \left\| \widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0 \right\|^2 + \left\| \widehat{\sigma}^2 - \sigma_0^2 \right\|^2 \right] \\ = & \sqrt{n} \left\{ O_P(n^{-2/3}) + O_P(1) \left\| \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \right\|^2 + O_P(1) \left\| \widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0 \right\|^2 + O_P(1) \left\| \widehat{\sigma}^2 - \sigma_0^2 \right\|^2 \right\} \\ = & O_P \left(n^{-1/6} + \sqrt{n} \left\| \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \right\|^2 + \sqrt{n} \left\| \widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0 \right\|^2 + \sqrt{n} \left\| \widehat{\sigma}^2 - \sigma_0^2 \right\|^2 \right). \end{aligned}$$

Thus,

$$\begin{aligned} \mathbb{G}_n \left\{ \boldsymbol{l}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}}) \right\} &= -\sqrt{n} \mathbb{P} \left\{ \boldsymbol{l}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \boldsymbol{l}_{\boldsymbol{\theta}\mathcal{A}}(\widehat{\mathcal{A}} - \mathcal{A}_0) \right\} \\ &\quad + O_P \left(n^{-1/6} + \sqrt{n} \left\| \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \right\|^2 + \sqrt{n} \left\| \widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0 \right\|^2 + \sqrt{n} \left\| \widehat{\sigma}^2 - \sigma_0^2 \right\|^2 \right), \end{aligned}$$

and

$$\begin{aligned}\mathbb{G}_n\left\{l_{\mathcal{A}}(\widehat{\boldsymbol{\theta}},\widehat{\mathcal{A}})(\boldsymbol{h})\right\} &= -\sqrt{n}\mathbb{P}\left\{l_{\mathcal{A}\boldsymbol{\theta}}(\boldsymbol{h})(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)+l_{\mathcal{A}\mathcal{A}}(\boldsymbol{h},\widehat{\mathcal{A}}-\mathcal{A}_0)\right\}\\ &\quad +O_P\left(n^{-1/6}+\sqrt{n}\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right\|^2+\sqrt{n}\left\|\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0\right\|^2+\sqrt{n}\left\|\widehat{\sigma}^2-\sigma_0^2\right\|^2\right),\end{aligned}$$

where $\boldsymbol{l}_{\boldsymbol{\theta}\boldsymbol{\theta}}$ is the second derivative of $\boldsymbol{l}(\boldsymbol{\theta},\mathcal{A})$ with respect to $\boldsymbol{\theta}$, $\boldsymbol{l}_{\boldsymbol{\theta}\mathcal{A}}(\boldsymbol{h})$ is the derivative of $\boldsymbol{l}_{\boldsymbol{\theta}}$ along the submodel $d\mathcal{A}_{\epsilon,h}$, $l_{\mathcal{A}\boldsymbol{\theta}}(\boldsymbol{h})$ is the derivative of $l_{\mathcal{A}}(\boldsymbol{h})$ with respect to $\boldsymbol{\theta}$, and $l_{\mathcal{A}\mathcal{A}}(\boldsymbol{h},\widehat{\mathcal{A}}-\mathcal{A}_0)$ is the derivative of $l_{\mathcal{A}}(\boldsymbol{h})$ along the submodel $d\mathcal{A}_0+\epsilon d(\widehat{\mathcal{A}}-\mathcal{A}_0)$. All derivatives are evaluated at $(\boldsymbol{\theta}_0,\mathcal{A}_0)$.

Let $\boldsymbol{h}^*=(\boldsymbol{h}_1^*,\boldsymbol{h}_2^*)$ denote the least favorable direction such that $l_{\mathcal{A}}^*l_{\mathcal{A}}(\boldsymbol{h}^*)=l_{\mathcal{A}}^*\boldsymbol{l}_{\boldsymbol{\theta}}$, where $\boldsymbol{h}_1^*$ is $(2p+1)$-dimensional vector of functions in $L_2(\mu)$, $\boldsymbol{h}_2^*$ is $(2p+1)$-dimensional vector of functions in $L_2[0,\tau]$, and $l_{\mathcal{A}}^*$ is the adjoint operator of $l_{\mathcal{A}}$. We first show the existence of $\boldsymbol{h}^*$. Let $\mathcal{Q}=L_2(\mu)\times L_2[0,\tau]$. We equip $\mathcal{Q}$ with an inner product defined as

$$<\boldsymbol{h}^{(1)},\boldsymbol{h}^{(2)}>=\int_{\mathcal{U}}h_1^{(1)}h_1^{(2)}d\mu(t)+\int_0^\tau h_2^{(1)}h_2^{(2)}dA_0(t),$$

where $\boldsymbol{h}^{(1)}=(h_1^{(1)},h_2^{(1)})$ and $\boldsymbol{h}^{(2)}=(h_1^{(2)},h_2^{(2)})$. On the same space, we define

$$\begin{aligned}\|\boldsymbol{h}\| &= \mathbb{P}\left\{l_{\mathcal{A}}\left(\boldsymbol{\theta}_0,\mathcal{A}_0\right)(\boldsymbol{h})^2\right\}^{1/2}\\ &= \mathbb{P}\left[\left\{\sum_{m=0}^M\delta_m\int_0^\tau H_1(t,U_m,U_{m+1},\mathcal{O};\boldsymbol{\theta}_0,\mathcal{A}_0)h_1(t)d\Lambda_0(t)\right.\right.\\ &\quad\left.\left.+\Delta h_2(Y)A_0'(Y)+\int_0^\tau H_2(t,\mathcal{O};\boldsymbol{\theta}_0,\mathcal{A}_0)h_2(t)dA_0(t)\right\}^2\right]^{1/2}\end{aligned}$$

for $\boldsymbol{h}=(h_1,h_2)$. It is easy to show that $\|\cdot\|$ is a seminorm on $\mathcal{Q}$. Furthermore, if $\|\boldsymbol{h}\|=0$, then $\mathbb{P}\{l_{\mathcal{A}}(\boldsymbol{\theta}_0,\mathcal{A}_0)(\boldsymbol{h})^2\}=0$. Thus, with probability 1, $l_{\mathcal{A}}(\boldsymbol{\theta}_0,\mathcal{A}_0)(\boldsymbol{h})=0$. By the arguments in the proof of Lemma 3.3, $h_1(t)=0$ for $t\in\mathcal{U}$ and $h_2(t)=0$ for $t\in[0,\tau]$. Clearly, $\|\boldsymbol{h}\|\le c<\boldsymbol{h},\boldsymbol{h}>^{1/2}$ for some constant c by the Cauchy-Schwarz inequality. According to the bounded inverse theorem in Banach spaces, we have $<\boldsymbol{h},\boldsymbol{h}>^{1/2}\le\widetilde{c}\|\boldsymbol{h}\|$ for another constant $\widetilde{c}$. By the Lax-Milgram theorem (Zeidler, 1995), $\boldsymbol{h}^*$ exists and satisfies

$$\int_0^\tau\mathbb{P}\left\{\sum_{m=0}^M\delta_mH_1(t,U_m,U_{m+1},\mathcal{O};\boldsymbol{\theta}_0,\mathcal{A}_0)H_1(s,U_m,U_{m+1},\mathcal{O};\boldsymbol{\theta}_0,\mathcal{A}_0)\right\}\boldsymbol{h}_1^*(s)d\Lambda_0(s)$$

$$= \mathbb{P}\left\{\sum_{m=0}^{M}\delta_m H_1(t, U_m, U_{m+1}, \mathcal{O}; \boldsymbol{\theta}_0, \mathcal{A}_0)\boldsymbol{l}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, \mathcal{A}_0)\right\}$$

for $t \in \mathcal{U}$, and

$$\begin{aligned}
&\int_0^\tau \mathbb{P}\left[\{\Delta I(Y=t) + H_2(t, \mathcal{O}; \boldsymbol{\theta}_0, \mathcal{A}_0)\}\{\Delta I(Y=s) + H_2(s, \mathcal{O}; \boldsymbol{\theta}_0, \mathcal{A}_0)\}\right]\boldsymbol{h}_2^*(s)dA_0(s) \\
= \;& \mathbb{P}\left[\{\Delta I(Y=t) + H_2(t, \mathcal{O}; \boldsymbol{\theta}_0, \mathcal{A}_0)\}\boldsymbol{l}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, \mathcal{A}_0)\right]
\end{aligned}$$

for $t \in [0, \tau]$. Differentiation of the first integral equations with respect to t at $t = t_1 \in \mathcal{U}$ yields

$$q_{11}(t_1)\boldsymbol{h}_1^*(t_1) + \sum_{k=1}^{2}\int_{t_1}^{\tau} q_{12}(s, t_1)\boldsymbol{h}_k^*(s)ds + \int_0^{t_1} q_{13}(s, t_1)\boldsymbol{h}_k^*(s)ds = \boldsymbol{q}_{14}(t_1)$$

and differentiating of the second integral equation with respect to t at $t = t_2 \in [0, \tau]$ gives

$$q_{21}(t_2)\boldsymbol{h}_2^*(t_2) + \sum_{k=1}^{2}\int_{t_2}^{\tau} q_{22}(s, t_2)\boldsymbol{h}_k^*(s)ds + \int_0^{t_2} q_{23}(s, t_2)\boldsymbol{h}_k^*(s)ds = \boldsymbol{q}_{24}(t_2),$$

where $q_{11}(t_1) > 0$, $q_{21}(t_2) > 0$, and q_{kj} $(k = 1, 2; j = 1, 2, 3)$ and $\boldsymbol{q}_{k4}$ $(k = 1, 2)$ are continuously differentiable functions. Thus, $\boldsymbol{h}^*$ can be expanded to be a continuously differentiable function in $[0, \tau] \times [0, \tau]$ with bounded total variations. It then follows that

$$\begin{aligned}
&\mathbb{G}_n\left\{\boldsymbol{l}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}})\right\} - \mathbb{G}_n\left\{l_{\mathcal{A}}(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}})(\boldsymbol{h}^*)\right\} \\
= \;& -\sqrt{n}\mathbb{P}\left\{\boldsymbol{l}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \boldsymbol{l}_{\boldsymbol{\theta}\mathcal{A}}(\widehat{\mathcal{A}} - \mathcal{A}_0)\right\} + \sqrt{n}\mathbb{P}\left\{l_{\mathcal{A}\boldsymbol{\theta}}(\boldsymbol{h}^*)(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + l_{\mathcal{A}\mathcal{A}}(\boldsymbol{h}^*, \widehat{\mathcal{A}} - \mathcal{A}_0)\right\} \\
&+ O_P\left(n^{-1/6} + \sqrt{n}\left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right\|^2 + \sqrt{n}\left\|\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\right\|^2 + \sqrt{n}\left\|\widehat{\sigma}^2 - \sigma_0^2\right\|^2\right) \\
= \;& \sqrt{n}\mathbb{P}\left[\{\boldsymbol{l}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, \mathcal{A}_0) - l_{\mathcal{A}}(\boldsymbol{\theta}_0, \mathcal{A}_0)(\boldsymbol{h}^*)\}^{\otimes 2}\right]\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \\
&+ O_P\left(n^{-1/6} + \sqrt{n}\left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right\|^2 + \sqrt{n}\left\|\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\right\|^2 + \sqrt{n}\left\|\widehat{\sigma}^2 - \sigma_0^2\right\|^2\right).
\end{aligned}$$

Using similar arguments as in the proof of Lemma 3.2, we can show that $\boldsymbol{l}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, \mathcal{A}_0) - l_{\mathcal{A}}(\boldsymbol{\theta}_0, \mathcal{A}_0)(\boldsymbol{h}^*)$ belongs to a Donsker class. Next, we show that the matrix $\mathbb{P}[\{\boldsymbol{l}_{\boldsymbol{\theta}} - l_{\mathcal{A}}(\boldsymbol{h}^*)\}^{\otimes 2}]$ is invertible. If the matrix is singular, then there exists a vector $\boldsymbol{v} \equiv (\boldsymbol{v}_1, \boldsymbol{v}_2, v_3)^{\mathrm{T}}$ with $\boldsymbol{v}_1, \boldsymbol{v}_2 \in \mathbb{R}^p$ and $v_3 \in \mathbb{R}$ such that $\boldsymbol{v}^{\mathrm{T}}E[\{\boldsymbol{l}_{\boldsymbol{\theta}} - l_{\mathcal{A}}(\boldsymbol{h}^*)\}^{\otimes 2}]\boldsymbol{v} = 0$. It follows that, with probability 1, the score function along the submodel $\{\boldsymbol{\theta}_0 + \epsilon \boldsymbol{v}, \mathcal{A}_\epsilon(\boldsymbol{v}^{\mathrm{T}} \boldsymbol{h}^*)\}$ is zero. That is,

$$\int_b K_1(b, \mathcal{O}; \boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, \mathcal{A}_0) \left[\sum_{m=0}^{M} \delta_m \int_0^\tau Q_1(t, U_m, U_{m+1}, b, \mathcal{O}; \boldsymbol{\beta}_0, \Lambda_0) \left\{ \boldsymbol{v}_1^{\mathrm{T}} \boldsymbol{X}(t) - \boldsymbol{v}^{\mathrm{T}} \boldsymbol{h}_1^*(t) \right\} d\Lambda_0(t) \right.$$

$$+ \Delta \left\{ \boldsymbol{v}_2^{\mathrm{T}} \boldsymbol{X}(Y) - \boldsymbol{v}^{\mathrm{T}} \boldsymbol{h}_2^*(Y) A_0'(Y) \right\} + \int_0^\tau Q_2(t, Y, b, \mathcal{O}; \boldsymbol{\gamma}_0, A_0) \left\{ \boldsymbol{v}_2^{\mathrm{T}} \boldsymbol{X}(t) - \boldsymbol{v}^{\mathrm{T}} \boldsymbol{h}_2^*(t) \right\} dA_0(t)$$

$$\left. - v_3 \frac{\phi'(b, \sigma_0^2)}{\phi(b; \sigma_0^2)} \right] \phi(b; \sigma_0^2) db = 0$$

with probability 1. For any $t \in [0, \tau]$, we let $\Delta = 0$ and integrate Y from 0 to t to obtain

$$\begin{aligned}
&\int_b \left[\sum_{m=0}^{M} \delta_m \int_0^\tau Q_1(s, U_m, U_{m+1}, b, \mathcal{O}; \boldsymbol{\beta}_0, \Lambda_0) \left\{ \boldsymbol{v}_1^{\mathrm{T}} \boldsymbol{X}(s) - \boldsymbol{v}^{\mathrm{T}} \boldsymbol{h}_1^*(s) \right\} d\Lambda_0(s) \right. \\
&\left. + H' \left\{ \int_0^t e^{\boldsymbol{\gamma}_0^{\mathrm{T}} \boldsymbol{X}(s) + b} dA_0(s) \right\} \int_0^t e^{\boldsymbol{\gamma}_0^{\mathrm{T}} \boldsymbol{X}(s) + b} \left\{ \boldsymbol{v}_2^{\mathrm{T}} \boldsymbol{X}(s) - \boldsymbol{v}^{\mathrm{T}} \boldsymbol{h}_2^*(s) \right\} dA_0(s) \right] \\
&\times \left\{ \sum_{m=0}^{M} \delta_m \left(\exp\left[-G \left\{ \int_0^{U_m} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}(s) + b} d\Lambda_0(s) \right\} \right] - \exp\left[-G \left\{ \int_0^{U_{m+1}} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}(s) + b} d\Lambda_0(s) \right\} \right] \right) \right\} \\
&\times \exp\left[-H \left\{ \int_0^t e^{\boldsymbol{\gamma}_0^{\mathrm{T}} \boldsymbol{X}(s) + b} dA_0(s) \right\} \right] \phi(b, \sigma_0^2) db = 0.
\end{aligned}$$

For any $m \in \{0, \ldots, M\}$, we sum over all possible $\delta_{m'}$ with $m' = m, \ldots, M$ to obtain

$$\begin{aligned}
&\int_b \left[G' \left\{ \int_0^{U_m} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}(s) + b} d\Lambda_0(s) \right\} \int_0^{U_m} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}(s) + b} \left\{ \boldsymbol{v}_1^{\mathrm{T}} \boldsymbol{X}(s) - \boldsymbol{v}^{\mathrm{T}} \boldsymbol{h}_1^*(s) \right\} d\Lambda_0(s) \right. \\
&\left. + H' \left\{ \int_0^t e^{\boldsymbol{\gamma}_0^{\mathrm{T}} \boldsymbol{X}(s) + b} dA_0(s) \right\} \int_0^t e^{\boldsymbol{\gamma}_0^{\mathrm{T}} \boldsymbol{X}(s) + b} \left\{ \boldsymbol{v}_2^{\mathrm{T}} \boldsymbol{X}(s) - \boldsymbol{v}^{\mathrm{T}} \boldsymbol{h}_2^*(s) \right\} dA_0(s) - v_3 \frac{\phi'(b, \sigma_0^2)}{\phi(b; \sigma_0^2)} \right] \\
&\times \exp\left[-G \left\{ \int_0^{U_m} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}(s) + b} d\Lambda_0(s) \right\} - H \left\{ \int_0^t e^{\boldsymbol{\gamma}_0^{\mathrm{T}} \boldsymbol{X}(s) + b} dA_0(s) \right\} \right] \phi(b, \sigma_0^2) db = 0.
\end{aligned}$$

By Condition 8, $\boldsymbol{v}_1^{\mathrm{T}} \boldsymbol{X}(t) - \boldsymbol{v}^{\mathrm{T}} \boldsymbol{h}_1^*(t) = 0$ for any $t \in \mathcal{U}$, $\boldsymbol{v}_2^{\mathrm{T}} \boldsymbol{X}(t) - \boldsymbol{v}^{\mathrm{T}} \boldsymbol{h}_2^*(t) = 0$ for any $t \in [0, \tau]$, and $v_3 = 0$. It then follows from Condition 5 that $\boldsymbol{v}_1 = \boldsymbol{v}_2 = \mathbf{0}$. Hence, the matrix $E\big[\big\{\boldsymbol{l}_\theta - l_{\mathcal{A}}(\boldsymbol{h}^*)\big\}^{\otimes 2}\big]$ is invertible.

Because the matrix $\mathbb{P}[\{\boldsymbol{l}_{\boldsymbol{\theta}} - l_{\mathcal{A}}(\boldsymbol{h}^*)\}^{\otimes 2}]$ is invertible, $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = O_P(n^{-1/2})$, and

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = \left(\mathbb{P} \left[\{ \boldsymbol{l}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, \mathcal{A}_0) - l_{\mathcal{A}}(\boldsymbol{\theta}_0, \mathcal{A}_0)(\boldsymbol{h}^*) \}^{\otimes 2} \right] \right)^{-1} \mathbb{G}_n \left\{ \boldsymbol{l}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}}) - l_{\mathcal{A}}(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}})(\boldsymbol{h}^*) \right\} + o_P(1).$$

The influence function for $\widehat{\boldsymbol{\theta}}$ is the efficient influence function, such that $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ converges weakly to a zero-mean random vector whose covariance matrix attains the semiparametric efficiency bound.

3.6.3 Some Useful Lemmas

Lemma 3.1. *Under Conditions 1–6, the classes of functions*

$$\widetilde{\mathcal{H}}_1 \equiv \left\{ \int_b K_1(b, \mathcal{O}; \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathcal{A}) \phi(b; \sigma^2) db : \boldsymbol{\theta} \in \Theta, \Lambda \in \mathcal{D}_\infty, A \in \mathcal{D}_{M_A} \right\}$$

and

$$\widetilde{\mathcal{H}}_2 \equiv \left\{ \int_b K_1(b, \mathcal{O}; \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathcal{A}) K_2(t, b, \mathcal{O}; \boldsymbol{\gamma}, A) \phi(b; \sigma^2) db : \boldsymbol{\theta} \in \Theta, t \in [0, \tau], \Lambda \in \mathcal{D}_\infty, A \in \mathcal{D}_{M_A} \right\}$$

are $\mathbb{P}$*-Glivenko-Cantelli, where* M_A *is a finite constant.*

Proof. Let $K_G^{(l)}$ denote the lth derivative of $\exp\{-G(\cdot)\}$ for $l = 1, 2$, and let $K_H^{(l)}$ denote the lth derivative of $\exp\{-H(\cdot)\}$ for $l = 1, 2, 3$. We define

$$J(t, \boldsymbol{X}, b; \boldsymbol{\beta}, \Lambda) = \frac{\int_0^t e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}(s) + b} d\Lambda(s)}{\Lambda(\tau)},$$

where $\boldsymbol{\beta} \in \mathcal{B}_1$ and $\Lambda \in \mathcal{D}_\infty$. The class of functions $\{e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}(s) + b} : \boldsymbol{\beta} \in \mathcal{B}_1\}$, with $\boldsymbol{X}$ and b as random variables, is a VC class with VC-index V. Thus, the class $\mathcal{J} \equiv \{J(t, \boldsymbol{X}, b; \boldsymbol{\beta}, \Lambda) : \boldsymbol{\beta} \in \mathcal{B}_1, \Lambda \in \mathcal{D}_\infty\}$ is a convex hull of the VC-class with the $L_2(\mathbb{P})$-bracketing number $O\{\exp(\epsilon^{-2V/(V+2)})\}$.

For any $(\boldsymbol{\beta}_{(1)}, \Lambda_{(1)})$ and $(\boldsymbol{\beta}_{(2)}, \Lambda_{(2)})$ in $\mathcal{B}_1 \times \mathcal{D}_\infty$, $t \in \mathcal{U}$, and any positive constant M_Λ, if $\Lambda_{(1)}(\tau) > M_\Lambda$ and $\Lambda_{(2)}(\tau) > M_\Lambda$, then

$$\begin{aligned}
& \left| \exp\left[-G\left\{ \int_0^t e^{\boldsymbol{\beta}_{(1)}^{\mathrm{T}} \boldsymbol{X}(s) + b} d\Lambda_{(1)}(s) \right\} \right] - \exp\left[-G\left\{ \int_0^t e^{\boldsymbol{\beta}_{(2)}^{\mathrm{T}} \boldsymbol{X}(s) + b} d\Lambda_{(2)}(s) \right\} \right] \right| \\
\leq\ & \exp\left[-G\left\{ \int_0^t e^{\boldsymbol{\beta}_{(1)}^{\mathrm{T}} \boldsymbol{X}(s) + b} d\Lambda_{(1)}(s) \right\} \right] + \exp\left[-G\left\{ \int_0^t e^{\boldsymbol{\beta}_{(2)}^{\mathrm{T}} \boldsymbol{X}(s) + b} d\Lambda_{(2)}(s) \right\} \right] \\
\leq\ & 2 \exp\left[-G\left\{ M_\Lambda e^{-\widetilde{M} + |b|} \right\} \right].
\end{aligned}$$

If $\Lambda_{(1)}(\tau) \leq M_\Lambda$ and $\Lambda_{(2)}(\tau) \leq M_\Lambda$, then

$$\begin{aligned}
& \left| \exp\left[-G\left\{ \int_0^t e^{\boldsymbol{\beta}_{(1)}^{\mathrm{T}} \boldsymbol{X}(s) + b} d\Lambda_{(1)}(s) \right\} \right] - \exp\left[-G\left\{ \int_0^t e^{\boldsymbol{\beta}_{(2)}^{\mathrm{T}} \boldsymbol{X}(s) + b} d\Lambda_{(2)}(s) \right\} \right] \right| \\
\leq\ & \max_{\boldsymbol{\beta} \in \mathcal{B}_1, \Lambda \in \mathcal{D}_\infty, \Lambda(\tau) \leq M_\Lambda} \left| K_G^{(1)} \{ \Lambda(\tau) J(t, \boldsymbol{X}, b; \boldsymbol{\beta}, \Lambda) \} \right|
\end{aligned}$$

$$
\begin{aligned}
& \times \left\{ \left| J(t, \boldsymbol{X}, b; \boldsymbol{\beta}_{(1)}, \Lambda_{(1)}) - J(t, \boldsymbol{X}, b; \boldsymbol{\beta}_{(2)}, \Lambda_{(2)}) \right| M_\Lambda \right. \\
& \left. + J(t, \boldsymbol{X}, b; \boldsymbol{\beta}_{(1)}, \Lambda_{(1)}) \left| \Lambda_{(1)}(\tau) - \Lambda_{(2)}(\tau) \right| \right\} \\
\leq \; & \max_{\boldsymbol{\beta} \in \mathcal{B}_1, \Lambda \in \mathcal{D}_\infty, \Lambda(\tau) \leq M_\Lambda} \left| K_G^{(1)} \left\{ \Lambda(\tau) J(t, \boldsymbol{X}, b; \boldsymbol{\beta}, \Lambda) \right\} \right| \\
& \times \left\{ \left| J(t, \boldsymbol{X}, b; \boldsymbol{\beta}_{(1)}, \Lambda_{(1)}) - J(t, \boldsymbol{X}, b; \boldsymbol{\beta}_{(2)}, \Lambda_{(2)}) \right| M_\Lambda + e^{\widetilde{M} + |b|} \left| \Lambda_{(1)}(\tau) - \Lambda_{(2)}(\tau) \right| \right\}.
\end{aligned}
$$

In the remaining scenario, we assume, without loss of generality, that $\Lambda_{(1)}(\tau) \leq M_\Lambda$ and $\Lambda_{(2)}(\tau) > M_\Lambda$. Then,

$$
\begin{aligned}
& \left| \exp\left[-G\left\{ \int_0^t e^{\boldsymbol{\beta}_{(1)}^{\mathrm{T}} \boldsymbol{X}(s) + b} d\Lambda_{(1)}(s) \right\} \right] - \exp\left[-G\left\{ \int_0^t e^{\boldsymbol{\beta}_{(2)}^{\mathrm{T}} \boldsymbol{X}(s) + b} d\Lambda_{(2)}(s) \right\} \right] \right| \\
\leq \; & \left| \exp\left[-G\left\{ \Lambda_{(1)}(\tau) J(t, \boldsymbol{X}, b; \boldsymbol{\beta}_{(1)}, \Lambda_{(1)}) \right\} \right] - \exp\left[-G\left\{ M_\Lambda J(t, \boldsymbol{X}, b; \boldsymbol{\beta}_{(1)}, \Lambda_{(1)}) \right\} \right] \right| \\
& + \left| \exp\left[-G\left\{ M_\Lambda J(t, \boldsymbol{X}, b; \boldsymbol{\beta}_{(1)}, \Lambda_{(1)}) \right\} \right] - \exp\left[-G\left\{ \Lambda_{(2)}(\tau) J(t, \boldsymbol{X}, b; \boldsymbol{\beta}_{(2)}, \Lambda_{(2)}) \right\} \right] \right| \\
\leq \; & \max_{\boldsymbol{\beta} \in \mathcal{B}_1, \Lambda \in \mathcal{D}_\infty, \Lambda(\tau) \leq M_\Lambda} \left| K_G^{(1)} \left\{ \Lambda(\tau) J(t, \boldsymbol{X}, b; \boldsymbol{\beta}, \Lambda) \right\} \right| \left(e^{\widetilde{M} + |b|} |\Lambda_{(1)}(\tau) - M_\Lambda| \right) \\
& + 2 \exp\left\{ -G\left(M_\Lambda e^{-\widetilde{M} + |b|} \right) \right\}.
\end{aligned}
$$

Because there exist M_Λ / ϵ ϵ-brackets to cover $[0, M_\Lambda]$, the above results imply that there exist $O\left\{ \exp\left(\epsilon^{-2V/(V+2)} \right) \right\} \times M_\Lambda / \epsilon$ brackets

$$
\{ J(t, \boldsymbol{X}, b; \boldsymbol{\beta}_{(1)}, \Lambda_{(1)}), J(t, \boldsymbol{X}, b; \boldsymbol{\beta}_{(2)}, \Lambda_{(2)}) \} \times \{ \Lambda_{(1)}(\tau), \Lambda_{(2)}(\tau) \}
$$

such that

$$
\begin{aligned}
& \left| \exp\left[-G\left\{ \int_0^t e^{\boldsymbol{\beta}_{(1)}^{\mathrm{T}} \boldsymbol{X}(s) + b} d\Lambda_{(1)}(s) \right\} \right] - \exp\left[-G\left\{ \int_0^t e^{\boldsymbol{\beta}_{(2)}^{\mathrm{T}} \boldsymbol{X}(s) + b} d\Lambda_{(2)}(s) \right\} \right] \right| \\
\leq \; & \max_{\boldsymbol{\beta} \in \mathcal{B}_1, \Lambda \in \mathcal{D}_\infty, \Lambda(\tau) \leq M_\Lambda} \left| K_G^{(1)} \left\{ \Lambda(\tau) J(t, \boldsymbol{X}, b; \boldsymbol{\beta}, \Lambda) \right\} \right| \left(M_\Lambda + e^{\widetilde{M} + |b|} \right) \epsilon + 2 \exp\left\{ -G\left(e^{-\widetilde{M} + |b|} M_\Lambda \right) \right\}.
\end{aligned}
$$

By Condition 6,

$$
\max_{\boldsymbol{\beta} \in \mathcal{B}_1, \Lambda \in \mathcal{D}_\infty, \Lambda(\tau) \leq M_\Lambda} \left| K_G^{(1)} \left\{ \Lambda(\tau) J(t, \boldsymbol{X}, b; \boldsymbol{\beta}, \Lambda) \right\} \right| = O_P(1).
$$

It also follows from Condition 6 that

$$
\left| \exp\left[-G\left\{ \int_0^t e^{\boldsymbol{\beta}_{(1)}^{\mathrm{T}} \boldsymbol{X}(s) + b} d\Lambda_{(1)}(s) \right\} \right] - \exp\left[-G\left\{ \int_0^t e^{\boldsymbol{\beta}_{(2)}^{\mathrm{T}} \boldsymbol{X}(s) + b} d\Lambda_{(2)}(s) \right\} \right] \right|
$$

$$\leq O_P\left(M_\Lambda + e^{\widetilde{M}+|b|}\right)\epsilon + 2O_P(M_\Lambda^{-1/\rho}).$$

We choose $M_\Lambda = \epsilon^{-\rho/(\rho+1)}$ such that

$$\left|\exp\left[-G\left\{\int_0^t e^{\boldsymbol{\beta}_{(1)}^{\mathrm{T}}\boldsymbol{X}(s)+b}d\Lambda_{(1)}(s)\right\}\right] - \exp\left[-G\left\{\int_0^t e^{\boldsymbol{\beta}_{(2)}^{\mathrm{T}}\boldsymbol{X}(s)+b}d\Lambda_{(2)}(s)\right\}\right]\right|$$

$$\leq O_P\left(\epsilon^{1/(\rho+1)} + \epsilon\right)\left(1 + e^{\widetilde{M}+|b|}\right).$$

We redefine ϵ as $\epsilon^{\rho+1}$ such that there exist $O\left\{\epsilon^{-(2\rho+1)}\exp\left(\epsilon^{-2V(\rho+1)/(V+2)}\right)\right\}$ ϵ-brackets to cover $\{\exp[-G\{\int_0^t e^{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}(s)+b}d\Lambda(s)\}] : \boldsymbol{\beta} \in \mathcal{B}_1, \Lambda \in \mathcal{D}_\infty\}$ in $L_2(\mathbb{P})$.

