Semiparametric Regression Models for Recurrent and Terminal Event Data

by
Xiaoyan Wang

A dissertation submitted to the faculty of 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, Gillings School of Global Public Health.

Chapel Hill
2010

Approved by:

Dr. Jianwen Cai, Advisor

Dr. Haibo Zhou, Advisor

Dr. Amy Herring, Committee Member

Dr. Susan Hogan, Committee Member

Dr. Gary Koch, Committee Member

Dr. Donglin Zeng, Committee Member

Abstract

XIAOYAN WANG: Semiparametric Regression Models for Recurrent and Terminal Event Data.
(Under the direction of Dr. Jianwen Cai and Dr. Haibo Zhou.)

Recurrent events are common in many clinical or observational studies. It is often of interest to evaluate the effects of risk factors on the frequencies of recurrent events. The recurrence of serious events are usually subject to censoring due to the death of a subject which is likely to be informative.

In this dissertation, we first consider an accelerated failure time marginal rate model for the cumulative number of the recurrent events over time, while taking the terminal events into account. The marginal approach does not require specifying the dependence structure between the recurrent events and the terminal events, and the mean function incorporates the facts that subjects who die cannot experience any further recurrent events. We develop an estimating procedure for both the regression parameters and the mean function by applying the inverse probability of censoring weighting technique(Robins and Rotnitzky(1992)). The proposed estimators are consistent and asymptotically normal. We investigate the finite-sample properties of the proposed estimators through simulation studies and provide an application to recurrent hospitalization data taken from the Studies of Left Ventricular Dysfunction (SOLVD) Treatment Trial data.

Second, we propose a proportional rate model for the recurrent event given the subjects are still alive. Again the dependence between the recurrent event process and the terminal event is unspecified. We consider two estimating procedures for the regression coefficients and the mean function of recurrent events. Asymptotic properties of the proposed estimators are derived. Simulation studies are conducted to assess the finite

sample properties of the proposed estimators and show that they perform well under the sample sizes considered. The proposed method are illustrated to the SOLVD Prevention Trial data.

Third, we deal with the problem of missing covariates under the proposed proportional rate model. Under the MCAR assumption, we obtain consistent and asymptotically normal estimators by modifying the estimating equation of the proportional rate model. Extensive simulation studies are conducted to evaluate the finite sample properties of the proposed estimators. We also compare the efficiency of our method with the complete-case analysis and the full data analysis. The proposed method is applied to the aforementioned SOLVD data.

Acknowledgments

I would like to gratefully and sincerely thank my dissertation advisor, Dr. Jianwen Cai for her insightful guidance, patience, encouragement and support during my study at UNC-Chapel Hill. Her mentorship is paramount not only in the development of this dissertation but also in my personal life. For everything you have done for me, Dr. Cai, I thank you very much.

I would like to thank my co-advisor, Dr. Haibo Zhou for his constant help in statistical consulting. The experience I have gained as a graduate research assistant during last three years are invaluable. I am deeply grateful for his financial support as well.

I also would like to thank my committee members, Drs. Donglin Zeng, Amy Herring, Gary Koch and Susan Hogan for their constructive comments and valuable suggestions. Especially, I want to thank Dr. Donglin Zeng for many helpful discussions.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

In many longitudinal medical studies, subjects often experience recurrent events. Often, the recurrence of serious events is associated with a terminating event such as death. For example, the recurrent events could be multiple occurrences of rejection episodes in patients receiving kidney transplants, while the occurrence of death or the complete loss of function of the transplanted kidney terminates the observation of a patient's rejection episodes. Another example is in HIV/AIDS study, the occurrence of AIDS-defining events can be interrupted by death. It is frequently of interest to evaluate the effects of covariate such as treatment, on the basis of both recurrent events and the terminal events.

In the absence of a terminal event, abundant semiparametric regression methods under the Cox(1972) model framework for recurrent events have been proposed by many authors. Prentice, Williams and Peterson(PWP,1981) extended the Cox model to multiple event data based on the inter-event times. Andersen and Gill (AG,1982) related event intensity process to the covariates through a Cox-type formulation, treating each subject as a counting process with independent increments. Models based on marginal distribution of multivariate failure time data have been studied by Wei, Lin and Weissfeld(WLW,1989) and Lee, Wei and Amato(LWA,1992). Relaxing the independent increment assumption of Poisson-type process, Pepe and Cai(1993) proposed a new approach by modelling the conditional rate function, while Lawless and Nadeau(1995) developed methods to estimate the cumulative rate function. Lin, Wei, Yang and Ying(2000) studied semiparametric regression models for the mean and rate functions of recurrent events and provided robust inference procedures. Some accelerated failure time(AFT) models were also studied. Lin and Wei(1992) provided an approach of analyzing multiple events data by formulating the marginal distribution of the time to each type of event with a univariate accelerated failure time model. A few years later, Lin, Wei and Ying(1998) presented another extension of the AFT model for the mean function of the counting process for recurrent events which was more natural and efficient. Pan(2001) incorporated frailty into an AFT model and Huang(2000) proposed a two-sample multistate accelerated sojourn time model. Chang(2004) considered a AFT model on the sojourn time between recurrent events, while Strawderman(2005,2006) studied accelerated gap time models, assuming gap times are independent and dependent, respectively.

When a terminal event is present, the assumption that observation of recurrent events can only be terminated by censoring is violated, as is the assumption that censoring times are known and independent of the recurrent event process. For subjects who experience the terminal event while under observation, censoring times are known only if it can be assumed that they equal to potential censoring time; that is, only if the reason for censoring is, with certainty, administrative censoring.

Despite the progress in the methods for analyzing recurrent event data without a terminal event, methodologies to address analysis of data involving recurrent events with a terminal event are limited. Li and Lagakos(1997) proposed two modifications of the Wei, Lin and Weissfeld(1989) method. Cook and Lawless(1997) studied models for recurrent event intensity process conditional on survival at specific points in time and developed corresponding non-parametric methods. Ghosh and Lin(2002) considered a marginal multiplicative means model for recurrent events in the presence of a terminal events, acknowledging the fact that the death precludes further recurrence. Wang, Qing and

Chiang(2001) modelled the occurrence of recurrent events by a multiplicative intensity model, treating both the censoring and latent variables as nuisance parameters. A few joint models have been studied by various authors. Liu, Wolfe and Huang(2004) postulated a joint semiparametic joint model for the intensity function of both recurrent events and death by a shared gamma frailty. Ghosh and Lin(2003) formulated the marginal distributions of the recurrent events and dependent censoring through a joint model in a AFT model framework. Huang and Wang(2003) also proposed two nested joint models for the recurrent events and failure time.

In this dissertation, we will investigate semiparametric regression models for recurrent events when taking the terminating events into account. Our first proposed model is an accelerated failure time marginal rate model. In our second method, we consider a conditional proportional rate model approach. The third topic deals with the problem of missing covariates under the conditional proportional rates model. We will investigate the large sample properties of our proposed estimators, as well as finite sample behavior through simulation studies and real life data applications.

Chapter 2

Literature Review

In this section, we will go over some important semiparametric failure time models that are developed in recent years.

2.1 Univariate Failure Time Model

For the ith subject, let T_i be the failure time , C_i be the censoring time, and $X_i = T_i \wedge C_i = \min(T_i, C_i)$, $i = 1, \cdots, n$. Define $\Delta_i = I(T_i \leq C_i)$, where $I(.)$ is the indicator function.

2.1.1 Cox Proportional Hazard Model

The Cox proportional hazards model(Cox, 1972,1975) is the most popular regression model for assessing the effect of covariates in survival data analysis. In this model, the hazard function for the ith subject with the $p \times 1$ covariate vector $\boldsymbol{Z}_i$, $i = 1, \cdots, n$, at time t, is given by:

$$\lambda_i(t|\boldsymbol{Z}_i) = \lambda_0(t)e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_i(t)},$$

where $\lambda_0(t)$ is the unspecified and unknown baseline hazard function and $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown regression coefficients.

The parameter estimation and statistical inferences are based on maximizing partial likelihood function

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \left\{ \frac{e^{\boldsymbol{\beta}^T \boldsymbol{Z}_i(X_i)}}{\sum_{j\in\mathcal{R}(X_i)} e^{\boldsymbol{\beta}^T \boldsymbol{Z}_j(X_i)}} \right\}^{\Delta_i},$$

where $\mathcal{R}(t)$ is the risk set prior to time t. One can obtain the score function by taking the first derivative of the $\log L(\boldsymbol{\beta})$

$$\boldsymbol{U}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left\{ \boldsymbol{Z}_i - \frac{\sum_{j\in\mathcal{R}(X_i)} \boldsymbol{Z}_j e^{\boldsymbol{\beta}^T \boldsymbol{Z}_j}}{\sum_{j\in\mathcal{R}(X_i)} e^{\boldsymbol{\beta}^T \boldsymbol{Z}_j}} \right\} \Delta_i.$$

A Newton-Raphson iterative procedure can be used to obtain the maximum partial likelihood estimator of $\boldsymbol{\beta}_0$, denoted by $\hat{\boldsymbol{\beta}}_n$, by solving $\boldsymbol{U}(\boldsymbol{\beta}) = \mathbf{0}$. The information matrix is the negative of the second derivative of the $\log L(\boldsymbol{\beta})$,

$$\boldsymbol{I}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left[\frac{\sum_{j\in\mathcal{R}(X_i)} \boldsymbol{Z}_j^{\otimes 2} e^{\boldsymbol{\beta}^T \boldsymbol{Z}_j}}{\sum_{j\in\mathcal{R}(X_i)} e^{\boldsymbol{\beta}^T \boldsymbol{Z}_j}} - \left\{ \frac{\sum_{j\in\mathcal{R}(X_i)} \boldsymbol{Z}_j e^{\boldsymbol{\beta}^T \boldsymbol{Z}_j}}{\sum_{j\in\mathcal{R}(X_i)} e^{\boldsymbol{\beta}^T \boldsymbol{Z}_j}} \right\}^{\otimes 2} \right] \Delta_i,$$

where $\boldsymbol{a}^{\otimes 2} = \boldsymbol{a}\boldsymbol{a}^T$ for a column vector $\boldsymbol{a}$.

Large-sample properties of parameter estimators are justified by Andersen and Gill(1982) through the theory of martingales or empirical processes(Tsiatis,1981).

2.1.2 Accelerated Failure Time Model

A useful alternative to the Cox proportional hazard models is the accelerated failure time(AFT) model, which assumes that the covariates are linearly related to the logarithm of the survival time(Kalbfleisch and Prentice, 2002; Cox and Oakes, 1984, pp.64-5). As pointed out by Sir D.R.Cox(Reid, 1994, p.450), "accelerated life models are in many ways more appealing (than the proportional hazards model) because of their quite direct physical interpretation".

In the univariate case, the accelerated failure time (AFT) model is of the same form as the usual linear regression model:

$$\log T_i = -\boldsymbol{\beta}_0^T \boldsymbol{Z}_i + \epsilon_i,$$

where $\boldsymbol{\beta}_0$ is the unknown true $p \times 1$ parameter of interest and ϵ_i, $i = 1, \cdots, n$ are unobservable independent random errors with a common but completely unspecified distribution function. The AFT model describes the covariates through their effect on either expanding or contracting the time scale.

Several approaches have been proposed for the estimation and inference on the AFT model in the literature. Rank-based methods were studied by Tsiatis(1990), Wei et al.(1990), Lai and Ying(1991a,1991b), Lin and Geyer(1992), Ying(1993), among many others. Least squares based and M-estimation methods were investigated by Buckley and James(1979), Ritov(1990) and Lai and Ying(1991b), among many others.

2.2 Multivariate Failure Time Model

For the ith subject($i = 1, \cdots, n$), let $N_i(t) = \int_0^t dN_i(s)$ represent the number of events in $[0, t]$, where $dN_i(s)$ denotes the number of events in the small time interval $[s, s + ds]$. Let C_i be the censoring time and $T_{i1}, \cdots, T_{im_i}$ be the events times, $\boldsymbol{Z}_i(s)$ be the possible time-dependent covariate vector. We assume that C_i is independent of $\{dN_i(t); t \geq 0\}$ in the presence of covariates. Subject i's observed time with respect to the kth event is denoted as $X_{ik} = T_{ik} \wedge C_i = \min(T_{ik}, C_i)$. Also define $\Delta_{ik} = I(T_{ik} < C_i)$, where $I(.)$ is the indicator function.

2.2.1 Multiplicative Conditional Models

The conditional intensity function is defined as:

$$\lambda_i(t|\mathcal{N}_i(t)) = \lim_{\delta \to 0} \frac{1}{\delta} P\left(dN_i(t) = 1|\mathcal{N}_i(t)\right),$$

where $\mathcal{N}_i(t) = \{N_i(s); s \in [0, t)\}$, the ith subject's event history at time $t-$. For Poisson processes, $\lambda_i(t|\mathcal{N}_i(t)) = \lambda_i(t)$.