For any $(\boldsymbol{\gamma}_{(1)}, A_{(1)})$ and $(\boldsymbol{\gamma}_{(2)}, A_{(2)})$ in $\mathcal{B}_2 \times \mathcal{D}_{M_A}$, we notice that for $l = 0, 1, 2$,

$$\begin{aligned}
&\left|K_H^{(l)}\left\{\int_0^Y e^{\boldsymbol{\gamma}_{(1)}{}^{\mathrm{T}}\boldsymbol{X}(s)+b}dA_{(1)}(s)\right\} - K_H^{(l)}\left\{\int_0^Y e^{\boldsymbol{\gamma}_{(2)}{}^{\mathrm{T}}\boldsymbol{X}(s)+b}dA_{(2)}(s)\right\}\right| \\
\leq\ & \sup_{x\geq 0}\left|K_H^{(l+1)}(x)\right|\left|\int_0^Y e^{\boldsymbol{\gamma}_{(1)}{}^{\mathrm{T}}\boldsymbol{X}(s)+b}dA_{(1)}(s) - \int_0^Y e^{\boldsymbol{\gamma}_{(2)}{}^{\mathrm{T}}\boldsymbol{X}(s)+b}dA_{(2)}(s)\right| \\
\leq\ & \sup_{x\geq 0}\left|K_H^{(l+1)}(x)\right|\left\{C^* e^{\widetilde{M}+|b|}\left\|\boldsymbol{\gamma}_{(1)} - \boldsymbol{\gamma}_{(2)}\right\| + \left|\int_0^Y e^{\boldsymbol{\gamma}_{(1)}{}^{\mathrm{T}}\boldsymbol{X}(s)+b}d\left(A_{(1)} - A_{(2)}\right)(s)\right|\right\} \\
\leq\ & \sup_{x\geq 0}\left|K_H^{(l+1)}(x)\right| e^{\widetilde{M}+|b|}\left\{C^*\left\|\boldsymbol{\gamma}_{(1)} - \boldsymbol{\gamma}_{(2)}\right\| + \left|A_{(1)}(Y) - A_{(2)}(Y)\right| + \int_0^\tau \left|A_{(1)}(s) - A_{(2)}(s)\right| ds\right\},
\end{aligned}$$

where the last inequality follows from the integration by parts. By Theorem 2.7.5 of van der Vaart and Wellner (1996), the bracketing number for $\mathcal{B}_2 \times \mathcal{D}_{M_A}$ is of order $O\{\exp(\epsilon^{-1})\}$. Thus, the bracketing number for $\widetilde{\mathcal{H}}_1$ is of order $O\{\exp(\epsilon^{-2V(\rho+1)/(V+2)} + \epsilon^{-1})\epsilon^{-(2\rho+1)}\}$. Therefore, the class $\widetilde{\mathcal{H}}_1$ is Glivenko-Cantelli.

To show that the class $\widetilde{\mathcal{H}}_2$ is Glivenko-Cantelli, we note that

$$\int_b K_1\left(b, \mathcal{O}; \boldsymbol{\beta}, \gamma, \mathcal{A}\right) K_2(t, b, \mathcal{O}; \boldsymbol{\gamma}, \mathcal{A})\phi(b; \sigma^2)db = I(Y \geq t)\int_b K_3\left(t, \mathcal{O}; \boldsymbol{\theta}, \mathcal{A}\right)\phi(b; \sigma^2)db,$$

where

$$\begin{aligned}
K_3(t, \mathcal{O}; \boldsymbol{\theta}, \mathcal{A}) \quad = \quad & K_1\left(b, \mathcal{O}; \boldsymbol{\beta}, \gamma, \mathcal{A}\right) e^{\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{X}(t)+b}\left[\Delta\frac{H''\left\{\int_0^Y e^{\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{X}(s)+b}dA(s)\right\}}{H'\left\{\int_0^Y e^{\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{X}(s)+b}dA(s)\right\}}\right. \\
& \left. - H'\left\{\int_0^Y e^{\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{X}(s)+b}dA(s)\right\}\right].
\end{aligned}$$

By the above arguments for $\widetilde{\mathcal{H}}_1$, the class $\{\int_b K_3\left(t, \mathcal{O}; \boldsymbol{\theta}, \mathcal{A}\right)\phi(b;\sigma^2)db : \boldsymbol{\theta} \in \Theta, t \in [0,\tau], \Lambda \in \mathcal{D}_\infty, A \in D_{M_A}\}$ is Glivenko-Cantelli. Because $I(Y \geq t)$ is Glivenko-Cantelli, $\widetilde{\mathcal{H}}_2$ is Glivenko-Cantelli by the preservation of the Glivenko-Cantelli property under the product. □

Lemma 3.2. *Under Conditions 1–6, the classes of functions*

$$\mathcal{H}_1 \equiv \left\{ \int_b K_1(b, \mathcal{O}; \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathcal{A})\phi(b;\sigma^2)db : \boldsymbol{\theta} \in \Theta, \Lambda \in \mathcal{D}_{M_\Lambda}, A \in \mathcal{D}_{M_A} \right\}$$

and

$$\mathcal{H}_2 \equiv \left\{ \int_b K_1(b, \mathcal{O}; \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathcal{A})K_2(t, b, \mathcal{O}; \boldsymbol{\gamma}, A)\phi(b;\sigma^2)db : \boldsymbol{\theta} \in \Theta, t \in [0,\tau], \Lambda \in \mathcal{D}_{M_\Lambda}, A \in \mathcal{D}_{M_A} \right\}$$

are $\mathbb{P}$-Donsker, where M_A and M_Λ are finite constants.

Proof. As in the proof of Lemma 3.1, for any $(\boldsymbol{\beta}_{(1)}, \Lambda_{(1)})$ and $(\boldsymbol{\beta}_{(2)}, \Lambda_{(2)})$ in $\mathcal{B}_1 \times \mathcal{D}_{M_\Lambda}$ and $t \in \mathcal{U}$, we have

$$\begin{aligned} &\left| K_G^{(l)}\left\{ \int_0^t e^{\boldsymbol{\beta}_{(1)}^{\mathrm{T}}\boldsymbol{X}(s)+b} d\Lambda_{(1)}(s) \right\} - K_G^{(l)}\left\{ \int_0^t e^{\boldsymbol{\beta}_{(2)}^{\mathrm{T}}\boldsymbol{X}(s)+b} d\Lambda_{(2)}(s) \right\} \right| \\ \leq\;& \sup_{x\geq 0}\left| K_G^{(l+1)}(x) \right| e^{\widetilde{M}+|b|}\left\{ C^*\left\| \boldsymbol{\beta}_{(1)} - \boldsymbol{\beta}_{(2)} \right\| + \left|\Lambda_{(1)}(t) - \Lambda_{(2)}(t)\right| + \int_0^\tau \left|\Lambda_{(1)}(s) - \Lambda_{(2)}(s)\right| ds \right\} \end{aligned}$$

for $l = 0, 1$. Thus,

$$\begin{aligned} &\left| K_1(b, \mathcal{O}; \boldsymbol{\beta}_{(1)}, \boldsymbol{\gamma}_{(1)}, , \mathcal{A}_{(1)}) - K_1(b, \mathcal{O}; \boldsymbol{\beta}_{(2)}, \boldsymbol{\gamma}_{(2)}, \mathcal{A}_{(2)}) \right| \\ \leq\;& \widetilde{C}e^{2\widetilde{M}+|b|}\left\{ \left\| \boldsymbol{\beta}_{(1)} - \boldsymbol{\beta}_{(2)} \right\| + \left\| \boldsymbol{\gamma}_{(1)} - \boldsymbol{\gamma}_{(2)} \right\| + \left|\Lambda_{(1)}(L) - \Lambda_{(2)}(L)\right| + \left|\Lambda_{(1)}(R) - \Lambda_{(2)}(R)\right| \right. \\ & \left. + \left|A_{(1)}(Y) - A_{(2)}(Y)\right| + \int_0^\tau \left|\Lambda_{(1)}(s) - \Lambda_{(2)}(s)\right| ds + \int_0^\tau \left|A_{(1)}(s) - A_{(2)}(s)\right| ds \right\}, \end{aligned}$$

where $\widetilde{C}$ is a constant. By similar arguments as in the proof of Lemma 3.1, the bracketing numbers for $\mathcal{H}_1$ and $\mathcal{H}_2$ are of order $O\{\exp(\epsilon^{-1})\}$. Thus, $\mathcal{H}_1$ and $\mathcal{H}_2$ are $\mathbb{P}$-Donsker. □

Lemma 3.3. *Under Conditions 1–6,*

$$E\left[\sum_{m=0}^{M} \left\{ \widehat{\Lambda}(U_m) - \Lambda_0(U_m) \right\}^2 + \left\{ \widehat{A}(Y) - A_0(Y) \right\}^2 \right]$$

$$= \quad O_P(n^{-2/3}) + O\left(\left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right\|^2 + \|\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\|^2 + \left\|\widehat{\sigma}^2 - \sigma_0^2\right\|^2\right).$$

Proof. By Theorem 1, $\widehat{\mathcal{A}}$ is consistent for $\mathcal{A}_0$. Thus, there exists a finite constant M_Λ such that $\widehat{\Lambda}(\tau) \le M_\Lambda$. By the Donsker results in Lemma 3.2, $m(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}})$ is then in a Donsker class. Note that

$$\int_0^\delta \sqrt{1 + \log N_{[]}(\epsilon, \mathcal{M}, L_2(\mathbb{P}))} d\epsilon \le O(\delta^{1/2}).$$

In addition, by Lemma 1.3 of van der Geer (2000) and the mean-value theorem,

$$\mathbb{P}\left\{m\left(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}}\right) - m\left(\boldsymbol{\theta}_0, \widetilde{\mathcal{A}}\right)\right\} \le -cH^2\left\{(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}}), (\boldsymbol{\theta}_0, \widetilde{\mathcal{A}})\right\},$$

where c is a positive constant, and $H\{(\theta, \mathcal{A}), (\theta_0, \widetilde{\mathcal{A}})\}$ is the Hellinger distance

$$\left(\int \left[\exp\left\{\frac{L(\theta, \mathcal{A})}{2}\right\} - \exp\left\{\frac{L(\theta_0, \widetilde{\mathcal{A}})}{2}\right\}\right]^2 d\mu\right)^{1/2}$$

with respect to the dominating measure μ. By Theorem 3.4.1 of van der Vaart and Wellner (1996), there exists r_n with $r_n^2\phi(1/r_n) \sim n^{1/2}$ such that $H\{(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}}), (\theta_0, \widetilde{\mathcal{A}})\} = O_P(1/r_n)$. In particular, we choose r_n in the order of $n^{1/3}$ such that $H\{(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}}), (\boldsymbol{\theta}_0, \widetilde{\mathcal{A}})\} = O_P(n^{-1/3})$.

By the mean-value theorem,

$$\begin{aligned}
& E\left(\left[\int_b \left\{\sum_{m=0}^M \delta_m \left(\exp\left[-G\left\{\int_0^{U_m} e^{\widehat{\boldsymbol{\beta}}^{\mathrm{T}} \boldsymbol{X}(s)+b} d\widehat{\Lambda}(s)\right\}\right] - \exp\left[-G\left\{\int_0^{U_{m+1}} e^{\widehat{\boldsymbol{\beta}}^{\mathrm{T}} \boldsymbol{X}(s)+b} d\widehat{\Lambda}(s)\right\}\right]\right)\right\}\right.\right. \\
& \left[e^{\widehat{\boldsymbol{\gamma}}^{\mathrm{T}} \boldsymbol{X}(Y)+b} \widehat{A}\{Y\} H'\left\{\int_0^Y e^{\widehat{\boldsymbol{\gamma}}^{\mathrm{T}} \boldsymbol{X}(s)+b} d\widehat{A}(s)\right\}\right]^\Delta \exp\left[-H\left\{\int_0^Y e^{\widehat{\boldsymbol{\gamma}}^{\mathrm{T}} \boldsymbol{X}(s)+b} d\widehat{A}(s)\right\}\right] \psi(b, \widehat{\sigma}^2) db \\
& - \int_b \left\{\sum_{m=0}^M \delta_m \left(\exp\left[-G\left\{\int_0^{U_m} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}(s)+b} d\Lambda_0(s)\right\}\right] - \exp\left[-G\left\{\int_0^{U_{m+1}} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}(s)+b} d\Lambda_0(s)\right\}\right]\right)\right\} \\
& \left.\left.\left[e^{\boldsymbol{\gamma}_0^{\mathrm{T}} \boldsymbol{X}(Y)+b} \widetilde{A}\{Y\} H'\left\{\int_0^Y e^{\boldsymbol{\gamma}_0^{\mathrm{T}} \boldsymbol{X}(s)+b} d\widetilde{A}(s)\right\}\right]^\Delta \exp\left[-H\left\{\int_0^Y e^{\boldsymbol{\gamma}_0^{\mathrm{T}} \boldsymbol{X}(s)+b} d\widetilde{A}(s)\right\}\right] \psi(b, \sigma_0^2) db \right\}^2\right] \\
= \quad & O_P(n^{-2/3}).
\end{aligned}$$

Consequently, using the mean-value theorem again, we have

$$
\begin{aligned}
& O_P(n^{-2/3}) + O(1)\left\|\widehat{\sigma}^2-\sigma_0^2\right\|^2 + O(1)\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right\|^2 \\
\geq\ & E\left\{\left(\int_b\left[\left\{\sum_{m=0}^{M}\delta_m\left(\exp\left[-G\left\{\int_0^{U_m} e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}(s)+b}d\widehat{\Lambda}(s)\right\}\right]-\exp\left[-G\left\{\int_0^{U_{m+1}} e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}(s)+b}d\widehat{\Lambda}(s)\right\}\right]\right)\right\}\right.\right.\right. \\
& \left[e^{\boldsymbol{\gamma}_0^{\mathrm{T}}\boldsymbol{X}(Y)+b}\widehat{A}\{Y\}H'\left\{\int_0^{Y} e^{\boldsymbol{\gamma}_0^{\mathrm{T}}\boldsymbol{X}(s)+b}d\widehat{A}(s)\right\}\right]^{\Delta}\exp\left[-H\left\{\int_0^{Y} e^{\boldsymbol{\gamma}_0^{\mathrm{T}}\boldsymbol{X}(s)+b}d\widehat{A}(s)\right\}\right] \\
& -\int_b\left\{\sum_{m=0}^{M}\delta_m\left(\exp\left[-G\left\{\int_0^{U_m} e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}(s)+b}d\Lambda_0(s)\right\}\right]-\exp\left[-G\left\{\int_0^{U_{m+1}} e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}(s)+b}d\Lambda_0(s)\right\}\right]\right)\right\} \\
& \left.\left.\left.\left[e^{\boldsymbol{\gamma}_0^{\mathrm{T}}\boldsymbol{X}(Y)+b}\widetilde{A}\{Y\}H'\left\{\int_0^{Y} e^{\boldsymbol{\gamma}_0^{\mathrm{T}}\boldsymbol{X}(s)+b}d\widetilde{A}(s)\right\}\right]^{\Delta}\exp\left[-H\left\{\int_0^{Y} e^{\boldsymbol{\gamma}_0^{\mathrm{T}}\boldsymbol{X}(s)+b}d\widetilde{A}(s)\right\}\right]\right\}\psi(b,\sigma_0^2)db\right]^2\right) \\
\geq\ & c_0 E\left(\left[\int_b K_1(b,\mathcal{O};\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0,\mathcal{A}_0)\left\{\sum_{m=0}^{M}\delta_m\int_0^{\tau} Q_1(t,U_m,U_{m+1},b,\mathcal{O};\boldsymbol{\beta}_0,\Lambda_0)d(\widehat{\Lambda}-\Lambda_0)(t)\right.\right.\right. \\
& \left.\left.\left.+\Delta\left(\widehat{A}\{Y\}-\widetilde{A}\{Y\}\right)+\int_0^{\tau} Q_2(t,Y,b,\mathcal{O};\boldsymbol{\gamma}_0,A_0)d(\widehat{A}-\widetilde{A})(t)\right\}\phi(b;\sigma_0^2)db\right]^2\right)
\end{aligned}
$$

for some positive constant c_0. We define a norm in $\mathcal{V}\equiv BV(\mathcal{U})\times BV[0,\tau]$ such that for any $\boldsymbol{f}\equiv(f_1,f_2)^{\mathrm{T}}\in\mathcal{V}$,

$$
\|\boldsymbol{f}\|_1=\left[E\left\{\sum_{m=0}^{M} f_1(U_m)^2+f_2(Y)^2\right\}\right]^{1/2}.
$$

In addition, we define a seminorm

$$
\begin{aligned}
\|\boldsymbol{f}\|_2 \quad = \quad & E\left(\left[\int_b K_1(b,\mathcal{O};\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0,\mathcal{A}_0)\left\{\sum_{m=0}^{M}\delta_m\int_0^{\tau} Q_1(t,U_m,U_{m+1},b,\mathcal{O};\boldsymbol{\beta}_0,\Lambda_0)df_1(t)\right.\right.\right. \\
& \left.\left.\left.+\Delta f_2(Y)+\int_0^{\tau} Q_2(t,Y,b,\mathcal{O};\boldsymbol{\gamma}_0,A_0)df_2(t)\right\}\phi(b;\sigma_0^2)db\right]^2\right)^{1/2}.
\end{aligned}
$$

Note that if $\|\boldsymbol{f}\|_2=0$ for some $\boldsymbol{f}\in\mathcal{V}$, then

$$
\begin{aligned}
& \int_b K_1(b,\mathcal{O};\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0,\mathcal{A}_0)\left\{\sum_{m=0}^{M}\delta_m\int_0^{\tau} Q_1(t,U_m,U_{m+1},b,\mathcal{O};\boldsymbol{\beta}_0,\Lambda_0)df_1(t)\right. \\
& \qquad \left.+\Delta f_2(Y)+\int_0^{\tau} Q_2(t,Y,b,\mathcal{O};\boldsymbol{\gamma}_0,A_0)df_2(t)\right\}\phi(b;\sigma_0^2)db=0 \qquad (3.8)
\end{aligned}
$$

with probability 1.

For $\Delta=0$, we set $Y=\tau$ in (4.11) to obtain an equation. For $\Delta=1$, we integrate Y from 0 to τ

in (4.11) to obtain another equation. We add the two equations to obtain

$$\int_b \left\{ \sum_{m=0}^{M} \delta_m \left(\exp\left[-G\left\{ \int_0^{U_m} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}(s)+b} d\Lambda_0(s) \right\} \right] - \exp\left[-G\left\{ \int_0^{U_{m+1}} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}(s)+b} d\Lambda_0(s) \right\} \right] \right) \right\}$$

$$\times \left\{ \sum_{m=0}^{M} \delta_m \int_0^{\tau} Q_1(t, U_m, U_{m+1}, b, \mathcal{O}; \boldsymbol{\beta}_0, \Lambda_0) df_1(t) \right\} \phi(b; \sigma_0^2) db = 0.$$

For any $m \in \{0, \ldots, M\}$, we sum over all possible $\delta_{m'}$ with $m' = m, \ldots, M$ to obtain

$$\int_b \left(G' \left\{ \int_0^{U_m} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}(t)+b} d\Lambda_0(t) \right\} \exp\left[-G\left\{ \int_0^{U_m} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}(s)+b} d\Lambda_0(s) \right\} \right] \right.$$

$$\left. \times \left\{ \int_0^{U_m} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}(t)+b} df_1(t) \right\} \right) \psi(b, \sigma_0^2) db = 0.$$

Because m is arbitrary, we can replace U_m with $t \in \mathcal{U}$. By Condition 8,

$$G' \left\{ \int_0^{t} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}(s)+b} d\Lambda_0(s) \right\} \left\{ \int_0^{t} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}(s)+b} df_1(s) \right\} = 0$$

with probability one. The term $G' \left\{ \int_0^{U_m} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}(t)+b} d\Lambda_0(t) \right\}$ is strictly greater than zero. Therefore, $f_1 = 0$. In addition, we sum over (4.11) with all possible δ_m for $m = 0, \ldots, M$ to obtain

$$\int_b \left[e^{\boldsymbol{\gamma}_0^{\mathrm{T}} \boldsymbol{X}(Y)+b} A_0\{Y\} H' \left\{ \int_0^{Y} e^{\boldsymbol{\gamma}_0^{\mathrm{T}} \boldsymbol{X}(s)+b} dA_0(s) \right\} \right]^{\Delta} \exp\left[-H\left\{ \int_0^{Y} e^{\boldsymbol{\gamma}_0^{\mathrm{T}} \boldsymbol{X}(s)+b} dA_0(s) \right\} \right]$$

$$\times \left\{ \int_0^{\tau} Q_2(t, Y, b, \mathcal{O}; \boldsymbol{\gamma}_0, A_0) df_2(t) \right\} \phi(b; \sigma_0^2) db = 0.$$

We let $\Delta = 0$ and integrate Y from 0 to t to obtain

$$\int_b \left(H' \left\{ \int_0^{t} e^{\boldsymbol{\gamma}_0^{\mathrm{T}} \boldsymbol{X}(s)+b} dA_0(s) \right\} \exp\left[-H\left\{ \int_0^{t} e^{\boldsymbol{\gamma}_0^{\mathrm{T}} \boldsymbol{X}(s)+b} dA_0(s) \right\} \right] \right.$$

$$\left. \times \left\{ \int_0^{t} e^{\boldsymbol{\gamma}_0^{\mathrm{T}} \boldsymbol{X}(s)+b} df_2(s) \right\} \right) \psi(b, \sigma_0^2) db = 0.$$

By Condition 8,

$$H' \left\{ \int_0^{t} e^{\boldsymbol{\gamma}_0^{\mathrm{T}} \boldsymbol{X}(s)+b} dA_0(s) \right\} \left\{ \int_0^{t} e^{\boldsymbol{\gamma}_0^{\mathrm{T}} \boldsymbol{X}(s)+b} df_2(s) \right\} = 0$$

with probability one. The term $H'\left\{\int_0^{U_m} e^{\boldsymbol{\gamma}_0^{\mathrm{T}}\boldsymbol{X}(t)+b} dA_0(t)\right\}$ is strictly greater than zero. Therefore, we obtain $f_2 = 0$, implying that $\|\cdot\|_2$ is a norm in $\mathcal{V}$.

By the Cauchy-Schwarz inequality, for any $\boldsymbol{f} \in \mathcal{V}$,

$$\begin{aligned}
\|\boldsymbol{f}\|_2 \leq & \left(E\left[\int_b K_1(b, \mathcal{O}; \boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, \mathcal{A}_0)\left\{\sum_{m=0}^{M} \delta_m \int_0^{\tau} Q_1(t, U_m, U_{m+1}, b, \mathcal{O}; \boldsymbol{\beta}_0, \Lambda_0) d(t)\right.\right.\right. \\
& \left.\left. + \int_0^{\tau} Q_2(t, Y, b, \mathcal{O}; \boldsymbol{\gamma}_0, A_0) d(t)\right\} \phi(b; \sigma_0^2) db\right]^2 E\left.\left\{\sum_{m=0}^{M} f_1(U_m)^2 + f_2(Y)^2\right\}\right)^{1/2} \\
\leq & \ c_1 \|\boldsymbol{f}\|_1,
\end{aligned}$$

where c_1 is a finite constant. By the bounded inverse theorem in the Banach space, we have $\|\boldsymbol{f}\|_2 \geq c_1' \|\boldsymbol{f}\|_1$ for some constant c_1'. Therefore,

$$O_P(n^{-2/3}) + O\left(\left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right\|^2 + \|\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\|^2 + \left\|\widehat{\sigma}^2 - \sigma_0^2\right\|^2\right)$$

$$\geq c_0 c_1' E\left[\sum_{m=0}^{M}\left\{\widehat{\Lambda}(U_m) - \Lambda_0(U_m)\right\}^2 + \left\{\widehat{A}(Y) - A_0(Y)\right\}^2\right].$$

The lemma thus holds. $\square$

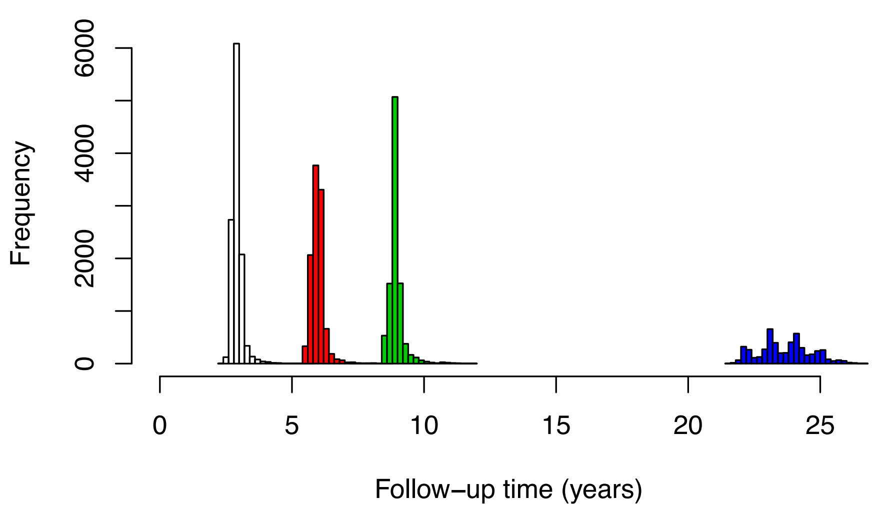

Figure 3.3: Frequency of examinations over follow-up time. The white, red, green, and blue histograms pertain, respectively, to the second, third, fourth, and fifth examinations.

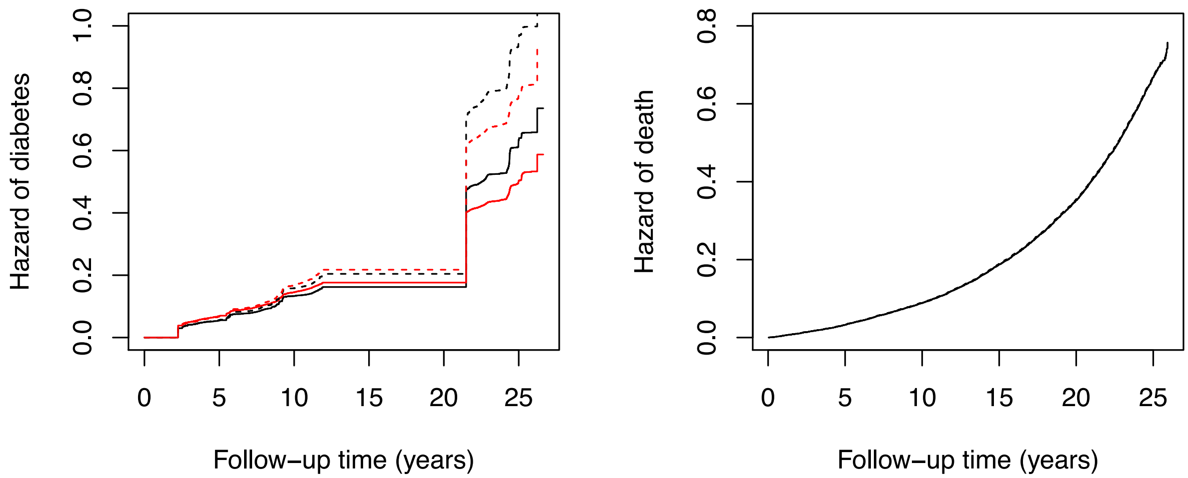

Figure 3.4: Estimation of baseline cumulative hazard functions, where baseline is defined for an African-American male residing in Forsyth County, NC, aged 54 years, body mass index 27 kg/m^2, glucose value 98 mg/dl, systolic blood pressure 118 mmHg, and diastolic blood pressure 73 mmHg. The red solid and dashed curves pertain to the proportional hazards and proportional odds models, respectively, from the naive method. The black solid and dashed curves pertain to the proportional hazards and proportional odds models, respectively, from the proposed method, where the dropout time is modeled by the proportional hazards model.

CHAPTER 4: SEMIPARAMETRIC REGRESSION ANALYSIS OF MULTIPLE RIGHT- AND INTERVAL-CENSORED EVENTS

4.1 Introduction

Many clinical and epidemiological studies are concerned with multiple diseases, which may be symptomatic or asymptomatic. Time to the development of a symptomatic disease is right-censored if the disease does not occur during the follow-up, whereas time to the development of an asymptomatic disease is typically interval-censored because the disease occurrence can only be monitored periodically using biomarkers. In the Atherosclerosis Risk in Communities (ARIC) study (The ARIC Investigators, 1989), for instance, subjects were followed for up to 27 years for symptomatic cardiovascular diseases, such as myocardial infarction (MI) and stroke, through reviews of hospital records; they were also examined over five clinic visits, with the first four at approximately 3-year intervals, for occurrences of asymptomatic diseases, such as diabetes and hypertension.

There is a large body of literature on right-censored data (Kalbfleisch and Prentice, 1980) and also a growing body of literature on interval-censored data (Huang, 1996; Huang and Wellner, 1997; Zhang, Sun, Zhao and Sun, 2005; Chang et al., 2007; Wen and Chen, 2013; Chen, Chen, Lin and Tong, 2014; Zeng et al., 2016). However, the existing literature has treated right-censored and interval-censored data separately. Joint modelling of the two types of data would allow investigators to evaluate the effects of covariates on both types of events and to predict the occurrence of a symptomatic disease given the history of asymptomatic diseases.

In this paper, we relate potentially time-dependent covariates to the joint distribution of multiple right- and interval-censored events through semiparametric proportional hazards models with random effects. Specifically, we assume a shared random effect for the interval-censored events, which affects the right-censored events with unknown coefficients. We assume an additional shared random effect for the right-censored events to capture their own dependence. The proposed models are reminiscent of selection models for joint modeling of survival and longitudinal data (Hogan and Laird, 1997).

We estimate the model parameters through nonparametric maximum likelihood estimation, under which the baseline hazard functions are completely nonparametric. We develop a simple EM algorithm that converges stably for arbitrary sample sizes, even with time-dependent covariates. We show that the resulting estimators are consistent and the parametric components are asymptotically normal and asymptotically efficient. We also show that the covariance matrix of the parametric components can be estimated consistently with profile likelihood or nonparametric bootstrap. We pay special attention to the estimation of the conditional cumulative incidence function, which can be used to predict disease occurrence dynamically by updating the event history. Finally, we assess the performance of the proposed numerical and inferential procedures through extensive simulation studies and provide a substantive application to the ARIC data on diabetes, hypertension, stroke, MI, and death.

4.2 Methods

4.2.1 Data, Models, and Likelihood

Suppose that there are K_1 asymptomatic events occurring at times $T_1, \ldots, T_{K_1}$ and K_2 symptomatic events occurring at times $T_{K_1+1}, \ldots, T_K$, where $K = K_1 + K_2$. Let $\boldsymbol{X}_k(\cdot)$ be a p-vector of possibly time-dependent external covariates for the event time T_k. For $k = 1, \ldots, K_1$, the hazard function of T_k conditional on covariate $\boldsymbol{X}_k$ and random effect b_1 is given by

$$\lambda_k(t; \boldsymbol{X}_k, b_1) = e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_k(t) + b_1} \lambda_k(t), \tag{4.1}$$

where $\boldsymbol{\beta}$ is a set of unknown regression parameters, $\lambda_k(\cdot)$ is an arbitrary baseline hazard function, and b_1 is a latent normal random variable with mean zero and variance σ_1^2. For $k = K_1 + 1, \ldots, K$, the hazard function of T_k conditional on covariates $\boldsymbol{X}_k$ and random effects b_1 and b_2 is given by

$$\lambda_k(t; \boldsymbol{X}_k, b_1, b_2) = e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_k(t) + \gamma_k b_1 + b_2} \lambda_k(t), \tag{4.2}$$

where $\lambda_k(\cdot)$ is an arbitrary baseline hazard function, $\boldsymbol{\gamma} \equiv (\gamma_{K_1+1}, \ldots, \gamma_K)^{\mathrm{T}}$ is a set of unknown coefficients, and b_2 is a latent normal random variable with mean zero and variance σ_2^2. Write $\boldsymbol{\Sigma} = (\sigma_1^2, \sigma_2^2)$. By letting $\boldsymbol{X}_k$ depend on k, models (4.1) and (4.2) allow the regression parameters to be different among the K events by appropriate definitions of dummy variables; see Lin (1994).