The Andersen-Gill(Andersen and Gill, 1982) model generalizes the Cox proportional model by relating the recurrent event intensity process to the covariate through the formulation:

$$\lambda_{ik}(t|\boldsymbol{Z}_i(t)) = \lambda_0(t) e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_i(t)},$$

for the kth event, $k = 1, \cdots, K$. Events for the same subject share a common baseline intensity function $\lambda_0(t)$ and common regression parameters. In this model, each subject is treated as a counting process with independent increment over the time interval $[0, 1]$. The at-risk process is defined as $Y_{ik}(t) = I(X_{i,k-1} < t \le X_{ik})$. The estimating equation for $\boldsymbol{\beta}_0$ is given by:

$$\boldsymbol{U}_n^{AG}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \int_0^t \{\boldsymbol{Z}_i - \boldsymbol{E}(s; \boldsymbol{\beta})\} \, dN_i(s),$$

where $\boldsymbol{E}(s; \boldsymbol{\beta}) = \boldsymbol{S}^{(1)}(s; \boldsymbol{\beta})/S^{(0)}(s; \boldsymbol{\beta})$, $\boldsymbol{S}^{(r)}(s; \boldsymbol{\beta}) = n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K} Y_{ik}(s) \boldsymbol{Z}_i(s)^{\otimes r} e^{\boldsymbol{\beta}^T \boldsymbol{Z}_i(s)}$, and, for a column vector $\mathbf{a}$, $\mathbf{a}^{\otimes r} = 1$, $\mathbf{a}^{\otimes 1} = \mathbf{a}$, $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^T$. The estimator of $\boldsymbol{\beta}_0$, $\hat{\boldsymbol{\beta}}_n^{AG}$, can be obtained by iteratively solving $\boldsymbol{U}_n^{AG}(\boldsymbol{\beta}) = \mathbf{0}$. The Breslow-Aalen(Breslow, 1974) estimate of the cumulative baseline hazards is given by:

$$\hat{\Lambda}_0(t; \hat{\boldsymbol{\beta}}_n^{AG}) = \sum_{i=1}^{n} \int_0^t \frac{dN_i(s)}{n S^{(0)}(s; \hat{\boldsymbol{\beta}}_n^{AG})}.$$

Using the martingales theory, one can justify that under certain regularity conditions,

$\hat{\boldsymbol{\beta}}_n^{AG} \xrightarrow{P} \boldsymbol{\beta}_0$, $n^{1/2}(\hat{\boldsymbol{\beta}}_n^{AG} - \boldsymbol{\beta}_0)$ is asymptotically normal with mean zero and a variance-covariance matrix that can be consistently estimated by $\hat{\boldsymbol{A}}_n(\hat{\boldsymbol{\beta}}_n^{AG})$, with

$$\hat{\boldsymbol{A}}_n(\hat{\boldsymbol{\beta}}_n^{AG}) = n^{-1} \sum_{i=1}^{n} \int_0^1 \left\{ \frac{\boldsymbol{S}^{(2)}(s; \hat{\boldsymbol{\beta}}_n^{AG})}{S^{(0)}(s; \hat{\boldsymbol{\beta}}_n^{AG})} - \boldsymbol{E}(s; \hat{\boldsymbol{\beta}}_n^{AG})^{\otimes 2} \right\} dN_i(s).$$

Another method for analyzing multivariate survival data is the frailty model. The underlying logic of frailty models is that some subjects are more or less prone to experiencing the events than are others. The frailty can be thought of as a random variable which induces dependence among the multiple event times. The most popular frailty model is to incorporate frailty into a Cox model. Conditional on the random effect, the intensity function of each subject for recurrences is,

$$\lambda_{ik}\{t|W_i\} = W_i \lambda_0(t) Y_{ik}(t) e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ik}(t)}.$$

Commonly used distributions for W_i are gamma(Clayton and Cuzick, 1985), positive stable(Hougaard, 1986), inverse Gaussian(Hougaard, 1986) and log-normal(McGilchrist and Aisbett 1991) distributions, with gamma being by far the most frequent due to the flexibility of this distribution.

Clayton and Cuzick(1985) considered estimation of the frailty parameter and covariate effects using a modified EM algorithm in the bivariate case, assuming the frailty follows a gamma distribution with mean one and an unknown variance. Nielsen et al.(1992) developed a counting process approach to estimate the regression parameters, the variance of the frailty, and the underlying intensity function.

2.2.2 Multiplicative Marginal Hazards Models

Wei, Lin and Weissfeld(WLW,1989) proposed modelling the marginal hazards as an alternative to modelling the conditional intensity function. The marginal distribution for the kth event time of the ith subject is modelled by using the Cox specification for the hazard function

$$\lambda_{ik}(t|\boldsymbol{Z}_{ik}) = \lambda_{0k}(t)e^{\boldsymbol{\beta}_k^T \boldsymbol{Z}_{ik}(t)},$$

for $k = 1, \cdots, K$. The kth-event partial likelihood function is given by

$$PL_k(\boldsymbol{\beta}_k) = \prod_{i=1}^{n} \left\{ \frac{e^{\boldsymbol{\beta}_k^T \boldsymbol{Z}_{ik}(X_{ik})}}{\sum_{j=1}^{n} Y_{jk}(X_{ik}) e^{\boldsymbol{\beta}_k^T \boldsymbol{Z}_{jk}(X_{ik})}} \right\}^{\Delta_{ik}},$$

and the score function is

$$\boldsymbol{U}_{k:n}(\boldsymbol{\beta}_k) = \sum_{i=1}^{n} \int_0^n \left\{ \boldsymbol{Z}_{ik} - \boldsymbol{E}_k(s; \boldsymbol{\beta}_k) \right\} dN_{ik}(s),$$

where $\Delta_{ik} = I(T_{ik} \leq C_i)$, $N_{ik}(t) = I(X_{ik} \leq t, \Delta_{ik} = 1)$, $\boldsymbol{E}_k(s; \boldsymbol{\beta}_k) = \boldsymbol{S}_k^{(1)}(s; \boldsymbol{\beta}_k)/S_k^{(0)}(s; \boldsymbol{\beta}_k)$, $\boldsymbol{S}_k^{(r)}(s; \boldsymbol{\beta}_k) = n^{-1} \sum_{i=1}^{n} Y_{ik}(s) \boldsymbol{Z}_{ik}(s)^{\otimes r} e^{\boldsymbol{\beta}_k^T \boldsymbol{Z}_{ik}(s)}$ and $Y_{ik}(t) = I(X_{ik} \geq t)$. When the marginal models are correctly specified, the solution to the above score function, $\hat{\boldsymbol{\beta}}_{k:n}$, is consistent for $\boldsymbol{\beta}_k$ and it follows that $n^{1/2}(\hat{\boldsymbol{\beta}}_{k:n} - \boldsymbol{\beta}_k)$ converges in distribution to a p-dimensional zero-mean normal vector and covariance matrix in the form of $\boldsymbol{A}_k^{-1} \boldsymbol{B}_k \boldsymbol{A}_k^{-1}$. The consistent estimators of $\boldsymbol{A}_k$ and $\boldsymbol{B}_k$ can be obtained through

$$\widehat{\boldsymbol{A}}_{k:n} = n^{-1} \sum_{i=1}^{n} \int_0^t \left\{ \frac{\boldsymbol{S}_k^{(2)}(s; \boldsymbol{\beta}_k)}{S_k^{(0)}(s; \boldsymbol{\beta}_k)} - \boldsymbol{E}_k(s; \boldsymbol{\beta}_k)^{\otimes 2} \right\} dN_{ik}(s),$$

and

$$\widehat{\boldsymbol{B}}_{k:n} = n^{-1} \sum_{i=1}^{n} \left\{ \int_0^t \{\boldsymbol{Z}_{ik}(s) - \boldsymbol{E}_k(s; \boldsymbol{\beta}_k)\} d\hat{M}_{ik}(s; \boldsymbol{\beta}_k) \right\}^{\otimes 2}$$

with $d\hat{M}_{ik}(s; \boldsymbol{\beta}_k) = dN_{ik}(s) - Y_{ik}(s) e^{\boldsymbol{\beta}_k^T \boldsymbol{Z}_{ik}(s)} d\hat{\Lambda}_{0k}(s; \boldsymbol{\beta}_k)$. This sandwich-type variance estimator is robust. Inferences regarding $\boldsymbol{\beta}_k$ are valid asymptotically regardless of the true intra-subject correlation structure. The WLW methods and its derivatives (Wei et al. 1990; Lee et al., 1992; Liang et al., 1993; Cai and Prentice, 1995,1997) are robust and well-developed theoretically. They are viewed as "population average" models, providing no information on inter-relationships among event times.

Lee, Wei and Amato(LWA, 1992) suggested a Cox-type regression analysis for highly stratified failure time data. In their approach, the assumption of the independence of all the failure times in the Cox proportional hazards model is relaxed to handle the correlated data structure due to natural or artificial grouping.

For highly stratified data, define T_{ij} be the potential failure time for the jth member in the ith stratum and C_{ij} is the potential censoring time and assumed to be independent of T_{ij}, given the $p \times 1$ vector of covaraites $\boldsymbol{Z}_{ij}$. The observed data are (X_{ij}, Δ_{ij}), where $X_{ij} = T_{ij} \wedge C_{ij}$, $\Delta_{ij} = I(T_{ij} \leq C_{ij})$, $j = 1, \cdots, L_i$; $i = 1, \cdots, n$.

The marginal hazard function has the common baseline hazard $\lambda_0(t)$,

$$\lambda(t|\boldsymbol{Z}_{ij}) = \lambda_0(t)e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ij}},$$

where $\boldsymbol{\beta}$ is a $p \times 1$ vector of the regression parameters. Define $Y_{ij} = I(X_{ij} \geq t)$, the at-risk indicator function at time t. The pseudo-partial likelihood function is

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n}\prod_{j=1}^{L_i}\left[\frac{e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ij}(X_{ij})}}{\sum_{l=1}^{n}\sum_{k=1}^{L_i} Y_{lk}(X_{ij})e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{lk}(X_{ij})}}\right]^{\Delta_{ij}}.$$

The estimator of $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}_{LWA}$, can be obtained by maximizing the above likelihood function. The consistency and asymptotic normality of $\hat{\boldsymbol{\beta}}_{LWA}$ are proved and a valid covariance matrix estimate for $\hat{\boldsymbol{\beta}}_{LWA}$ is provided in the paper.

The intensity models conditional on $\mathcal{N}_i(t)$ require the correlation structure be fully specified and are prone to misspecification. Intensity models are especially useful in prediction. Marginal models focus on fixed covariate effects, while the correlation between event times is not modelled. The estimators are robust to correlation structure between events but less efficient and also sensitive to censoring assumptions.

2.2.3 Multivariate Accelerated Failure Time Model

Lin and Wei(1992) studied the AFT model for the multivariate case. Let T_{ik} be the failure time of the ith subject at the kth event, $i = 1, \cdots, n$, $k = 1, \cdots, K$. Conditional on the $p \times 1$ covariate vector $\boldsymbol{Z}_{ik}$, the accelerated failure time model is

$$\log T_{ik} = \boldsymbol{\beta}_k^T \boldsymbol{Z}_{ik} + \epsilon_{ik},$$

where $\boldsymbol{\beta}_k$ is a $p \times 1$ vector of regression coefficients and $\epsilon_i = (\epsilon_{i1}, \cdots, \epsilon_{ik})$, $i = 1, \cdots, n$, are independent identically distributed with an unspecified distribution function.

The weighted score function based on rank statistics is

$$\boldsymbol{U}_k(\boldsymbol{\beta}) = \sum_{i=1}^{n} \Delta_{ik}\phi_{ik} \left[\boldsymbol{Z}_{ik} - \frac{\sum_{j=1}^{n} \boldsymbol{Z}_{ik} I\left\{e_{jk}(\boldsymbol{\beta}) \geq e_{ik}(\boldsymbol{\beta})\right\}}{\sum_{j=1}^{n} I\left\{e_{jk}(\boldsymbol{\beta}) \geq e_{ik}(\boldsymbol{\beta})\right\}} \right],$$

where $I(.)$ is the indicator function and ϕ is a twice continuously differentiable function on $[0, 1]$. If $\phi(.) = 1$, then it reduces to the Cox partial likelihood score function. If $\phi(u) = 1 - u$, then it corresponds to the Peto-Prentice generalization of the Wilcoxon statistic(Peto and Peto, 1972; Prentice, 1978). This method essentially formulates the marginal distribution for the time to each event with a univariate AFT model and derives the joint distribution for the regression parameter estimators of the marginal models. It can only handle a small equal number of events per subject, and does not provide global estimation of the underlying mean function.

Pan(2001) studied the AFT model with frailties. The model takes the form

$$\log T_{ik} = \boldsymbol{\beta}^T \boldsymbol{Z}_{ik} + \epsilon_{ik}, i = 1, \cdots, n; k = 1, \cdots, m_i,$$

where $\boldsymbol{Z}_{ik}$ are the observed covariates and $\boldsymbol{\beta}$ is the unknown regression coefficient vector. The marginal independence approach by Lee *et al*(1993) estimates $\boldsymbol{\beta}$ by assuming the random vector $\epsilon = (\epsilon_{i1}, \cdots, \epsilon_{in_i})^T$, $i = 1, \cdots n$, are independent of each other. Pan(2001), however, proposed to use frailties to explicitly model the possible correlation among failure times. Suppose that the hazard function of ϵ_{ik}, conditional on a random variable W_i, is

$$h_{ij}(t|W_i) = W_i h_0(t),$$

where h_0 is an unknown baseline hazard function, and W_i are the frailties or random-effects. Conditional on W_i, ϵ_{ik} are all independent. As usual W_i are assumed to be from a gamma distribution with unit mean and variance σ^2. Simulation results shows improved efficiency over the marginal independence approach.

Huang(2000) proposed a two-sample multistate accelerated sojourn times model, which yields a class of estimators that are consistent and asymptotically normal. Let $g = 0, 1$ be the control and treatment groups in a medical study. The K sojourn times are defined as $S_{g1} = T_{g1}$, $S_{g2} = T_{g2} - T_{g1}$, $\cdots$, $S_{gK} = T_{g,K} - T_{g,K-1}$. Under the assumption that the censoring time is independent of the sojourn times, the proposed multistate accelerated sojourn times model postulates that $(S_{11}, \cdots S_{1K})^T$ has the same distribution as $\{e^{\beta_1} S_{01}, \cdots, e^{\beta_K} S_{0K}\}^T$ for some $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_K)^T$. This model has a simple interpretation in that e^{β_k} is the scale change of the kth state resulting from the treatment. Estimation of $\boldsymbol{\beta}$ is carried out by solving a system of estimating functions

$$\boldsymbol{U}(\boldsymbol{\beta}) = \{U_1(\beta_1), \cdots, U_K(\beta_K)\}^T,$$

where

$$U_k(\boldsymbol{\beta}_k) = \int_{s=0}^{L_{1k}} \int_{u=0}^{L_{0k}} w_k(s, u; \boldsymbol{\beta}_k) \left\{ \hat{F}_{1k}(s, u; -\boldsymbol{\beta}_k) - \hat{F}_{01}(s, u; \boldsymbol{\beta}_k) \right\} du\, ds,\ k > 1,$$

where $\boldsymbol{\beta}_k = (\beta_1, \cdots, \beta_k)^T$, $k = 1, \cdots, K$. Here $L(\beta_1) = e^{-\beta_1/2} L_{11} \wedge e^{\beta_1/2} L_{01}$ for positive constant $L_{gk} < \sup\{t : Pr(T_{gk} > t) > 0\}$. $\hat{F}_{gk}$ is the Huang-Louis estimator(Huang and Louis, 1998) of $F_{gk}(t, u; \boldsymbol{\beta}_k) = Pr\{T^0_{gk} \leq t, \sum_{j=1}^{k} \exp(\beta_j) S^0_{gj} \leq u\}$. The complement of Kaplan-Meier estimators is a special case of the Huang-Louis estimator. w_k is a positive weight function, $k = 1, \cdots, K$. The estimator $\hat{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}$ is defined as the zero-crossing of $\boldsymbol{U}(\boldsymbol{\beta})$.

2.3 Semiparametric Models for Recurrent Event data

Recurrent event data can be treated as a special case of multivariate survival data, where the occurrence times of the recurrent events from a subject are ordered and only the last recurrence time is subject to right censoring. Moreover, the number of recurrent events is informative for the recurrence time distribution while the size of a cluster for standard correlated data is generally not. Methods that exploit the special structure of recurrent event data have been studied in recent years.

2.3.1 Multiplicative Conditional Models

Prentice, Williams and Peterson(PWP model,1981) extended the Cox model by considering two classes of stratified proportional intensity models with time-dependent strata. In their models, the intensity function for subject i at time t for the kth recurrent event, conditional on $\mathcal{N}_i(t)$, takes the forms:

$$
\begin{aligned}
\lambda_{ik}(t) &= Y_{ik}(t)\lambda_{0k}(t)e^{\boldsymbol{\beta}_k^T \boldsymbol{Z}_{ik}(t)}, \\
\lambda_{ik}(t) &= Y_{ik}(t)\lambda_{0k}(t - T_{i,k-1})e^{\boldsymbol{\beta}_k^T \boldsymbol{Z}_{ik}(t)}
\end{aligned}
$$

for total and gap times, respectively. The at-risk processes are defined as $Y_{ik}(t) = I(X_{i,k-1} \leq t < X_{ik})$ for total time model and $Y_{ik}(t) = I(X_{ik} \geq X_{i,k-1} + t)$ for gap times model. The PWP models allow event-specific baseline intensity functions and regression parameters. Same as the AG model, parameters are estimated through partial likelihood. For example, for the total time model, the $p_k \times 1$ regression parameters, $\boldsymbol{\beta}_k$, is estimated as the solution to the estimating equation $\boldsymbol{U}^{PT}(\boldsymbol{\beta}_k) = \mathbf{0}_{p_k \times 1}$, where

$$\boldsymbol{U}^{PT}(\boldsymbol{\beta}_k) = \sum_{i=1}^{n} \int_0^t \left\{ \boldsymbol{Z}_{ik}(s) - \boldsymbol{E}_k^{PT}(s; \boldsymbol{\beta}_k) \right\} dN_{ik}(s),$$

for $k = 1, \cdots, K$, where $\boldsymbol{E}_k^{PT}(s; \boldsymbol{\beta}_k) = \boldsymbol{Q}^{(1)}(s; \boldsymbol{\beta}_k)/Q^{(0)}(s; \boldsymbol{\beta}_k)$, $\boldsymbol{Q}^{(r)}(s; \boldsymbol{\beta}_k) =$
$n^{-1} \sum_{i=1}^{n} Y_{ik}(s) \boldsymbol{Z}_{ik}(s)^{\otimes r} e^{\boldsymbol{\beta}_k^T \boldsymbol{Z}_{ik}(s)}$ and $N_{ik}(t) = I(T_{ik} \leq t, \Delta_{ik} = 1)$.

The PWP model provides both event-specific estimates and an overall estimate. The overall estimate is obtained by fitting the single covariate vector $\boldsymbol{Z}_i$ to the model. The event-specific estimates are obtained by fitting event-specific covariates to the model, such that $\mathbf{Z}_i = (Z_{i1}, 0, \cdots, 0)^T, \mathbf{Z}_i = (0, Z_{i2}, \cdots, 0)^T, \cdots, \mathbf{Z}_i = (0, 0, \cdots, Z_{ik})^T$ for $k = 1, 2, \cdots, K$.

Chang and Wang(1999) proposed the following semiparametric hazards model for recurrent event data

$$\lambda_{ij}(t) = \lambda_{0j}(t - T_{i,j-1}) e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{i1}(t) + \boldsymbol{\gamma}_j^T \boldsymbol{Z}_{i2}(t)}.$$

Here the baseline hazard function $\lambda_{0j}(.)$ is an arbitrary function depending on the jth recurrent event. The common $p \times 1$ vector, $\boldsymbol{\beta}$, is the structural parameter of primary interest, and the $q \times 1$ episode-specific parameters $\boldsymbol{\gamma}_j$ may or may not be of interest, depending on the application.

In the absence of $\boldsymbol{\beta}$, the model is reduced to

$$\lambda_{ij}(t) = \lambda_{0j}(t - T_{i,j-1}) e^{\boldsymbol{\gamma}_j^T \boldsymbol{Z}_{i2}(t)},$$

which has been considered by Prentice et al(1981). This model is attractive if the primary interest is on the covariate effects over different episodes. In contrast, if we set $\boldsymbol{\gamma}_j = \mathbf{0}$, the model becomes

$$\lambda_{ij}(t) = \lambda_{0j}(t - T_{i,j-1})e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{i1}(t)}.$$

This would be useful when the covariate effects remain constant for distinct episodes. It was studied in depth independently by Chang and Hsiung(1994) and Chang(1995). Let $K_i(t)$ be the number of events occuring in the interval $[0, t]$, C_i the censoring time, and $K = \max_{1 \le i \le n}\{K_i(C_i)\}$. Define $y_{(1),j} < y_{(2),j} < \cdots < y_{(d_j),j}$ to be the d_j-ordered distinct recurrence times for the jth recurrent event, $\boldsymbol{Z}_{(i),j}$ be the covariate history up to and including $t_{(i),j}$, the ordered event time corresponding to $y_{(i),j}$. The partial likelihood, $\mathcal{L}_{pl}$, is defined as

$$\mathcal{L}_{pl} = \prod_{j=1}^{K}\prod_{i=1}^{d_j} p_{i-1,j}(\boldsymbol{Z}_{(i),j}|y_{(i),j}),$$

where

$$p_{i-1,j}(\boldsymbol{Z}_{(i),j}|y_{(i),j}) = \frac{e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{(i),1}(t_{(i),j}) + \boldsymbol{\gamma}_j^T \boldsymbol{Z}_{(i),2}(t_{(i),j})}}{\sum_{\{l \in R_j(y_{(i),j})\}} e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_{l1}(t_{l,j-1}+y_{(i),j}) + \boldsymbol{\gamma}_j^T \boldsymbol{Z}_{l2}(t_{l,j-1}+y_{(i),j})}}.$$

Here $R_{(i),j}$ is the risk set defined at $y_{(i),j}$ for the jth recurrent event among those who have had $j-1$ recurrent events.

The score function for $\boldsymbol{\beta}_0$, subject to fixed $\boldsymbol{\gamma}_j$, is expressed as

$$\boldsymbol{U}_{(i),j}(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_j) = \partial \log p_{i-1,j}(\boldsymbol{Z}_{(i),j}|y_{(i),j})/\partial \boldsymbol{\beta}_0,$$

and the score function for $\boldsymbol{\gamma}_j$ for given $\boldsymbol{\beta}_0$ is

$$\boldsymbol{V}_{(i),j}(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_j) = \partial \log p_{i-1,j}(\boldsymbol{Z}_{(i),j}|y_{(i),j})/\partial \boldsymbol{\gamma}_j.$$

A profile likelihood approach is used to estimate $\boldsymbol{\beta}_0$. The major feature of the profile likelihood approach is that all the data are used in the estimation procedures. Using the counting process(Andersen et al., 1993) and martingale(Fleming and Harrington, 1991) theory, Chang and Wang(1999) proved that, under certain regularity conditions, the estimator of $\boldsymbol{\beta}_0$, $\hat{\boldsymbol{\beta}}_0$, is consistent and asymptotically normal, even in the situations where the episode-specific parameters, $\boldsymbol{\gamma}_j$, cannot be estimated consistently.

2.3.2 Multiplicative Marginal Rates/Means Models

Pepe and Cai(1993) proposed a method that was considered to be intermediate between the conditional intensity and marginal hazard approaches by modelling the rate function $\{r_{i1}(t), r_{i2}(t), \cdots\}$, where $r_{ik}(t)$ is defined as

$$r_{ik}(t) = \lim_{\delta \to 0} \frac{1}{\delta} P\left(t < T_{ik} \leq t + \delta | T_{ik} \geq t, T_{i,k-1} < t\right).$$

$r_{ik}(t)$ is the conditional rate of occurrence of the kth event among subjects at risk at time t who have already experienced $(k-1)$ events. The authors modelled the conditional rate with the Cox form, $r_{ik}(t) = r_{0k}(t)e^{\boldsymbol{\beta}_k^T \boldsymbol{Z}_i(t)}$, where $\{r_{0k}(t), k = 1, 2, \cdots\}$ are arbitrary non-negative functions. The estimator of $\boldsymbol{\beta}_k$, $\hat{\boldsymbol{\beta}}_{k:n}$, can be obtained by solving $\boldsymbol{U}_n^k(\boldsymbol{\beta}_k) = 0$. The kth component of the score function is defined as,

$$\boldsymbol{U}_n^k(\boldsymbol{\beta}_k) = \sum_{i=1}^{n} \int_0^t \boldsymbol{Z}_i(s) \left\{ dN_{ik}(s) - Y_{ik}(s)\hat{r}_{0k}(s)e^{\boldsymbol{\beta}_k^T \boldsymbol{Z}_i(s)} ds \right\}.$$

The estimators of the baseline rate functions are given by $\hat{r}_{0k}(t; \hat{\boldsymbol{\beta}}_{k:n}) = n^{-1} \sum dN_{ik}(t)/S_k^{(0)}(t; \hat{\boldsymbol{\beta}}_{k:n})$. A large sample theory for the estimation of the regression parameters was also established.

Lawless and Nadeau(1995) originally proposed the marginal means/rates models. They considered the following semi and fully parametric models:

$$E[dN_i(t)] = m_0(t)g\left(t; \boldsymbol{\beta}_0, \boldsymbol{Z}_i(t)\right),$$

$$E[dN_i(t)] = m_0(t;\alpha)g\left(t;\boldsymbol{\beta}_0,\boldsymbol{Z}_i(t)\right),$$

where $m_0(t)$ is an unspecified non-negative function, $m_0(t;\alpha)$ is a known function with unknown parameter α and $g(.) \geq 0$ is a pre-specified link function.

Lin et al.(2000) provided a rigorous formulation of the marginal means/rate model, and developed inference procedures for the continuous time setting. Lawless and Nadeau(1995) only considered discrete time case and provided no large sample results for the continuous time setting . A semi-parametric continuous time model with a Cox-type link function was proposed. The model is defined through assumption:

$$E[dN_i(t)|\boldsymbol{Z}_i(t)] = \lambda_0(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}_i(t)}ds$$

A proportional rate model was proposed:

$$E[dN_i(t)|\boldsymbol{Z}_i(t)] = d\mu_i(t) = e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}_i(t)}d\mu_0(t)$$

Note that although $d\mu_i(t)$ is always a rate function, $\mu_i(t) = \int_0^t d\mu_i(s)$ represents a mean function only if $\boldsymbol{Z}_i(.)$ consists only of external covariates.
With respect to parameter estimation, $\boldsymbol{\beta}_0$ is estimated by $\hat{\boldsymbol{\beta}}_n$, the solution to $\boldsymbol{U}_n(\boldsymbol{\beta},\tau) = \mathbf{0}_{p\times 1}$, with

$$\boldsymbol{U}_n(\boldsymbol{\beta};t) = \sum_{i=1}^n \int_0^t \{\boldsymbol{Z}_i(s) - \boldsymbol{E}(s;\boldsymbol{\beta})\}dN_i(s),$$

where $\boldsymbol{E}(t;\boldsymbol{\beta}) = \boldsymbol{S}^{(1)}(t;\boldsymbol{\beta})/\ S^{(0)}(t;\boldsymbol{\beta})$, and $\boldsymbol{S}^{(k)}(t;\boldsymbol{\beta}) = n^{-1}\sum_{i=1}^n Y_i(t)\boldsymbol{Z}_i(t)^{\otimes k}e^{\boldsymbol{\beta}^T\boldsymbol{Z}_i(t)}$. The baseline mean is estimated by the Breslow-type estimator

$$\hat{\mu}_0(t;\hat{\boldsymbol{\beta}}_n) = \sum_{i=1}^n \int_0^t \frac{dN_i(s)}{nS^{(0)}(s;\boldsymbol{\beta})}.$$

Lin et al.(2000) show that, under the proportional mean model, $n^{1/2}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \xrightarrow{D}$

$N_p(\mathbf{0}_{p\times 1}, \boldsymbol{A}(\boldsymbol{\beta}_0)^{-1}\boldsymbol{B}(\boldsymbol{\beta}_0)\boldsymbol{A}(\boldsymbol{\beta}_0)^{-1})$, where $\boldsymbol{A}_n(\boldsymbol{\beta}) = -\partial \boldsymbol{U}_n(\boldsymbol{\beta})/\partial \boldsymbol{\beta}^T$ and $\boldsymbol{A}(\boldsymbol{\beta}_0)$ is its limiting value evaluated at $\boldsymbol{\beta}_0$ and $\boldsymbol{B}(\boldsymbol{\beta}) = E[(\int_0^t \{\boldsymbol{Z}_1(s) - \boldsymbol{E}(s;\boldsymbol{\beta})\} dM_1(s;\boldsymbol{\beta}))^{\otimes 2}]$. Procedures of constructing simultaneous confidence bands for the mean function were established as well as graphical and numerical techniques for checking the adequacy of the fitted models.

2.3.3 Accelerated Failure Time Models

Lin, Wei and Ying(1998) provided another extension of the AFT model to accommodate recurrent events when no terminal event is present. This approach is more natural and efficient for handling recurrent events, especially when the numbers of events vary substantially among subjects or the number of events is large. Their proposed model retains the direct physical interpretation of the original AFT model in that the role of the covariates is to accelerate or decelerate the time to each occurrence.

Let T_{ik}, $i = 1, \cdots, n$ and $k = 1, 2, \cdots$, be the kth event time for the ith subjects. Assuming that the subjects are independent, define $N_i^*(t)$ as the number of events that have occurred on the ith subject by time t in the absence of censoring. That is,

$$N_i^*(t) = \sum_{k=1}^{\infty} I(T_{ik} \leq t),$$

where $I(.)$ is the indicator function. We assume the following accelerated failure time(AFT) marginal means model for recurrent events:

$$E[N_i^*(t)|\boldsymbol{Z}_i] = \mu_0(te^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_i}),$$

where $\boldsymbol{Z}_i$ is a $p \times 1$ vector of covariates for the ith subject, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown regression parameters, and $\mu_0(.)$ is an unspecified continuous function. Let $\tilde{T}_{ik}(\boldsymbol{\beta}) = T_{ik}e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_i}$ and $\tilde{N}_i^*(t;\boldsymbol{\beta}) = \sum_{k=1}^{\infty} I\{\tilde{T}_{ik}(\boldsymbol{\beta}) \leq t\}$.