We implicitly assume that K_1 and K_2 are greater than one; otherwise, some of the parameters need to be fixed to ensure identifiability. For example, if $K_1 = K_2 = 1$, we require $\sigma_2^2 = 0$ and $\gamma_1 = 1$; if $K_1 > 1$ and $K_2 = 1$, we require $\sigma_2^2 = 0$; and if $K_1 = 1$ and $K_2 > 1$, we require one of the γ_k's to be 1.

Remark 4.1. *The random effects b_1 and b_2 characterize the underlying health conditions for the asymptomatic and symptomatic events, respectively. The random effect for the asymptomatic events affects the kth symptomatic event through the unknown coefficient γ_k. For example, in the ARIC study, b_1 represents the common pathways for diabetes and hypertension, such as obesity, inflammation, oxidative stress, and insulin resistance, which also serve as potential risk factors for MI, stroke, and death. The random effect b_2 represents the underlying propensity for major cardiovascular diseases and death.*

Suppose that the asymptomatic event time T_k $(k = 1, \ldots, K_1)$ is monitored at a sequence of positive time points $U_{k1} < \cdots < U_{k,M_k}$ and is known to lie in the interval $(L_k, R_k]$, where $L_k = \max\{U_{kl} : U_{kl} < T_k, l = 0, \ldots, M_k\}$, and $R_k = \min\{U_{kl} : U_{kl} \geq T_k, l = 1, \ldots, M_k + 1\}$, with $U_{k0} = 0$ and $U_{k,M_k+1} = \infty$. Let C_k denote the censoring time on the symptomatic event time T_k $(k = K_1 + 1, \ldots, K)$ such that we observe $Y_k = \min(T_k, C_k)$ and $\Delta_k = I(T_k \leq C_k)$, where $I(\cdot)$ is the indicator function. For a random sample of n subjects, the data consist of $\{\mathcal{O}_i : i = 1, \ldots, n\}$, where

$$\mathcal{O}_i = \{L_{ik}, R_{ik}, \boldsymbol{X}_{ik}(\cdot) : k = 1, \ldots, K_1\} \cup \{Y_{ik}, \Delta_{ik}, \boldsymbol{X}_{ik}(\cdot) : k = K_1 + 1, \ldots, K\}.$$

We assume that $\{U_{ikl} : k = 1, \ldots, K_1; l = 1, \ldots, M_{ik}\}$ and $\{C_{ik} : k = K_1 + 1, \ldots, K\}$ are independent of $\{T_{ik} : k = 1, \ldots, K\}$ and $\boldsymbol{b}_i \equiv (b_{i1}, b_{i2})$ conditional on $\{\boldsymbol{X}_{ik}(\cdot) : k = 1, \ldots, K\}$. Then, the likelihood concerning the parameters $\boldsymbol{\theta} \equiv (\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\Sigma})$ and $\mathcal{A} \equiv (\Lambda_1, \ldots, \Lambda_K)$ is

$$\prod_{i=1}^{n} \int_{\boldsymbol{b}_i} \prod_{k=1}^{K_1} \left[\exp\left\{ -\int_0^{L_{ik}} e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_{ik}(s) + b_{i1}} d\Lambda_k(s) \right\} - \exp\left\{ -\int_0^{R_{ik}} e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_{ik}(s) + b_{i1}} d\Lambda_k(s) \right\} \right]$$

$$\times \prod_{k=K_1+1}^{K} \left[\left\{ e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_{ik}(Y_{ik}) + \gamma_k b_{i1} + b_{i2}} \lambda_k(Y_{ik}) \right\}^{\Delta_{ik}} \exp\left\{ -\int_0^{Y_{ik}} e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_{ik}(s) + \gamma_k b_{i1} + b_{i2}} d\Lambda_k(s) \right\} \right] \psi(\boldsymbol{b}_i; \boldsymbol{\Sigma}) d\boldsymbol{b}_i,$$

where $\psi(\boldsymbol{b}_i; \boldsymbol{\Sigma}) = \prod_{j=1}^2 \phi(b_{ij}; \sigma_j^2)$, $\phi(b_{ij}; \sigma_j^2) = (2\pi\sigma_j^2)^{-1/2} \exp\{-b_{ij}^2/(2\sigma_j^2)\}$, $\Lambda_k(t) = \int_0^t \lambda_k(s) ds$, and $\exp\{-\int_0^\infty e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_{ik}(s) + b_{i1}} d\Lambda_k(s)\} = 0$.

4.2.2 Estimation Procedure

We adopt the nonparametric maximum likelihood estimation approach. For $k = 1, \ldots, K_1$, let $0 = t_{k0} < t_{k1} < t_{k2} < \cdots < t_{k,m_k} < \infty$ be the ordered sequence of all L_{ik} and R_{ik} with $R_{ik} < \infty$. For $k = K_1 + 1, \ldots, K$, let $0 = t_{k0} < t_{k1} < t_{k2} < \cdots < t_{k,m_k} < \infty$ be the ordered sequence of all Y_{ik} with $\Delta_{ik} = 1$. The estimator for Λ_k ($k = 1, \ldots, K$) is a step function that jumps only at $t_{k1}, \ldots, t_{k,m_k}$ with respective jump sizes $\boldsymbol{\lambda}_k \equiv (\lambda_{k1}, \ldots, \lambda_{k,m_k})$. We maximize the objective function

$$L_n(\boldsymbol{\theta}, \mathcal{A}) = \prod_{i=1}^{n} \int_{\boldsymbol{b}_i} \left\{ \prod_{k=1}^{K_1} g_{ik}^{(1)}(b_{i1}; \boldsymbol{\beta}, \boldsymbol{\lambda}_k) \right\} \left\{ \prod_{k=K_1+1}^{K} g_{ik}^{(2)}(\boldsymbol{b}_i; \boldsymbol{\beta}, \boldsymbol{\lambda}_k) \right\} \psi(\boldsymbol{b}_i; \boldsymbol{\Sigma}) d\boldsymbol{b}_i,$$

over $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_K$, where

$$g_{ik}^{(1)}(b_{i1}; \boldsymbol{\beta}, \boldsymbol{\lambda}_k) = \exp\left(- \sum_{t_{kl} \le L_{ik}} e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_{ikl} + b_{i1}} \lambda_{kl} \right) - I(R_{ik} < \infty) \exp\left(- \sum_{t_{kl} \le R_{ik}} e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_{ikl} + b_{i1}} \lambda_{kl} \right),$$

$$g_{ik}^{(2)}(\boldsymbol{b}_i; \boldsymbol{\beta}, \boldsymbol{\lambda}_k) = \left[\Lambda_k\{Y_{ik}\} e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_{ik}(Y_{ik}) + \gamma_k b_{i1} + b_{i2}} \right]^{\Delta_{ik}} \exp\left(- \sum_{t_{kl} \le Y_{ik}} e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_{ikl} + \gamma_k b_{i1} + b_{i2}} \lambda_{kl} \right),$$

$\boldsymbol{X}_{ikl} = \boldsymbol{X}_{ik}(t_{kl})$ for $k = 1, \ldots, K$ and $l = 1, \ldots, m_k$, and $\Lambda_k\{Y_{ik}\}$ is the jump size of Λ_k at Y_{ik}.

Direct maximization of the objective function is difficult due to the lack of analytical expressions for $\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_K$. We introduce latent Poisson random variables to form a likelihood equivalent to the objective function such that the maximum likelihood estimators can be easily obtained via a simple EM algorithm. For $k = 1, \ldots, K_1$, we denote $R_{ik}^* = I(R_{ik} = \infty) L_{ik} + I(R_{ik} < \infty) R_{ik}$ and introduce independent Poisson random variables W_{ikl} ($l = 1, \ldots, m_k, t_{kl} \le R_{ik}^*$) with means $\lambda_{kl} \exp(\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_{ikl} + b_{i1})$. Conditional on b_{i1}, the likelihood function of $\{W_{ikl}; l = 1, \ldots, m_k, t_{kl} \le R_{ik}^*\}$ is

$$\prod_{l=1, t_{kl} \le R_{ik}^*}^{m_k} \left\{ \frac{1}{W_{ikl}!} \left(\lambda_{kl} e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_{ikl} + b_{i1}} \right)^{W_{ikl}} \exp\left(-\lambda_{kl} e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_{ikl} + b_{i1}} \right) \right\}.$$

Let $A_{ik} = \sum_{t_{kl} \le L_{ik}} W_{ikl}$ and $B_{ik} = I(R_{ik} < \infty) \sum_{L_{ik} < t_{kl} \le R_{ik}} W_{ikl}$. The observed-data likelihood for $A_{ik} = 0$ and $B_{ik} > 0$ given b_{i1} is equal to

$$\exp\left(-\sum_{t_{kl}\le L_{ik}} e^{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_{ikl}+b_{i1}}\lambda_{kl}\right) - I(R_{ik}<\infty)\exp\left(-\sum_{t_{kl}\le R_{ik}} e^{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_{ikl}+b_{i1}}\lambda_{kl}\right),$$

which is the same as $g_{ik}^{(1)}(b_{i1};\boldsymbol{\beta},\boldsymbol{\lambda}_k)$. Therefore, the objective function $L_n(\boldsymbol{\theta},\mathcal{A})$ can be viewed as the observed-data likelihood for $\{A_{ik}=0, B_{ik}>0 : i=1,\ldots,n; k=1,\ldots,K_1\}\cup\{Y_{ik},\Delta_{ik} : i = 1,\ldots,n; k=K_1+1,\ldots,K\}$ with $(W_{ikl},\boldsymbol{b}_i)$ $(i=1,\ldots,n; k=1,\ldots,K_1; l=1,\ldots,m_k, t_{kl}\le R^*_{ik})$ as latent variables. In view of the foregoing results, we propose an EM algorithm treating W_{ikl} and $\boldsymbol{b}_i$ as missing data.

In the M-step, we maximize the conditional expectation of the complete-data log-likelihood given the observed data so as to update the parameters. Specifically, we update $\boldsymbol{\beta}$ by solving the equation

$$\sum_{i=1}^{n}\left\{\sum_{k=1}^{K_1}\sum_{l=1}^{m_k}\widehat{E}\left(W_{ikl}\right)I(t_{kl}\le R^*_{ik})\left[\boldsymbol{X}_{ikl}-\frac{\sum_{j=1}^{n}\boldsymbol{X}_{jkl}I(t_{kl}\le R^*_{jk})\widehat{E}\left\{\exp\left(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_{jkl}+b_{j1}\right)\right\}}{\sum_{j=1}^{n}I(t_{kl}\le R^*_{jk})\widehat{E}\left\{\exp\left(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_{jkl}+b_{j1}\right)\right\}}\right]\right.$$
$$\left.+\sum_{k=K_1+1}^{K}\Delta_{ik}\left(\boldsymbol{X}_{ik}(Y_{ik})-\frac{\sum_{j=1}^{n}I(Y_{jk}\ge Y_{ik})\boldsymbol{X}_{jk}(Y_{ik})\widehat{E}\left[\exp\left\{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_{jk}(Y_{ik})+\gamma_k b_{j1}+b_{j2})\right\}\right]}{\sum_{j=1}^{n}I(Y_{jk}\ge Y_{ik})\widehat{E}\left[\exp\left\{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_{jk}(Y_{ik})+\gamma_k b_{j1}+b_{j2}\right\}\right]}\right)\right\}$$
$$=\mathbf{0},$$

where $\widehat{E}(\cdot)$ denotes the conditional expectation given the observed data $\widetilde{\mathcal{O}}_i$ $(i=1,\ldots,n)$, with $\widetilde{\mathcal{O}}_i=\{A_{ik}=0,B_{ik}>0,\boldsymbol{X}_{ik}(\cdot):k=1,\ldots,K_1\}\cup\{Y_{ik},\Delta_{ik},\boldsymbol{X}_{ik}(\cdot):k=K_1+1,\ldots,K\}$. We update γ_k by solving the equation

$$\sum_{i=1}^{n}\Delta_{ik}\left(\widehat{E}\left(b_{i1}\right)-\frac{\sum_{j=1}^{n}I(Y_{jk}\ge Y_{ik})\widehat{E}\left[b_{j1}\exp\left\{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_{jk}(Y_{ik})+\gamma_k b_{j1}+b_{j2})\right\}\right]}{\sum_{j=1}^{n}I(Y_{jk}\ge Y_{ik})\widehat{E}\left[\exp\left\{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_{jk}(Y_{ik})+\gamma_k b_{j1}+b_{j2}\right\}\right]}\right)=0.$$

We update $\boldsymbol{\lambda}_k$ using

$$\lambda_{kl}=\frac{\sum_{i=1}^{n}I(t_{kl}\le R^*_{ik})\widehat{E}\left(W_{ikl}\right)}{\sum_{i=1}^{n}I(t_{kl}\le R^*_{ik})\widehat{E}\left\{\exp\left(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_{ikl}+b_{i1}\right)\right\}}$$

for $k=1,\ldots,K_1$ and $l=1,\ldots,m_k$ and

$$\lambda_{kl}=\frac{\sum_{i=1}^{n}\Delta_{ik}I(Y_{ik}=t_{kl})}{\sum_{i=1}^{n}I(Y_{ik}\ge t_{kl})\widehat{E}\left\{\exp\left(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_{ikl}+\gamma_k b_{i1}+b_{i2}\right)\right\}}$$

for $k = K_1+1, \ldots, K$ and $l = 1, \ldots, m_k$. Finally, we update σ_j^2 by $\sigma_j^2 = \sum_{i=1}^n \widehat{E}(b_{ij}^2)/n$ for $j = 1, 2$.

In the E-step, we evaluate the conditional expectation of W_{ikl} ($k = 1, \ldots, K_1; l = 1, \ldots, m_k, t_{kl} \leq R_{ik}^*$) and the other terms of $\boldsymbol{b}_i$ given the observed data $\widetilde{\mathcal{O}}_i$ for $i = 1, \ldots, n$. Specifically, the conditional expectation of W_{ikl} ($k = 1, \ldots, K_1; l = 1, \ldots, m_k, t_{kl} \leq R_{ik}^*$) given $\widetilde{\mathcal{O}}_i$ and $\boldsymbol{b}_i$ is

$$I(L_{ik} < t_{kl} \leq R_{ik} < \infty) \frac{\lambda_{kl} \exp\left(\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_{ikl} + b_{i1}\right)}{1 - \exp\left(-\sum_{L_{ik} < t_{kl'} \leq R_{ik}} \lambda_{kl'} e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_{ikl'} + b_{i1}}\right)}.$$

Note that the density of $\boldsymbol{b}_i$ given $\widetilde{\mathcal{O}}_i$ is proportional to $\{\prod_{k=1}^{K_1} g_{ik}^{(1)}(b_{i1}; \boldsymbol{\beta}, \boldsymbol{\lambda}_k)\}$ $\times \{\prod_{k=K_1+1}^{K} g_{ik}^{(2)}(\boldsymbol{b}_i; \boldsymbol{\beta}, \boldsymbol{\lambda}_k)\} \psi(\boldsymbol{b}_i; \boldsymbol{\Sigma})$. We evaluate the conditional expectation of W_{ikl} and the other terms through numerical integration over $\boldsymbol{b}_i$ with Gauss-Hermite quadratures.

We iterate between the E-step and M-step until convergence. In the M-step, the high-dimensional nuisance parameters λ_{kl} ($k = 1, \ldots, K; l = 1, \ldots, m_k$) are calculated explicitly, such that inversion of high-dimensional matrices is avoided. We denote the final estimators for $\boldsymbol{\theta}$ and $\mathcal{A}$ as $\widehat{\boldsymbol{\theta}} \equiv (\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\Sigma}})$ and $\widehat{\mathcal{A}} \equiv (\widehat{\Lambda}_1, \ldots, \widehat{\Lambda}_K)$.

4.2.3 Asymptotic Theory

We establish the asymptotic properties of $(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}})$ under the following regularity conditions.

Condition 1. The true value of $\boldsymbol{\theta}$, denoted by $\boldsymbol{\theta}_0 \equiv (\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, \boldsymbol{\Sigma}_0)$, belongs to the interior of a known compact set $\Theta \equiv \mathcal{B} \times \mathcal{G} \times \mathcal{S}$, where $\mathcal{B} \subset \mathbb{R}^p$, $\mathcal{G} \subset \mathbb{R}^{K_2}$, and $\mathcal{S} \subset (0, \infty) \times (0, \infty)$.

Condition 2. The true value $\Lambda_{k0}(\cdot)$ of $\Lambda_k(\cdot)$ is strictly increasing and continuously differentiable with $\Lambda_{k0}(0) = 0$.

Condition 3. For $k = 1, \ldots, K_1$, the monitoring times have finite support $\mathcal{U}_k$ with the least upper bound τ_k. The number of potential monitoring times M_k is positive with $E(M_k) < \infty$. There exists a positive constant η such that $\Pr\{\min_{1 \leq k \leq K_1, 0 \leq m < M_k}(U_{k,m+1} - U_{km}) \geq \eta | M_k, \boldsymbol{X}_k\} = 1$. In addition, there exists a probability measure μ_k in $\mathcal{U}_k$ such that the bivariate distribution function of $(U_{km}, U_{k,m+1})$ conditional on $(M_k, \boldsymbol{X}_k)$ is dominated by $\mu_k \times \mu_k$ and its Radon-Nikodym derivative, denoted by $\widetilde{f}_{km}(u, v; M_k, \boldsymbol{X}_k)$, can be expanded to a positive and twice-continuously differentiable function in the set $\{(u, v) : 0 \leq u \leq \tau_k, 0 \leq v \leq \tau_k, v - u \geq \eta\}$.

Condition 4. For $k = K_1+1, \ldots, K$, let τ_k denote the study duration time and $\mathcal{U}_k = [0, \tau_k]$. There exists a positive constant δ such that $\Pr(C_k \geq \tau_k | \boldsymbol{X}_k) = \Pr(C_k = \tau_k | \boldsymbol{X}_k) \geq \delta$ almost surely.

Condition 5. With probability 1, $\boldsymbol{X}_k(\cdot)$ has bounded total variation in $\mathcal{U}_k$. If there exists a constant vector $\boldsymbol{a}_1$ and a deterministic function $a_{2k}(t)$ such that $\boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{X}_k(t)+a_{2k}(t)=0$ for any $t\in\mathcal{U}_k$ and any $k\in\{1,\ldots,K\}$ with probability 1, then $\boldsymbol{a}_1=\mathbf{0}$ and $a_{2k}(t)=0$ for any $t\in\mathcal{U}_k$ and any $k\in\{1,\ldots,K\}$.

Remark 4.2. *Conditions 1, 2, and 5 are standard conditions for failure time regression with time-dependent covariates. Condition 3 pertains to the joint distribution of monitoring times of the asymptomatic events; it requires that two adjacent monitoring times are separated by at least η; otherwise, the data may contain exact observations such that different theoretical treatment is needed. The dominating measure μ_k is chosen as the Lebesgue measure if the monitoring times are continuous random variables and as the counting measure if monitorings occur only at a finite number of time points. The number of potential monitoring times M_k can be fixed or random, is possibly different among study subjects and event types, and is allowed to depend on covariates. Condition 4 implies that there is a positive probability for the kth symptomatic event to be observed in the time interval $[0,\tau_k]$.*

We state the strong consistency of $(\widehat{\boldsymbol{\theta}},\widehat{\mathcal{A}})$ and the weak convergence of $\widehat{\boldsymbol{\theta}}$ in two theorems.

Theorem 4.1. *Under Conditions 1–5, $\|\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\|\to_{a.s.}0$, and $\|\widehat{\Lambda}_k-\Lambda_{k0}\|_{l^\infty(\mathcal{U}_k)}\to_{a.s.}0$, where $\|\cdot\|_{l^\infty(\mathcal{U}_k)}$ denotes the supremum norm on $\mathcal{U}_k$ for $k=1,\ldots,K$.*

Theorem 4.2. *Under Conditions 1–5, $n^{1/2}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)$ converges weakly to a $(p+K_2+2)$-dimensional zero-mean normal random vector with a covariance matrix that attains the semiparametric efficiency bound.*

The proofs of all theorems are provided in Section 4.6.

We propose two approaches to estimate the covariance matrix of $\widehat{\boldsymbol{\theta}}$. The first approach makes use of the profile likelihood (Murphy and Van der Vaart, 2000). Specifically, we define the profile log-likelihood function

$$pl_n(\boldsymbol{\theta})=\max_{\mathcal{A}\in\mathcal{C}_1\times\cdots\times\mathcal{C}_K}\log L_n(\boldsymbol{\theta},\mathcal{A}),$$

where $\mathcal{C}_k$ is the set of step functions with non-negative jumps at t_{kl} $(k=1,\ldots,K;l=1,\ldots,m_k)$.

We estimate the covariance matrix of $\widehat{\boldsymbol{\theta}}$ by the inverse of

$$\sum_{i=1}^{n} \begin{pmatrix} \dfrac{pl_i(\widehat{\boldsymbol{\theta}} + h_n e_1) - pl_i(\widehat{\boldsymbol{\theta}})}{h_n} \\ \vdots \\ \dfrac{pl_i(\widehat{\boldsymbol{\theta}} + h_n e_{p+K_2+2}) - pl_i(\widehat{\boldsymbol{\theta}})}{h_n} \end{pmatrix}^{\otimes 2},$$

where pl_i is the ith subject's contribution to pl_n, e_j is the jth canonical vector in $\mathbb{R}^{p+K_2+2}$, $\boldsymbol{a}^{\otimes 2} = \boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}$, and h_n is a constant of order $n^{-1/2}$. To evaluate the profile likelihood, we use the EM algorithm of Section 2.2 but only update $\Lambda_1, \ldots, \Lambda_K$ in the M-step.

Alternatively, we approximate the asymptotic distribution of $\widehat{\boldsymbol{\theta}}$ by bootstrapping the observations. In particular, we draw a simple random sample of size n with replacement from the observed data $\{\mathcal{O}_i : i = 1, \ldots, n\}$. Let $\widehat{\boldsymbol{\theta}}^*$ be the estimator of $\boldsymbol{\theta}$ in the bootstrap sample. The empirical distribution of $\widehat{\boldsymbol{\theta}}^*$ can be used to approximate the distribution of $\widehat{\boldsymbol{\theta}}$. Confidence intervals for $\boldsymbol{\theta}_0$ can be constructed by the Wald method (with the variance of $\widehat{\boldsymbol{\theta}}^*$) or from the empirical percentiles of $\widehat{\boldsymbol{\theta}}^*$. The following theorem states the asymptotic properties of $\widehat{\boldsymbol{\theta}}^*$, thereby validating the bootstrap procedure.

Theorem 4.3. *Under Conditions 1–5, the conditional distribution of* $n^{1/2}(\widehat{\boldsymbol{\theta}}^* - \widehat{\boldsymbol{\theta}})$ *given the data converges weakly to the asymptotic distribution of* $n^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$.

4.2.4 Dynamic Prediction

Given the fitted joint model, we can predict future events by updating the event history. For a subject with covariate $\boldsymbol{X}$, let $\mathcal{O}(t)$ denote the event history at time $t > 0$, which includes the interval-censored observations of the asymptomatic events $\{L_k(t), R_k(t) : k = 1, \ldots, K_1\}$, and the right-censored observations of the symptomatic events $\{Y_k(t), \Delta_k(t) : k = K_1 + 1, \ldots, K\}$.

If no event history is available, the density of the random effect $\boldsymbol{b}$ can be estimated by $\psi(\boldsymbol{b}; \widehat{\boldsymbol{\Sigma}})$. We estimate the survival function of T_k, denoted by $P(T_k \geq t | \boldsymbol{X})$, by

$$\int_{\boldsymbol{b}} s_k(t; \boldsymbol{X}, \boldsymbol{b}) \psi(\boldsymbol{b}; \widehat{\boldsymbol{\Sigma}}) d\boldsymbol{b},$$

where $s_k(t; \boldsymbol{X}, \boldsymbol{b})$ is the conditional survival probability given $\boldsymbol{b}$ that takes the form

$$s_k(t; \boldsymbol{X}, \boldsymbol{b}) = \begin{cases} \exp\left\{-\int_0^t e^{\widehat{\boldsymbol{\beta}}^{\mathrm{T}} \boldsymbol{X}_k(u) + b_1} d\widehat{\Lambda}_k(u)\right\} & k = 1, \ldots, K_1 \\ \exp\left\{-\int_0^t e^{\widehat{\boldsymbol{\beta}}^{\mathrm{T}} \boldsymbol{X}_k(u) + \widehat{\gamma}_k b_1 + b_2} d\widehat{\Lambda}_k(u)\right\} & k = K_1 + 1, \ldots, K \end{cases},$$

and the integral is evaluated by numerical integration with Gauss-Hermite quadratures. In some studies, one of the symptomatic event is terminal (e.g., death) such that its occurrence precludes the development of other events. Without loss of generality, we assume the Kth event is terminal and estimate the cumulative incidence function of the kth event ($k = 1, \ldots, K-1$), i.e., $P(T_k \leq t, T_k \leq T_K | \boldsymbol{X})$, by

$$\int_{\boldsymbol{b}} \left[\{1 - s_k(t; \boldsymbol{X}, \boldsymbol{b})\} s_K(t; \boldsymbol{X}, \boldsymbol{b}) + \int_0^t \{1 - s_k(u; \boldsymbol{X}, \boldsymbol{b})\} s_K(u; \boldsymbol{X}, \boldsymbol{b}) e^{\widehat{\boldsymbol{\beta}}^{\mathrm{T}} \boldsymbol{X}_K(u) + \widehat{\gamma}_K b_1 + b_2} d\widehat{\Lambda}_K(u) \right]$$
$$\times \psi(\boldsymbol{b}; \widehat{\boldsymbol{\Sigma}}) d\boldsymbol{b}.$$

At time $t_0 > 0$, we update the posterior density of $\boldsymbol{b}$ given the event history $\mathcal{O}(t_0)$ so as to perform dynamic prediction. Note that the posterior density of $\boldsymbol{b}$ is proportional to

$$\begin{aligned} J(\boldsymbol{b}; t_0, \boldsymbol{X}) \equiv & \prod_{k=1}^{K_1} \{s_k(L_k(t_0); \boldsymbol{X}, \boldsymbol{b}) - s_k(R_k(t_0); \boldsymbol{X}, \boldsymbol{b})\} \\ & \times \prod_{k=K_1+1}^{K} \left(s_k(Y_k(t_0); \boldsymbol{X}, \boldsymbol{b}) \left[\widehat{\Lambda}_k\{Y_k(t_0)\} e^{\widehat{\boldsymbol{\beta}}^{\mathrm{T}} \boldsymbol{X}_k\{Y_k(t_0)\} + \widehat{\gamma}_k b_1 + b_2} \right]^{\Delta_k(t_0)} \right) \psi(\boldsymbol{b}; \widehat{\boldsymbol{\Sigma}}). \end{aligned}$$

If the subject has not developed the kth event or the terminal event by time t_0, i.e., $Y_k(t_0) = Y_K(t_0) = t_0$ and $\Delta_k(t_0) = \Delta_K(t_0) = 0$, we estimate the conditional cumulative incidence function of the kth event, $P(T_k \leq t, T_k \leq T_K | \mathcal{O}(t_0), \boldsymbol{X})$, by

$$\begin{aligned} \int_{\boldsymbol{b}} \frac{J(\boldsymbol{b}; t_0, \boldsymbol{X})}{s_k(t_0; \boldsymbol{X}, \boldsymbol{b}) s_K(t_0; \boldsymbol{X}, \boldsymbol{b}) \int_{\boldsymbol{b}'} J(\boldsymbol{b}'; t_0, \boldsymbol{X}) d\boldsymbol{b}'} & \left[\{s_k(t_0; \boldsymbol{X}, \boldsymbol{b}) - s_k(t; \boldsymbol{X}, \boldsymbol{b})\} s_K(t; \boldsymbol{X}, \boldsymbol{b}) \right. \\ + \int_s^t \{s_k(t_0; \boldsymbol{X}, \boldsymbol{b}) - s_k(u; \boldsymbol{X}, \boldsymbol{b})\} s_K(u; \boldsymbol{X}, \boldsymbol{b}) & \left. e^{\widehat{\boldsymbol{\beta}}^{\mathrm{T}} \boldsymbol{X}_K(u) + \widehat{\gamma}_K b_1 + b_2} d\widehat{\Lambda}_K(u) \right] \psi(\boldsymbol{b}; \widehat{\boldsymbol{\Sigma}}) d\boldsymbol{b}. \end{aligned}$$

In practice, it is desirable to identify subjects who are at increased risk as the event history is accumulating. In the same vein as the risk score under the standard proportional hazards model, we use the risk score $\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{X}_k(t_0)+\widehat{\gamma}_k\widehat{b}_1(t_0)+\widehat{b}_2(t_0)$ to dynamically predict the kth event ($k=K_1+1,\ldots,K$), where $\widehat{\boldsymbol{b}}(t_0)\equiv(\widehat{b}_1(t_0),\widehat{b}_2(t_0))$ is a suitable estimator of $\boldsymbol{b}$ given the event history $\mathcal{O}(t_0)$. The estimator $\widehat{\boldsymbol{b}}(t_0)$ can be the posterior mean or mode of $\boldsymbol{b}$ or an imputed value from the posterior distribution. For example, the risk score using the posterior mean is given by

$$\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{X}_k(t_0)+\frac{\int_{\boldsymbol{b}}(\widehat{\gamma}_k b_1+b_2)J(\boldsymbol{b};t_0,\boldsymbol{X})d\boldsymbol{b}}{\int_{\boldsymbol{b}}J(\boldsymbol{b};t_0,\boldsymbol{X})d\boldsymbol{b}}.$$

The risk score quantifies the subject-specific risk and can be very useful to both individual patients and clinicians when making decisions about lifestyle modifications and preventive medical treatments.