Then the marginal means model is equivalent to

$$E\{\tilde{N}_i^*(t;\boldsymbol{\beta}_0)\} = \mu_0(t).$$

Due to censoring, $\{N_i^*(s) : s > 0\}$ is incompletely observed. Let C_i be the censoring time for the ith subject, which is assumed to be independent of T_{ik}, $k = 1, 2, \cdots$, conditional on $\boldsymbol{Z}_i$, and set $\tilde{C}_i(\boldsymbol{\beta}) = C_i e^{\boldsymbol{\beta}^T \boldsymbol{Z}_i}$. Define

$$N_i(t) = \sum_{k=1}^{\infty} I(T_{ik} \leq t \wedge C_i),$$

where $a \wedge b = \min(a, b)$. On the transformed time scale, the observed counting process becomes:

$$\tilde{N}_i^*(t;\boldsymbol{\beta}) \doteq \sum_{k=1}^{\infty} I\{\tilde{T}_{ik}(\boldsymbol{\beta}) \leq t \wedge \tilde{C}_i(\boldsymbol{\beta})\}$$

Motivated by the partial likelihood score function for the proportional intensity Poisson process model(Andersen & Gill, 1982) and the weighted rank estimating functions for the univariate accelerated failure time model(Prentice, 1978; Tsiatis, 1990; Wei *et al.*, 1990), the authors propose to estimate $\boldsymbol{\beta}_0$ by working with the following estimating function:

$$\boldsymbol{U}_n(t;\boldsymbol{\beta}) = \sum_{i=1}^{n} \int_0^t Q(s;\boldsymbol{\beta})\{\boldsymbol{Z}_i - \bar{\boldsymbol{Z}}(s;\boldsymbol{\beta})\} d\tilde{N}_i(s;\boldsymbol{\beta}),$$

where $Q(t;\boldsymbol{\beta})$ is a specified weight function, $d\tilde{N}_i(t;\boldsymbol{\beta}) = \tilde{Y}_i(t;\boldsymbol{\beta}) d\tilde{N}_i^*(t;\boldsymbol{\beta})$, and $\bar{\boldsymbol{Z}}(s;\boldsymbol{\beta}) = \frac{\sum_{i=1}^n \tilde{Y}_i(s;\boldsymbol{\beta})\boldsymbol{Z}_i}{\tilde{Y}(s;\boldsymbol{\beta})}$, with $\tilde{Y}_i(t;\boldsymbol{\beta}) = I(\tilde{C}_i(\boldsymbol{\beta}) \geq t)$ and $\tilde{Y}(t;\boldsymbol{\beta}) = \sum_{i=1}^n \tilde{Y}_i(t;\boldsymbol{\beta})$.
$\boldsymbol{U}(\boldsymbol{\beta})$ is referred to as the log-rank estimating function if $Q = 1$ and as the Gehan estimating function if $Q(s;\boldsymbol{\beta}) = n^{-1}\tilde{Y}(s;\boldsymbol{\beta})$.

The above estimating function is a piecewise constant function of $\boldsymbol{\beta}$. The estimator of $\boldsymbol{\beta}_0$ is defined to be $\hat{\boldsymbol{\beta}}_n$, the minimizer of $\|\boldsymbol{U}_n(t;\boldsymbol{\beta})\|$, where $\|\mathbf{a}\| = (\mathbf{a}^T\mathbf{a})^{1/2}$. If the number of covariates is small, a direct grid search or the golden section method may be used to obtain $\hat{\boldsymbol{\beta}}$. For high-dimensional covariate vectors, more advanced numerical methods, such as the method of simulated annealing(Kirkpatrick *et al.*, 1983; Lin and Geyer, 1992) may be more efficient.

Given $\hat{\boldsymbol{\beta}}$, $\mu_0(t)$ is estimated by the Nelson-Aalen-type estimator $\hat{\mu}_0(t; \hat{\boldsymbol{\beta}})$, where

$$\hat{\mu}_0(t; \boldsymbol{\beta}) = \sum_{i=1}^{n} \int_0^t \frac{d\tilde{N}_i(s; \boldsymbol{\beta})}{\tilde{Y}(s; \boldsymbol{\beta})}$$

By employing the modern empirical process theory, the authors proved the consistency of $\hat{\boldsymbol{\beta}}_n$ and $\hat{\mu}_0(t; \hat{\boldsymbol{\beta}})$. They also showed that $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ is asymptotically zero-mean normal and $n^{1/2}\{\hat{\mu}_0(t; \hat{\boldsymbol{\beta}}) - \mu_0(t)\}$ converges weakly to a zero-mean Gaussian process.

Chang(2004) considered an accelerated failure time model on the sojourn time between recurrent events. This model assumes that the covariate effect and the subject-specific frailty are additive on the logarithm of sojourn time, and the covariate effect maintains the same over distinct episodes, while the distributions of the frailty and the random error in the model are unspecified. The model is expressed as

$$\log O_{ij} = w_i + \boldsymbol{\beta}^T \boldsymbol{Z}_i + \epsilon_{ij}\,;\; i = 1, \cdots, n,$$

where O_{ij}'s are the sojourn times, $\boldsymbol{\beta}$ is a $p \times 1$ vector of common covariate, w_i is the frailty and $\{\epsilon_{ij}, j \geq 1\}$ are i.i.d random errors with an unspecified distribution. Given $\boldsymbol{Z}_i$, the censoring time C_i is assumed to be conditionally independent of both the frailty and random errors. Two estimation procedures are proposed. The first one is based on the multiple sojourn times O_{ij}'s. Let $\tilde{O}_{ij}(\boldsymbol{\beta}) = O_{ij} e^{-\boldsymbol{\beta}^T \boldsymbol{Z}_i}$, $\tilde{X}_{ij}(\boldsymbol{\beta}) = \min\{\tilde{O}_{ij}(\boldsymbol{\beta}), C_i e^{-\boldsymbol{\beta}^T \boldsymbol{Z}_i} - \sum_{l=1}^{j-1} \tilde{X}_{il}(\boldsymbol{\beta})\}$, k_i be the number of observed events for the ith subject, $k_i^* = \max\{1, k_i\}$ and $\delta_{ij} = I(\sum_{l=1}^{j} O_{il} \leq C_i)$. The estimating function can be expressed as

$$\boldsymbol{U}_u(\boldsymbol{\beta}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{1}{k_i^*} \sum_{j=1}^{k_i^*} \delta_{ij} I(k_i \geq 1) \left[\boldsymbol{Z}_i - \frac{\boldsymbol{S}_1\{\boldsymbol{\beta}, \tilde{X}_{ij}(\boldsymbol{\beta})\}}{S_0\{\boldsymbol{\beta}, \tilde{X}_{ij}(\boldsymbol{\beta})\}} \right],$$

where

$$S_0(\boldsymbol{\beta}, x) = \frac{1}{n}\sum_{l=1}^{n}\frac{1}{k_l^*}\sum_{k=1}^{k_l^*} I(\tilde{X}_{lk}(\boldsymbol{\beta}) \geq x) \text{ and } \boldsymbol{S}_1(\boldsymbol{\beta}, x) = \frac{1}{n}\sum_{l=1}^{n}\frac{1}{k_l^*}\sum_{k=1}^{k_l^*} \boldsymbol{Z}_l I(\tilde{X}_{lk}(\boldsymbol{\beta}) \geq x).$$

The solution of the step function $\boldsymbol{U}_u(\boldsymbol{\beta}) = 0$, $\hat{\boldsymbol{\beta}}_u$, is an estimator of $\boldsymbol{\beta}$. It can be shown that $\hat{\boldsymbol{\beta}}_u$ is consistent and $\sqrt{n}(\hat{\boldsymbol{\beta}}_u - \boldsymbol{\beta})$ is asymptotically normal with mean 0 and covariance matrix that can be estimated by a resampling technique developed by Parzen, Wei and Ying(1994).

The second estimation procedure is based on the ordinal nature of recurrent events. Consider the transformed total time of the jth event, $\tilde{T}_{ij}(\boldsymbol{\beta}) = \sum_{l=1}^{j} O_{il} e^{-\boldsymbol{\beta}^T \boldsymbol{Z}_i}$ for $j \geq 1$. Let $\tilde{Y}_{ij}(\boldsymbol{\beta}) = \min\{\tilde{T}_{ij}(\boldsymbol{\beta}), C_i e^{-\boldsymbol{\beta}^T \boldsymbol{Z}_i}\}$, $K_n = \max_{i=1}^{n}\{k_i\}$. The alternative unbiased estimating function is

$$\boldsymbol{U}_s(\boldsymbol{\beta}) = \frac{1}{\sqrt{n}}\sum_{j=1}^{K_n}\sum_{i=1}^{n}\delta_{ij}\left[\boldsymbol{Z}_i - \frac{\sum_{l=1}^{n}\boldsymbol{Z}_l I\{\tilde{Y}_{lj}(\boldsymbol{\beta}) \geq \tilde{Y}_{ij}(\boldsymbol{\beta})\}}{\sum_{l=1}^{n} I\{\tilde{Y}_{lj}(\boldsymbol{\beta}) \geq \tilde{Y}_{ij}(\boldsymbol{\beta})\}}\right].$$

The estimate of $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}_s$, satisfies $\boldsymbol{U}_s(\boldsymbol{\beta}) = 0$. $\sqrt{n}(\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta})$ is asymptotically normal with mean 0 and covariance matrix that can be estimated by the similar resampling procedure of $\hat{\boldsymbol{\beta}}_u$.

Chang(2004) also studied bivariate recurrent events alternating over time. The dependence among the sojourn times for the type-1 and type-2 events are captured by two correlated frailties. Both the frailties and covariate effects influence the bivariate sojourn times through two bivariate AFT models. The two rank-based estimating methods developed for univariate recurrent event data previously can be extended for parameter estimations based on the transformed bivariate recurrent event data.

Strawderman(2006) studied an accelerated gap time model for dependent gap time data. This semiparametric intensity model relaxes the gap time independence assumption by incorporating gamma frailty into the accelerated gap time model previously developed by Strawderman(2005) for independent gap time data. Suppose for the ith subject, the jth gap time is defined as $O_{ij} = T_{ij} - T_{i,j-1}$, where $j \geq 1$ and $T_{i0} = 0$. Let W_i be a Gamma$(\eta_0^{-1}, \eta_0^{-1})$ random variable with unit mean and variance η_0. Conditionally on W_i, suppose $V_{i1}, V_{i2}, \cdots$ are independent and identically distributed random variables with hazard function $W_i\lambda_0(.)$, where $\lambda_0(.)$ is positive and $\int_0^t \lambda_0(u)du < \infty$ for $t < \infty$. Given $p \times 1$ covariate vector $\boldsymbol{Z}_i$ and W_i, $O_{i1}, O_{i2}, \cdots$ are independent random variables with $O_{ij} = V_{ij}e^{-\boldsymbol{\beta}_0^T\boldsymbol{Z}_i}$, for each $j \geq 1$.

Let C_i be the censoring time for the ith subject, $T_{im} = \sum_{j=1}^m O_{ij}$. Define $N_i(t) = \max(m : T_{im} < (t \wedge C_i))$, $Y_i(t) = I(C_i \geq t)$, and $a \wedge b = \min(a, b)$. Let N_i denote the number of events experienced by subject i. To accommodate censoring it is convenient to formulate the model using intensities. Given the observed covariates $\{\boldsymbol{Z}_1, \cdots, \boldsymbol{Z}_n\}$ and unobserved frailties $\{W_1, \cdots, W_n\}$, the process $(N_i(t), i = 1, \cdots, n)$ form a multivariate counting process, with the ith component having intensity

$$\lambda_i(t) = W_i\lambda_0\{e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}_i}R_i(t)\}e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}_i}Y_i(t),$$

where $R_i(t) = t - T_{i,N_i(t-)}$. The model assumes that censoring is noninformative, conditionally on the frailty and covariates.

An "Expectation Substitution"(ES) algorithm that is similar to EM algorithm is used to estimate $\boldsymbol{\psi} = (\boldsymbol{\beta}_0, \eta_0, \Lambda_0)$, where $\Lambda_0(.) = \int_0^{\cdot} \lambda_0(u)du$. Specifically, let ES$(\eta)$ denote the E- and S-steps. For fixed $\eta = \hat{\eta}_{[0]}$, we run ES$(\hat{\eta}_{[0]})$ to get estimates $\hat{\Lambda}_{[1]}$, which can be obtained from

$$\hat{\Lambda}(t|\boldsymbol{\beta}) = \sum_{i=1}^{n}\sum_{j=1}^{N_i} \frac{I(X_{ij(\boldsymbol{\beta})} \leq t)}{\sum_{r=1}^n \sum_{s=1}^{N_r+1} \hat{W}_r I(X_{rs}(\boldsymbol{\beta}) \geq X_{ij}(\boldsymbol{\beta}))}$$

and $\hat{\boldsymbol{\beta}}_{[1]}$ is the zero-crossing or the minimizer of the following estimating function

$$\boldsymbol{U}_H(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{N_i} H(X_{ij}(\boldsymbol{\beta})|\boldsymbol{\beta}) \left\{ \boldsymbol{Z}_i - \hat{\mathbf{E}}(X_{ij}(\boldsymbol{\beta})|\boldsymbol{\beta}) \right\}.$$

Here $X_{ij}(\boldsymbol{\beta}) = O_{ij}e^{\boldsymbol{\beta}^T\boldsymbol{Z}_i}$ for $j \le N_i$, $X_{ij}(\boldsymbol{\beta}) = (C_i - T_{i,N_i})e^{\boldsymbol{\beta}^T\boldsymbol{Z}_i}$ for $j = N_i + 1$,

$$\hat{W}_i = \frac{1 + \eta\sum_{j=1}^{N_i} I(O_{ij}e^{\boldsymbol{\beta}^T\boldsymbol{Z}_i} \le t)}{1 + \eta\sum_{j=1}^{N_i+1}\hat{\Lambda}(X_{ij}(\hat{\boldsymbol{\beta}}))},\ \hat{\mathbf{E}}(t|\boldsymbol{\beta}) = \frac{\sum_{k=1}^{n}\sum_{r=1}^{N_k+1}\hat{W}_k\boldsymbol{Z}_k I(X_{kr}(\boldsymbol{\beta}) \ge t)}{\sum_{k=1}^{n}\sum_{r=1}^{N_k+1}\hat{W}_k I(X_{kr}(\boldsymbol{\beta}) \ge t)},$$

and $H(t|\boldsymbol{\beta})$ is a data-dependent weight function. Then, $\hat{\eta}_{[1]}$ is determined by maximizing the observed data log likelihood

$$\begin{aligned} l_p(\eta) \ = \ & \sum_{i=1}^{n}\left(\sum_{j=1}^{N_i}\log[1 + \eta(j-1)] - \left(1 + \eta\sum_{j=1}^{N_i} I(O_{ij}e^{\boldsymbol{\beta}^T\boldsymbol{Z}_i} \le t)\right)\cdot\right. \\ & \left.\left\{\eta^{-1}\log\left[1 + \eta\sum_{j=1}^{N_i+1}\Lambda(X_{ij}(\boldsymbol{\beta}))\right]\right\}\right). \end{aligned}$$

Although there is no guarantee of local or global convergence of an ES algorithm, the conditions needed for convergence are typically satisfied in a limiting sense, that is, implied by smoothness conditions needed for establishing consistency and weak convergence of the estimators. It's also important to use a good set of starting values to achieve local convergence.

2.4 Semiparametric Models for Recurrent and Terminal Events data

The models mentioned in the previous section focus on recurrent events data that are not terminated by death. In many medical studies, however, there exists terminal events which preclude the occurrence of further recurrent events. Therefore, it is desirable to investigate models that deal with recurrent events while taking the terminal events into account.

2.4.1 Multiplicative Conditional Models

A multiplicative intensity model is considered by Wang, Qin and Chiang(2001) as the underlying model for nonparametric estimation of the cumulative rate function and is extended to a regression model if taking the covariates into account. In their paper, the occurrence of recurrent events is modelled by a subject-specific nonstationary Poisson process via a latent variable. Both the censoring and latent variables are treated as nuisance parameters.

Let $N(t)$ be the number of recurrent events occurring at or before t, $t \geq 0$, and W be a nonnegative valued latent variable. Define the unconditional rate function of a continuous recurrent event process at t, $t \geq 0$, as

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

and the cumulative rate function(CRF) as $\Lambda(t) = \int_0^t \lambda(u)du$. The first model assumes that conditioning on $W = w$, $N(t)$ is a nonstationary Poisson process with the intensity function $W\lambda_0(t)$, where $\lambda_0(t)$ is a continuous baseline intensity function. It's also assumed that conditioning on the latent variable, $N(.)$ is independent of C, the censoring time. When considering the association between $N(.)$ and $p \times 1$ vector of covariate $\boldsymbol{Z}$, $N(t)$ is a nonstationary Poisson process with the intensity function $W\lambda_0(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}}$, where $\boldsymbol{\beta}$ is a $p \times 1$ vector of parameters, $\lambda_0(t)$ is a continuous baseline function, and the latent variable W satisfies $E[W|\boldsymbol{Z}] = 1$. Given $(W, \boldsymbol{Z})$, $N(.)$ is independent of C. The second model implies the marginal proportional rate function

$$\lambda(t|\boldsymbol{Z}) = \lambda_0(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}}.$$

In order to adjust the bias in the risk sets due to dependent censoring, conditional likelihood approaches are used in the estimation procedures of cumulative rate function $\Lambda(t)$ and parameter $\boldsymbol{\beta}$. For subject i, $i = 1, \cdots, n$, let C_i denote the observed censoring time and let $T_{i1} \leq \cdots \leq T_{i,m_i}$ be the observed event time with m_i defined as the index of the last event occurring at or before C_i. Under regularity conditions, the cumulative distribution function can be estimated as

$$\hat{F}(t) = \prod_{s_{(l)}>t} \left(1 - \frac{d_{(l)}}{N_{(l)}}\right),$$

where $\{s_{(l)}\}$ are the ordered and distinct values of the event time T_{ij}, $d_{(l)}$ is the number of events occurring at $s_{(l)}$, and $N_{(l)}$ is the total number of events with event time and censoring time satisfying $T_{ij} \leq s_{(l)} \leq C_i$.

Define the density function

$$f(t) = \frac{\lambda_0(t)I(0 \leq t \leq T_0)}{\Lambda_0(T_0)} = \frac{W_i\lambda_0(t)I(0 \leq t \leq T_0)}{W_i\Lambda_0(T_0)},$$

and $F(t)$ as the corresponding CDF. Then the CRF $\Lambda(t)$ is related to F through the formulation $\Lambda(t) = F(t)\Lambda(T_0)$, where the parameter $\Lambda(T_0)$ is interpreted as the average number of recurrent events occurring in $[0, T_0]$. Conditioning on (C_i, W_i), the number of the observed events, m_i, has expected value $W_i\Lambda_0(C_i)$. We have,

$$\begin{aligned} E\left[m_iF^{-1}(C_i)\right] &= E\left[E\left[m_iF^{-1}(C_i)|C_i, W_i\right]\right] \\ &= E\left[W_i\Lambda_0(C_i)F^{-1}(C_i)\right] \\ &= E\left[W_i\Lambda_0(T_0)\right] = \Lambda(T_0). \end{aligned}$$

Therefore, the estimator of cumulative rate function can be constructed as

$$\hat{\Lambda}(t) = \hat{F}(t)\hat{\Lambda}(T_0), \text{ where } \hat{\Lambda}(T_0) = \frac{1}{n}\sum_{i=1}^{n} m_i \hat{F}^{-1}(C_i).$$

Asymptotic properties of $\hat{\Lambda}(t)$ and $\hat{\Lambda}(T_0)$ are studied. Similarly, conditioning on $(\boldsymbol{Z}_i, C_i, W_i)$, the expected value of m_i is $W_i \Lambda_0(C_i)\exp(\boldsymbol{\beta}^T \boldsymbol{Z}_i)$. Thus

$$\begin{aligned} E\left[m_i F^{-1}(C_i)|\boldsymbol{Z}_i\right] &= E\left[E\left[m_i F^{-1}(C_i)|C_i, W_i\right]|\boldsymbol{Z}_i\right] \\ &= E\left[W_i e^{\boldsymbol{\beta}^T \boldsymbol{Z}_i}\Lambda_0(C_i)F^{-1}(C_i)|\boldsymbol{Z}_i\right] \\ &= e^{\boldsymbol{\beta}^T \boldsymbol{Z}_i}\Lambda_0(T_0)E\left[W_i|\boldsymbol{Z}_i\right] = e^{\boldsymbol{\beta}^T \boldsymbol{Z}_i}\boldsymbol{\beta}_0 \text{ with } \boldsymbol{\beta}_0 = \Lambda_0(T_0). \end{aligned}$$

A class of unbiased estimating equations can be derived as follows:

$$n^{-1}\sum_{i=1}^{n} q_i \bar{\boldsymbol{Z}}_i (m_i \hat{F}^{-1}(C_i) - e^{\boldsymbol{\gamma}^T \bar{\boldsymbol{Z}}_i}) = \boldsymbol{0},$$

where $\bar{\boldsymbol{Z}}_i = (1, \boldsymbol{Z}_i^T)^T$, $\boldsymbol{\gamma} = (\ln \Lambda_0(T_0), \boldsymbol{\beta}^T)^T$, and q_i is a weight function depending on $(\boldsymbol{Z}_i, \boldsymbol{\gamma}, F)$. It can be shown that the solution of above equations, $\hat{\boldsymbol{\gamma}}$, is asymptotically normal.

Liu, Wolfe and Huang(2004) considered a joint semiparametric model for the intensity functions of both recurrent events and death by a shared gamma frailty. The observation for subject i is $\mathbf{O}_i(t) = \{Y_i(u), N_i^D(u), N_i^R(u), 0 \le u \le t\}$, where $Y_i(t) = I(X_i \ge t)$ be the at-risk indicator, $N_i^D(t) = I(X_i \le t, \Delta_i = 1)$ is the observed death indicator by time t and $N_i^R(t)$ is the observed number of recurrent events by time t. Assuming

$$P(dN_i^D(t) = 1|\mathcal{F}_{t^-}) = Y_i(t)\lambda_i(t)dt$$

and

$$P(dN_i^R(t) = 1|\mathcal{F}_{t^-}, D \geq t) = Y_i(t)r_i(t)dt.$$

The authors extended the model in Huang and Wolfe(2002) to the new setting as

$$r_i(t) = \nu_i e^{\boldsymbol{\beta}^T \boldsymbol{Z}_i} r_0(t),$$

$$\lambda_i(t) = \nu_i^\gamma e^{\boldsymbol{\alpha}^T \boldsymbol{Z}_i} \lambda_0(t).$$

The common frailty parameter ν follows a gamma distribution with unit mean and variance θ. The full likelihood is proportional to the product of these two terms

$$\exp\left\{-\int_0^\infty Y_i(t)\nu_i e^{\boldsymbol{\beta}^T \boldsymbol{Z}_i} dR_0(t)\right\} \times \prod_j \left[\nu_i e^{\boldsymbol{\beta}^T \boldsymbol{Z}_i} dR_0(t_{ij})\right]^{\delta_{ij}},$$

and

$$\exp\left\{-\int_0^\infty Y_i(t)\nu_i^\gamma e^{\boldsymbol{\alpha}^T \boldsymbol{Z}_i} d\Lambda_0(t)\right\} \times \left[\nu_i^\gamma e^{\boldsymbol{\alpha}^T \boldsymbol{Z}_i} d\Lambda_0(x_i)\right]^{\Delta_{ij}}.$$

The log likelihood gives the likelihood for the complete data with known frailties, which is more readily maximized than the observed data likelihood. This makes the EM algorithm a natural choice for parameter estimation.

2.4.2 Multiplicative Marginal Mean/Rate Models

Cook and Lawless(1997) study methods of analyzing recurrent and terminal events based on mean and rate functions for the recurrent events. Assuming there is just one type of recurrent event and one type of terminal event, define $N_i(t)$ as the number of recurrent events experienced by subject i up to time t and T_i as the failure time. $\boldsymbol{Z}_i = (z_{1i}, \cdots, z_{pi})^T$ is the $p \times 1$ fixed covariates vector for the ith subject. Subject i is observed for a fixed study time τ_i unless they experience failure first, and τ_i is independent of the recurrent events and the failure time.

The mean and rate functions for recurrent events are defined as $M_i(t) = E\{N_i(t)|\boldsymbol{Z}_i\}$ and $m_i(t) = M_i'(t)$, respectively. The authors considered two approaches to specify the effect of failure time on the recurrent events. One way is to consider rate of occurrence functions

$$r_i(s;t) = (d/ds)E\{N_i(s)|T_i = t, \boldsymbol{Z}_i\}, s \leq t$$

with associated mean functions

$$R_i(s;t) = \int_0^s r_i(u;t)du = E\{N_i(s)|T_i = t, \boldsymbol{Z}_i\}, i = 1, \cdots, n.$$

The second approach considers rate of occurrence functions:

$$m_i(s;t) = (d/ds)E\{N_i(s)|T_i \geq t, \boldsymbol{Z}_i\}, s \leq t$$

associated mean functions:

$$M_i(s;t) = \int_0^s m_i(u;t)du = E\{N_i(s)|T_i \geq t, \boldsymbol{Z}_i\}, i = 1, \cdots, n.$$

The first approach is natural, but inconvenient to deal with censored failure ties. The second rate function turns out to deal with censored failure ties rather easily.

Suppose that n subjects all have the same rate function, $m_i(s;t) = m(s;t), i = 1, \cdots, n$. Letting $\delta_i(t) = I(T_i \geq t)I(\tau_i \geq t)$, noting that $E\{dN_i(s)|\delta_i(t) = 1\} = m(s;t)$, the nonparametric estimates based on the second rate function are

$$\hat{m}(s;t) = \frac{\sum\limits_{i=1}^{n} \delta_i(t)dN_i(s)}{\sum\limits_{i=1}^{n} \delta_i(t)}$$

and

$$\hat{M}(s;t) = \sum_{u=1}^{s} \hat{m}(u;t) = \frac{\sum_{i=1}^{n} \delta_i(t) N_i(s)}{\sum_{i=1}^{n} \delta_i(t)}.$$

Variance of $\hat{M}(s;t)$ is estimated consistently by

$$\hat{V}(s;t) = \sum_{i=1}^{n} \{\frac{\delta_i(t)}{\delta(t)}\}^2 \{N_i(s) - \hat{M}(s;t)\}^2$$

Semi-parametric models can be set up based on the first rate function and parameter estimates are obtained by utilizing the relation of the two rate functions

$$m_i(s;t) = \int_0^{\infty} r_i(s;v) \frac{f_i(v)}{\bar{F}_i(t)} dv, s \le t,$$

where $F_i(t)$ is the CDF of the failure time T_i, $\bar{F}_i(t) = 1 - F_i(t)$ and $f_i(t) = F_i'(t)$.

Li and Lagakos(1997) discuss the use of the Wei-Lin-Weissfeld (WLW) approach with data involving recurrent events and terminal event. For each individual, let T_k be time from start until kth recurrent event or death, whichever comes first and the indicator variable

$$a_k = \begin{cases} 1 & \text{if the subject experieces at least } k \text{ recurrent events} \\ 2 & \text{otherwise} \end{cases}$$

a_k indicates whether death occurs after($a_k = 1$) or before ($a_k = 2$) the kth recurrent event. Thus a subject's experience is specified by $(T_1, a_1, T_2, a_2, \cdots)$, the joint distribution of the T_k and a_k. One way to describe this joint distribution is with the cause-specific hazard function, where $j = 1$ denotes recurrent event and $j = 2$ denotes death

$$\mu_{1j}(t) = \lim_{h \to 0} \frac{1}{h} P(t \le T_1 < t + h, a_1 = j | T_1 \ge t)$$

and, for $k > 1$

$$
\begin{aligned}
& \mu_{kj}(t|t_1, t_2, \cdots t_{k-1}) \\
= \ & \lim_{h \to 0} \frac{1}{h} P(t \le T_k < t + h, a_k = j | T_1 = t_1, a_1 = 1, \cdots, T_{k-1} = t_{k-1}, a_{k-1} = 1, T_k \ge t).
\end{aligned}
$$

The marginal hazard function for T_k is given by

$$h_k(t) = \lim_{h \to 0} \frac{1}{h} P(t \le T_k < t + h | T_k \ge t), \ \ k \ge 1;$$

the marginal cause-specific hazard for a kth recurrent event at time t is

$$h_{k1}(t) = \lim_{h \to 0} \frac{1}{h} P(t \le T_k < t + h, a_k = 1 | T_k \ge t), \ \ k \ge 1;$$

the hazard for a kth recurrent event or death at time t is

$$h_{kp}(t) = \lim_{h \to 0} \frac{1}{h} P(t \le T_k < t + h, a_{k-1} = 1 | T_k \ge t), \ \ k \ge 1;$$

and the conditional hazard function for T_k, given a kth recurrent event occurs, is

$$h_{k|p}(t) = \lim_{h \to 0} \frac{1}{h} P(t \le T_k < t + h | a_k = 1, T_k \ge t), \ \ k \ge 1.$$

The authors discussed various scenarios of defining a subject's K "events" which lead to inferences about different probabilistic aspects of the process. When applying the WLW method, different hazard functions are selected, depending on how the scenario defines an "observed time" and "censoring indicator" for each of the K failure times. In particular, two choices of scenarios seem to be the most natural. The first one treats death as a censoring event. The marginal cause-specific hazard function and the hazard function for death are used in the WLW analysis. In the second method, time until the minimum of recurrent event or death is analyzed based on the marginal hazard function for T_k. The first method focuses on comparison of groups with respect to recurrent event process, while the second method makes no distinction between the recurrent and terminal events.

Ghosh and Lin(2002) consider the marginal mean function for the cumulative number of recurrent events over time, acknowledging the fact that death stops further recurrences. Essentially, a proportional means model is assumed to hold marginally across survival status.