4.3 Simulation Studies

We conducted simulation studies to assess the performance of the proposed methods. We considered one time-independent covariate $X_1\sim Unif(0,1)$ and one time-dependent covariate $X_2(t)=I(t\leq V)B_1+I(t>V)B_2$, where B_1 and B_2 are independent Bernoulli(0.5), $V\sim Unif(0,\tau)$, and $\tau=4$. We considered two asymptomatic events and two symptomatic events. We set $\boldsymbol{X}_k=e_k\otimes(X_1,X_2)^{\mathrm{T}}$, where e_k is the kth canonical vector in $\mathbb{R}^4$, and $\otimes$ denotes the Kronecker product. We set $\boldsymbol{\beta}=(0.5,0.4,0.5,-0.2,-0.5,0.5,-0.5,0.5)^{\mathrm{T}}$, $\Lambda_1(t)=0.5t$, $\Lambda_k(t)=\log\{1+t/(k-1)\}$ for $k=2,3,4$, $\gamma_3=\gamma_4=0.25$, and $\sigma_1^2=\sigma_2^2=1$. Both symptomatic events were censored by $C\sim Unif(2\tau/3,\tau)$. The series of monitoring times were generated sequentially, with $U_m=U_{m-1}+0.1+\mathrm{Unif}(0,0.5)$ for $m\geq 1$ and $U_0=0$. The last monitoring time is the largest U_m that is smaller than C. We set $n=100$ or 200 and simulated 2,000 replicates. For each dataset, we applied the proposed EM algorithm by setting the initial value of $\boldsymbol{\beta}$ to $\mathbf{0}$, the initial values of γ_k and σ_k^2 to 1 and the initial value of λ_{kl} to $1/m_k$. We used 20 quadrature points for integration with respect to each random effect and set the convergence threshold to 10^{-3}. For variance estimation, we set $h_n=5/\sqrt{n}$ for profile likelihood and used 100 bootstrap samples.

Table 4.1 summarizes the simulation results. The biases for all parameter estimators are small, especially for $n=200$. Both the profile-likelihood and bootstrap variance estimators for $\widehat{\boldsymbol{\beta}}$ are accurate, especially for $n=200$. Both variance estimators for $\widehat{\boldsymbol{\gamma}}$ tend to overestimate the true variabilities, but the coverage probabilities of the confidence intervals get closer to the nominal level as sample size increases. The profile-likelihood variance estimators for $\widehat{\sigma}_1^2$ and $\widehat{\sigma}_2^2$ overestimate the true variabilities, while the bootstrap variance estimators for $\widehat{\sigma}_1^2$ and $\widehat{\sigma}_2^2$ accurately reflect the true

Table 4.1: Summary statistics for the simulation studies without a terminal event

	$n = 100$						$n = 200$					
			Profile		Bootstrap				Profile		Bootstrap	
	Bias	SE	SEE	CP	SEE	CP	Bias	SE	SEE	CP	SEE	CP
β_{11}	0.006	0.585	0.597	0.961	0.627	0.967	0.027	0.405	0.399	0.947	0.412	0.953
β_{12}	0.029	0.327	0.321	0.941	0.348	0.960	0.019	0.222	0.216	0.949	0.228	0.953
β_{21}	0.015	0.623	0.609	0.946	0.648	0.963	0.014	0.410	0.409	0.951	0.424	0.958
β_{22}	−0.005	0.341	0.329	0.940	0.355	0.962	−0.002	0.225	0.222	0.951	0.233	0.961
β_{31}	−0.022	0.617	0.635	0.957	0.610	0.949	−0.004	0.416	0.428	0.960	0.420	0.948
β_{32}	−0.002	0.319	0.338	0.965	0.322	0.949	0.009	0.221	0.229	0.958	0.222	0.947
β_{41}	−0.012	0.623	0.651	0.969	0.629	0.955	0.006	0.449	0.440	0.947	0.431	0.942
β_{42}	0.004	0.330	0.348	0.967	0.332	0.950	−0.001	0.231	0.235	0.955	0.229	0.945
γ_1	−0.012	0.227	0.252	0.979	0.260	0.971	−0.012	0.159	0.171	0.962	0.170	0.960
γ_2	−0.013	0.237	0.260	0.976	0.266	0.972	−0.016	0.162	0.173	0.966	0.177	0.963
σ_1^2	0.062	0.445	0.751	0.978	0.493	0.956	0.031	0.317	0.482	0.982	0.318	0.946
σ_2^2	−0.102	0.413	0.510	0.993	0.482	0.971	−0.062	0.297	0.335	0.987	0.312	0.974

SE and SEE denote, respectively, the empirical standard error and mean standard error estimator. CP stands for the empirical coverage probability of the 95% confidence interval based on the Wald method for the profile-likelihood approach and the 95% symmetric confidence interval for the bootstrap approach. For γ_1, γ_2, σ_1^2, and σ_2^2, bias and SEE are based on the median instead of the mean, and SE is based on the mean absolute deviation. For σ_1^2 and σ_2^2, the confidence intervals are based on the log transformation.

variabilities. Figure 4.1(a) shows the estimation of the baseline survival functions with sample size $n = 200$. The estimators are virtually unbiased.

We considered a second setup with an additional terminal event. We set $\boldsymbol{X}_k = e_k \otimes (X_1, X_2)^{\mathrm{T}}$, where e_k is the kth canonical vector in $\mathbb{R}^5$. In addition, we set $\boldsymbol{\beta} = (0.5, 0.4, 0.5, -0.2, -0.5, 0.5, -0.5, 0.5, 0.3, -0.2)^{\mathrm{T}}$, $\Lambda_5(t) = \log(1 + t/4)$, and $\gamma_5 = 0.25$. The terminal event was also censored by C. The results are shown in Table 4.2 and Figure 4.1(b). The conclusions are similar to the case of no terminal event.

We assessed the performance of dynamic prediction based on the conditional cumulative incidence function in the setting with a terminal event. Suppose that at the first monitoring time $t_0 = 1$, event 2 has occurred but events 1, 3, and 4 have not. Figure 4.2 shows the estimation of the baseline cumulative incidence functions (pertaining to $\boldsymbol{X} = \boldsymbol{0}$) for events 3 and 4 given the event history at time $t_0 = 1$. The estimators slightly underestimate the true values at the right tail, but the biases get smaller as n increases.

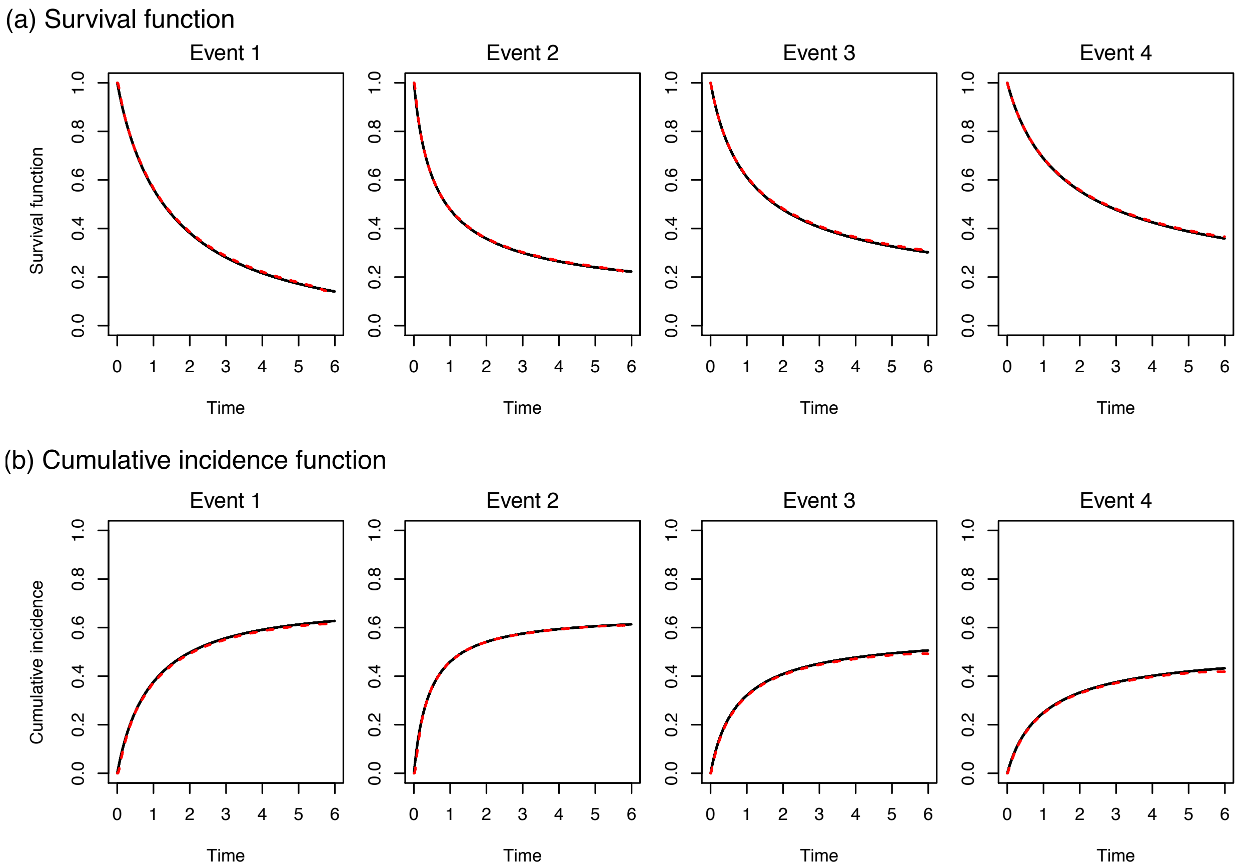

Figure 4.1: Estimation of (a) the baseline survival function and (b) the baseline cumulative incidence function based on $n = 200$. The solid black curve and dashed red curve pertain, respectively, to the true value and mean estimate from the proposed method.

Table 4.2: Summary statistics for the simulation studies with a terminal event

	$n = 100$						$n = 200$					
			Profile		Bootstrap				Profile		Bootstrap	
	Bias	SE	SEE	CP	SEE	CP	Bias	SE	SEE	CP	SEE	CP
β_{11}	0.058	0.825	0.797	0.951	0.879	0.968	0.024	0.505	0.515	0.959	0.544	0.969
β_{12}	0.031	0.445	0.436	0.955	0.502	0.980	0.020	0.291	0.286	0.946	0.310	0.964
β_{21}	0.042	0.780	0.794	0.960	0.881	0.980	0.022	0.521	0.517	0.955	0.549	0.961
β_{22}	−0.019	0.452	0.437	0.948	0.498	0.978	−0.013	0.295	0.288	0.944	0.311	0.961
β_{31}	−0.019	0.675	0.724	0.968	0.694	0.962	−0.014	0.457	0.477	0.957	0.468	0.954
β_{32}	0.007	0.353	0.393	0.973	0.372	0.956	0.005	0.245	0.260	0.963	0.253	0.957
β_{41}	−0.027	0.716	0.775	0.971	0.749	0.959	−0.020	0.483	0.505	0.963	0.500	0.960
β_{42}	0.008	0.385	0.421	0.972	0.405	0.968	0.014	0.268	0.276	0.953	0.271	0.948
β_{51}	0.019	0.631	0.680	0.970	0.659	0.959	−0.004	0.440	0.451	0.959	0.446	0.955
β_{52}	−0.009	0.339	0.361	0.969	0.343	0.950	0.007	0.230	0.240	0.957	0.234	0.950
γ_1	0.004	0.313	0.331	0.972	0.490	0.985	−0.002	0.225	0.226	0.967	0.264	0.977
γ_2	0.009	0.374	0.410	0.982	0.549	0.980	0.008	0.261	0.277	0.977	0.311	0.980
γ_3	0.019	0.335	0.371	0.980	0.502	0.980	0.003	0.244	0.252	0.974	0.281	0.974
σ_1^2	0.134	0.598	0.933	0.969	0.753	0.950	0.077	0.424	0.573	0.970	0.453	0.945
σ_2^2	−0.129	0.390	0.519	0.993	0.532	0.974	−0.047	0.289	0.332	0.988	0.327	0.981

See the Note to Table 4.1.

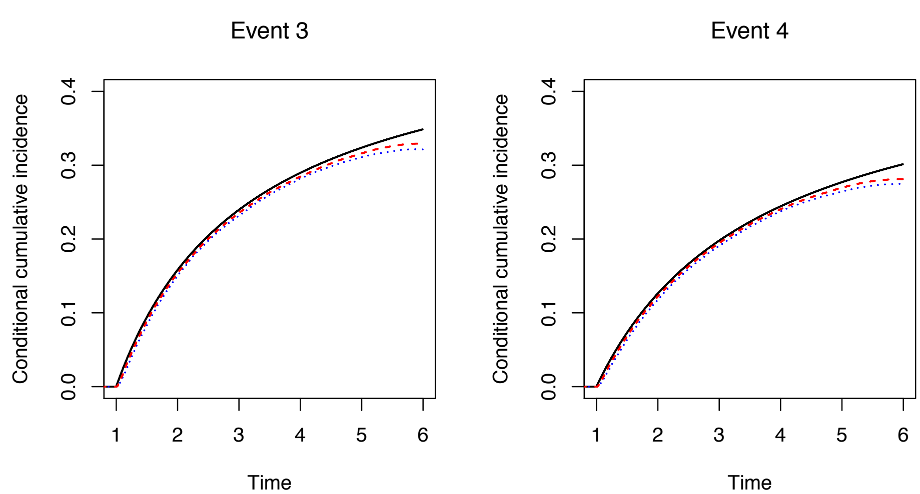

Figure 4.2: Estimation of the baseline cumulative incidence function conditional on the event history. The solid black curve, dotted blue curve, and dashed red curve pertain, respectively, to the true value and the mean estimates from the proposed method with $n = 100$ and $n = 200$.

4.4 ARIC Study

ARIC is a perspective epidemiological cohort study conducted in four U.S. communities: Forsyth County, NC; Jackson, MS; Minneapolis, MN; and Washington County, MD. A total of 15,792 participants received a baseline examination between 1987 and 1989 and four subsequent examinations in 1990-1992, 1993-1995, 1996-1998, and 2011-2013. At each examination, medical data were collected, such that interval-censored observations for diabetes and hypertension were obtained. The participants were also followed for cardiovascular diseases through reviews of hospital records, such that potentially right-censored observations on MI, stroke, and death were collected.

We related the disease incidence to race, sex, and five baseline risk factors: age, body mass index (BMI), glucose level, systolic blood pressure, and smoking status. Since the Jackson cohort is composed of black subjects only, and neither Minneapolis nor Washington County cohorts contain black subjects, we included the cohort×race indicators as predictors. We excluded subjects with prevalent cases at baseline or missing covariate values to obtain a total of 8,728 subjects.

Table 4.3 shows the proportions of incidence cases, non-cases during follow-up, and the observations with no information (i.e., no observations at the scheduled visits) for the asymptomatic events. Less than 20% of the subjects have developed diabetes, while approximately half of the subjects have developed hypertension during the study. Table 4.4 shows the proportions of incidence cases and non-cases for the symptomatic events. A small proportion of subjects have developed MI or stroke during the study.

Table 4.3: Distribution of observations for the asymptomatic events in the ARIC study

Event	Incidence Case	Non-case During Follow-up	No Information
Diabetes	1508 (17.3%)	6771 (77.6%)	449 (5.1%)
Hypertension	4081 (46.8%)	4202 (48.1%)	445 (5.1%)

Table 4.4: Distribution of observations for the symptomatic events in the ARIC study

Event	Incidence Case	Non-case
MI	726 (8.3%)	8002 (91.7%)
Stroke	445 (5.1%)	8283 (94.9%)
Death	2503 (28.7%)	6225 (71.3%)

We jointly modeled the asymptomatic and symptomatic events in the ARIC study with equations (4.1) and (4.2). Table 4.5 and 4.6 show the estimation results for the regression parameters. Several

Table 4.5: Estimation results for the regression parameters of the asymptomatic events in the ARIC study

	Diabetes			Hypertension		
Covariate	Estimate	Std error	p-value	Estimate	Std error	p-value
Forsyth County, white	−0.5332	0.1817	0.0033	−0.5032	0.0615	<0.0001
Jackson, black	−0.1356	0.1806	0.4530	−0.1075	0.0673	0.1104
Minneapolis, white	−0.9415	0.1802	<0.0001	−0.5747	0.0579	<0.0001
Washington County, white	−0.3778	0.1778	0.0336	−0.3798	0.0592	<0.0001
Age	−0.0093	0.0057	0.1025	0.0166	0.0036	<0.0001
Male	−0.0655	0.0593	0.2694	−0.2329	0.0396	<0.0001
BMI	0.0911	0.0059	<0.0001	0.0254	0.0044	<0.0001
Glucose	0.1075	0.0033	<0.0001	0.0004	0.0023	0.8744
Systolic blood pressure	0.0096	0.0026	0.0003	0.0780	0.0022	<0.0001
Smoker	0.4576	0.0674	<0.0001	0.3134	0.0468	<0.0001

The blacks in Forsyth County form the reference group for the cohort×race variables.

characteristics and baseline risk factors are found to be predictive of the events. Older subjects have higher risks of hypertension, MI, stroke, and death than younger subjects. Males have lower risk of hypertension but higher risks of MI, stroke, and death than females. Smokers have significantly higher risks for all events than non-smokers. In addition, higher baseline BMI increases the risks of diabetes, hypertension, and MI; higher baseline glucose level increases the risks of diabetes, stroke, and death; and higher baseline value of systolic blood pressure increases the risks of all considered events.

Table 4.6: Estimation results for the regression parameters of the symptomatic events in the ARIC study

	MI			Stroke			Death		
Covariate	Estimate	Std error	p-value	Estimate	Std error	p-value	Estimate	Std error	p-value
Forsyth County, white	0.0467	0.2477	0.8504	0.1308	0.3688	0.7228	−0.2475	0.1049	0.0183
Jackson, black	−0.3121	0.2681	0.2444	0.6622	0.3755	0.0778	0.1871	0.1118	0.0941
Minneapolis, white	−0.1052	0.2476	0.6710	0.0507	0.3688	0.8907	−0.3262	0.1040	0.0017
Washington County, white	0.1953	0.2457	0.4266	0.5013	0.3653	0.1700	−0.1194	0.1032	0.2471
Age	0.0805	0.0078	<0.0001	0.1121	0.0099	<0.0001	0.1465	0.0054	<0.0001
Male	0.9279	0.0901	<0.0001	0.4050	0.1071	0.0002	0.6108	0.0545	<0.0001
BMI	0.0273	0.0101	0.0068	−0.0010	0.0123	0.9356	0.0080	0.0060	0.1847
Glucose	0.0059	0.0046	0.2007	0.0215	0.0057	0.0002	0.0104	0.0030	0.0006
Systolic blood pressure	0.0135	0.0036	0.0002	0.0192	0.0047	<0.0001	0.0089	0.0022	0.0001
Smoker	1.2378	0.0888	<0.0001	1.0023	0.1127	<0.0001	1.3045	0.0599	<0.0001

See the Note to Table 4.6.

Table 4.7: Estimation results for the random effects in the ARIC study

Parameter	Estimate	Std error	p-value
γ_{MI}	0.7145	0.1258	<0.0001
γ_{Stroke}	0.9045	0.1450	<0.0001
γ_{Death}	0.7184	0.1026	<0.0001
σ_1^2	0.5801	0.1215	<0.0001
σ_2^2	1.1465	0.1165	<0.0001

The estimation results for the remaining parametric components are shown in Table 4.7. The variance components σ_1^2 and σ_2^2 are significantly larger than zero, indicating strong correlation among the asymptomatic events and among the symptomatic events. The parameters γ_{MI}, γ_{Stroke}, and γ_{Death} are also significantly larger than zero, reflecting the strong positive dependence of the symptomatic events on the asymptomatic events.

To evaluate the performance of the proposed prediction methods, we randomly divided the study cohort into training and testing sets with equal numbers of subjects. We analyzed the training set to obtain parameter estimates, based on which we calculated the risk scores for subjects in the testing set, where the posterior means of the random effects were used. Specifically, at examinations 2, 3, and 4, we calculated the risk scores of MI or stroke for subjects who have not developed the disease. We evaluated the performance of the prediction using C-index (Uno et al., 2011) and compared it with that of the risk scores based on the standard proportional hazards model. The values of the C-index based on twenty randomly divided training/test tests are shown in Figure 4.3. The proposed risk score performs better than the risk score of the standard proportional hazards model at all examinations for all symptomatic events.

Figure 4.4 shows the estimated conditional cumulative incidence functions of MI and stroke for two smokers and two non-smokers who have different event histories at year 3 but with the same values of other risk factors. The risks of MI and stroke are considerably higher for the smokers than the non-smokers with the same event history. The estimated conditional probabilities for the subjects who have developed both diabetes and hypertension are higher than those who have not developed diabetes or hypertension.

Figures 4.5(a) and 4.5(b) illustrate the estimation of the conditional cumulative incidence functions of stroke given different event histories. We estimated the cumulative incidence functions at baseline and then updated them at two examinations at year 3 and year 6. The development of

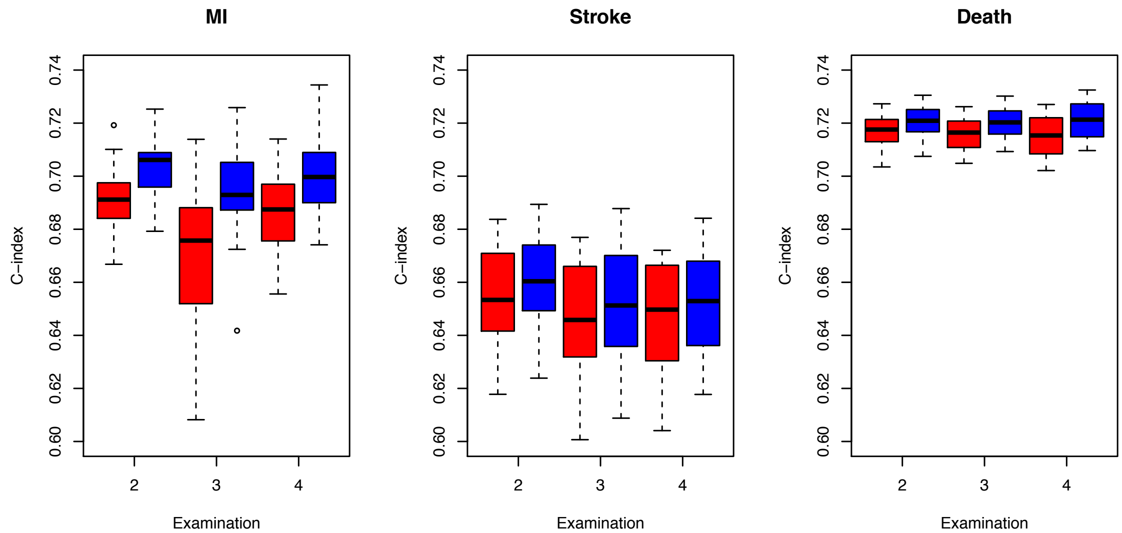

Figure 4.3: Boxplots of the estimates of the C-index at each examination in the ARIC study. The red and blue boxes pertain to the standard proportional hazards model and the proposed joint model, respectively.

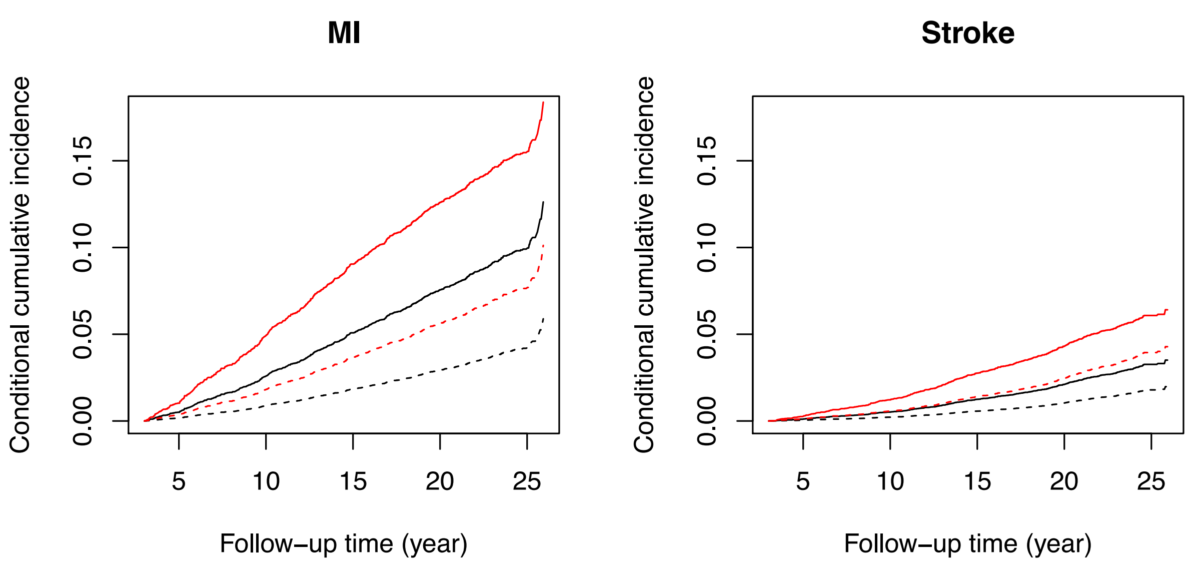

Figure 4.4: Estimation of the conditional cumulative incidence functions of MI and stroke for a 50-year-old white female residing in Forsyth County, NC, with BMI 40 kg/m^2, glucose 98 mg/dl, and systolic blood pressure 113 mmHg. The solid curves pertain to smokers, while the dashed curves pertain to non-smokers. The black curves pertain to subjects who have not developed diabetes or hypertension by year 3. The red curves pertain to subjects who have developed both diabetes and hypertension by year 3.

diabetes, hypertension, and MI substantially increases the incidence of stroke, whereas the history of no diabetes, hypertension, or MI over the first six years entails lower incidence of stroke. For comparison, we show in Figure 4.5(c) the estimated cumulative incidence function of stroke under the univariate model of Fine and Gray (1999), which does not condition on the event history and thus reflects the population average. This estimate lies between the two previous conditional estimates, as expected.

4.5 Discussion

In this chapter, we formulated the joint distribution of multiple right- and interval-censored events with proportional hazards models with random effects. We characterized the correlation structure of the asymptomatic and symptomatic events through two independent random effects and used unknown coefficients to capture the effects of the asymptomatic events on the symptomatic events. To our knowledge, no such modelling approach has been previously adopted.

We studied efficient nonparametric maximum likelihood estimation of the proposed joint model and established the asymptotic properties of the estimators through innovative use of modern empirical process theory. We showed the Glivenko-Cantelli and Donsker properties for the classes of functions of interest by carefully evaluating their bracketing numbers. The estimators of the cumulative baseline hazard functions for the symptomatic and asymptomatic events converge at different ($n^{1/2}$ and $n^{1/3}$) rates, such that separate treatments were required in the proofs.

We proposed nonparametric bootstrap for variance estimation as an alternative to the conventional profile-likelihood approach. We established the validity of the bootstrap procedure and showed through simulation studies that bootstrap yields more accurate estimators of the variabilities for the variance components. To our knowledge, bootstrap with interval-censored data has not been rigorously studied. In large studies, bootstrap may be overly time-consuming. It would be worthwhile to develop other versions of bootstrap, such as subsampling bootstrap, to reduce computational burden.

ARIC is one of many epidemiological cohort studies with multiple symptomatic and asymptomatic events. Such events are also available in electronic health records. Indeed, other types of outcomes, such as longitudinal repeated measures and recurrent events, may also be available. The proposed joint model can be extended to accommodate additional multivariate outcomes and improve dynamic prediction.

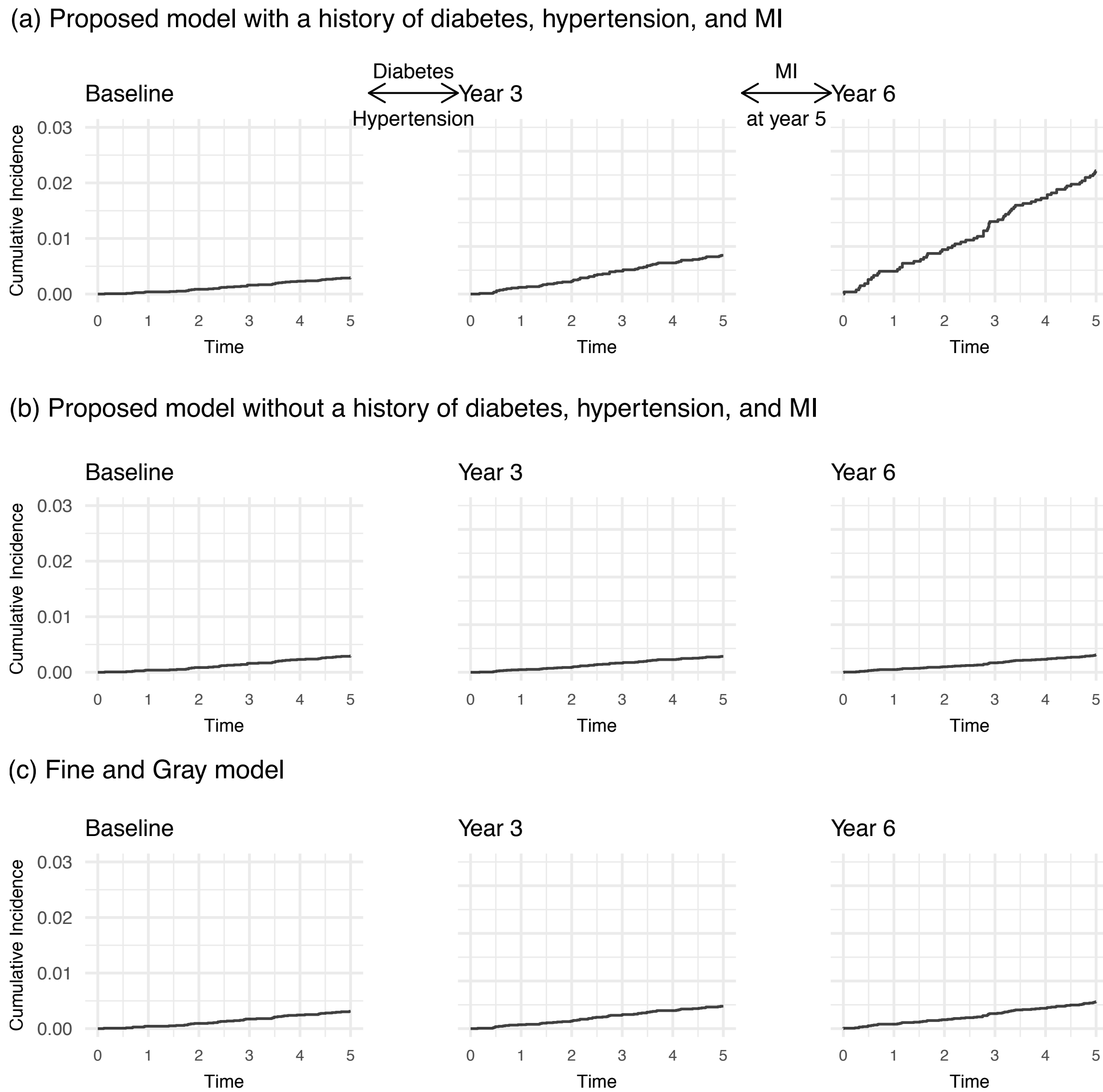

Figure 4.5: Estimation of the cumulative incidence of stroke for a 50-year-old white female smoker residing in Forsyth County, NC, with BMI 40 kg/m^2, glucose 98 mg/dl, and systolic blood pressure 113 mmHg: (a) proposed model with MI developed at year 5 and diabetes and hypertension developed between baseline and year 3; (b) proposed model without MI, diabetes or hypertension by year 6; and (c) Fine and Gray model.

4.6 Technical Details

Let $\mathbb{P}_n$ denote the empirical measure for n independent subjects, $\mathbb{P}$ denote the true probability measure, and $\mathbb{G}_n \equiv \sqrt{n}(\mathbb{P}_n - \mathbb{P})$ denote the empirical process. The proofs of Theorems 4.1, 4.2, and 4.3 make use of three lemmas, which are stated and proved in Section 4.6.4.