For the ith subject, let $N_i^*(t)$ be the number of recurrent events over time interval $[0,t]$, D_i be the failure time, $\boldsymbol{Z}_i$ be a $p \times 1$ vector of covariates, C_i denote the censoring time and assume that $N_i^*(.)$ is independent of C_i conditional on $\boldsymbol{Z}_i(.)$. Notice that only the minimum of D_i and C_i is observed, we write $X_i = D_i \wedge C_i$, $\delta_i = I(D_i \leq C_i)$ and $N_i(t) = N_i^*(t \wedge C)$, where $a \wedge b = \min(a,b)$ and $I(.)$ is the indicator function. For a random sample of n subjects, the data consist of $\{N_i(.), X_i, \delta_i, \boldsymbol{Z}_i(.)\}$, $i = 1, \cdots, n$.

Define the marginal mean function as $\mu_{\boldsymbol{Z}_i}(t) = E\{N_i^*(t)|\boldsymbol{Z}_i\}$ and formulate it through the semiparametric models

$$\mu_{\boldsymbol{Z}_i}(t) = e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_i}\mu_0(t),$$

where $\mu_0(.)$ is an unspecified continuous function, and $\boldsymbol{\beta}_0$ is a $p \times 1$ vector of unknown regression parameters.

To accommodate time-varying covariates, the rate function

$$d\mu_{\boldsymbol{Z}_i}(t) = E\{dN_i^*(t)|\boldsymbol{Z}_i(s) : s \geq 0\},$$

where $\boldsymbol{Z}(.)$ is a p-dimensional external covariate process (Kalbfleisch and Prentice(2002), is also considered. We have

$$d\mu_{\boldsymbol{Z}_i}(t) = e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_i(t)} d\mu_0(t),$$

Notice that $\mu_{\boldsymbol{Z}_i}(t) = \int_0^t e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_i(s)} d\mu_0(s) = e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_i}\mu_0(t)$ if covariates are all time-invariant.

If the censoring times C_i,$i = 1, \cdots, n$ are known, one can use the following estimating function(Lin, Wei, Yang, Ying, 2000) to estimate $\boldsymbol{\beta}_0$:

$$\boldsymbol{U}(\boldsymbol{\beta}) = \sum_{i=1}^n \int_0^{\tau^*} \left\{ \boldsymbol{Z}_i(t) - \frac{\sum_{j=1}^n I(C_i \geq t)\boldsymbol{Z}_j(t) e^{\boldsymbol{\beta}^T \boldsymbol{Z}_j(t)}}{\sum_{j=1}^n I(C_i \geq t) e^{\boldsymbol{\beta}^T \boldsymbol{Z}_j(t)}} \right\} I(C_i \geq t) dN_i(t),$$

where τ^* is a constant such that $P(C_i \geq \tau^*) > 0$, $i = 1, \cdots, n$. In most applications, there is potential loss to follow-up. If the ith subject dies before he/she is censored, then C_i is unknown and the above estimating function can not be evaluated. Two modifications are considered to replace $I(C_i \geq t)$, $i = 1, \cdots, n$, which are observable and have the same expectation as $I(C_i \geq t)$.

The first method is related to the inverse probability of censoring weighting (IPCW) technique (Robins and Rotnitzky(1992)) by considering the quantity $W_i^C(t) = I(C_i \geq D_i \wedge t)G(t|\boldsymbol{Z}_i)/G(X_i \wedge t|\boldsymbol{Z}_i)$, where $G(t|\boldsymbol{Z})$ is the survival function of C conditional on $\boldsymbol{Z}(.)$. Note that $E\{W_i^C(t)|\boldsymbol{Z}_i\} = G(t|\boldsymbol{Z}_i)$.

By formulating $G(t|\boldsymbol{Z})$ through the Cox proportional hazards model(Cox, 1972), one can estimate $\hat{G}(t|\boldsymbol{Z})$ and thus approximate $W_i^C(t)$ by $\widehat{W}_i^C(t) = I(C_i \geq D_i \wedge t)\hat{G}(t|\boldsymbol{Z}_i)/\hat{G}(X_i \wedge t|\boldsymbol{Z}_i)$. The modified estimating function for $\boldsymbol{\beta}_0$:

$$\boldsymbol{U}^C(\boldsymbol{\beta}) = \sum_{i=1}^n \int_0^{\tau} \left\{ \boldsymbol{Z}_i(t) - \frac{\sum_{j=1}^n \widehat{W}_i^C(t)\boldsymbol{Z}_j(t) e^{\boldsymbol{\beta}^T \boldsymbol{Z}_j(t)}}{\sum_{j=1}^n \widehat{W}_i^C(t) e^{\boldsymbol{\beta}^T \boldsymbol{Z}_j(t)}} \right\} \widehat{W}_i^C(t) dN_i(t),$$

where τ is a constant such that $P(X_i \geq \tau) > 0$, $i = 1, \cdots, n$.

The corresponding estimator of the baseline mean function $\mu_0(.)$ is given by the Brewlow-type estimator

$$\hat{\mu}_0^C(t) = \sum_{i=1}^n \int_0^t \frac{\widehat{W}_i^C(s) dN_i(s)}{\sum_{j=1}^n \widehat{W}_i^C(s) e^{\widehat{\boldsymbol{\beta}}_C^T \boldsymbol{Z}_j(s)}}, 0 \leq t \leq \tau,$$

where $\widehat{\boldsymbol{\beta}}_C$ is the solution to $\boldsymbol{U}^C(\boldsymbol{\beta}) = 0$.

Under regularity conditions similar to those of Andersen and Gill (1982, Thm 4.1), the authors prove that $\hat{\boldsymbol{\beta}}_C \xrightarrow{a.s.} \boldsymbol{\beta}_0$ and $n^{1/2}(\hat{\boldsymbol{\beta}}_C - \boldsymbol{\beta}_0)$ converges in distribution to a zero-mean normal random vector with a covariance matrix that can be consistently estimated. Asymptotic properties of $\hat{\mu}_0^C(t)$ are also studied. The process $n^{1/2}(\hat{\mu}_0^C(t) - \mu_0(t)), 0 \leq t \leq \tau$, converges weakly to a zero-mean Gaussian process whose covariance function can be consistently estimated.

The second modification in Ghosh and Lin(2002) paper involves modeling the survival distribution which is referred to as inverse probability of survival weighting (IPSW). It shares the spirit of the IPCW technique. Since X_i is always observable, we substitute $W_i^D(t)/I(X_i \geq t)/S(t|\boldsymbol{Z}_i)$ for $I(C_i \geq t)$ in the original estimating function, where $S(t|\boldsymbol{Z}_i) \equiv P(D_i \geq t|\boldsymbol{Z}_i)$. Similarly, assuming that D and C are independent conditional on $\boldsymbol{Z}(.)$, $E\{W_i^D(t)|\boldsymbol{Z}_i\} = G(t|\boldsymbol{Z}_i)$. By specifying the proportional hazard model for the survival distribution, one can get estimate $\hat{S}(t|\boldsymbol{Z})$ and thus approximate $W_i^D(t)$ by $\widehat{W}_i^D(t) \equiv I(X_i \geq t)/\hat{S}(t|\boldsymbol{Z})$.

The modified estimating function is

$$\boldsymbol{U}^D(\boldsymbol{\beta}) = \sum_{i=1}^{n} \int_0^{\tau} \left\{ \boldsymbol{Z}_i(t) - \frac{\sum_{j=1}^{n} \widehat{W}_i^D(t)\boldsymbol{Z}_j(t)e^{\boldsymbol{\beta}^T \boldsymbol{Z}_j(t)}}{\sum_{j=1}^{n} \widehat{W}_i^D(t)e^{\boldsymbol{\beta}^T \boldsymbol{Z}_j(t)}} \right\} \widehat{W}_i^D(t)dN_i(t),$$

where τ is a constant such that $P(X_i \geq \tau) > 0$, $i = 1, \cdots, n$.

Let $\hat{\boldsymbol{\beta}}_D$ be the solution to $\boldsymbol{U}^D(\boldsymbol{\beta}) = 0$. The Brewlow-type estimator of the baseline mean function $\mu_0(.)$ is given by

$$\hat{\mu}_0^D(t) = \sum_{i=1}^{n} \int_0^{t} \frac{\widehat{W}_i^D(s)dN_i(s)}{\sum_{j=1}^{n} \widehat{W}_i^D(s)e^{\widehat{\boldsymbol{\beta}}_D^T \boldsymbol{Z}_j(s)}}, 0 \leq t \leq \tau.$$

Imposing the same regularity conditions, asymptotic properties of these estimators are established. It can be shown that $\hat{\boldsymbol{\beta}}_D \xrightarrow{a.s.} \boldsymbol{\beta}_0$ and $n^{1/2}(\hat{\boldsymbol{\beta}}_D - \boldsymbol{\beta}_0)$ converges in distribution to a zero-mean normal random vector with a covariance matrix that can be consistently estimated. Asymptotic properties of $\hat{\mu}_0^D(t)$ are also studied. Weak convergence of the process $n^{1/2}(\hat{\mu}_0^D(t) - \mu_0(t))$ to a zero-mean Gaussian process is also proven in a similar manner.

2.4.3 Accelerated Failure Times Models(joint models)

Ghosh and Lin(2003) formulated the marginal distributions of the recurrent event process and dependent censoring time through a joint model, while leaving the dependence structure and distributional form unspecified. Suppose that there exist unknown constant vectors $\boldsymbol{\eta}_0$ and $\boldsymbol{\beta}_0$ such that, for given $\boldsymbol{Z}_i$ and t, the bivariate random vectors $\{D_i e^{\boldsymbol{\eta}_0^T \boldsymbol{Z}_i}, N_i^*(te^{-\boldsymbol{\beta}_0^T \boldsymbol{Z}_i})\}^T$,$i = 1, \cdots, n$ have a common but unspecified joint distribution,

$$\begin{bmatrix} D_i e^{\boldsymbol{\eta}_0^T \boldsymbol{Z}_i} \\ N_i^*(te^{-\boldsymbol{\beta}_0^T \boldsymbol{Z}_i}) \end{bmatrix} = \begin{bmatrix} D_0 \\ N_0^*(t) \end{bmatrix},$$

where $\{D_0, N_0^*(t)\}^T$ has an arbitrary bivariate distributions. Two major forms of dependent censoring were considered. The first case is when subjects withdraw from the study for reasons related to the recurrent event process. The second case is death. Under the joint model, the marginal distribution for the dependent censoring time satisfies the familiar accelerated failure time model for the univariate case, while that of the recurrent event times satisfy the AFT model for the counting processes(Lin, Wei, and Ying, 1998). Interest focuses on estimation of $\boldsymbol{\theta}_0 = (\boldsymbol{\eta}_0^T, \boldsymbol{\beta}_0^T)^T$.

The existing methods for the AFT model can be used to estimate $\boldsymbol{\eta}_0$. Let $\tilde{X}_i(\boldsymbol{\eta}) = X_i e^{\boldsymbol{\eta}_0^T \boldsymbol{Z}_i}$, $\tilde{T}_{ik}(\boldsymbol{\beta}) = T_{ik} e^{\boldsymbol{\beta}^T \boldsymbol{Z}_i}$ and $\tilde{X}_i(\boldsymbol{\beta}) = X_i e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}_i}$, $i = 1, \cdots, n; k = 1, 2, \cdots$. The log-rank estimating function is

$$\boldsymbol{U}_1(\boldsymbol{\eta}) = \sum_{i=1}^{n} I(D_i \le C_i) \left[\boldsymbol{Z}_i - \frac{\sum_{j=1}^{n} I\left\{\tilde{X}_j(\boldsymbol{\eta}) \ge \tilde{X}_i(\boldsymbol{\eta})\right\} \boldsymbol{Z}_j}{\sum_{j=1}^{n} I\left\{\tilde{X}_j(\boldsymbol{\eta}) \ge \tilde{X}_i(\boldsymbol{\eta})\right\}} \right].$$

$\hat{\boldsymbol{\eta}}$ is the solution to $\boldsymbol{U}_1(\boldsymbol{\eta}) = \mathbf{0}$. It can be shown that $\hat{\boldsymbol{\eta}}$ is consistent and asymptotically normal(Tsiatis, 1990).

For fixed $\boldsymbol{\eta}$, one can estimate $\boldsymbol{\beta}$ using the estimating function

$$\boldsymbol{U}_2(\boldsymbol{\beta}; \boldsymbol{\eta}) = \sum_{i=1}^{n} \int_0^{\infty} \left\{ \boldsymbol{Z}_i - \bar{\boldsymbol{Z}}^{(2)}(t; \boldsymbol{\theta}) \right\} dN_{2i}(t; \boldsymbol{\theta}), \text{ where}$$

$\bar{\boldsymbol{Z}}^{(2)}(t; \boldsymbol{\theta}) = \sum_{j=1}^{n} I\{\tilde{X}_i(\boldsymbol{\theta}) \ge t\} \boldsymbol{Z}_j / \sum_{j=1}^{n} I\{\tilde{X}_i(\boldsymbol{\theta}) \ge t\}$, $N_{2i}(t; \boldsymbol{\theta}) = \sum_{k=0}^{\infty} I\left\{\tilde{T}_{ik}(\boldsymbol{\beta}) \le t \wedge \tilde{X}_i(\boldsymbol{\theta})\right\}$. Given $\hat{\boldsymbol{\eta}}$, $\hat{\boldsymbol{\beta}}$ is a zero-crossing of $\boldsymbol{U}_2(\boldsymbol{\beta}; \boldsymbol{\eta})$. Let $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\eta}})^T$, large sample properties of $\hat{\boldsymbol{\theta}}$, such as consistency and asymptotic normality, have been established.