4.6.1 Proof of Theorem 4.1

We first show the existence of the estimator $(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}})$. Let $\widetilde{M} = \sum_{k=1}^{K} \sup_{t \in \mathcal{U}_k} \sup_{\boldsymbol{X}_k(t), \boldsymbol{\beta}} |\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_k(t)| + \sum_{k=K_1+1}^{K} |\gamma_k|$. For any $(\boldsymbol{\theta}, \mathcal{A})$ in the parameter space, the integrand in the ith term of $l_n(\boldsymbol{\theta}, \mathcal{A})$ is bounded by

$$O(1) \prod_{k=K_1+1}^{K} \left[\left(\Lambda_k\{Y_{ik}\} e^{\widetilde{M}|\boldsymbol{b}_i|} \right)^{\Delta_{ik}} \left\{ 1 + \int_0^{Y_{ik}} e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_{ik}(s) + \gamma_k b_{i1} + b_{i2}} d\Lambda_k(s) \right\}^{-\Delta_{ik}} \right] \psi(\boldsymbol{b}_i; \boldsymbol{\Sigma}).$$

Thus, $l_n(\boldsymbol{\theta}, \mathcal{A})$ attains the maximum for finite values of Λ_k for $k = K_1+1, \ldots, K$, so the estimator $(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}})$ exists by allowing $\widehat{\Lambda}_k(\tau_k) = \infty$ for $k = 1, \ldots, K_1$.

We shall prove that $\limsup_n \widehat{\Lambda}_k(\tau_k - \epsilon) < \infty$ with probability 1 for any $\epsilon > 0$ and $k = 1, \ldots, K_1$ and that $\limsup_n \widehat{\Lambda}_k(\tau_k) < \infty$ with probability 1 for $k = K_1+1, \ldots, K$. By definition, $l_n(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}}) - l_n(\boldsymbol{\theta}, \mathcal{A}) \geq 0$ for any $(\boldsymbol{\theta}, \mathcal{A})$ in the parameter space. We wish to show that if $\limsup_n \widehat{\Lambda}_k(\tau_k - \epsilon) = \infty$ for some $\epsilon > 0$ for $k = 1, \ldots, K_1$ or $\widehat{\Lambda}_k(\tau_k) = \infty$ for $k = K_1+1, \ldots, K$, then this difference must be negative, which is a contradiction. The key is to construct a suitable function in the parameter space that converges uniformly to $\mathcal{A}_0$.

For $k = 1, \ldots, K_1$, we define the step function $\widetilde{\Lambda}_k$ with $\widetilde{\Lambda}_k(t) = \Lambda_{k0}(t)$ for $t = t_{k1}, \ldots, t_{k,m_k}$ such that it converges uniformly to Λ_{k0}. For $k = K_1+1, \ldots, K$, we construct function $\widetilde{\Lambda}_k$ by imitating $\widehat{\Lambda}_k$. Specifically, by differentiating $l_n(\boldsymbol{\theta}, \mathcal{A})$ with respect to $\Lambda_k\{Y_{ik}\}$ and setting the derivative to 0, we find that $\widehat{\Lambda}_k$ satisfies the equation

$$\frac{\Delta_{ik}}{\widehat{\Lambda}_k\{Y_{ik}\}} = \sum_{j=1}^{n} \frac{\int_{\boldsymbol{b}} J_1\left(\boldsymbol{b}, \mathcal{O}_j; \widehat{\boldsymbol{\beta}}, \widehat{\gamma}, \widehat{\mathcal{A}}\right) J_{2k}\left(Y_{ik}, \boldsymbol{b}, \mathcal{O}_j; \widehat{\boldsymbol{\beta}}, \widehat{\gamma}_k\right) \phi\left(\boldsymbol{b}; \widehat{\boldsymbol{\Sigma}}\right) d\boldsymbol{b}}{\int_{\boldsymbol{b}} J_1\left(\boldsymbol{b}, \mathcal{O}_j; \widehat{\boldsymbol{\beta}}, \widehat{\gamma}, \widehat{\mathcal{A}}\right) \phi\left(\boldsymbol{b}; \widehat{\boldsymbol{\Sigma}}\right) d\boldsymbol{b}}, \tag{4.3}$$

where

$$J_1(\boldsymbol{b}, \mathcal{O}; \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathcal{A}) = \prod_{k=1}^{K_1} \left[\exp\left\{ -\int_0^{L_k} e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_k(s) + b_1} d\Lambda_k(s) \right\} - \exp\left\{ -\int_0^{R_k} e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_k(s) + b_1} d\Lambda_k(s) \right\} \right]$$

$$\times \prod_{K_1+1}^{K} \left[\left\{ e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_k(Y_k) + \gamma_k b_1 + b_2} \right\}^{\Delta_k} \exp\left\{ - \int_0^{Y_k} e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_k(s) + \gamma_k b_1 + b_2} d\Lambda_k(s) \right\} \right],$$

and

$$J_{2k}\left(t, \boldsymbol{b}, \mathcal{O}; \boldsymbol{\beta}, \gamma_k\right) = -I(Y_k \geq t) e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_k(t) + \gamma_k b_1 + b_2}.$$

We replace $\widehat{\boldsymbol{\theta}}$ and $\widehat{\mathcal{A}}$ on the right side of equation (4.3) by $\boldsymbol{\theta}_0$ and $\mathcal{A}_0$, respectively, to obtain a similar function. We denote the solution as $\widetilde{\Lambda}_k$. By the Glivenko-Cantelli result in Lemma 1, $\widetilde{\Lambda}_k$ converges uniformly to Λ_{k0} in $\mathcal{U}_k$ for $k = K_1+1, \ldots, K$. We denote $\widetilde{\mathcal{A}} = (\widetilde{\Lambda}_1, \ldots, \widetilde{\Lambda}_K)$.

Clearly, $n^{-1}\left\{ l_n(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}}) - l_n(\boldsymbol{\theta}_0, \widetilde{\mathcal{A}}) \right\} \geq 0$. Let $\delta_{ikm} = I(U_{ikm} < T_{ik} \leq U_{ik,m+1})$ for $i = 1, \ldots, n$, $k = 1, \ldots, K_1$, and $m = 0, \ldots, M_{ik}$, where $U_{ik,M_{ik}+1} = \infty$. By the fact that $e^{-|x|}(1+y) \leq 1 + e^x y \leq e^{|x|}(1+y)$, we obtain

$$\begin{aligned}
0 \;&\leq\; n^{-1} l_n(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}}) - n^{-1} l_n(\boldsymbol{\theta}_0, \widetilde{\mathcal{A}}) \\
&\leq\; O(1) + n^{-1} \sum_{i=1}^{n} \sum_{k=K_1+1}^{K} \log\left(n\widehat{\Lambda}_k\{Y_{ik}\} \right) \\
&\quad + n^{-1} \sum_{i=1}^{n} \left[\log \int_{\boldsymbol{b}} \prod_{k=K_1+1}^{K} \left\{ \frac{e^{\widehat{\boldsymbol{\beta}}^{\mathrm{T}} \boldsymbol{X}_{ik}(Y_{ik}) + \gamma_k b_{i1} + b_{i2}}}{1 + \int_0^{Y_{ik}} e^{\widehat{\boldsymbol{\beta}}^{\mathrm{T}} \boldsymbol{X}_{ik}(t) + \gamma_k b_{i1} + b_{i2}} d\widehat{\Lambda}_k(t)} \right\}^{\Delta_{ik}} \phi\left(\boldsymbol{b}; \widehat{\boldsymbol{\Sigma}} \right) d\boldsymbol{b} \right] \\
&\leq\; O(1) + n^{-1} \sum_{i=1}^{n} \sum_{k=K_1+1}^{K} \log\left(n\widehat{\Lambda}_k\{Y_{ik}\} \right) \\
&\quad + n^{-1} \sum_{i=1}^{n} \left(\log \int_{\boldsymbol{b}} \prod_{k=K_1+1}^{K} \left[\frac{e^{\widetilde{M}\|\boldsymbol{b}\|}}{e^{-\widetilde{M}\|\boldsymbol{b}\|} \left\{ 1 + \widehat{\Lambda}_k(Y_{ik}) \right\}} \right]^{\Delta_{ik}} \phi\left(\boldsymbol{b}; \widehat{\boldsymbol{\Sigma}} \right) d\boldsymbol{b} \right) \\
&\leq\; O(1) + n^{-1} \sum_{i=1}^{n} \sum_{k=K_1+1}^{K} \log\left(n\widehat{\Lambda}_k\{Y_{ik}\} \right) - n^{-1} \sum_{i=1}^{n} \sum_{k=K_1+1}^{K} \left[\Delta_{ik} \log\left\{ 1 + \widehat{\Lambda}_k(Y_{ik}) \right\} \right].
\end{aligned}$$

We first show that $\limsup_n \widehat{\Lambda}_k(\tau_k) < \infty$ using the partitioning idea of Murphy (1994). Specifically, we construct a sequence $u_{k0} = \tau_k > u_{k1} > \cdots > u_{k,Q_k} = 0$. Then,

$$\begin{aligned}
& n^{-1} \sum_{i=1}^{n} \sum_{k=K_1+1}^{K} \log\left(n\widehat{\Lambda}_k\{Y_{ik}\} \right) - n^{-1} \sum_{i=1}^{n} \sum_{k=K_1+1}^{K} \left[\Delta_{ik} \log\left\{ 1 + \widehat{\Lambda}_k(Y_{ik}) \right\} \right] \\
\leq\;& O(1) + \sum_{k=K_1+1}^{K} \sum_{q=0}^{Q_k-1} n^{-1} \sum_{i=1}^{n} I(Y_{ik} \in [u_{k,q+1}, u_{kq})) \log\left(n\widehat{\Lambda}_k\{Y_{ik}\} \right)
\end{aligned}$$

$$
\begin{aligned}
&- \sum_{k=K_1+1}^{K} n^{-1} \sum_{i=1}^{n} I(Y_{ik} = \tau_k)\Delta_{ik} \log\left\{1 + \widehat{\Lambda}_k(\tau_k)\right\} \\
&- \sum_{k=K_1+1}^{K} \sum_{q=0}^{Q_k-1} n^{-1} \sum_{i=1}^{n} \Delta_{ik} I\left(Y_{ik} \in [u_{k,q+1}, u_{kq})\right) \log\left\{1 + \widehat{\Lambda}_k(u_{k,q+1})\right\},
\end{aligned}
$$

which is further bounded by

$$
\begin{aligned}
&-(2n)^{-1} \sum_{i=1}^{n} \sum_{k=K_1+1}^{K} \Delta_{ik} I(Y_{ik} = \tau_k) \log\left\{1 + \widehat{\Lambda}_k(\tau_k)\right\} \\
&- \sum_{k=K_1+1}^{K} \left\{(2n)^{-1} \sum_{i=1}^{n} \Delta_{ik} I(Y_{ik} = \tau_k) - n^{-1} \sum_{i=1}^{n} \Delta_{ik} I\left(Y_{ik} \in [u_1, u_0)\right)\right\} \log\left\{1 + \widehat{\Lambda}_k(\tau_k)\right\} \\
&- \sum_{k=K_1+1}^{K} \sum_{q=1}^{Q_k-1} \left\{n^{-1} \sum_{i=1}^{n} \Delta_{ik} I\left(Y_{ik} \in [u_{kq}, u_{k,q-1})\right) - n^{-1} \sum_{i=1}^{n} \Delta_{ik} I\left(Y_{ik} \in [u_{k,q+1}, u_{kq})\right)\right\} \\
&\times \log\left\{1 + \widehat{\Lambda}_k(u_{kq})\right\}.
\end{aligned}
$$

Note that u_{kq} is chosen such that the coefficients in front of $\log\{1 + \widehat{\Lambda}_k(u_{kq})\}$ are all negative when n is large enough. Thus, the corresponding terms cannot diverge to ∞. However, if $\widehat{\Lambda}_k(\tau_k)$ diverges to ∞, then the first term diverges to $-\infty$. We conclude that there exists some $M^* < \infty$ such that $\max_{K_1+1 \le k \le K} \limsup_n \widehat{\Lambda}_k(\tau_k) \le M^*$ for $k = K_1 + 1, \ldots, K$.

We denote $\widetilde{\mathcal{A}}^* = (\widetilde{\Lambda}_1, \ldots, \widetilde{\Lambda}_{K_1}, \widehat{\Lambda}_{K_1+1}, \ldots, \widehat{\Lambda}_K)$. Then,

$$
\begin{aligned}
0 &\le n^{-1} l_n(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}}) - n^{-1} l_n(\boldsymbol{\theta}_0, \widetilde{\mathcal{A}}^*) \\
&\le O(1) + n^{-1} \sum_{i=1}^{n} \left\{\log \int_{\boldsymbol{b}} \prod_{k=1}^{K_1} \left[\exp\left\{-e^{\widetilde{M}\|\boldsymbol{b}\|} \widehat{\Lambda}_k(U_{ik,M_{ik}})\right\}\right]^{\delta_{ik,M_{ik}}} \phi\left(\boldsymbol{b}; \widehat{\boldsymbol{\Sigma}}\right) d\boldsymbol{b}\right\} \\
&\le O(1) + n^{-1} \sum_{i=1}^{n} \left\{\log \int_{\|\boldsymbol{b}\| \le 1} \prod_{k=1}^{K_1} \left(\exp\left[-G\left\{e^{\widetilde{M}\|\boldsymbol{b}\|} \widehat{\Lambda}_k(U_{i,M_{ik}})\right\}\right]\right)^{\delta_{i,M_{ik}}} \phi\left(\boldsymbol{b}; \widehat{\boldsymbol{\Sigma}}\right) d\boldsymbol{b}\right\} \\
&\quad + n^{-1} \sum_{i=1}^{n} \left\{\log \int_{\|\boldsymbol{b}\| > 1} \phi\left(\boldsymbol{b}; \widehat{\boldsymbol{\Sigma}}\right) d\boldsymbol{b}\right\} \\
&\le O(1) - n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K_1} \delta_{ik,M_{ik}} e^{\widetilde{M}} \widehat{\Lambda}_k(U_{ik,M_{ik}}).
\end{aligned}
$$

Therefore, for $k = 1, \ldots, K_1$, $\limsup_n \widehat{\Lambda}_k(\tau_k - \epsilon) < \infty$ with probability 1 for any $\epsilon > 0$. By choosing a sequence of ϵ decreasing to 0, it then follows from Helly's selection lemma that along a subsequence, $\widehat{\Lambda}_k \to \Lambda_{k*}$ pointwise on any interior set of $\mathcal{U}_k$ and $\widehat{\boldsymbol{\theta}} \to \boldsymbol{\theta}_* \equiv (\boldsymbol{\beta}_*, \boldsymbol{\gamma}_*)$. We denote $\mathcal{A}_* = (\Lambda_{1*}, \ldots, \Lambda_{K*})$.

We now show that $\boldsymbol{\theta}_* = \boldsymbol{\theta}_0$ and $\mathcal{A}_* = \mathcal{A}_0$. First, we consider the differentiability of Λ_{k*} for $k = K_1 + 1, \ldots, K$. By the definition of $\widetilde{\Lambda}_k$, $\widehat{\Lambda}_k(t)$ is absolutely continuous with respect to $\widetilde{\Lambda}_k(t)$, and

$$\widehat{\Lambda}_k(t) = \int_0^t \frac{\mathbb{P}_n \nu_k\left(s, \mathcal{O}; \boldsymbol{\theta}_0, \mathcal{A}_0\right)}{\left|\mathbb{P}_n \nu_k\left(s, \mathcal{O}; \widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}}\right)\right|} d\widetilde{\Lambda}_k(s), \tag{4.4}$$

where

$$\nu_k(t, \mathcal{O}; \boldsymbol{\theta}, \mathcal{A}) = \frac{\int_{\boldsymbol{b}} J_1\left(b, \mathcal{O}; \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathcal{A}\right) J_{2k}(t, \boldsymbol{b}, \mathcal{O}; \boldsymbol{\beta}, \gamma_k) \psi(\boldsymbol{b}; \boldsymbol{\Sigma}) d\boldsymbol{b}}{\int_{\boldsymbol{b}} J_1\left(b, \mathcal{O}; \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathcal{A}\right) \psi(\boldsymbol{b}; \boldsymbol{\Sigma}) d\boldsymbol{b}}.$$

To take limits on the two sides of equation (4.4), we first show that the denominator of the integrand is uniformly bounded away from zero. It follows from the Glivenko-Cantelli property in Lemma 4.1 that

$$\sup_{t \in \mathcal{U}_k} \left|\mathbb{P}_n \nu_k\left(t, \mathcal{O}; \boldsymbol{\theta}_0, \mathcal{A}_0\right) - \mathbb{P} \nu_k\left(t, \mathcal{O}; \boldsymbol{\theta}_0, \mathcal{A}_0\right)\right| \to_{a.s.} 0$$

and

$$\sup_{t \in \mathcal{U}_k} \left|\mathbb{P}_n \nu_k\left(t, \mathcal{O}; \widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}}\right) - \mathbb{P} \nu_k\left(t, \mathcal{O}; \boldsymbol{\theta}_*, \mathcal{A}_*\right)\right| \to_{a.s.} 0.$$

Note that for any $\epsilon > 0$,

$$\limsup_n \widehat{\Lambda}_k(\tau_k) \geq \int_0^{\tau_k} \frac{\mathbb{P} \nu_k\left(t, \mathcal{O}; \boldsymbol{\theta}_0, \mathcal{A}_0\right)}{\epsilon + \left|\mathbb{P} \nu_k\left(t, \mathcal{O}; \boldsymbol{\theta}_*, \mathcal{A}_*\right)\right|} d\Lambda_{k0}(s).$$

Let $\epsilon \to 0$. By the Monotone Convergence Theorem,

$$\int_0^{\tau_k} \frac{\mathbb{P} \nu_k\left(t, \mathcal{O}; \boldsymbol{\theta}_0, \mathcal{A}_0\right)}{\left|\mathbb{P} \nu_k\left(t, \mathcal{O}; \boldsymbol{\theta}_*, \mathcal{A}_*\right)\right|} d\Lambda_{k0}(t) < \infty.$$

We claim that $\min_{t \in \mathcal{U}_k} |\mathbb{P} \nu_k(t, \mathcal{O}; \boldsymbol{\theta}_*, \mathcal{A}_*)| > 0$. If this inequality does not hold, then there exists some $t_* \in \mathcal{U}_k$ such that $\mathbb{P} \nu_k(t_*, \mathcal{O}; \boldsymbol{\theta}_*, \mathcal{A}_*) = 0$. The function $\mathbb{P} \nu_k(t_*, \mathcal{O}; \boldsymbol{\theta}_*, \mathcal{A}_*)$ is right-differentiable almost everywhere. Thus, there exists $\delta > 0$ such that for $t \in (t_*, t_* + \delta)$,

$$\left|\mathbb{P} \nu_k\left(t, \mathcal{O}; \boldsymbol{\theta}_*, \mathcal{A}_*\right)\right| = \left|\mathbb{P} \nu_k\left(t, \mathcal{O}; \boldsymbol{\theta}_*, \mathcal{A}_*\right) - \mathbb{P} \nu_k\left(t_*, \mathcal{O}; \boldsymbol{\theta}_*, \mathcal{A}_*\right)\right| \leq O(1)|t - t_*|$$

almost everywhere. Hence,

$$\int_{t_*}^{t_*+\delta} \frac{1}{|t-t_*|} d\Lambda_{k0}(t) < \infty,$$

which is a contradiction. By taking the limits on both sides of (4.4), we conclude that $\Lambda_{k*}(t)$ is absolutely continuous with respect to $\Lambda_{k0}(t)$, so that $\Lambda_{k*}(t)$ is differentiable with respect to t. In addition, $d\widehat{\Lambda}_k(t)/d\widetilde{\Lambda}_k(t)$ converges to $d\Lambda_{k*}(t)/d\Lambda_{k0}(t)$ uniformly in t.

Define

$$m(\boldsymbol{\theta}, \mathcal{A}) = \log\left\{\frac{L(\boldsymbol{\theta}, \mathcal{A}) + L(\boldsymbol{\theta}_0, \widetilde{\mathcal{A}})}{2}\right\}$$

and

$$\mathcal{M} = \left\{m(\boldsymbol{\theta}, \mathcal{A}) : \boldsymbol{\theta} \in \Theta, \mathcal{A} \in \mathcal{D}_{1,\infty} \times \cdots \times \mathcal{D}_{K_1,\infty} \times \mathcal{D}_{K_1+1,M} \times \cdots \times \mathcal{D}_{K,M}\right\},$$

where $L(\boldsymbol{\theta}, \mathcal{A})$ is the objective function for a single subject, and $\mathcal{D}_{k,c} = \{\Lambda : \Lambda$ is increasing with $\Lambda(0) = 0, \Lambda(\tau_k) \le c\}$. By the concavity of the log function,

$$\mathbb{P}_n m(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}}) \ge \frac{1}{2}\left\{\mathbb{P}_n \log L(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}}) + \mathbb{P}_n \log L(\boldsymbol{\theta}_0, \widetilde{\mathcal{A}})\right\} \ge \mathbb{P}_n l(\boldsymbol{\theta}_0, \widetilde{\mathcal{A}}) = \mathbb{P}_n m(\boldsymbol{\theta}_0, \widetilde{\mathcal{A}}).$$

It follows from Lemma 4.1 that the class $\mathcal{M}$ is Glivenko-Cantelli. Thus,

$$\begin{aligned}
0 &\le \mathbb{P}_n m\left(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}}\right) - \mathbb{P}_n m\left(\boldsymbol{\theta}_0, \widetilde{\mathcal{A}}\right) \\
&= \mathbb{P}\left\{m\left(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}}\right) - m\left(\boldsymbol{\theta}_0, \widetilde{\mathcal{A}}\right)\right\} + o_P(1) \\
&= \mathbb{P}\log\left[\frac{L(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}}) + L(\boldsymbol{\theta}_0, \widetilde{\mathcal{A}})}{2L(\boldsymbol{\theta}_0, \widetilde{\mathcal{A}})}\right] + o_P(1) \\
&= \mathbb{P}\log\left\{\frac{1}{2} + \frac{\prod_{k=K_1+1}^K \widehat{\Lambda}_k\{Y_k\}^{\Delta_k} \int_{\boldsymbol{b}} J_1\left(\boldsymbol{b}, \mathcal{O}; \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}}, \widehat{\mathcal{A}}\right) \psi(\boldsymbol{b}; \widehat{\boldsymbol{\Sigma}}) d\boldsymbol{b}}{2\prod_{k=K_1+1}^K \widetilde{\Lambda}_k\{Y_k\}^{\Delta_k} \int_{\boldsymbol{b}} J_1\left(\boldsymbol{b}, \mathcal{O}; \boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, \widetilde{\mathcal{A}}\right) \psi(\boldsymbol{b}; \boldsymbol{\Sigma}_0) d\boldsymbol{b}}\right\} + o_P(1) \\
&\to \mathbb{P}\left[\log\left\{\frac{1}{2} + \frac{\prod_{k=K_1+1}^K \Lambda'_{k*}(Y_k)^{\Delta_k} \int_{\boldsymbol{b}} J_1\left(\boldsymbol{b}, \mathcal{O}; \boldsymbol{\beta}_*, \boldsymbol{\gamma}_*, \mathcal{A}_*\right) \psi(\boldsymbol{b}; \boldsymbol{\Sigma}_*) d\boldsymbol{b}}{2\prod_{k=K_1+1}^K \Lambda'_{k0}(Y_k)^{\Delta_k} \int_{\boldsymbol{b}} J_1\left(\boldsymbol{b}, \mathcal{O}; \boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, \mathcal{A}_0\right) \psi(\boldsymbol{b}; \boldsymbol{\Sigma}_0) d\boldsymbol{b}}\right\}\right],
\end{aligned}$$

such that the negative Kullback-Leibler information is positive. Therefore,

$$\int_{\boldsymbol{b}} \prod_{k=1}^{K_1}\left(\sum_{m=0}^{M_k} \delta_{km}\left[\exp\left\{-\int_0^{U_{km}} e^{\boldsymbol{\beta}_*^{\mathrm{T}} \boldsymbol{X}_k(s) + b_1} d\Lambda_{k*}(s)\right\} - \exp\left\{-\int_0^{U_{k,m+1}} e^{\boldsymbol{\beta}_*^{\mathrm{T}} \boldsymbol{X}_k(s) + b_1} d\Lambda_{k*}(s)\right\}\right]\right)$$

$$\times \prod_{k=K_1+1}^{K} \left[\left\{ e^{\boldsymbol{\beta}_*^{\mathrm{T}} \boldsymbol{X}_k(Y_k)+\gamma_{k*}b_1+b_2} \Lambda_{k*}(Y_k) \right\}^{\Delta_k} \exp\left\{ -\int_0^{Y_k} e^{\boldsymbol{\beta}_*^{\mathrm{T}} \boldsymbol{X}_k(s)+\gamma_{k*}b_1+b_2} d\Lambda_{k*}(s) \right\} \right] \psi(\boldsymbol{b}; \boldsymbol{\Sigma}_*) d\boldsymbol{b}$$

$$= \int_{\boldsymbol{b}} \prod_{k=1}^{K_1} \left(\sum_{m=0}^{M_k} \delta_{km} \left[\exp\left\{ -\int_0^{U_{km}} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(s)+b_1} d\Lambda_{k0}(s) \right\} - \exp\left\{ -\int_0^{U_{k,m+1}} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(s)+b_1} d\Lambda_{k0}(s) \right\} \right] \right)$$

$$\times \prod_{k=K_1+1}^{K} \left[\left\{ e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(Y_k)+\gamma_{k0}b_1+b_2} \Lambda_{k0}(Y_k) \right\}^{\Delta_k} \exp\left\{ -\int_0^{Y_k} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(s)+\gamma_{k0}b_1+b_2} d\Lambda_{k0}(s) \right\} \right] \psi(\boldsymbol{b}; \boldsymbol{\Sigma}_0) d\boldsymbol{b}$$

with probability 1. For any $k \in \{1, \ldots, K_1\}$ and $m \in \{0, \ldots, M_k\}$, we set $\delta_{km'} = 1$ in the above equation for $m' = m, \ldots, M_k$ and take the sum of the resulting equations to obtain

$$\begin{aligned}
& \int_{\boldsymbol{b}} \prod_{k=1}^{K_1} \exp\left\{ -\int_0^{U_{km}} e^{\boldsymbol{\beta}_*^{\mathrm{T}} \boldsymbol{X}_k(s)+b_1} d\Lambda_{k*}(s) \right\} \prod_{k=K_1+1}^{K} \left[\left\{ e^{\boldsymbol{\beta}_*^{\mathrm{T}} \boldsymbol{X}_k(Y_k)+\gamma_{k*}b_1+b_2} \Lambda_{k*}(Y_k) \right\}^{\Delta_k} \right. \\
& \left. \exp\left\{ -\int_0^{Y_k} e^{\boldsymbol{\beta}_*^{\mathrm{T}} \boldsymbol{X}_k(s)+\gamma_{k*}b_1+b_2} d\Lambda_{k*}(s) \right\} \right] \psi(\boldsymbol{b}; \boldsymbol{\Sigma}_*) d\boldsymbol{b} \\
= \quad & \int_{\boldsymbol{b}} \prod_{k=1}^{K_1} \exp\left\{ -\int_0^{U_{km}} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(s)+b_1} d\Lambda_{k0}(s) \right\} \prod_{k=K_1+1}^{K} \left[\left\{ e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(Y_k)+\gamma_{k0}b_1+b_2} \Lambda_{k0}(Y_k) \right\}^{\Delta_k} \right. \\
& \left. \exp\left\{ -\int_0^{Y_k} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(s)+\gamma_{k0}b_1+b_2} d\Lambda_{k0}(s) \right\} \right] \psi(\boldsymbol{b}; \boldsymbol{\Sigma}_0) d\boldsymbol{b}.
\end{aligned}$$

Because m is arbitrary, we can replace U_{km} in the above equation by any $t_k \in \mathcal{U}_k$. For $k = K_1+1, \ldots, K$, we set $\Delta_k = 1$ and integrate Y_k from 0 to $t_k \in \mathcal{U}_k$ to obtain

$$\begin{aligned}
& \int_{\boldsymbol{b}} \exp\left\{ -\sum_{k=1}^{K_1} \int_0^{t_k} e^{\boldsymbol{\beta}_*^{\mathrm{T}} \boldsymbol{X}_k(s)+b_1} d\Lambda_{k*}(s) - \sum_{k=K_1+1}^{K} \int_0^{t_k} e^{\boldsymbol{\beta}_*^{\mathrm{T}} \boldsymbol{X}_k(s)+\gamma_{k*}b_1+b_2} d\Lambda_{k*}(s) \right\} \psi(\boldsymbol{b}; \boldsymbol{\Sigma}_*) d\boldsymbol{b} \\
= & \int_{\boldsymbol{b}} \exp\left\{ -\sum_{k=1}^{K_1} \int_0^{t_k} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(s)+b_1} d\Lambda_{k0}(s) - \sum_{k=K_1+1}^{K} \int_0^{t_k} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(s)+\gamma_{k0}b_1+b_2} d\Lambda_{k0}(s) \right\} \psi(\boldsymbol{b}; \boldsymbol{\Sigma}_0) d\boldsymbol{b}.
\end{aligned} \tag{4.5}$$

For any $k = 1, \ldots, K_1$, we set $t_{k'} = 0$ for $k' \neq k$ in (4.5) to obtain

$$\begin{aligned}
& \int_{b_1} \exp\left\{ -e^{b_1} \int_0^{t_k} e^{\boldsymbol{\beta}_*^{\mathrm{T}} \boldsymbol{X}_k(s)} d\Lambda_{k*}(s) \right\} \phi(b_1; \sigma_{1*}^2) db_1 \\
= \quad & \int_{b_1} \exp\left\{ -e^{b_1} \int_0^{t_k} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(s)} d\Lambda_{k0}(s) \right\} \phi(b_1; \sigma_{10}^2) db_1.
\end{aligned}$$

By the arguments in the proof of Theorem 1 of Elbers and Ridder (1982), we find $\sigma_{1*}^2 = \sigma_{10}^2$ and

$$\int_0^{t_k} e^{\boldsymbol{\beta}_*^{\mathrm{T}} \boldsymbol{X}_k(s)} d\Lambda_{k*}(s) = \int_0^{t_k} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(s)} d\Lambda_{k0}(s). \tag{4.6}$$

We differentiate both sides with respect to t_k and take the logarithm to obtain

$$\boldsymbol{\beta}_*^{\mathrm{T}} \boldsymbol{X}_k(t_k) + \log \lambda_{k*}(t_k) = \boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(t_k) + \log \lambda_{k0}(t_k) \tag{4.7}$$

for $t_k \in \mathcal{U}_k$ and $k = 1, \ldots, K_1$. For $k = K_1 + 1, \ldots, K$, we set $t_{k'} = 0$ for $k' \notin \{1, k\}$ in (4.5) to obtain

$$\int_{\boldsymbol{b}} \exp\left\{-e^{b_1} \int_0^{t_1} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_1(s)} d\Lambda_{10}(s) - e^{\gamma_{k*} b_1 + b_2} \int_0^{t_k} e^{\boldsymbol{\beta}_*^{\mathrm{T}} \boldsymbol{X}_k(s)} d\Lambda_{k*}(s)\right\} \phi(b_1; \sigma_{10}^2) \phi(b_2; \sigma_{2*}^2) d\boldsymbol{b}$$

$$= \int_{\boldsymbol{b}} \exp\left\{-e^{b_1} \int_0^{t_1} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_1(s)} d\Lambda_{10}(s) - e^{\gamma_{k0} b_1 + b_2} \int_0^{t_k} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(s)} d\Lambda_{k0}(s)\right\} \phi(b_1; \sigma_{10}^2) \phi(b_2; \sigma_{20}^2) d\boldsymbol{b}.$$

We let $b_{3k*} = \gamma_{k*} b_1 + b_2$ and $b_{3k0} = \gamma_{k0} b_1 + b_2$ to obtain

$$\begin{aligned}
& \int_{b_1} \exp\left\{-e^{b_1} \int_0^{t_1} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_1(s)} d\Lambda_{10}(s)\right\} \\
& \times \left[\int_{b_{3k*}} \exp\left\{-e^{b_{3k*}} \int_0^{t_k} e^{\boldsymbol{\beta}_*^{\mathrm{T}} \boldsymbol{X}_k(s)} d\Lambda_{k*}(s)\right\} \phi(b_{3k*} - \gamma_{k*} b_1; \sigma_{2*}^2) db_{3k*}\right] db_1 \\
= & \int_{b_1} \exp\left\{-e^{b_1} \int_0^{t_1} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_1(s)} d\Lambda_{10}(s)\right\} \\
& \times \left[\int_{b_{3k0}} \exp\left\{-e^{b_{3k0}} \int_0^{t_k} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(s)} d\Lambda_{k0}(s)\right\} \phi(b_{3k0} - \gamma_{k0} b_1; \sigma_{20}^2) db_{3k0}\right] db_1.
\end{aligned}$$

We apply the inverse Laplace transform to both sides to obtain

$$\begin{aligned}
& \int_{b_{3k*}} \exp\left\{-e^{b_{3k*}} \int_0^{t_k} e^{\boldsymbol{\beta}_*^{\mathrm{T}} \boldsymbol{X}_k(s)} d\Lambda_{k*}(s)\right\} \phi(b_{3k*} - \gamma_{k*} b_1; \sigma_{2*}^2) db_{3k*} \\
= & \int_{b_{3k0}} \exp\left\{-e^{b_{3k0}} \int_0^{t_k} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(s)} d\Lambda_{k0}(s)\right\} \phi(b_{3k0} - \gamma_{k0} b_1; \sigma_{20}^2) db_{3k0}
\end{aligned}$$

for any b_1. By the arguments in the proof of Theorem 1 of Elbers and Ridder (1982), we find

$\sigma^2_{2*} = \sigma^2_{20}$, $\gamma_{k*} = \gamma_{k0}$, and

$$\int_0^{t_k} e^{\boldsymbol{\beta}_*^{\mathrm{T}} \boldsymbol{X}_k(s)} d\Lambda_{k*}(s) = \int_0^{t_k} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(s)} d\Lambda_{k0}(s) \tag{4.8}$$

for $k = K_1 + 1, \ldots, K$. We differentiate both sides with respect to t_k and take the logarithm to obtain

$$\boldsymbol{\beta}_*^{\mathrm{T}} \boldsymbol{X}_k(t_k) + \log \lambda_{k*}(t_k) = \boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(t_k) + \log \lambda_{k0}(t_k) \tag{4.9}$$

for $t_k \in \mathcal{U}_k$ and $k = K_1 + 1, \ldots, K$. By Condition 5, (4.7), and (4.9), $\boldsymbol{\beta}_* = \boldsymbol{\beta}_0$ and $\lambda_{k*}(t_k) = \lambda_{k^*}(t_k)$ for $k = 1, \ldots, K$ and $t_k \in \mathcal{U}_k$. We let $\boldsymbol{X}_k(t) = 0$ by redefining $\boldsymbol{X}_k(t)$ to centre at a deterministic function in the support of $\boldsymbol{X}_k(t)$ in (4.6) and (4.8) to obtain $\Lambda_{k*}(t_k) = \Lambda_{k^*}(t_k)$ for $k = 1, \ldots, K$ and $t_k \in \mathcal{U}_k$. We conclude that $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| \to 0$ and $|\widehat{\Lambda}_k(t_k) - \Lambda_{k0}(t_k)| \to 0$ for any $t_k \in \mathcal{U}_k$. Because $\mathcal{A}_0$ is continuous, $\widehat{\mathcal{A}}$ converges uniformly to $\mathcal{A}_0$ on $\prod_k \mathcal{U}_k$.

4.6.2 Proof of Theorem 4.2

Let

$$H_{1k}(t, u, v, \boldsymbol{b}, \mathcal{O}; \boldsymbol{\theta}, \mathcal{A}) = \frac{J_1(\boldsymbol{b}, \mathcal{O}; \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathcal{A}) Q_1(t, u, v, b_1, \boldsymbol{X}_k; \boldsymbol{\beta}, \Lambda_k) \psi(\boldsymbol{b}; \boldsymbol{\Sigma})}{\int_{\boldsymbol{b}'} J_1(\boldsymbol{b}', \mathcal{O}; \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathcal{A}) \psi(\boldsymbol{b}'; \boldsymbol{\Sigma}) d\boldsymbol{b}'}$$

for $k = 1, \ldots, K_1$, and

$$H_{2k}(t, \boldsymbol{b}, \mathcal{O}; \boldsymbol{\theta}, \mathcal{A}) = \frac{J_1(\boldsymbol{b}, \mathcal{O}; \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathcal{A}) Q_2(t, Y_k, \boldsymbol{b}, \boldsymbol{X}_k; \boldsymbol{\beta}, \gamma_k) \psi(\boldsymbol{b}; \boldsymbol{\Sigma})}{\int_{\boldsymbol{b}'} J_1(\boldsymbol{b}', \mathcal{O}; \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathcal{A}) \psi(\boldsymbol{b}'; \boldsymbol{\Sigma}) d\boldsymbol{b}'}$$

for $k = K_1 + 1, \ldots, K$, where

$$\begin{aligned}
& Q_1(t, u, v, b_1, \boldsymbol{X}_k; \boldsymbol{\beta}, \Lambda_k) \\
= \; & e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_k(t) + b_1} \left[\frac{I(v \geq t) \exp\left\{ -\int_0^v e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_k(s) + b_1} d\Lambda_k(s) \right\}}{\exp\left\{ -\int_0^u e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_k(s) + b_1} d\Lambda_k(s) \right\} - \exp\left\{ -\int_0^v e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_k(s) + b_1} d\Lambda_k(s) \right\}} \right. \\
& \left. - \frac{I(u \geq t) \exp\left\{ -\int_0^u e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_k(s) + b_1} d\Lambda_k(s) \right\}}{\exp\left\{ -\int_0^u e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_k(s) + b_1} d\Lambda_k(s) \right\} - \exp\left\{ -\int_0^v e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_k(s) + b_1} d\Lambda_k(s) \right\}} \right],
\end{aligned}$$

and

$$Q_2(t, u, \boldsymbol{b}, \boldsymbol{X}_k; \boldsymbol{\beta}, \gamma_k) = -I(u \geq t) e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_k(t) + \gamma_k b_1 + b_2}.$$

Then, the score function for $\boldsymbol{\theta}$ is $\boldsymbol{l_\theta}(\boldsymbol{\theta},\mathcal{A}) = (\boldsymbol{l_\beta}(\boldsymbol{\theta},\mathcal{A})^{\mathrm{T}}, l_{\gamma_{K_1+1}}(\boldsymbol{\theta},\mathcal{A}),\ldots,l_{\gamma_K}(\boldsymbol{\theta},\mathcal{A}), l_{\sigma_1^2}(\boldsymbol{\theta},\mathcal{A}), l_{\sigma_2^2}(\boldsymbol{\theta},\mathcal{A}))^{\mathrm{T}}$, where

$$\begin{aligned}
\boldsymbol{l_\beta}(\boldsymbol{\theta},\mathcal{A}) &= \sum_{k=1}^{K_1}\sum_{m=0}^{M_k}\delta_{km}\int_0^{\tau_k}\int_{\boldsymbol{b}} H_{1k}(t,U_{km},U_{k,m+1},\boldsymbol{b},\mathcal{O};\boldsymbol{\theta},\mathcal{A})d\boldsymbol{b}\boldsymbol{X}_k(t)d\Lambda_k(t), \\
&+ \sum_{k=K_1+1}^{K}\left\{\Delta_k\boldsymbol{X}_k(Y_k) + \int_0^{\tau_k}\int_{\boldsymbol{b}} H_{2k}(t,\boldsymbol{b},\mathcal{O};\boldsymbol{\theta},\mathcal{A})d\boldsymbol{b}\boldsymbol{X}_k(t)d\Lambda_k(t)\right\},
\end{aligned}$$

$$l_{\gamma_k}(\boldsymbol{\theta},\mathcal{A}) = \Delta_k\frac{\int_{\boldsymbol{b}} b_1 J_1(\boldsymbol{b},\mathcal{O};\boldsymbol{\beta},\boldsymbol{\gamma},\mathcal{A})\psi(\boldsymbol{b};\boldsymbol{\Sigma})d\boldsymbol{b}}{\int_{\boldsymbol{b}} J_1(\boldsymbol{b},\mathcal{O};\boldsymbol{\beta},\boldsymbol{\gamma},\mathcal{A})\psi(\boldsymbol{b};\boldsymbol{\Sigma})d\boldsymbol{b}} + \int_0^{\tau_k}\int_{\boldsymbol{b}} b_1 H_{2k}(t,\boldsymbol{b},\mathcal{O};\boldsymbol{\theta},\mathcal{A})d\boldsymbol{b}d\Lambda_k(t),$$

$$l_{\sigma_j^2}(\boldsymbol{\theta},\mathcal{A}) = \frac{\int J_1(\boldsymbol{b},\mathcal{O};\boldsymbol{\beta},\boldsymbol{\gamma},\mathcal{A})\phi'_{\sigma_j^2}(b_j;\sigma_j^2)\phi(b_{3-j};\sigma_{3-j}^2)d\boldsymbol{b}}{\int J_1(\boldsymbol{b},\mathcal{O};\boldsymbol{\beta},\boldsymbol{\gamma},\mathcal{A})\psi(\boldsymbol{b};\boldsymbol{\Sigma})d\boldsymbol{b}}$$

for $j=1,2$, and $\phi'_{\sigma_j^2}(b_j;\sigma_j^2)$ is the derivative of $\phi(b_j;\sigma_j^2)$ with respect to σ_j^2. The score operator for $\mathcal{A}$ along the submodel $d\mathcal{A}_{\epsilon,\boldsymbol{h}} = ((1+\epsilon h_1)d\Lambda_1,\ldots,(1+\epsilon h_K)d\Lambda_K)^{\mathrm{T}}$ for $\boldsymbol{h}=(h_1,\ldots,h_K)$ with $h_k\in L_2(\mu_k)$ for $k=1,\ldots,K_1$ and $h_k\in BV_1(\mathcal{U}_k)$ for $k=K_1+1,\ldots,K$ is

$$\begin{aligned}
l_{\mathcal{A}}(\boldsymbol{\theta},\mathcal{A})(\boldsymbol{h}) &= \sum_{k=1}^{K_1}\sum_{m=0}^{M_k}\delta_{km}\int_0^{\tau_k}\int_{\boldsymbol{b}} H_{1k}(t,U_{km},U_{k,m+1},\boldsymbol{b},\mathcal{O};\boldsymbol{\theta},\mathcal{A})d\boldsymbol{b}h_k(t)d\Lambda_k(t) \\
&+ \sum_{k=K_1+1}^{K}\left\{\Delta_k h_k(Y_k) + \int_0^{\tau_k}\int_{\boldsymbol{b}} H_{2k}(t,\boldsymbol{b},\mathcal{O};\boldsymbol{\theta},\mathcal{A})d\boldsymbol{b}h_k(t)d\Lambda_k(t)\right\},
\end{aligned}$$

where $BV_1(\mathcal{B})$ denotes the set of functions on $\mathcal{B}$ with total variation bounded by 1.

Clearly,

$$\mathbb{G}_n\left\{\boldsymbol{l_\theta}(\widehat{\boldsymbol{\theta}},\widehat{\mathcal{A}})\right\} = -\sqrt{n}\mathbb{P}\left\{\boldsymbol{l_\theta}(\widehat{\boldsymbol{\theta}},\widehat{\mathcal{A}}) - \boldsymbol{l_\theta}(\boldsymbol{\theta}_0,\mathcal{A}_0)\right\},$$

and

$$\mathbb{G}_n\left\{l_{\mathcal{A}}(\widehat{\boldsymbol{\theta}},\widehat{\mathcal{A}})(\boldsymbol{h})\right\} = -\sqrt{n}\mathbb{P}\left\{l_{\mathcal{A}}(\widehat{\boldsymbol{\theta}},\widehat{\mathcal{A}})(\boldsymbol{h}) - l_{\mathcal{A}}(\boldsymbol{\theta}_0,\mathcal{A}_0)(\boldsymbol{h})\right\}.$$

We apply the Taylor series expansions at $(\boldsymbol{\theta}_0,\mathcal{A}_0)$ to the right sides of the above two equations. In light of Lemma 4.3, the second-order terms are bounded by

$$\begin{aligned}
&O_P(1)\sqrt{n}E\left[\sum_{k=1}^{K_1}\sum_{m=0}^{M_k}\left\{\widehat{\Lambda}_k(U_{km})-\Lambda_{k0}(U_{km})\right\}^2 + \sum_{k=K_1+1}^{K}\left\{\widehat{\Lambda}_k(Y_k)-\Lambda_{k0}(Y_k)\right\}^2\right. \\
&\left.+\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right\|^2 + \|\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0\|^2 + \left\|\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}_0\right\|^2\right]
\end{aligned}$$

$$
\begin{aligned}
&= \sqrt{n}\left\{O_P(n^{-2/3}) + O_P(1)\left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right\|^2 + O_P(1)\left\|\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\right\|^2 + O_P(1)\left\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_0\right\|^2\right\} \\
&= O_P\left(n^{-1/6} + \sqrt{n}\left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right\|^2 + \sqrt{n}\left\|\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\right\|^2 + \sqrt{n}\left\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_0\right\|^2\right).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\mathbb{G}_n\left\{\boldsymbol{l}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}})\right\} &= -\sqrt{n}\mathbb{P}\left\{\boldsymbol{l}_{\boldsymbol{\theta}\boldsymbol{\theta}}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) + \boldsymbol{l}_{\boldsymbol{\theta}\mathcal{A}}(\widehat{\mathcal{A}} - \mathcal{A}_0)\right\} \\
&\quad + O_P\left(n^{-1/6} + \sqrt{n}\left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right\|^2 + \sqrt{n}\left\|\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\right\|^2 + \sqrt{n}\left\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_0\right\|^2\right),
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbb{G}_n\left\{l_{\mathcal{A}}(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}})(\boldsymbol{h})\right\} &= -\sqrt{n}\mathbb{P}\left\{l_{\mathcal{A}\boldsymbol{\theta}}(\boldsymbol{h})\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) + l_{\mathcal{A}\mathcal{A}}(\boldsymbol{h}, \widehat{\mathcal{A}} - \mathcal{A}_0)\right\} \\
&\quad + O_P\left(n^{-1/6} + \sqrt{n}\left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right\|^2 + \sqrt{n}\left\|\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\right\|^2 + \sqrt{n}\left\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_0\right\|^2\right),
\end{aligned}
$$

where $\boldsymbol{l}_{\boldsymbol{\theta}\boldsymbol{\theta}}$ is the second derivative of $\boldsymbol{l}(\boldsymbol{\theta}, \mathcal{A})$ with respect to $\boldsymbol{\theta}$, $\boldsymbol{l}_{\boldsymbol{\theta}\mathcal{A}}(\boldsymbol{h})$ is the derivative of $\boldsymbol{l}_{\boldsymbol{\theta}}$ along the submodel $d\mathcal{A}_{\epsilon,h}$, $l_{\mathcal{A}\boldsymbol{\theta}}(\boldsymbol{h})$ is the derivative of $l_{\mathcal{A}}(\boldsymbol{h})$ with respect to $\boldsymbol{\theta}$, and $l_{\mathcal{A}\mathcal{A}}(\boldsymbol{h}, \widehat{\mathcal{A}} - \mathcal{A}_0)$ is the derivative of $l_{\mathcal{A}}(\boldsymbol{h})$ along the submodel $d\mathcal{A}_0 + \epsilon d(\widehat{\mathcal{A}} - \mathcal{A}_0)$. All derivatives are evaluated at $(\boldsymbol{\theta}_0, \mathcal{A}_0)$.

If the least favorable direction exists, we denote it as $\boldsymbol{h}^* = (\boldsymbol{h}_1^*, \ldots, \boldsymbol{h}_K^*)$, where $\boldsymbol{h}_k^*$ $(k = 1, \ldots, K_1)$ is $(p + K_2 + 2)$-dimensional vector of functions in $L_2(\mu_k)$ and $\boldsymbol{h}_k^*$ $(k = K_1 + 1, \ldots, K)$ is $(p + K_2 + 2)$-dimensional vector of functions in $L_2(\mathcal{U}_k)$. We first show the existence of $\boldsymbol{h}^*$, which is the solution to $l_{\mathcal{A}}^* l_{\mathcal{A}}(\boldsymbol{h}^*) = l_{\mathcal{A}}^* \boldsymbol{l}_{\boldsymbol{\theta}}$ with $l_{\mathcal{A}}^*$ as the adjoint operator of $l_{\mathcal{A}}$. Let $\mathcal{Q} = \prod_{k=1}^{K_1} L_2(\mu_k) \times \prod_{k=K_1+1}^{K} L_2(\mathcal{U}_k)$. We equip $\mathcal{Q}$ with an inner product defined as

$$
< \boldsymbol{h}^{(1)}, \boldsymbol{h}^{(2)} >= \sum_{k=1}^{K_1} \int_{\mathcal{U}_k} h_k^{(1)} h_k^{(2)} d\mu_k(t) + \sum_{k=K_1+1}^{K} \int_0^{\tau_k} h_k^{(1)} h_k^{(2)} d\Lambda_{k0}(t),
$$

where $\boldsymbol{h}^{(1)} = (h_1^{(1)}, \ldots, h_K^{(1)})$ and $\boldsymbol{h}^{(2)} = (h_1^{(2)}, \ldots, h_K^{(2)})$. On the same space, we define

$$
\begin{aligned}
\|\boldsymbol{h}\| &= \mathbb{P}\left\{l_{\mathcal{A}}\left(\boldsymbol{\theta}_0, \mathcal{A}_0\right)(\boldsymbol{h})^2\right\}^{1/2} \\
&= \mathbb{P}\left(\left[\sum_{k=1}^{K_1} \sum_{m=0}^{M_k} \delta_{km} \int_0^{\tau_k} \int_{\boldsymbol{b}} H_{1k}(t, U_{km}, U_{k,m+1}, \boldsymbol{b}, \mathcal{O}; \boldsymbol{\theta}_0, \mathcal{A}_0) d\boldsymbol{b} h_k(t) d\Lambda_{k0}(t)\right.\right.
\end{aligned}
$$

$$+\sum_{k=K_1+1}^{K}\left\{\Delta h_k(Y_k)+\int_0^{\tau_k}\int_{\boldsymbol{b}} H_{2k}(t,\boldsymbol{b},\mathcal{O};\boldsymbol{\theta}_0,\mathcal{A}_0)d\boldsymbol{b}h_k(t)d\Lambda_{k0}(t)\right\}\Bigg]^2\Bigg)^{1/2}$$

for $\boldsymbol{h}=(h_1,\ldots,h_K)$. It is easy to show that $\|\cdot\|$ is a seminorm on $\mathcal{Q}$. Furthermore, if $\|\boldsymbol{h}\|=0$, then $\mathbb{P}\{l_{\mathcal{A}}(\boldsymbol{\theta}_0,\mathcal{A}_0)(\boldsymbol{h})^2\}=0$. Thus, with probability 1, $l_{\mathcal{A}}(\boldsymbol{\theta}_0,\mathcal{A}_0)(\boldsymbol{h})=0$. By the arguments in the proof of Lemma 4.3, $h_k(t_k)=0$ for $t_k\in\mathcal{U}_k$ for $k=1,\ldots,K$. Clearly, $\|\boldsymbol{h}\|\le c<\boldsymbol{h},\boldsymbol{h}>^{1/2}$ for some constant c by the Cauchy-Schwarz inequality. According to the bounded inverse theorem in Banach spaces, we have $<\boldsymbol{h},\boldsymbol{h}>^{1/2}\le\widetilde{c}\|\boldsymbol{h}\|$ for another constant $\widetilde{c}$. By the Lax-Milgram theorem (Zeidler, 1995), $\boldsymbol{h}^*$ exists and satisfies that for any $t_k\in\mathcal{U}_k$,

$$\int_0^{\tau_k}\mathbb{P}\left\{\sum_{m=0}^{M_k}\delta_{km}\int_{\boldsymbol{b}}H_{1k}(t_k,U_{km},U_{k,m+1},\boldsymbol{b},\mathcal{O};\boldsymbol{\theta}_0,\mathcal{A}_0)d\boldsymbol{b}\int_{\boldsymbol{b}}H_{1k}(s,U_{km},U_{k,m+1},\boldsymbol{b},\mathcal{O};\boldsymbol{\theta}_0,\mathcal{A}_0)d\boldsymbol{b}\right\}$$
$$\times\boldsymbol{h}_k^*(s)d\Lambda_{k0}(s)$$
$$=\mathbb{P}\left\{\sum_{m=0}^{M_k}\delta_{km}\int_{\boldsymbol{b}}H_{1k}(t_k,U_{km},U_{k,m+1},\boldsymbol{b},\mathcal{O};\boldsymbol{\theta}_0,\mathcal{A}_0)d\boldsymbol{b}\boldsymbol{l}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0,\mathcal{A}_0)\right\}\tag{4.10}$$

for $k=1,\ldots,K_1$ and

$$\int_0^{\tau_k}\mathbb{P}\left(\left[I(t_k\le C_k)\exp\{-\Lambda_k(t_k)\}+\int_{\boldsymbol{b}}H_{2k}(t_k,\boldsymbol{b},\mathcal{O};\boldsymbol{\theta}_0,\mathcal{A}_0)d\boldsymbol{b}\right]\int_{\boldsymbol{b}}H_{2k}(s,\boldsymbol{b},\mathcal{O};\boldsymbol{\theta}_0,\mathcal{A}_0)d\boldsymbol{b}\right)$$
$$\times\boldsymbol{h}_k^*(s)d\Lambda_{k0}(s)+\mathbb{P}\left[I(t_k\le C_k)\exp\{-\Lambda_k(t_k)\}\right]\boldsymbol{h}_k^*(t_k)$$
$$=\mathbb{P}\left\{E\{\boldsymbol{l}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0,\mathcal{A}_0)|T_k=t_k\}I(t_k\le C_k)\exp\{-\Lambda_k(t_k)\}+\int_{\boldsymbol{b}}H_{2k}(t_k,\boldsymbol{b},\mathcal{O};\boldsymbol{\theta}_0,\mathcal{A}_0)d\boldsymbol{b}\boldsymbol{l}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0,\mathcal{A}_0)\right\}$$

for $k=K_1+1,\ldots,K$. We differentiate (4.10) with respect to t_k to obtain

$$q_{k1}(t_k)\boldsymbol{h}_k^*(t_k)+\sum_{k'=1}^{K}\int_{t_k}^{\tau_k}q_{k2}(s,t_k)\boldsymbol{h}_{k'}^*(s)ds+\int_0^{t_k}q_{k3}(s,t_k)\boldsymbol{h}_k^*(s)ds=\boldsymbol{q}_{k4}(t_k),$$

where $q_{k1}(t_k)>0$ and q_{kj} $(k=1,\ldots,K;j=1,2,3)$ and $\boldsymbol{q}_{k4}$ $(k=1,\ldots,K)$ are continuously differentiable functions. Thus, $\boldsymbol{h}^*$ can be expanded to be a continuously differentiable function in $[0,\tau_k]^K$ with bounded total variation. It then follows that

$$\mathbb{G}_n\left\{\boldsymbol{l}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}},\widehat{\mathcal{A}})\right\}-\mathbb{G}_n\left\{l_{\mathcal{A}}(\widehat{\boldsymbol{\theta}},\widehat{\mathcal{A}})(\boldsymbol{h}^*)\right\}$$

$$
\begin{aligned}
=& -\sqrt{n}\mathbb{P}\left\{\boldsymbol{l}_{\boldsymbol{\theta}\boldsymbol{\theta}}\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)+\boldsymbol{l}_{\boldsymbol{\theta}\mathcal{A}}(\widehat{\mathcal{A}}-\mathcal{A}_0)\right\}+\sqrt{n}\mathbb{P}\left\{l_{\mathcal{A}\boldsymbol{\theta}}(\boldsymbol{h}^*)\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)+l_{\mathcal{A}\mathcal{A}}(\boldsymbol{h}^*,\widehat{\mathcal{A}}-\mathcal{A}_0)\right\}\\
&+O_P\left(n^{-1/6}+\sqrt{n}\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right\|^2+\sqrt{n}\left\|\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0\right\|^2+\sqrt{n}\left\|\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}_0\right\|^2\right)\\
=&\sqrt{n}\mathbb{P}\left[\left\{\boldsymbol{l}_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}_0,\mathcal{A}_0\right)-l_{\mathcal{A}}\left(\boldsymbol{\theta}_0,\mathcal{A}_0\right)\left(\boldsymbol{h}^*\right)\right\}^{\otimes 2}\right]\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)\\
&+O_P\left(n^{-1/6}+\sqrt{n}\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right\|^2+\sqrt{n}\left\|\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0\right\|^2+\sqrt{n}\left\|\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}_0\right\|^2\right).
\end{aligned}
$$

Using arguments in the proof of Lemma 4.2, we can show that $\boldsymbol{l}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0,\mathcal{A}_0)-l_{\mathcal{A}}(\boldsymbol{\theta}_0,\mathcal{A}_0)(\boldsymbol{h}^*)$ belongs to a Donsker class. Next, we show that the matrix $\mathbb{P}[\{\boldsymbol{l}_{\boldsymbol{\theta}}-l_{\mathcal{A}}(\boldsymbol{h}^*)\}^{\otimes 2}]$ is invertible. If the matrix is singular, then there exists a vector $\boldsymbol{v}\equiv(\boldsymbol{v}_1,\boldsymbol{v}_2,v_3,v_4)^{\mathrm{T}}$ with $\boldsymbol{v}_1\in\mathbb{R}^p$, $\boldsymbol{v}_2\equiv(v_{2,K_1+1},\ldots,v_{2K})\in\mathbb{R}^{K_2}$, and $v_3,v_4\in\mathbb{R}$ such that $\boldsymbol{v}^{\mathrm{T}}E[\{\boldsymbol{l}_{\boldsymbol{\theta}}-l_{\mathcal{A}}(\boldsymbol{h}^*)\}^{\otimes 2}]\boldsymbol{v}=0$. It follows that, with probability 1, the score function along the submodel $\{\boldsymbol{\theta}_0+\epsilon\boldsymbol{v},\mathcal{A}_\epsilon(\boldsymbol{v}^{\mathrm{T}}\boldsymbol{h}^*)\}$ is zero. That is,

$$
\begin{aligned}
&\int_{\boldsymbol{b}}\Bigg(\sum_{k=1}^{K_1}\sum_{m=0}^{M_k}\delta_{km}\int_0^{\tau_k}Q_1(t,U_{km},U_{k,m+1},b_1,\boldsymbol{X}_k;\boldsymbol{\beta}_0,\Lambda_{k0})\left\{\boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{X}_k(t)-\boldsymbol{v}^{\mathrm{T}}\boldsymbol{h}_k^*(t)\right\}d\Lambda_{k0}(t)\\
&-v_3\frac{\phi'_{\sigma_{10}^2}(b_1,\sigma_{10}^2)}{\phi(b_1;\sigma_{10}^2)}-v_4\frac{\phi'_{\sigma_{20}^2}(b_2,\sigma_{20}^2)}{\phi(b_2;\sigma_{20}^2)}+\sum_{k=K_1+1}^{K}\Bigg[\Delta_k\left\{\boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{X}_k(Y_k)+v_{2k}b_1-\boldsymbol{v}^{\mathrm{T}}\boldsymbol{h}_k^*(Y_k)\Lambda'_{k0}(Y_k)\right\}\\
&+\int_0^{\tau_k}Q_2(t,Y_k,\boldsymbol{b},\boldsymbol{X}_k;\boldsymbol{\beta}_0,\gamma_{k0})\left\{\boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{X}_k(t)+v_{2k}b_1-\boldsymbol{v}^{\mathrm{T}}\boldsymbol{h}_k^*(t)\right\}d\Lambda_{k0}(t)\Bigg]\Bigg)\\
&\times J_1(\boldsymbol{b},\mathcal{O};\boldsymbol{\beta}_0,\mathcal{A}_0)\psi(\boldsymbol{b};\boldsymbol{\Sigma}_0)d\boldsymbol{b}=0
\end{aligned}
$$

with probability 1. For any $t_k\in\mathcal{U}_k$ for $k=K_1+1,\ldots,K$, we let $\Delta_k=0$ and set $Y_k=t_k$ to obtain

$$
\begin{aligned}
&\int_{\boldsymbol{b}}\Bigg[\sum_{k=1}^{K_1}\sum_{m=0}^{M_k}\delta_{km}\int_0^{\tau_k}Q_1(s,U_{km},U_{k,m+1},b_1,\boldsymbol{X}_k;\boldsymbol{\beta}_0,\Lambda_{k0})\left\{\boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{X}_k(s)-\boldsymbol{v}^{\mathrm{T}}\boldsymbol{h}_k^*(s)\right\}d\Lambda_{k0}(s)\\
&-v_3\frac{\phi'_{\sigma_{10}^2}(b_1,\sigma_{10}^2)}{\phi(b_1;\sigma_{10}^2)}-v_4\frac{\phi'_{\sigma_{20}^2}(b_2,\sigma_{20}^2)}{\phi(b_2;\sigma_{20}^2)}\\
&+\sum_{k=K_1+1}^{K}\int_0^{t_k}e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_k(s)+\gamma_{k0}b_1+b_2}\left\{\boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{X}_k(s)+v_{2k}b_1-\boldsymbol{v}^{\mathrm{T}}\boldsymbol{h}_k^*(s)\right\}d\Lambda_{k0}(s)\Bigg]\\
&\times\left(\prod_{k=1}^{K_1}\sum_{m=0}^{M_k}\delta_{km}\left[\exp\left\{-\int_0^{U_{km}}e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_k(s)+b_1}d\Lambda_{k0}(s)\right\}-\exp\left\{-\int_0^{U_{k,m+1}}e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_k(s)+b_1}d\Lambda_{k0}(s)\right\}\right]\right)\\
&\times\prod_{k=K_1+1}^{K}\exp\left\{-\int_0^{t}e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_k(s)+\gamma_{k0}b_1+b_2}d\Lambda_{k0}(s)\right\}\psi(\boldsymbol{b};\boldsymbol{\Sigma}_0)d\boldsymbol{b}=0.
\end{aligned}
$$

For any $k=1,\ldots,K_1$ and $m_k\in\{0,\ldots,M_k\}$, we sum over all possible δ_{k,m'_k} with $m'_k=m_k,\ldots,M_k$

to obtain

$$\int_{\boldsymbol{b}} \left[\sum_{k=1}^{K_1} \int_0^{U_{k,m_k}} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(s)+b_1} \left\{ \boldsymbol{v}_1^{\mathrm{T}} \boldsymbol{X}_k(s) - \boldsymbol{v}^{\mathrm{T}} \boldsymbol{h}_k^*(s) \right\} d\Lambda_{k0}(s) - v_3 \frac{\phi'_{\sigma_{10}^2}(b_1, \sigma_{10}^2)}{\phi(b_1; \sigma_{10}^2)} - v_4 \frac{\phi'_{\sigma_{20}^2}(b_2, \sigma_{20}^2)}{\phi(b_2; \sigma_{20}^2)} \right.$$

$$\left. + \sum_{k=K_1+1}^{K} \int_0^{t_k} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(s)+\gamma_{k0} b_1 + b_2} \left\{ \boldsymbol{v}_1^{\mathrm{T}} \boldsymbol{X}_k(s) + v_{2k} b_1 - \boldsymbol{v}^{\mathrm{T}} \boldsymbol{h}_k^*(s) \right\} d\Lambda_{k0}(s) \right]$$

$$\times \exp \left\{ - \sum_{k=1}^{K_1} \int_0^{U_{k,m_k}} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(s)+b_1} d\Lambda_{k0}(s) - \sum_{k=K_1+1}^{K} \int_0^{t_k} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(s)+\gamma_{k0} b_1 + b_2} d\Lambda_{k0}(s) \right\} \psi(\boldsymbol{b}; \boldsymbol{\Sigma}_0) d\boldsymbol{b} = 0.$$

Because m_k is arbitrary, we can replace U_{k,m_k} in the above equation by any $t_k \in \mathcal{U}_k$. We apply the inverse Laplace transform to obtain

$$\sum_{k=1}^{K_1} \int_0^{t_k} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(s)+b_1} \left\{ \boldsymbol{v}_1^{\mathrm{T}} \boldsymbol{X}_k(s) - \boldsymbol{v}^{\mathrm{T}} \boldsymbol{h}_k^*(s) \right\} d\Lambda_{k0}(s) - v_3 \frac{\phi'_{\sigma_{10}^2}(b_1, \sigma_{10}^2)}{\phi(b_1; \sigma_{10}^2)} - v_4 \frac{\phi'_{\sigma_{20}^2}(b_2, \sigma_{20}^2)}{\phi(b_2; \sigma_{20}^2)}$$
$$+ \sum_{k=K_1+1}^{K} \int_0^{t_k} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(s)+\gamma_{k0} b_1 + b_2} \left\{ \boldsymbol{v}_1^{\mathrm{T}} \boldsymbol{X}_k(s) + v_{2k} b_1 - \boldsymbol{v}^{\mathrm{T}} \boldsymbol{h}_k^*(s) \right\} d\Lambda_{k0}(s) = 0$$

for any b_1 and b_2. Therefore, $\boldsymbol{v}_2 = \boldsymbol{0}$, $v_3 = v_4 = 0$, and

$$\sum_{k=1}^{K} \int_0^{t_k} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(s)+b_1} \boldsymbol{v}_1^{\mathrm{T}} \left\{ \boldsymbol{X}_k(s) - \boldsymbol{h}_k^*(s) \right\} d\Lambda_{k0}(s) = 0.$$

We differentiate both sides with respect to t_k to obtain $\boldsymbol{v}_1^{\mathrm{T}} \{ \boldsymbol{X}_k(t_k) - \boldsymbol{h}_k^*(t_k) \} = 0$ for $t_k \in \mathcal{U}_k$ and $k = 1, \ldots, K$. By Condition 5, $\boldsymbol{v}_1 = \boldsymbol{0}$. Hence, the matrix $E\big[\{\boldsymbol{l}_\theta - l_{\mathcal{A}}(\boldsymbol{h}^*)\}^{\otimes 2}\big]$ is invertible.