Huang and Wang(2003) also modelled the recurrent events and the failure time jointly. For the ith subject, $i = 1, \cdots, n$, let D_i be the failure time, $R(t)$ be the number of recurrent events occurring before or at time t, C_{D_i} be the censoring time for failure, and C_{R_i} be the censoring time for recurrent events. It is also assumed that C_{D_i} may exceed C_{R_i} with $Pr(C_{D_i} \ge C_{R_i}) = 1$. Define $X_{D_i} = D_i \wedge C_{D_i}$, $\Delta_i = I(D_i \le C_{D_i})$, $X_{R_i} = D_i \wedge C_{R_i}$, $Q_i(t) = R(t \wedge D_i \wedge C_{R_i})$, $\forall t$. The first joint model A for D_i and $R(D_i)$ is

$$\begin{cases} \log D_i = \boldsymbol{\beta}_D^T \boldsymbol{Z}_i + \log D_0 \\ E\{R(D_i) | D_i, \boldsymbol{Z}_i\} = e^{\boldsymbol{\beta}_R^T \boldsymbol{Z}_i} E\{R_0(D_0) | D_0\} \\ \{D_i, R(D_i)\} \perp (C_{D_i}, C_{R_i}) | \boldsymbol{Z}_i \end{cases},$$

where $\boldsymbol{Z}_i$ is a $p \times 1$ vector of covariates, $\boldsymbol{\beta}_D$ is a $p \times 1$ parameter vector for failure time, $\boldsymbol{\beta}_R$ is a $p \times 1$ parameter vector for recurrent events, D_0 is the baseline failure time, and $R_0(t)$ is the baseline $R(t)$. The second model B imposes the same effects at every point of the baseline process $[D_0, E\{R_0(t)|D_0\}]$,

$$\begin{cases} \log D_i = \boldsymbol{\beta}_D^T \boldsymbol{Z}_i + \log D_0 \\ E[R\{te^{\boldsymbol{\beta}_D^T \boldsymbol{Z}_i}\}|D_i, \boldsymbol{Z}_i] = e^{\boldsymbol{\beta}_R^T \boldsymbol{Z}_i} E\{R_0(t)|D_0\}, \, \forall t \in [0, D_0] \\ \{D_i, R(t)\} \perp (C_{D_i}, C_{R_i})|\boldsymbol{Z}_i. \end{cases} \cdot$$

Model B is nested in model A by adopting stronger assumptions on both the covariate effects and the censoring mechanism. Meanwhile, model B makes a greater use of the available data.

Define at-risk process $Y_{D_i}(t; \boldsymbol{\beta}_D) = I(X_{D_i} \geq te^{\boldsymbol{\beta}_D^T \boldsymbol{Z}_i})$ and counting process $N_{D_i}(t; \boldsymbol{\beta}_D) = I(X_{D_i} \leq te^{\boldsymbol{\beta}_D^T \boldsymbol{Z}_i})\Delta_i$. Estimation of $\boldsymbol{\beta}_D$ can be carried out through the weighted log-rank estimating function(Tsiatis, 1990)

$$\boldsymbol{U}_D(\boldsymbol{\beta}_D) = n^{-1} \sum_{i=1}^{n} \int_0^\infty \phi_D(s; \boldsymbol{\beta}_D) \left\{ \boldsymbol{Z}_i - \frac{\sum_{j=1}^n Y_{D_j}(s; \boldsymbol{\beta}_D) \boldsymbol{Z}_j}{\sum_{j=1}^n Y_{D_j}(s; \boldsymbol{\beta}_D)} \right\} dN_{D_i}(s; \boldsymbol{\beta}_D),$$

where $\phi_D(t; \boldsymbol{\beta}_D)$ is a nonnegative weight function with special case $\phi_D(t; \boldsymbol{\beta}_D) = 1$ and $\phi_D(t; \boldsymbol{\beta}_D) = n^{-1} \sum_{i=1}^n Y_{D_i}(t; \boldsymbol{\beta}_D)$ correspond to the long-rank and Gehan estimating functions, respectively.

Under Model A, define at-risk process $Y_{R_i}(t; \boldsymbol{\beta}_R) = I(X_{R_i} \geq te^{\boldsymbol{\beta}_D^T \boldsymbol{Z}_i})$, counting process $N^A_{R_i}(t; \boldsymbol{\beta}_D) = Q_i(X_{R_i}) I(X_{R_i} \leq te^{\boldsymbol{\beta}_D^T \boldsymbol{Z}_i}, X_{D_i} = X_{R_i})\Delta_i$ and $\phi^A_R(.)$ be a nonnegative weight function. A class of estimating functions for $\boldsymbol{\beta}_R$ with given $\boldsymbol{\beta}_D$,

$$\boldsymbol{U}^A_R(\boldsymbol{\beta}_R, \boldsymbol{\beta}_D) = n^{-1} \sum_{i=1}^{n} \int_0^\tau \phi^A_R(s; \boldsymbol{\beta}_R, \boldsymbol{\beta}_D) \left\{ \boldsymbol{Z}_i - \frac{\sum_{j=1}^n Y_{R_i}(s; \boldsymbol{\beta}_R, \boldsymbol{\beta}_D) \boldsymbol{Z}_j e^{\boldsymbol{\beta}_R^T \boldsymbol{Z}_i}}{\sum_{j=1}^n Y_{R_i}(s; \boldsymbol{\beta}_R, \boldsymbol{\beta}_D) e^{\boldsymbol{\beta}_R^T \boldsymbol{Z}_i}} \right\} dN^A_{D_i}(s; \boldsymbol{\beta}_D).$$

Under Model B, define counting process $N^B_{R_i}(t; \boldsymbol{\beta}_D) = Q_i(te^{\boldsymbol{\beta}_D^T \boldsymbol{Z}_i})$ and $\phi^B_R(.)$ be a nonnegative weight function. We have the following estimating function for $\boldsymbol{\beta}_R$ for given

$\boldsymbol{\beta}_D$,

$$\boldsymbol{U}_R^B(\boldsymbol{\beta}_R, \boldsymbol{\beta}_D) = n^{-1} \sum_{i=1}^{n} \int_0^{\tau} \phi_R^B(s; \boldsymbol{\beta}_R, \boldsymbol{\beta}_D) \left\{ \boldsymbol{Z}_i - \frac{\sum_{j=1}^n Y_{R_i}(s; \boldsymbol{\beta}_R, \boldsymbol{\beta}_D) \boldsymbol{Z}_j e^{\boldsymbol{\beta}_R^T \boldsymbol{Z}_i}}{\sum_{j=1}^n Y_{R_i}(s; \boldsymbol{\beta}_R, \boldsymbol{\beta}_D) e^{\boldsymbol{\beta}_R^T \boldsymbol{Z}_i}} \right\} dN_{D_i}^B(s; \boldsymbol{\beta}_D).$$

For both models, it can be shown that the estimator of $\boldsymbol{\beta} = (\boldsymbol{\beta}_D^T, \boldsymbol{\beta}_R^T)^T$, $\hat{\boldsymbol{\beta}}$, is strongly consistent, and $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is asymptotically normal with mean $\mathbf{0}$ and vairance-covariance matrix that can be consistently estimated by either a modified sandwich variance estimation(Huang, 2002) or bootstrap(Efron and Tibshirani, 1993).

2.5 Incomplete and Auxiliary Covariate Data in Failure Time Regression Analysis

In many applications, missing data are a frequent problem for statistical analysis of failure time data. For example, in a study concerning 97 patients given bone marrow transplants from female sibling donors, time to occurrence of acute graft versus host disease is outcome of interest. Covariates thought to be predictive of the risk of acute graft versus host disease are patient age, previous donor pregnancies, isolation in a laminar airflow room during the immediate posttransplant period, and the acute graft versus host disease prophylactic treatment received. Among these 97 patients, donor pregnancy status was missing for 31 patients. The primary focus in this study was to evaluate the effect of prophylactic regimen. An example of auxiliary covariate data arises from the Studies of Left Ventricular Dysfunction(SOLVD, 1991), where only 162 out of 1111 patients had left ventricular ejection fraction(EF) accurately measured with radionucleotide technique. However, a related nonstandardized measure of EF was ascertained for all 1111 patients. This nonstandardized measure of EF can be treated as an auxiliary covariate to assess if the true EF accurately measured is predictive of risk of heart failure.

To fix notation, suppose that n subjects have observation times X_i, where $X_i = \min(T_i, C_i)$, T_i be the failure times and C_i be the censoring times, $i = 1, \cdots, n$. Define $\delta_i = I(T_i \leq C_i)$ as the censoring indicator and $Y_i = I(T_i \geq t)$ as the at risk indicator, $i = 1, \ldots, n$. For each subject, let $\boldsymbol{Z}_i$ denote the true covariate with parameter $\boldsymbol{\beta}$ and $\boldsymbol{W}_i$ be auxiliary variable that are observed. Observed data can be divided into two parts, a simple random subsample called the validation sample $\boldsymbol{V}$ in which contains both the true covariate $\boldsymbol{Z}$ and auxiliary covariate $\boldsymbol{W}$ and a nonvalidation set $\bar{\boldsymbol{V}}$ in which only the auxiliary covariate $\boldsymbol{W}$ is observed. Therefore we have $(X_i, \delta_i, \boldsymbol{Z}_i, \boldsymbol{W}_i), i \in \boldsymbol{V}$ and $(X_i, \delta_i, \boldsymbol{W}_i), i \in \bar{\boldsymbol{V}}$.

To obtain asymptotically unbiased inference about the true parameter $\boldsymbol{\beta}_0$, a naive approach is to simply discard information in the nonvalidation sample and use a partial likelihood method based on the validation sample only. This approach can lead to substantial reduction in efficiency. Thus interests are focused on estimating $\boldsymbol{\beta}$ by using the information from subjects in $\bar{\boldsymbol{V}}$. A fully parametric maximum likelihood analysis can be used, however, it involves specifying both the baseline hazard λ_0 and the conditional distribution of $P_\theta(\boldsymbol{Z}|\boldsymbol{W})$. Although this approach is efficient under correctly specified models, it is not robust to model misspecification and hard to implement.

Prentice(1982) proposed a partial likelihood estimator based on the induced relative risk function after correcting for covariate measurement errors under a rare disease assumption. Suppose that an individual with covariate $\boldsymbol{Z}_i(.) = \{Z_{i1}(.), \cdots, Z_{ip}(.)\}$ has failure time intensity:

$$\lambda_i(t) = Y_i(t)\lambda_0(t)r\{\boldsymbol{\beta}, \boldsymbol{Z}_i(t)\} = Y_i(t)\lambda_0(t)r_i(t),$$

where $\boldsymbol{\beta}^T = (\beta_1, \cdots, \beta_p)$ is a $p \times 1$ relative risk parameter vector to be estimated and $\lambda_0(.) \geq 0$ is an unspecified baseline hazard function. Given only the auxiliary covariate $\boldsymbol{W}_i(.)$, the intensity conditional on $\boldsymbol{W}_i(t)$ rather than on $\boldsymbol{Z}_i(t)$, called the induced intensity process, is

$$\bar{\lambda}_i(t) = Y_i(t)\lambda_0(t)E\{r_i(t)|Y_i(t) = 1, \boldsymbol{W}_i(t)\} = Y_i(t)\lambda_0(t)e_i(t).$$

Therefore, for each individual i, the relative risk function given all observed covariate data is

$$R_i(t) = r_i(t)I(i \in \boldsymbol{V}) + e_i(t)I(i \in \bar{\boldsymbol{V}}).$$

An induced partial likelihood based on the whole sample was proposed

$$IPL(\boldsymbol{\beta}) = \prod_{i=1}^{n} \left\{ \frac{R_i(T_i)}{\sum_{j=1}^{n} Y_j(T_i)R_j(T_i)} \right\}^{\delta_i}, i = 1, \cdots, n.$$

Calculation of $e_i(t) = E\{r_i(t)|Y_i(t) = 1, \boldsymbol{W}_i(t)\}$ requires specification of the baseline hazard λ_0 and the conditional distribution of $\boldsymbol{Z}$ given $\boldsymbol{U}$. Under certain assumptions such as if the disease is rare, the dependence on λ_0 can be ignored and $e_i(t)$ can be replaced with $E\{r_i(t)|\boldsymbol{W}_i(t)\}$. However, violation of these assumptions can lead to seriously biased estimates(Hughes, 1993). Besides, one still need a complete specification of $f_\theta(\boldsymbol{Z}|\boldsymbol{W})$, which is rarely known.

Prentice(1982) approach was further developed by Pepe *et al.*(1989) using parametric modelling for the conditional distribution $f_\theta(\boldsymbol{Z}|\boldsymbol{W})$. Suppose the induced relative risk is known up to the relative risk parameter $\boldsymbol{\beta}$ and certain measurement error parameters $\boldsymbol{\theta}$. In order to get consistent estimation of $\boldsymbol{\beta}$ from the partial likelihood function, one needs to know $\boldsymbol{\theta}$. Hence a validation sample which includes information of both $\boldsymbol{Z}$ and $\boldsymbol{W}$ will be helpful. The proposed partial likelihood is

$$PL(\boldsymbol{\beta}, \boldsymbol{\theta}) = \prod_{i=1}^{n} \left\{ \frac{R_i(\boldsymbol{\beta}, \boldsymbol{\theta}, T_i)}{\sum_{j=1}^{n} Y_j(T_i)R_j(\boldsymbol{\beta}, \boldsymbol{\theta}, T_i)} \right\}^{\delta_i} \prod_{k \in \boldsymbol{V}} f_\theta(\boldsymbol{Z}_k|\boldsymbol{W}_k).$$

Following the same arguments as in Andersen and Gill(1982) and assuming some mild regularity conditions, it can be shown that the MLE $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})$ are consistent and $\{n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}), n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})\}$ is asymptotically normal with mean 0 and the covariance matrix that is the analogue of the inverse of the Fisher information matrix in the case of time-independent covariates. However, realistic models for $f_\theta(\boldsymbol{Z}|\boldsymbol{W})$ may be difficult to construct and inference can be rather sensitive to the choice of mismeasurement model(Carroll, Spiegelman, Lan, Gailey and Abbott 1984, Pepe *et al.* 1989). Computationally, implementation of these maximum likelihood methods involves numerical integration.