Because the matrix $\mathbb{P}[\{\boldsymbol{l}_{\boldsymbol{\theta}} - l_{\mathcal{A}}(\boldsymbol{h}^*)\}^{\otimes 2}]$ is invertible, $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = O_P(n^{-1/2})$, and

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) = \left(\mathbb{P}\left[\{\boldsymbol{l}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, \mathcal{A}_0) - l_{\mathcal{A}}(\boldsymbol{\theta}_0, \mathcal{A}_0)(\boldsymbol{h}^*)\}^{\otimes 2} \right] \right)^{-1} \mathbb{G}_n \left\{ \boldsymbol{l}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}}) - l_{\mathcal{A}}(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}})(\boldsymbol{h}^*) \right\} + o_P(1).$$

The influence function for $\widehat{\boldsymbol{\theta}}$ is the efficient influence function, such that $\sqrt{n}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)$ converges weakly to a zero-mean normal random vector whose covariance matrix attains the semiparametric efficiency bound.

4.6.3 Proof of Theorem 4.3

Let $\widehat{\mathcal{A}}^*$ be the estimator of $\mathcal{A}$ in the bootstrap sample. We denote $\widehat{\mathbb{P}}_n$ as the bootstrap empirical distribution and $\widehat{\mathbb{G}}_n = \sqrt{n}(\widehat{\mathbb{P}}_n - \mathbb{P}_n)$ as the bootstrap empirical process. Using arguments in the proof of Theorem 4.2, we can show that

$$\begin{aligned}
\widehat{\mathbb{G}}_n\left\{\boldsymbol{l}_{\boldsymbol{\theta}}\left(\widehat{\boldsymbol{\theta}},\widehat{\mathcal{A}}\right)\right\} &= \sqrt{n}\widehat{\mathbb{P}}_n\left\{\boldsymbol{l}_{\boldsymbol{\theta}}\left(\widehat{\boldsymbol{\theta}},\widehat{\mathcal{A}}\right)-\boldsymbol{l}_{\boldsymbol{\theta}}\left(\widehat{\boldsymbol{\theta}}^*,\widehat{\mathcal{A}}^*\right)\right\} \\
&= \sqrt{n}\widehat{\mathbb{P}}_n\left\{\boldsymbol{l}_{\boldsymbol{\theta}\boldsymbol{\theta}}\left(\widehat{\boldsymbol{\theta}}-\widehat{\boldsymbol{\theta}}^*\right)+\boldsymbol{l}_{\boldsymbol{\theta}\mathcal{A}}\left(\widehat{\mathcal{A}}-\widehat{\mathcal{A}}^*\right)\right\} \\
&\quad +O_P\left(n^{-1/6}+\sqrt{n}\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right\|^2+\sqrt{n}\left\|\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0\right\|^2+\sqrt{n}\left\|\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}_0\right\|^2\right)
\end{aligned}$$

and

$$\begin{aligned}
\widehat{\mathbb{G}}_n\left\{l_{\mathcal{A}}\left(\widehat{\boldsymbol{\theta}},\widehat{\mathcal{A}}\right)(\boldsymbol{h})\right\} &= \sqrt{n}\widehat{\mathbb{P}}_n\left\{l_{\mathcal{A}}\left(\widehat{\boldsymbol{\theta}},\widehat{\mathcal{A}}\right)(\boldsymbol{h})-l_{\mathcal{A}}\left(\widehat{\boldsymbol{\theta}}^*,\widehat{\mathcal{A}}^*\right)(\boldsymbol{h})\right\} \\
&= \sqrt{n}\widehat{\mathbb{P}}_n\left\{l_{\mathcal{A}\boldsymbol{\theta}}(\boldsymbol{h})\left(\widehat{\boldsymbol{\theta}}-\widehat{\boldsymbol{\theta}}^*\right)+l_{\mathcal{A}\mathcal{A}}\left(\boldsymbol{h},\widehat{\mathcal{A}}-\widehat{\mathcal{A}}^*\right)\right\} \\
&\quad +O_P\left(n^{-1/6}+\sqrt{n}\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right\|^2+\sqrt{n}\left\|\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0\right\|^2+\sqrt{n}\left\|\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}_0\right\|^2\right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\widehat{\mathbb{G}}_n\left\{\boldsymbol{l}_{\boldsymbol{\theta}}\left(\widehat{\boldsymbol{\theta}},\widehat{\mathcal{A}}\right)\right\}-\widehat{\mathbb{G}}_n\left\{l_{\mathcal{A}}\left(\widehat{\boldsymbol{\theta}},\widehat{\mathcal{A}}\right)(\boldsymbol{h}^*)\right\} \\
=\ &\sqrt{n}\widehat{\mathbb{P}}_n\left\{\boldsymbol{l}_{\boldsymbol{\theta}\boldsymbol{\theta}}\left(\widehat{\boldsymbol{\theta}}-\widehat{\boldsymbol{\theta}}^*\right)+\boldsymbol{l}_{\boldsymbol{\theta}\mathcal{A}}\left(\widehat{\mathcal{A}}-\widehat{\mathcal{A}}^*\right)\right\}-\sqrt{n}\widehat{\mathbb{P}}_n\left\{l_{\mathcal{A}\boldsymbol{\theta}}(\boldsymbol{h}^*)\left(\widehat{\boldsymbol{\theta}}-\widehat{\boldsymbol{\theta}}^*\right)+l_{\mathcal{A}\mathcal{A}}\left(\boldsymbol{h}^*,\widehat{\mathcal{A}}-\widehat{\mathcal{A}}^*\right)\right\} \\
&+O_P\left(n^{-1/6}+\sqrt{n}\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right\|^2+\sqrt{n}\left\|\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0\right\|^2+\sqrt{n}\left\|\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}_0\right\|^2\right) \\
=\ &\sqrt{n}\widehat{\mathbb{P}}_n\left[\left\{\boldsymbol{l}_{\boldsymbol{\theta}}\left(\widehat{\boldsymbol{\theta}},\widehat{\mathcal{A}}\right)-l_{\mathcal{A}}\left(\widehat{\boldsymbol{\theta}},\widehat{\mathcal{A}}\right)(\boldsymbol{h}^*)\right\}^{\otimes 2}\right]\left(\widehat{\boldsymbol{\theta}}-\widehat{\boldsymbol{\theta}}^*\right) \\
&+O_P\left(n^{-1/6}+\sqrt{n}\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right\|^2+\sqrt{n}\left\|\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0\right\|^2+\sqrt{n}\left\|\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}_0\right\|^2\right).
\end{aligned}$$

By the arguments in the proof of Theorem 4.2,

$$\begin{aligned}
\sqrt{n}\left(\widehat{\boldsymbol{\theta}}-\widehat{\boldsymbol{\theta}}^*\right) &= \left(\mathbb{P}\left[\left\{\boldsymbol{l}_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}_0,\mathcal{A}_0\right)-l_{\mathcal{A}}\left(\boldsymbol{\theta}_0,\mathcal{A}_0\right)(\boldsymbol{h}^*)\right\}^{\otimes 2}\right]\right)^{-1}\widehat{\mathbb{G}}_n\left\{\boldsymbol{l}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}},\widehat{\mathcal{A}})-l_{\mathcal{A}}(\widehat{\boldsymbol{\theta}},\widehat{\mathcal{A}})(\boldsymbol{h}^*)\right\}+o_P(1) \\
&= \left(\mathbb{P}\left[\left\{\boldsymbol{l}_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}_0,\mathcal{A}_0\right)-l_{\mathcal{A}}\left(\boldsymbol{\theta}_0,\mathcal{A}_0\right)(\boldsymbol{h}^*)\right\}^{\otimes 2}\right]\right)^{-1}\mathbb{G}_n\left\{\boldsymbol{l}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}},\widehat{\mathcal{A}})-l_{\mathcal{A}}(\widehat{\boldsymbol{\theta}},\widehat{\mathcal{A}})(\boldsymbol{h}^*)\right\}+o_P(1),
\end{aligned}$$

where the last equality follows from Theorem 3.6.1 of van der Vaart and Wellner (1996). Therefore, $\sqrt{n}(\widehat{\boldsymbol{\theta}}-\widehat{\boldsymbol{\theta}}^*)$ converges weakly to a zero-mean normal random vector, and $\sqrt{n}(\widehat{\boldsymbol{\theta}}-\widehat{\boldsymbol{\theta}}^*)$ and $\sqrt{n}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)$ have the same asymptotic distribution.

4.6.4 Some Useful Lemmas

Lemma 4.1. *Under Conditions 1−5, the classes of functions*

$$\widetilde{\mathcal{H}}_1 \equiv \left\{ \int_{\boldsymbol{b}} J_1(\boldsymbol{b}, \mathcal{O}; \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathcal{A}) \psi(\boldsymbol{b}; \boldsymbol{\Sigma}) d\boldsymbol{b} : \boldsymbol{\theta} \in \Theta, \mathcal{A} \in \mathcal{D}_1 \right\}$$

and

$$\widetilde{\mathcal{H}}_{2k} \equiv \left\{ \int_{\boldsymbol{b}} J_1(\boldsymbol{b}, \mathcal{O}; \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathcal{A}) J_{2k}(t, b, \mathcal{O}; \boldsymbol{\beta}, \gamma_k) \psi(\boldsymbol{b}; \boldsymbol{\Sigma}) d\boldsymbol{b} : \boldsymbol{\theta} \in \Theta, t \in \mathcal{U}_k, \mathcal{A} \in \mathcal{D}_1 \right\}$$

for $k = K_1{+}1, \ldots, K$ *are* $\mathbb{P}$*-Glivenko-Cantelli, where* $\mathcal{D}_1 = \mathcal{D}_{1,\infty} \times \cdots \times \mathcal{D}_{K_1,\infty} \times \mathcal{D}_{K_1+1,M} \times \cdots \times \mathcal{D}_{K,M}$ *and* M *is a finite constant.*

Proof. Define

$$W_k(t, \boldsymbol{X}, b_1; \boldsymbol{\beta}, \Lambda_k) = \frac{\int_0^t e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_k(s) + b_1} d\Lambda_k(s)}{\Lambda_k(\tau_k)}$$

for $k = 1, \ldots, K_1$, where $\boldsymbol{\beta} \in \mathcal{B}$ and $\Lambda_k \in \mathcal{D}_{k,\infty}$. The class of functions $\{e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_k(s) + b_1} : \boldsymbol{\beta} \in \mathcal{B}\}$, with $\boldsymbol{X}$ and b_1 as random variables, is a VC class with VC-index V. Thus, the class $\mathcal{W}_k \equiv \{W_k(t, \boldsymbol{X}, b_1; \boldsymbol{\beta}, \Lambda_k) : \boldsymbol{\beta} \in \mathcal{B}, \Lambda_k \in \mathcal{D}_{k,\infty}\}$ is a convex hull of the VC-class with the $L_2(\mathbb{P})$-bracketing number $O\{\exp(\epsilon^{-2V/(V+2)})\}$.

For any $(\boldsymbol{\beta}^{(1)}, \Lambda_k^{(1)})$ and $(\boldsymbol{\beta}^{(2)}, \Lambda_k^{(2)})$ in $\mathcal{B} \times \mathcal{D}_{k,\infty}$, $t_k \in \mathcal{U}_k$, and any positive constant M, if $\Lambda_k^{(1)}(\tau_k) > M$ and $\Lambda_k^{(2)}(\tau_k) > M$, then

$$\begin{aligned} & \left| \exp\left\{ -\int_0^{t_k} e^{\boldsymbol{\beta}^{(1)\mathrm{T}} \boldsymbol{X}_k(s) + b_1} d\Lambda_k^{(1)}(s) \right\} - \exp\left\{ -\int_0^{t_k} e^{\boldsymbol{\beta}^{(2)\mathrm{T}} \boldsymbol{X}_k(s) + b_1} d\Lambda_k^{(2)}(s) \right\} \right| \\ \leq \; & 2 \exp\left(-M e^{-\widetilde{M} - |b_1|} \right). \end{aligned}$$

If $\Lambda_k^{(1)}(\tau_k) \leq M$ and $\Lambda_k^{(2)}(\tau_k) \leq M$, then

$$\begin{aligned} & \left| \exp\left\{ -\int_0^{t_k} e^{\boldsymbol{\beta}^{(1)\mathrm{T}} \boldsymbol{X}_k(s) + b_1} d\Lambda_k^{(1)}(s) \right\} - \exp\left\{ -\int_0^{t_k} e^{\boldsymbol{\beta}^{(2)\mathrm{T}} \boldsymbol{X}_k(s) + b_1} d\Lambda_k^{(2)}(s) \right\} \right| \\ \leq \; & \sup_{\boldsymbol{\beta} \in \mathcal{B}, \Lambda_k \in \mathcal{D}_{k,\infty}, \Lambda_k(\tau_k) \leq M} \left| \exp\left\{ -\int_0^{t_k} e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}_k(s) + b_1} d\Lambda_k(s) \right\} \right| \\ & \times \left\{ \left| W_k\left(t, \boldsymbol{X}, b_1; \boldsymbol{\beta}^{(1)}, \Lambda_k^{(1)}\right) - W_k\left(t, \boldsymbol{X}, b_1; \boldsymbol{\beta}^{(2)}, \Lambda_k^{(2)}\right) \right| M \right. \\ & \left. + W_k\left(t, \boldsymbol{X}, b_1; \boldsymbol{\beta}^{(1)}, \Lambda_k^{(1)}\right) \left| \Lambda_k^{(1)}(\tau_k) - \Lambda_k^{(2)}(\tau_k) \right| \right\} \\ \leq \; & \left| W_k\left(t, \boldsymbol{X}, b_1; \boldsymbol{\beta}^{(1)}, \Lambda_k^{(1)}\right) - W_k\left(t, \boldsymbol{X}, b_1; \boldsymbol{\beta}^{(2)}, \Lambda_k^{(2)}\right) \right| M + e^{\widetilde{M} + |b_1|} \left| \Lambda_k^{(1)}(\tau_k) - \Lambda_k^{(2)}(\tau_k) \right|. \end{aligned}$$

In the remaining scenario, we assume, without loss of generality, that $\Lambda_k^{(1)}(\tau_k) \le M$ and $\Lambda_k^{(2)}(\tau_k) > M$. Then,

$$\begin{aligned}
& \left|\exp\left\{-\int_0^{t_k} e^{\boldsymbol{\beta}^{(1)\mathrm{T}}\boldsymbol{X}_k(s)+b_1} d\Lambda_k^{(1)}(s)\right\} - \exp\left\{-\int_0^{t_k} e^{\boldsymbol{\beta}^{(2)\mathrm{T}}\boldsymbol{X}_k(s)+b_1} d\Lambda_k^{(2)}(s)\right\}\right| \\
\le\ & \sup_{\boldsymbol{\beta}\in\mathcal{B},\Lambda_k\in\mathcal{D}_{k,\infty},\Lambda_k(\tau_k)\le M} \left|\exp\left\{-\int_0^{t_k} e^{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_k(s)+b_1} d\Lambda_k(s)\right\}\right| \\
& \times \left[\left|\exp\left\{-\Lambda_k^{(1)}(\tau_k) W_k\left(t,\boldsymbol{X},b_1;\boldsymbol{\beta}^{(1)},\Lambda_k^{(1)}\right)\right\} - \exp\left\{-MW_k\left(t,\boldsymbol{X},b_1;\boldsymbol{\beta}^{(1)},\Lambda_k^{(1)}\right)\right\}\right|\right. \\
& \left. + \left|\exp\left\{-MW_k\left(t,\boldsymbol{X},b_1;\boldsymbol{\beta}^{(1)},\Lambda_k^{(1)}\right)\right\} - \exp\left\{-\Lambda_k^{(2)}(\tau_k) W_k\left(t,\boldsymbol{X},b_1;\boldsymbol{\beta}^{(2)},\Lambda_k^{(2)}\right)\right\}\right|\right] \\
\le\ & \left(e^{\widetilde{M}+|b_1|}|\Lambda_k^{(1)}(\tau_k) - M|\right) + 2\exp\left(-Me^{-\widetilde{M}-|b_1|}\right).
\end{aligned}$$

Because there exist M/ϵ ϵ-brackets to cover $[0, M]$, the above results imply that there exist $O\left\{\exp\left(\epsilon^{-2V/(V+2)}\right)\right\} \times M/\epsilon$ brackets

$$\left\{W_k\left(t,\boldsymbol{X},b_1;\boldsymbol{\beta}^{(1)},\Lambda_k^{(1)}\right), W_k\left(t,\boldsymbol{X},b_1;\boldsymbol{\beta}^{(2)},\Lambda_k^{(2)}\right)\right\} \times \left\{\Lambda_k^{(1)}(\tau_k), \Lambda_k^{(2)}(\tau_k)\right\}$$

such that

$$\begin{aligned}
& \left|\exp\left\{-\int_0^{t_k} e^{\boldsymbol{\beta}^{(1)\mathrm{T}}\boldsymbol{X}_k(s)+b_1} d\Lambda_k^{(1)}(s)\right\} - \exp\left\{-\int_0^{t_k} e^{\boldsymbol{\beta}^{(2)\mathrm{T}}\boldsymbol{X}_k(s)+b_1} d\Lambda_k^{(2)}(s)\right\}\right| \\
\le\ & \left(M + e^{\widetilde{M}+|b_1|}\right)\epsilon + 2\exp\left(-e^{-\widetilde{M}-|b_1|}M\right).
\end{aligned}$$

Therefore, there exist $O\left\{\exp\left(\epsilon^{-2V/(V+2)}\right)/\epsilon\right\}$ ϵ-brackets to cover $\{\exp\{-\int_0^{t_k} e^{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_k(s)+b_1} d\Lambda_k(s)\} : \boldsymbol{\beta}\in\mathcal{B}, \Lambda_k \in \mathcal{D}_{k,\infty}\}$ in $L_2(\mathbb{P})$.

For any $(\boldsymbol{\beta}^{(1)}, \boldsymbol{\gamma}^{(1)}, \Lambda_k^{(1)})$ and $(\boldsymbol{\beta}^{(2)}, \boldsymbol{\gamma}^{(2)}, \Lambda_k^{(2)})$ in $\mathcal{B}\times\mathcal{G}\times\mathcal{D}_{k,M}$ for $k = K_1+1,\ldots,K$,

$$\begin{aligned}
& \left|\exp\left\{-\int_0^{Y_k} e^{\boldsymbol{\beta}^{(1)\mathrm{T}}\boldsymbol{X}_k(s)+\gamma_k^{(1)}b_1+b_2} d\Lambda_k^{(1)}(s)\right\} - \exp\left\{-\int_0^{Y_k} e^{\boldsymbol{\beta}^{(2)\mathrm{T}}\boldsymbol{X}_k(s)+\gamma_k^{(2)}b_1+b_2} d\Lambda_k^{(2)}(s)\right\}\right| \\
\le\ & \sup_{\boldsymbol{\beta}\in\mathcal{B},\boldsymbol{\gamma}\in\mathcal{G},\Lambda_k\in\mathcal{D}_{k,M}} \left|\exp\left\{-\int_0^{Y_k} e^{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_k(s)+\gamma_k b_1+b_2} d\Lambda_k(s)\right\}\right| \\
& \times \left|\int_0^{Y_k} e^{\boldsymbol{\beta}^{(1)\mathrm{T}}\boldsymbol{X}_k(s)+\gamma_k^{(1)}b_1+b_2} d\Lambda_k^{(1)}(s) - \int_0^{Y_k} e^{\boldsymbol{\beta}^{(2)\mathrm{T}}\boldsymbol{X}_k(s)+\gamma_k^{(2)}b_1+b_2} d\Lambda_k^{(2)}(s)\right| \\
\le\ & \left\{C^* e^{\widetilde{M}\|\boldsymbol{b}\|}\left(\left\|\boldsymbol{\beta}^{(1)}-\boldsymbol{\beta}^{(2)}\right\| + \left|\gamma_k^{(1)}-\gamma_k^{(2)}\right|\right) + \left|\int_0^{Y_k} e^{\boldsymbol{\beta}^{(1)\mathrm{T}}\boldsymbol{X}_k(s)+\gamma_k^{(1)}b_1+b_2} d\left(\Lambda_k^{(1)}-\Lambda_k^{(2)}\right)(s)\right|\right\} \\
\le\ & e^{\widetilde{M}\|\boldsymbol{b}\|}\left\{C^*\left\|\boldsymbol{\beta}^{(1)}-\boldsymbol{\beta}^{(2)}\right\| + C^*\left|\gamma_k^{(1)}-\gamma_k^{(2)}\right| + \left|\Lambda_k^{(1)}(Y_k)-\Lambda_k^{(2)}(Y_k)\right|\right.
\end{aligned}$$

$$+\int_0^{\tau_k}\left|\Lambda_k^{(1)}(s)-\Lambda_k^{(2)}(s)\right|ds\Big\},$$

where the last inequality follows from integration by parts. By Theorem 2.7.5 of van der Vaart and Wellner (1996), the bracketing number of $\mathcal{B}\times\mathcal{G}\times\mathcal{D}_{k,M}$ is of order $O\{\exp(\epsilon^{-1})\}$. Thus, the bracketing number of $\widetilde{\mathcal{H}}_1$ is of order $O\{\exp(\epsilon^{-2V/(V+2)}+\epsilon^{-1})\epsilon^{-1}\}$. Therefore, the class $\widetilde{\mathcal{H}}_1$ is Glivenko-Cantelli. Because $I(Y_k\geq t)$ is Glivenko-Cantelli, $\widetilde{\mathcal{H}}_{2k}$ is Glivenko-Cantelli by the preservation of the Glivenko-Cantelli property under the product. □

Lemma 4.2. *Under Conditions 1–5, the classes of functions*

$$\mathcal{H}_1\equiv\left\{\int_{\boldsymbol{b}}J_1(\boldsymbol{b},\mathcal{O};\boldsymbol{\beta},\boldsymbol{\gamma},\mathcal{A})\psi(\boldsymbol{b};\boldsymbol{\Sigma})d\boldsymbol{b}:\boldsymbol{\theta}\in\Theta,\mathcal{A}\in\mathcal{D}_2\right\}$$

and

$$\mathcal{H}_{2k}\equiv\left\{\int_{\boldsymbol{b}}J_1(\boldsymbol{b},\mathcal{O};\boldsymbol{\beta},\boldsymbol{\gamma},\mathcal{A})J_{2k}(t,b,\mathcal{O};\boldsymbol{\beta},\gamma_k)\psi(\boldsymbol{b};\boldsymbol{\Sigma})d\boldsymbol{b}:\boldsymbol{\theta}\in\Theta,t\in\mathcal{U}_k,\mathcal{A}\in\mathcal{D}_2\right\}$$

for $k=K_1+1,\ldots,K$ are $\mathbb{P}$-Donsker, where $\mathcal{D}_2=\mathcal{D}_{1,M}\times\cdots\times\mathcal{D}_{K,M}$ and M is a finite constant.

Proof. As in the proof of Lemma 4.1, for any $(\boldsymbol{\beta}^{(1)},\Lambda_k^{(1)})$ and $(\boldsymbol{\beta}^{(2)},\Lambda_k^{(2)})$ in $\mathcal{B}\times\mathcal{D}_{k,M}$ and $t_k\in\mathcal{U}_k$, we have

$$\begin{aligned}&\left|\exp\left\{-\int_0^{t_k}e^{\boldsymbol{\beta}^{(1)\mathrm{T}}\boldsymbol{X}_k(s)+b_1}d\Lambda_k^{(1)}(s)\right\}-\exp\left\{-\int_0^{t_k}e^{\boldsymbol{\beta}^{(2)\mathrm{T}}\boldsymbol{X}_k(s)+b_1}d\Lambda_k^{(2)}(s)\right\}\right|\\ \leq\;& e^{\widetilde{M}+|b_1|}\left\{C^*\left\|\boldsymbol{\beta}^{(1)}-\boldsymbol{\beta}^{(2)}\right\|+\left|\Lambda_k^{(1)}(t)-\Lambda_k^{(2)}(t)\right|+\int_0^{\tau_k}\left|\Lambda_k^{(1)}(s)-\Lambda_k^{(2)}(s)\right|ds\right\}.\end{aligned}$$

Thus,

$$\begin{aligned}&\left|J_1(\boldsymbol{b},\mathcal{O};\boldsymbol{\beta}^{(1)},\boldsymbol{\gamma}^{(1)},\mathcal{A}^{(1)})-J_1(\boldsymbol{b},\mathcal{O};\boldsymbol{\beta}^{(2)},\boldsymbol{\gamma}^{(2)},\mathcal{A}^{(2)})\right|\\ \leq\;&\widetilde{C}e^{2\widetilde{M}\|\boldsymbol{b}\|}\Bigg\{\left\|\boldsymbol{\beta}^{(1)}-\boldsymbol{\beta}^{(2)}\right\|+\left\|\boldsymbol{\gamma}^{(1)}-\boldsymbol{\gamma}^{(2)}\right\|+\sum_{k=1}^{K}\int_0^{\tau_k}\left|\Lambda_k^{(1)}(s)-\Lambda_k^{(2)}(s)\right|ds\\ &+\sum_{k=1}^{K_1}\left\{\left|\Lambda_k^{(1)}(L_k)-\Lambda_k^{(2)}(L_k)\right|+\left|\Lambda_k^{(1)}(R_k)-\Lambda_k^{(2)}(R_k)\right|\right\}+\sum_{k=K_1+1}^{K}\left|\Lambda_k^{(1)}(Y_k)-\Lambda_k^{(2)}(Y_k)\right|\Bigg\},\end{aligned}$$

where $\widetilde{C}$ is a constant. By the arguments in the proof of Lemma 4.1, the bracketing numbers of $\mathcal{H}_1$ and $\mathcal{H}_{2k}$ are of order $O\{\exp(\epsilon^{-1})\}$. Thus, $\mathcal{H}_1$ and $\mathcal{H}_{2k}$ are $\mathbb{P}$-Donsker. □

Lemma 4.3. *Under Conditions 1–5,*

$$
\begin{aligned}
& E\left[\sum_{k=1}^{K_1}\sum_{m=0}^{M_k}\left\{\widehat{\Lambda}_k(U_{km})-\Lambda_{k0}(U_{km})\right\}^2+\sum_{k=K_1+1}^{K}\left\{\widehat{\Lambda}_k(Y_k)-\Lambda_{k0}(Y_k)\right\}^2\right] \\
= \;& O_P(n^{-2/3})+O\left(\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right\|^2+\|\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0\|^2+\left\|\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}_0\right\|^2\right).
\end{aligned}
$$

Proof. By Theorem 4.1, $\widehat{\mathcal{A}}$ is consistent for $\mathcal{A}_0$. Thus, there exists a finite constant M such that $\widehat{\Lambda}(\tau_k)\le M$. By the Donsker results in Lemma 4.2, $m(\widehat{\boldsymbol{\theta}},\widehat{\mathcal{A}})$ is in a Donsker class. Note that

$$
\int_0^\delta \sqrt{1+\log N_{[]}(\epsilon,\mathcal{M},L_2(\mathbb{P}))}d\epsilon \le O(\delta^{1/2}).
$$

In addition, by Lemma 1.3 of van der Geer (2000) and the mean-value theorem,

$$
\mathbb{P}\left\{m\left(\widehat{\boldsymbol{\theta}},\widehat{\mathcal{A}}\right)-m\left(\boldsymbol{\theta}_0,\widetilde{\mathcal{A}}\right)\right\}\le -cH^2\left\{(\widehat{\boldsymbol{\theta}},\widehat{\mathcal{A}}),(\boldsymbol{\theta}_0,\widetilde{\mathcal{A}})\right\},
$$

where c is a positive constant, and $H\{(\theta,\mathcal{A}),(\theta_0,\widetilde{\mathcal{A}})\}$ is the Hellinger distance

$$
\left(\int\left[\exp\left\{\frac{L(\theta,\mathcal{A})}{2}\right\}-\exp\left\{\frac{L(\theta_0,\widetilde{\mathcal{A}})}{2}\right\}\right]^2 d\mu\right)^{1/2}
$$

with respect to the dominating measure μ. By Theorem 3.4.1 of van der Vaart and Wellner (1996), there exists r_n with $r_n^2\phi(1/r_n)\sim\sqrt{n}$ such that $H\{(\widehat{\boldsymbol{\theta}},\widehat{\mathcal{A}}),(\theta_0,\widetilde{\mathcal{A}})\}=O_P(1/r_n)$. In particular, we choose r_n in the order of $n^{1/3}$ such that $H\{(\widehat{\boldsymbol{\theta}},\widehat{\mathcal{A}}),(\boldsymbol{\theta}_0,\widetilde{\mathcal{A}})\}=O_P(n^{-1/3})$.

By the mean-value theorem,

$$
\begin{aligned}
E\Bigg[\Bigg\{ & \int_{\boldsymbol{b}}\prod_{k=1}^{K_1}\left(\sum_{m=0}^{M_k}\delta_{km}\left[\exp\left\{-\int_0^{U_{km}}e^{\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{X}_k(s)+b_1}d\widehat{\Lambda}_k(s)\right\}-\exp\left\{-\int_0^{U_{k,m+1}}e^{\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{X}_k(s)+b_1}d\widehat{\Lambda}_k(s)\right\}\right]\right) \\
& \prod_{k=K_1+1}^{K}\left[\left\{e^{\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{X}_k(Y_k)+\boldsymbol{\gamma}_k b_1+b_2}\widehat{\Lambda}_k\{Y_k\}\right\}^{\Delta_k}\exp\left\{-\int_0^{Y_k}e^{\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{X}_k(s)(s)+\widehat{\gamma}_k b_1+b_2}d\widehat{\Lambda}_k(s)\right\}\right]\psi(\boldsymbol{b};\widehat{\boldsymbol{\Sigma}})d\boldsymbol{b} \\
& -\int_{\boldsymbol{b}}\prod_{k=1}^{K_1}\left(\sum_{m=0}^{M_k}\delta_{km}\left[\exp\left\{-\int_0^{U_{km}}e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_k(s)+b_1}d\Lambda_{k0}(s)\right\}-\exp\left\{-\int_0^{U_{k,m+1}}e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_k(s)+b_1}d\Lambda_{k0}(s)\right\}\right]\right) \\
& \prod_{k=K_1+1}^{K}\left[\left\{e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_k(Y_k)+\gamma_{k0}b_1+b_2}\widetilde{\Lambda}_k\{Y_k\}\right\}^{\Delta_k}\exp\left\{-\int_0^{Y_k}e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_k(s)+\gamma_{k0}b_1+b_2}d\widetilde{\Lambda}_k(s)\right\}\right]\psi(\boldsymbol{b};\boldsymbol{\Sigma}_0)d\boldsymbol{b}\Bigg\}^2\Bigg]
\end{aligned}
$$

$$= O_P(n^{-2/3}).$$

Consequently, using the mean-value theorem again, we have

$$
\begin{aligned}
& O_P(n^{-2/3}) + O(1)\left\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_0\right\|^2 + O(1)\left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right\|^2 + O(1)\left\|\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\right\|^2 \\
\geq & E\left(\left[\int_{\boldsymbol{b}} \left\{ \prod_{k=1}^{K_1} \left(\sum_{m=0}^{M_k} \delta_{km} \left[\exp\left\{ -\int_0^{U_{km}} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(s) + b_1} d\widehat{\Lambda}_k(s) \right\} - \exp\left\{ -\int_0^{U_{k,m+1}} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(s) + b_1} d\widehat{\Lambda}_k(s) \right\} \right] \right) \right.\right.\\
& \prod_{k=K_1+1}^{K} \left[\left\{ e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(Y_k) + \gamma_{k0} b_1 + b_2} \widehat{\Lambda}_k\{Y_k\} \right\}^{\Delta_k} \exp\left\{ -\int_0^{Y_k} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(s) + \gamma_{k0} b_1 + b_2} d\widehat{\Lambda}_k(s) \right\} \right] \\
& - \prod_{k=1}^{K_1} \left(\sum_{m=0}^{M_k} \delta_{km} \left[\exp\left\{ -\int_0^{U_{km}} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(s) + b_1} d\Lambda_{k0}(s) \right\} - \exp\left\{ -\int_0^{U_{k,m+1}} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(s) + b_1} d\Lambda_{k0}(s) \right\} \right] \right) \\
& \left.\left. \prod_{k=K_1+1}^{K} \left[\left\{ e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(Y_k) + \gamma_{k0} b_1 + b_2} \widetilde{\Lambda}_k\{Y_k\} \right\}^{\Delta_k} \exp\left\{ -\int_0^{Y_k} e^{\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{X}_k(s) + \gamma_{k0} b_1 + b_2} d\widetilde{\Lambda}_k(s) \right\} \right] \right\} \psi(\boldsymbol{b}; \boldsymbol{\Sigma}_0) d\boldsymbol{b} \right]^2 \right) \\
\geq & c_0 E\left\{ \left(\int_{\boldsymbol{b}} J_1(\boldsymbol{b}, \mathcal{O}; \boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, \mathcal{A}_0) \left[\sum_{k=1}^{K_1} \sum_{m=0}^{M_k} \delta_{km} \int_0^{\tau_k} Q_1(t, U_{km}, U_{k,m+1}, b_1, \boldsymbol{X}_k; \boldsymbol{\beta}_0, \Lambda_{k0}) d(\widehat{\Lambda}_k - \Lambda_{k0})(t) \right.\right.\right. \\
& \left.\left.\left. + \sum_{k=K_1+1}^{K} \left\{ \Delta_k \left(\widehat{\Lambda}_k\{Y_k\} - \widetilde{\Lambda}_k\{Y_k\} \right) + \int_0^{\tau_k} Q_2(t, Y_k, \boldsymbol{b}, \boldsymbol{X}_k; \boldsymbol{\beta}_0, \gamma_{k0}) d(\widehat{\Lambda}_k - \widetilde{\Lambda}_k)(t) \right\} \right] \psi(\boldsymbol{b}; \boldsymbol{\Sigma}_0) d\boldsymbol{b} \right)^2 \right\}
\end{aligned}
$$

for some positive constant c_0. We define a norm in $\mathcal{V} \equiv \prod_{k=1}^{K} BV(\mathcal{U}_k)$ such that for any $\boldsymbol{f} \equiv (f_1, \ldots, f_K)^{\mathrm{T}} \in \mathcal{V}$,

$$\|\boldsymbol{f}\|_1 = \left[E\left\{ \sum_{k=1}^{K_1} \sum_{m=0}^{M} f_k(U_{km})^2 + \sum_{k=K_1+1}^{K} f_k(Y_k)^2 \right\} \right]^{1/2}.$$

In addition, we define a seminorm

$$
\begin{aligned}
\|\boldsymbol{f}\|_2 = & E\left\{ \left(\int_{\boldsymbol{b}} J_1(\boldsymbol{b}, \mathcal{O}; \boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, \mathcal{A}_0) \left[\sum_{k=1}^{K_1} \sum_{m=0}^{M_k} \delta_{km} \int_0^{\tau_k} Q_1(t, U_{km}, U_{k,m+1}, b_1, \boldsymbol{X}_k; \boldsymbol{\beta}_0, \Lambda_{k0}) df_k(t) \right.\right.\right. \\
& \left.\left.\left. + \sum_{k=K_1+1}^{K} \left\{ \Delta_k f_k(Y_k) + \int_0^{\tau_k} Q_2(t, Y_k, \boldsymbol{b}, \boldsymbol{X}_k; \boldsymbol{\beta}_0, \gamma_{k0}) df_k(t) \right\} \right] \psi(\boldsymbol{b}; \boldsymbol{\Sigma}_0) d\boldsymbol{b} \right)^2 \right\}^{1/2}.
\end{aligned}
$$

Note that if $\|\boldsymbol{f}\|_2 = 0$ for some $\boldsymbol{f} \in \mathcal{V}$, then

$$\int_{\boldsymbol{b}} J_1(\boldsymbol{b}, \mathcal{O}; \boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, \mathcal{A}_0) \left[\sum_{k=1}^{K_1} \sum_{m=0}^{M_k} \delta_{km} \int_0^{\tau_k} Q_1(t, U_{km}, U_{k,m+1}, b_1, \boldsymbol{X}_k; \boldsymbol{\beta}_0, \Lambda_{k0}) df_k(t) \right.$$

$$+\sum_{k=K_1+1}^{K}\left\{\Delta_k f_k(Y_k)+\int_0^{\tau_k} Q_2(t,Y_k,\boldsymbol{b},\boldsymbol{X}_k;\boldsymbol{\beta}_0,\gamma_{k0})df_k(t)\right\}\Bigg]\psi(\boldsymbol{b};\boldsymbol{\Sigma}_0)d\boldsymbol{b}=0 \tag{4.11}$$

with probability 1.

Consider $k=K_1+1,\ldots,K$. For $\Delta_k=0$, we set $Y_k=\tau_k$ in (4.11) to obtain an equation; for $\Delta_k=1$, we integrate Y_k from 0 to τ_k in (4.11) to obtain another equation. We add all the equations for $k=K_1+1,\ldots,K$ to obtain

$$\int_{b_1}\sum_{k=1}^{K_1}\Bigg\{\left(\sum_{m=0}^{M_k}\delta_{km}\left[\exp\left\{-\int_0^{U_{km}} e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_k(s)+b_1}d\Lambda_{k0}(s)\right\}-\exp\left\{-\int_0^{U_{k,m+1}} e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_k(s)+b_1}d\Lambda_{k0}(s)\right\}\right]\right)$$

$$\times\left\{\sum_{m=0}^{M_k}\delta_{km}\int_0^{\tau_k} Q_1(t,U_{km},U_{k,m+1},b_1,\boldsymbol{X}_k;\boldsymbol{\beta}_0,\Lambda_{k0})df_k(t)\right\}\Bigg\}\phi(b_1;\sigma_{10}^2)db_1=0.$$

For any $k\in\{1,\ldots,K_1\}$ and any $m_k\in\{0,\ldots,M_k\}$, we set $U_{k'm}=0$ for $k'\neq k$ and sum over all possible $\delta_{km'}$ with $m'=m_k,\ldots,M_k$ to obtain

$$\int_{b_1}\left\{\int_0^{U_{km}} e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_k(t)+b_1}df_k(t)\right\}\exp\left\{-\int_0^{U_{km}} e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_k(s)+b_1}d\Lambda_{k0}(s)\right\}\phi(b_1;\sigma_{10}^2)db_1=0.$$

Therefore, $f_k(U_{km})=0$ for $k=1,\ldots,K_1$. Because m is arbitrary, $f_k(t_k)=0$ for any $t_k\in\mathcal{U}_k$ for $k=1,\ldots,K_1$. In addition, we sum over (4.11) with all possible δ_{km} for $k=1,\ldots,K_1$ and $m=0,\ldots,M_k$ to obtain

$$\int_{\boldsymbol{b}}\sum_{k=K_1+1}^{K}\left[\left\{e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_k(Y_k)+\gamma_{k0}b_1+b_2}\Lambda_{k0}\{Y_k\}\right\}^{\Delta_k}\exp\left\{-\int_0^{Y_k} e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_k(s)+\gamma_{k0}b_1+b_2}d\Lambda_{k0}(s)\right\}\right.$$

$$\left.\times\left\{\int_0^{\tau_k} Q_2(t,Y_k,\boldsymbol{b},\boldsymbol{X}_k;\boldsymbol{\beta}_0,\gamma_{k0})df_k(t)\right\}\right]\psi(\boldsymbol{b};\boldsymbol{\Sigma}_0)d\boldsymbol{b}=0.$$

For $k=K_1+1,\ldots,k$, we let $\Delta_k=0$ and set $Y_k=t_k\in\mathcal{U}_k$ to obtain

$$\int_{\boldsymbol{b}}\sum_{k=K_1+1}^{K}\left[\left\{\int_0^{t_k} e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_k(s)+}df_k(s)\right\}\exp\left\{-\int_0^{t_k} e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_k(s)+\gamma_{k0}b_1+b_2}d\Lambda_{k0}(s)\right\}\right]\psi(\boldsymbol{b};\boldsymbol{\Sigma}_0)d\boldsymbol{b}=0.$$

We set $t_{k'}=0$ for $k'\neq k$ to obtain

$$\int_{\boldsymbol{b}}\left\{\int_0^{t_k} e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_k(s)+}df_k(s)\right\}\exp\left\{-\int_0^{t_k} e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}_k(s)+\gamma_{k0}b_1+b_2}d\Lambda_{k0}(s)\right\}\psi(\boldsymbol{b};\boldsymbol{\Sigma}_0)d\boldsymbol{b}=0.$$

Therefore, $f_k(t_k) = 0$ for any $t_k \in \mathcal{U}_k$ for $k = K_1 + 1, \ldots, K$. We obtain $\boldsymbol{f} = \mathbf{0}$, implying that $\|\cdot\|_2$ is a norm in $\mathcal{V}$.

By the Cauchy-Schwarz inequality, for any $\boldsymbol{f} \in \mathcal{V}$,

$$\begin{aligned}
\|\boldsymbol{f}\|_2 \leq & \left(E\left[\int_{\boldsymbol{b}} J_1(\boldsymbol{b}, \mathcal{O}; \boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, \mathcal{A}_0) \left\{ \sum_{k=1}^{K_1} \sum_{m=0}^{M_k} \delta_{km} \int_0^{\tau_k} Q_1(t, U_{km}, U_{k,m+1}, b_1, \boldsymbol{X}_k; \boldsymbol{\beta}_0, \Lambda_{k0}) dt \right.\right.\right. \\
& \left.\left. + \sum_{k=K_1+1}^{K} \int_0^{\tau_k} Q_2(t, Y_k, \boldsymbol{b}, \boldsymbol{X}_k; \boldsymbol{\beta}_0, \gamma_{k0}) dt \right\} \psi(\boldsymbol{b}; \boldsymbol{\Sigma}_0) d\boldsymbol{b} \right]^2 \\
& \left. \times E\left\{ \sum_{k=1}^{K_1} \sum_{m=0}^{M} f_k(U_{km})^2 + \sum_{k=K_1+1}^{K} f_k(Y_k)^2 \right\} \right)^{1/2} \\
\leq & \ c_1 \|\boldsymbol{f}\|_1,
\end{aligned}$$

where c_1 is a finite constant. By the bounded inverse theorem in the Banach space, we have $\|\boldsymbol{f}\|_2 \geq c_1' \|\boldsymbol{f}\|_1$ for some constant c_1'. Therefore,

$$O_P(n^{-2/3}) + O\left(\left\| \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \right\|^2 + \|\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\|^2 + \left\| \widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_0 \right\|^2 \right)$$

$$\geq c_0 {c_1'}^2 E\left[\sum_{k=1}^{K_1} \sum_{m=0}^{M_k} \left\{ \widehat{\Lambda}_k(U_{km}) - \Lambda_{k0}(U_{km}) \right\}^2 + \sum_{k=K_1+1}^{K} \left\{ \widehat{\Lambda}_k(Y_k) - \Lambda_{k0}(Y_k) \right\}^2 \right].$$

The lemma thus holds. $\square$

CHAPTER 5: EXTENSIONS AND FUTURE RESEARCH

5.1 Accelerated Failure Time Model with Interval-Censored Data

In Chapter 2, we studied the efficient estimation of the AFT model with PIC data that requires a non-negligible proportion of exact observations. The assumption is crucial in establishing the asymptotic properties and constructing the computation algorithm. Therefore, the proposed approach cannot be trivially applied to the interval-censored data where no exact observations are present.

Semiparametric regression analysis of interval-censored data without treating any observations as exact is extremely challenging. Although progress has been made on the semiparametric analysis of interval-censored data under the AFT model (Rabinowitz, Tsiatis and Aragon, 1995; Murphy, van der Vaart and Wellner, 1999; Shen, 2000; Betensky, Rabinowitz and Tsiatis, 2001; Tian and Cai, 2006), efficient estimation has not been explored. The similar idea of one-step efficient estimation, as proposed in Chapter 2 for PIC data, may be applied to obtain efficient estimators for the AFT model. Smoothing and approximations may be needed to obtain desirable numerical performance in finite samples.

5.2 Regression Analysis of Interval-Censored Data With Informative Examination Times

In some applications, the examination times are directly related to the event of interest, instead of through dropout. This may be the case if patients tend to visit their doctors more frequently when they are not feeling well. Zhang, Sun and Sun (2005), Chen et al. (2012), Chen, Wei, Hsu and Lee (2014), and Ma et al. (2015) studied this problem for current status data by assuming a frailty model or copula structure for the event of interest and the examination time. Zhang et al. (2007) considered the case of two examination times and modeled the first examination time, the gap time, and the event of interest through a proportional hazards frailty model. Zhao et al. (2015) considered the same type of data and assumed a copula model for the event of interest and the gap time. Wang et al. (2016) considered an arbitrary number of examination times and assumed a shared frailty model. All of these methods require parametric assumptions or approximations for the cumulative baseline hazard functions.

We can extend the proposed NPMLE to the aforementioned settings. In particular, for an arbitrary number of examination times, we can model the intensity of the examination process using a transformation model with frailty and modify the proposed EM algorithm to accommodate the recurrent examination process.

5.3 Regression Analysis of Panel Count Data

Panel count data arise in studies that concern recurrent events. Study subjects are observed only at discrete time points, such that only the number of recurrent events that occurred before each observation time is known. The regression analysis of panel count data, especially with the proportional mean model, has been studied in literature. In particular, Sun and Wei (2000) proposed a GEE-type procedure for the estimation of regression parameters. Wellner and Zhang (2007) considered two likelihood-based approaches: the pseudolikelihood estimator is fairly easy to compute, but it can be inefficient in certain cases (Wellner et al., 2004); and the algorithm for the more efficient nonparametric maximum likelihood estimation is computational intensive. Lu et al. (2009) modeled the baseline mean function with monotone B-splines and established the asymptotic properties of their spline-based estimators.

We may extend the proposed EM algorithm for interval-censored data to conduct NPMLE for the proportional mean model with panel count data. Specifically, we may introduce similar independent Poisson random variables such that the likelihood function can be viewed as the observed-data likelihood for the Poisson random variables. The algorithm can also be extended to accommodate the dependent observation process by modeling the examination process using another proportional mean model with frailty. The asymptotic theory for the panel count data, with independent or dependent examination process, can also be developed.

REFERENCES

Betensky, R. A., Rabinowitz, D., and Tsiatis, A. A. (2001), "Computationally Simple Accelerated Failure Time Regression for Interval Censored Data," *Biometrika*, 88, 703–711.

Bickel, P. J., Klaassen, C. A. J., Ritov, Y., and Wellner, J. A. (1993), *Efficient and Adaptive Estimation for Semiparametric Models*, New York: Springer.

Buckley, J., and James, I. (1979), "Linear Regression With Censored Data," *Biometrika*, 66, 429–436.

Cai, T., and Betensky, R. A. (2003), "Hazard Regression for Interval-Censored Data With Penalized Spline," *Biometrics*, 59, 570–579.

Chang, I. S., Wen, C. C., and Wu, Y. J. (2007), "A Profile Likelihood Theory for the Correlated Gamma-Frailty Model With Current Status Family Data," *Statistica Sinica*, 17, 1023–1046.

Chen, C. M., Lu, T. F. C., Chen, M. H., and Hsu, C. M. (2012), "Semiparametric Transformation Models for Current Status Data With Informative Censoring," *Biometrical Journal*, 54, 641–656.

Chen, C. M., Wei, J. C. C., Hsu, C. M., and Lee, M. Y. (2014), "Regression Analysis of Multivariate Current Status Data With Dependent Censoring: Application to Ankylosing Spondylitis Data," *Statistics in Medicine*, 33, 772–785.

Chen, M. H., Chen, L. C., Lin, K. H., and Tong, X. (2014), "Analysis of Multivariate Interval Censoring by Diabetic Retinopathy Study," *Communications in Statistics–Simulation and Computation*, 43, 1825–1835.

Efron, B., and Tibshirani, R. J. (1993), *An Introduction to the Bootstrap*, Boca Raton: Chapman & Hall/CRC.

Elbers, C., and Ridder, G. (1982), "True and Spurious Duration Dependence: the Identifiability of the Proportional Hazard Model," *The Review of Economic Studies*, 49, 403–409.

Fine, J. P., and Gray, R. J. (1999), "A Proportional Hazards Model for the Subdistribution of a Competing Risk," *Journal of the American Statistical Association*, 94, 496–509.

Fine, J. P., Jiang, H., and Chappell, R. (2001), "On Semi-Competing Risks Data," *Biometrika*, 88, 907–919.

Gehan, E. A. (1965), "A Generalized Wilcoxon Test for Comparing Arbitrarily Singly-Censored Samples," *Biometrika*, 52, 203–223.

Goggins, W. B., and Finkelstein, D. M. (2000), "A Proportional Hazards Model for Multivariate Interval-Censored Failure Time Data," *Biometrics*, 56, 940–943.

Gu, M. G., Sun, L., and Zuo, G. (2005), "A Baseline-free Procedure for Transformation Models Under Interval Censorship," *Lifetime Data Analysis*, 11, 473–488.

Hogan, J. W., and Laird, N. M. (1997), "Model-Based Approaches to Analyzing Incomplete Longitudinal and Failure Time Data," *Statistics in Medicine*, 16, 259–272.

Huang, J. (1995), "Maximum Likelihood Estimation for Proportional Odds Regression Model With Current Status Data," in *Lecture Notes-Monograph Series*, pp. 129–145.

Huang, J. (1996), "Efficient Estimation for the Proportional Hazards Model With Interval Censoring," *The Annals of Statistics*, 24, 540–568.

Huang, J. (1999), "Asymptotic Properties of Nonparametric Estimation Based on Partly Interval-Censored Data," *Statistica Sinica*, 9, 501–519.

Huang, J., and Rossini, A. (1997), "Sieve Estimation for the Proportional-odds Failure-time Regression Model With Interval Censoring," *Journal of the American Statistical Association*, 92, 960–967.

Huang, J., and Wellner, J. A. (1997), "Interval Censored Survival Data: a Review of Recent Progress," in *Proceedings of the First Seattle Symposium in Biostatistics*, Springer, pp. 123–169.

Jin, Z., Lin, D. Y., Wei, L., and Ying, Z. (2003), "Rank-based Inference for the Accelerated Failure Time Model," *Biometrika*, 90, 341–353.

Jin, Z., Lin, D. Y., and Ying, Z. (2006), "On Least-squares Regression With Censored Data," *Biometrika*, 93, 147–161.

Kalbfleisch, J. D., and Prentice, R. L. (1980), *The Statistical Analysis of Failure Time Data*, New York: Wiley.

Kim, J. S. (2003), "Maximum Likelihood Estimation for the Proportional Hazards Model With Partly Interval-Censored Data," *Journal of the Royal Statistical Society, Series B*, 65, 489–502.

Kosorok, M. R. (2007), *Introduction to Empirical Processes and Semiparametric Inference*, New York: Springer.

Lai, T. L., and Ying, Z. (1991), "Large Sample Theory of a Modified Buckley-James Estimator for Regression Analysis With Censored Data," *The Annals of Statistics*, 19, 1370–1402.

Lin, D., and Geyer, C. (1992), "Computational Methods for Semiparametric Linear Regression With Censored Data," *Journal of Computational and Graphical Statistics*, 1, 77–90.

Lin, D. Y. (1994), "Cox Regression Analysis of Multivariate Failure Time Data: the Marginal Approach," *Statistics in Medicine*, 13, 2233–2247.

Lin, Y., and Chen, K. (2013), "Efficient Estimation of the Censored Linear Regression Model," *Biometrika*, 100, 525–530.

Lu, M., Zhang, Y., and Huang, J. (2009), "Semiparametric Estimation Methods for Panel Count Data Using Monotone B-splines," *Journal of the American Statistical Association*, 104, 1060–1070.

Ma, L., Hu, T., and Sun, J. (2015), "Sieve Maximum Likelihood Regression Analysis of Dependent Current Status Data," *Biometrika*, 102, 731–738.

Murphy, S. A. (1994), "Consistency in a Proportional Hazards Model Incorporating a Random Effect," *The Annals of Statistics*, 22, 712–731.

Murphy, S. A., and Van der Vaart, A. W. (2000), "On Profile Likelihood," *Journal of the American Statistical Association*, 95, 449–465.

Murphy, S. A., van der Vaart, A. W., and Wellner, J. A. (1999), "Current Status Regression," *Mathematical Methods of Statistics*, 8, 407–425.

Prentice, R. L. (1978), "Linear Rank Tests With Right Censored Data," *Biometrika*, 65, 167–179.

Rabinowitz, D., Tsiatis, A., and Aragon, J. (1995), "Regression With Interval-Censored Data," *Biometrika*, 82, 501–513.

Ritov, Y. (1990), "Estimation in a Linear Regression Model With Censored Data," *The Annals of Statistics*, 18, 303–328.

Ritov, Y., and Wellner, J. A. (1988), "Censoring, Martingales, and the Cox Model," *Contemporary Mathematics*, 80, 191–219.

Rossini, A., and Tsiatis, A. (1996), "A Semiparametric Proportional Odds Regression Model for the Analysis of Current Status Data," *Journal of the American Statistical Association*, 91, 713–721.

Rudin, W. (1973), *Functional Analysis*, London: McGraw Hill.

Schick, A., and Yu, Q. (2000), "Consistency of the GMLE With Mixed Case Interval-Censored Data," *Scandinavian Journal of Statistics*, 27, 45–55.

Schuster, E. F. (1969), "Estimation of a Probability Density Function and its Derivatives," *The Annals of Mathematical Statistics*, 4, 1187–1195.

Schwartz, J. T. (1969), *Nonlinear Functional Analysis*, New York: Gordon & Breach.

Shen, X. (1998), "Proportional Odds Regression and Sieve Maximum Likelihood Estimation," *Biometrika*, 85, 165–177.

Shen, X. (2000), "Linear Regression With Current Status Data," *Journal of the American Statistical Association*, 95, 842–852.

Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, London: Chapman & Hall/CRC.

Sun, J., and Sun, L. (2005), "Semiparametric Linear Transformation Models for Current Status Data," *Canadian Journal of Statistics*, 33, 85–96.

Sun, J., and Wei, L. (2000), "Regression Analysis of Panel Count Data With Covariate-Dependent Observation and Censoring Times," *Journal of the Royal Statistical Society, Series B*, 62, 293–302.

Swanepoel, J. W. H. (1988), "Mean Integrated Squared Error Properties and Optimal Kernels When Estimating a Distribution Function," *Communications in Statistics–Theory and Methods*, 17, 3785–3799.

The ARIC Investigators (1989), "The Atherosclerosis Risk in Communities (ARIC) Study: Design and Objectives," *American Journal of Epidemiology*, 129, 687–702.

Tian, L., and Cai, T. (2006), "On the Accelerated Failure Time Model for Current Status and Interval Censored Data," *Biometrika*, 93, 329–342.

Tsiatis, A. A. (1990), "Estimating Regression Parameters Using Linear Rank Tests for Censored Data," *The Annals of Statistics*, 18, 354–372.

Turnbull, B. W. (1976), "The Empirical Distribution Function With Arbitrarily Grouped, Censored and Truncated Data," *Journal of the Royal Statistical Society, Series B*, 38, 290–295.

Uno, H., Cai, T., Pencina, M. J., D'Agostino, R. B., and Wei, L. (2011), "On the C-statistics for Evaluating Overall Adequacy of Risk Prediction Procedures With Censored Survival Data," *Statistics in Medicine*, 30, 1105–1117.

van der Geer, S. A. (2000), *Empirical Processes in M-estimation*, Cambridge University Press.

van der Vaart, A. W. (1998), *Asymptotic Statistics*, Cambridge: Cambridge University Press.

van der Vaart, A. W., and Wellner, J. A. (1996), *Weak Convergence and Empirical Processes*, New York: Springer.

Wang, P., Zhao, H., and Sun, J. (2016), "Regression Analysis of Case K Interval-Censored Failure Time Data in the Presence of Informative Censoring," *Biometrics*, 72, 1103–1112.

Wei, L.-J., Ying, Z., and Lin, D. (1990), "Linear Regression Analysis of Censored Survival Data Based on Rank Tests," *Biometrika*, 77, 845–851.

Wellner, J. A., and Zhang, Y. (2007), "Two Likelihood-based Semiparametric Estimation Methods for Panel Count Data With Covariates," *The Annals of Statistics*, 35, 2106–2142.

Wellner, J. A., Zhang, Y., and Liu, R. (2004), "A Semiparametric Regression Model for Panel Count Data: When do Pseudo-likelihood Estimators Become Badly Inefficient?," in *Proceedings of the Second Seattle Symposium in Biostatistics*, Springer, pp. 143–174.

Wen, C. C., and Chen, Y. H. (2013), "A Frailty Model Approach for Regression Analysis of Bivariate Interval-Censored Survival Data," *Statistica Sinica*, 23, 383–408.

Ying, Z. (1993), "A Large Sample Study of Rank Estimation for Censored Regression Data," *The Annals of Statistics*, 21, 76–99.

Zeidler, E. (1995), *Applied Functional Analysis – Applications to Mathematical Physics*, New York: Springer.

Zeng, D., and Lin, D. Y. (2007), "Efficient Estimation for the Accelerated Failure Time Model," *Journal of the American Statistical Association*, 102, 1387–1396.

Zeng, D., Mao, L., and Lin, D. (2016), "Maximum Likelihood Estimation for Semiparametric Transformation Models With Interval-Censored Data," *Biometrika*, 103, 253–271.

Zhang, Z., Sun, J., and Sun, L. (2005), "Statistical Analysis of Current Status Data With Informative Observation Times," *Statistics in Medicine*, 24, 1399–1407.

Zhang, Z., Sun, L., Sun, J., and Finkelstein, D. M. (2007), "Regression Analysis of Failure Time Data With Informative Interval Censoring," *Statistics in Medicine*, 26, 2533–2546.

Zhang, Z., Sun, L., Zhao, X., and Sun, J. (2005), "Regression Analysis of Interval-Censored Failure Time Data With Linear Transformation Models," *Canadian Journal of Statistics*, 33, 61–70.

Zhang, Z., and Zhao, Y. (2013), "Empirical Likelihood for Linear Transformation Models With Interval-Censored Failure Time Data," *Journal of Multivariate Analysis*, 116, 398–409.

Zhao, S., Hu, T., Ma, L., Wang, P., and Sun, J. (2015), "Regression Analysis of Informative Current Status Data With the Additive Hazards Model," *Lifetime Data Analysis*, 21, 241–258.

Zhao, X., Zhao, Q., Sun, J., and Kim, J. S. (2008), "Generalized Log-Rank Tests for Partly Interval-Censored Failure Time Data," *Biometrical Journal*, 50, 375–385.

Zhou, M. (2005), "Empirical Likelihood Analysis of the Rank Estimator for the Censored Accelerated Failure Time Model," *Biometrika*, 92, 492–498.