To circumvent the aforementioned issues, Pepe and Fleming(1991) suggested an easily implemented method that is nonparametic with respect to the mismeasurement process. Denote $\boldsymbol{O}$ as an outcome random variable. Assuming i.i.d.observations, the likelihood for $\boldsymbol{\beta}$ is

$$L(\boldsymbol{\beta}) = \prod_{i \in \boldsymbol{V}} P_{\boldsymbol{\beta}}(\boldsymbol{O}_i|\boldsymbol{Z}_i, \boldsymbol{W}_i) \prod_{j \in \bar{\boldsymbol{V}}} P_{\boldsymbol{\beta}}(\boldsymbol{O}_j|\boldsymbol{W}_j),$$

where

$$P_{\boldsymbol{\beta}}(\boldsymbol{O}|\boldsymbol{W}) = \int P_{\boldsymbol{\beta}}(\boldsymbol{O}|\boldsymbol{z}, \boldsymbol{W}) dP(\boldsymbol{z}|\boldsymbol{W}).$$

The proposed method is to estimate $P(\boldsymbol{Z}|\boldsymbol{W})$ empirically from the validation sample and, in turn, to estimate the likelihood component $P_{\boldsymbol{\beta}}(\boldsymbol{O}_j|\boldsymbol{W}_j)$ for the non-validation sample subjects. In the case when $\boldsymbol{W}_i$ is categorical, the empirical estimates of $P(\boldsymbol{Z}|\boldsymbol{W})$ are

$$\hat{P}(\boldsymbol{Z}_i|\boldsymbol{W}_i) = \frac{\sum_{j \in \boldsymbol{V}} I\{\boldsymbol{Z}_j \leq \boldsymbol{Z}_i, \boldsymbol{W}_j = \boldsymbol{W}_i\}}{\sum_{j \in \boldsymbol{V}} I\{\boldsymbol{W}_j = \boldsymbol{W}_i\}},$$

which leads to an unbiased estimate of $P_{\boldsymbol{\beta}}(\boldsymbol{O}_j|\boldsymbol{W}_j)$ for a non-validation sample member i,

$$\hat{P}_{\boldsymbol{\beta}}(\boldsymbol{O}_i|\boldsymbol{W}_i) = \int P_{\boldsymbol{\beta}}(\boldsymbol{O}_i|\boldsymbol{z}, \boldsymbol{W}_i) d\hat{P}(\boldsymbol{z}|\boldsymbol{W}_i) = \frac{\sum_{j \in \boldsymbol{V}} P_{\boldsymbol{\beta}}(\boldsymbol{O}_i|\boldsymbol{Z}_j, \boldsymbol{W}_i) I\{\boldsymbol{W}_j = \boldsymbol{W}_i\}}{\sum_{j \in \boldsymbol{V}} I\{\boldsymbol{W}_j = \boldsymbol{W}_i\}}.$$

Inference for $\boldsymbol{\beta}$ can be based on the estimated likelihood

$$\hat{L}(\boldsymbol{\beta}) = \prod_{i \in \boldsymbol{V}} P_{\boldsymbol{\beta}}(\boldsymbol{O}_i | \boldsymbol{Z}_i, \boldsymbol{W}_i) \prod_{j \in \bar{\boldsymbol{V}}} \hat{P}_{\boldsymbol{\beta}}(\boldsymbol{O}_j | \boldsymbol{W}_j).$$

It is shown that under regular conditions and in a neighborhood of the true $\boldsymbol{\beta}_0$, the solution $\hat{\boldsymbol{\beta}}$ to the estimated score equation $\frac{d}{d\boldsymbol{\beta}} \log \hat{L}(\boldsymbol{\beta}) = \mathbf{0}$ is consistent and $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ is asymptotically normal with mean $\mathbf{0}$ and variance-covariance matrix that can be consistently estimated.

Based on Prentice(1982) method, Zhou and Pepe(1995) proposed an estimated partial likelihood(EPL) method. The unknown induced relative risk $R_i(t) = r_i(t)I(i \in \boldsymbol{V}) + e_i(t)I(i \in \bar{\boldsymbol{V}})$ can be estimated using the information provided by the validation sample. In particular, if the auxiliary covariate $\boldsymbol{W}_i(t)$ is categorical, an unbiased estimate of $e_i(t)$ can be constructed as the average relative risk among subjects in the validation sample with covariate values equal to $W_i(t)$. That is,

$$\hat{e}_i(t) = \frac{\sum_{j \in \boldsymbol{V}} Y_j(t) I\{\boldsymbol{W}_j(t) = \boldsymbol{W}_i(t)\} r_j(t)}{\sum_{j \in \boldsymbol{V}} Y_j(t) I\{\boldsymbol{W}_j(t) = \boldsymbol{W}_i(t)\}}$$

If $\boldsymbol{W}_i(t)$ is continuous, a kernel type estimator of $e_i(t)$ can be calculated(Zhou and Wang, 2000) based on Nadaraya(1964) and Watson(1964),

$$\hat{e}_i^*(t) = \frac{\sum_{j=1}^{n_V} I\{X_j \geq t\} K_h\{\boldsymbol{W}_i(t) - \boldsymbol{W}_j(t)\} r_j(t)/n_V}{\sum_{j=1}^{n_V} I\{X_j \geq t\} K_h\{\boldsymbol{W}_i(t) - \boldsymbol{W}_j(t)\}/n_V},$$

where n_V is the sample size in the validation data. $K_h(.) = K(./h)$ and $h > 0$ is the bandwidth, $K(.)$ is a kernel function which is piecewise smooth and satisfies $\int K(u)du = 1$.

Writing $\hat{R}_i(t) = r_i(t)I(i \in \boldsymbol{V}) + \hat{e}_i(t)I(i \in \bar{\boldsymbol{V}})$ and $\hat{R}_i^*(t) = r_i(t)I(i \in \boldsymbol{V}) + \hat{e}_i^*(t)I(i \in$

$\bar{\boldsymbol{V}}$) an estimate of the induced partial likelihood is

$$EPL(\boldsymbol{\beta}) = \prod_{i=1}^{n} \left\{ \frac{\hat{R}_i(T_i)}{\sum_{i=1}^{n} Y_j(T_i)\hat{R}_j(T_i)} \right\}^{\delta_i}$$

for categorical $W_i(t)$. Denoting the kth in $\boldsymbol{\beta}$ derivatives by a superscript (k), the score equation is

$$\boldsymbol{U}_{EPL}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \delta_i \left\{ \frac{\hat{R}_i^{(1)}(T_i)}{\hat{R}_i(T_i)} - \frac{\sum_{j=1}^{n} Y_j(T_i)\hat{R}_j^{(1)}(T_i)}{\sum_{j=1}^{n} Y_j(T_i)\hat{R}_j(T_i)} \right\}.$$

The maximum estimated partial likelihood estimator, $\hat{\boldsymbol{\beta}}_{EPL}$, can be found by solving $\boldsymbol{U}_{EPL}(\boldsymbol{\beta}) = \mathbf{0}$ using Newton-Raphson iteration. For continuous auxiliary covariate $W_i(t)$, replace $\hat{R}_i(t)$ with $\hat{R}_i^*(t)$ in the above equations. It's shown that $n^{1/2}(\hat{\boldsymbol{\beta}}_{EPL} - \boldsymbol{\beta}_0)$ is asymptotically normal with variance-covariance $\Sigma_{EPL}(\boldsymbol{\beta})$ that can be consistently estimated by the sample quantities. These approaches(Zhou and Pepe, 1995; Zhou and Wang, 2000) do not require strong assumptions such as rare disease assumption and any parametric assumption regarding $f_\theta(\boldsymbol{Z}|\boldsymbol{W})$. They are more robust and leave $\lambda_0(t)$ and $f_\theta(\boldsymbol{Z}|\boldsymbol{W})$ completely arbitrary. They are applicable to both auxiliary data and missing data framework. The validation set can be chosen at the baseline under the ignorable missing value mechanism condition(Rubin, 1976). Information about $\boldsymbol{\beta}$ contained in the non-validation set is fully utilized.

Lin and Ying(1993) introduced an approach in the pure missing data context. Under the Cox proportional hazard model, the partial likelihood score function for $\boldsymbol{\beta}$ is

$$\boldsymbol{U}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \delta_i \left\{ \boldsymbol{Z}_i(X_i) - \bar{\boldsymbol{Z}}(\boldsymbol{\beta}, X_i) \right\},$$

where

$$\bar{\boldsymbol{Z}}(\boldsymbol{\beta}, t) = \frac{\sum_{l=1}^{n} Y_l(t) \exp\{\boldsymbol{\beta}^T \boldsymbol{Z}_l(t)\} \boldsymbol{Z}_l(t)}{\sum_{l=1}^{n} Y_l(t) \exp\{\boldsymbol{\beta}^T \boldsymbol{Z}_l(t)\}}$$

Note that $\bar{\boldsymbol{Z}}(\boldsymbol{\beta}, t)$ is the conditional expectation of $\boldsymbol{Z}_l(t)$ on $\{j : X_l \geq t\}$ with respect to a probability distribution proportional to $\exp\{\boldsymbol{\beta}^T \boldsymbol{Z}_l(t)\}$. The authors estimated $\bar{\boldsymbol{Z}}(\boldsymbol{\beta}, t)$ from the subjects who have complete measurements on all covariate components at time t.

Suppose that the data consist of i.i.d. random components $\{X_i, \delta_i, \boldsymbol{Z}_i(.), H_{0i}(.), \boldsymbol{H}_i(.)\}$, $i = 1, \cdots, n$. $\boldsymbol{Z}_i(.)$ may not be completely observed, $\boldsymbol{H}_i(.)$ is a $p \times p$ diagonal matrix with indicator functions $\{H_{1i}(.), \cdots, H_{pi}(.)\}$ as the diagonal elements. $H_{ji}(t) = I\{Z_{ji}(t) \text{ is available}\}$ and $H_{0i}(t) = I\{H_{ji}(t) = 1, j = 1, \cdots, p\}$. If the ith subject belongs to the validation sample $\boldsymbol{V}$, then $\boldsymbol{H}_i(.) = \boldsymbol{I}_p$ and $H_{0i}(.) = 1$.

Assuming that, conditional on $\{X_i \geq t\}$, the missing indicators $\{H_{ji}(t), j = 1, \cdots, p\}$ are independent of all other random variables, which corresponds to the missing completely at random(MCAR) assumption of Rubin(1976). The proposed partial likelihood score function is

$$\tilde{\boldsymbol{U}}_{APLE}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \delta_i \boldsymbol{H}_i(X_i) \left\{ \boldsymbol{Z}_i(X_i) - E(\boldsymbol{\beta}, X_i) \right\},$$

where $E(\boldsymbol{\beta}, t) = \boldsymbol{S}^{(1)}(\boldsymbol{\beta}, t)/S^{(0)}(\boldsymbol{\beta}, t)$,

and $\boldsymbol{S}^{(r)}(\boldsymbol{\beta}, t) = n^{-1} \sum_{i=1}^{n} H_{0i}(t) Y_i(t) \exp\{\boldsymbol{\beta}^T \boldsymbol{Z}_i(t)\} \boldsymbol{Z}_i(t)^{\otimes r}$. The approximate partial likelihood estimator(APLE) $\tilde{\boldsymbol{\beta}}$ is the root to the estimating equation $\{\tilde{\boldsymbol{U}}_{APLE}(\boldsymbol{\beta}) = \boldsymbol{0}\}$, which can be solved by the Newton-Raphson algorithm. The authors showed that $\tilde{\boldsymbol{\beta}}$ is consistent and asymptotically normal with a covariance matrix for which a simple and consistent estimator is provided. The proposed approach is more efficient than the analysis based on subjects with completely observed covariates.

A natural estimator of the cumulative baseline hazard function $\Lambda_0(t) = \int_0^t \lambda_0(u) du$ is

$$\tilde{\Lambda}(\tilde{\boldsymbol{\beta}}, t) = \sum_{i=1}^{n} \frac{I(X_i \leq t) \delta_i H_{0i}(X_i)}{n \boldsymbol{S}^{(0)}(\tilde{\boldsymbol{\beta}}, X_i)}.$$

If the number of subjects in the validation sample varies over time due to augmentations, then the following estimator is recommended:

$$\tilde{\Lambda}^*(\tilde{\boldsymbol{\beta}}, t) = \sum_{i=1}^{n} \frac{I(X_i \leq t)\delta_i \sum_{l=1}^{n} H_{0l}(X_i)Y_l(X_i)}{n\boldsymbol{S}^{(0)}(\tilde{\boldsymbol{\beta}}, X_i) \sum_{l=1}^{n} Y_l(X_i)}.$$

The asymptotic properties of these estimators are studied. It is shown that the process $n^{1/2}\{\tilde{\Lambda}(\tilde{\boldsymbol{\beta}}, .) - \Lambda_0(.)\}$ and the process $n^{1/2}\{\tilde{\Lambda}^*(\tilde{\boldsymbol{\beta}}, .) - \Lambda_0(.)\}$ converge weakly to zero-mean Gaussian processes with limiting covariances that can be consistently estimated.

Paik and Tsai (1997) proposed an imputation method that can be used when data are MAR. Chen and Little (1999) proposed a nonparametric maximum likelihood procedure when data are MAR. But their method works only when the missing covariates are all discrete or all normally distributed. Herring and Ibrahim (2001) developed a likelihood-based methodology for MAR covariates based on partial likelihood using an EM-type algorithm.

Chapter 3

Accelerated Failure Time Marginal Rate Model for Recurrent and Terminal Events Data

3.1 Introduction

Recurrent events are common in many clinical or observational studies. Examples include bladder tumor recurrences, repeated hospitalizations due to cardiovascular disease and AIDS associated opportunistic diseases in HIV-infected patients. In these studies it is often of interest to evaluate the effects of risk factors on the frequencies of recurrent events. The recurrence of serious events are usually subject to censoring due to the death of a subject which is likely to be informative. However, most of the existing literatures on recurrent event analysis(e.g., Anderson and Gill(1982); Prentice et al(1981); Wei, Lin and Weissfeld(1989); Pepe and Cai(1993), Lawless and Nadeau(1995); Lin, Wei and Ying(1998); Lin, Wei, Yang and Ying(2000)) take on the assumption that the terminal event is non-informative.

Methodologies to address analysis of data involving recurrent events with a terminal event have been studied in recent years. There are complete intensity approaches which involve jointly modelling both recurrent events and terminal event. Wang, Qing and Chiang(2001), Huang and Wang(2004), Liu, Wolfe and Huang(2004) and Ye, Kalbfleisch and Schaubel(2007) studied shared frailty models that assumed the proportional intensity model and proportional hazards model for the recurrent events and the terminal event, respectively. In their approaches, a common latent variable was used to model the association between the intensity of the recurrent event process and the hazard of the failure time. In additional to the intensity models, rate models have been studied by various authors. Cook and Lawless(1997) studied mean and rate models for recurrent process conditional on survival at specific time points. Ghosh and Lin(2000) proposed a nonparametric estimator for the rate function of the recurrent events. Ghosh and Lin(2002) further considered a proportional rate model and obtained consistent estimators for the regression coefficients from an inverse probability weighted estimating equation. Miloslavsky et al(2004) also studied the proportional rate model and adopted an modified version of the estimating equation of Ghosh and Lin(2002). In these models, the dependence between the recurrent event process and the terminal event is unspecified.

A useful and important alternative to the proportional rate model is the accelerated rate model. Chang(2000) proposed the accelerated failure time models on the gap time between the recurrent and terminal events. Ghosh and Lin (2003) studied the accelerated failure time models on the total time of the recurrent events and the terminal event. They developed inference procedures by censoring some of the originally uncensored data.

In this article, we focus on the accelerated failure time marginal rate models for the cumulative number of the recurrent events over time, while taking the terminal events into account. The marginal approach does not require us to specify the dependence structure between the recurrent events and the terminal events, and the mean function incorporates the facts that subjects who die cannot experience any further recurrent events. We develop an estimating procedure for both the regression parameters and the mean function by applying the inverse probability of censoring weighting technique(Robins and Rotnitzky(1992)). The remainder of this article is organized as follows. In Section 2, we introduce notations and presents the proposed methods. In Section 3, we describe the asymptotic properties of the resulting estimators and discuss the inference procedures on the mean function of recurrent events. In Section 4, we conduct simulation studies to evaluate the proposed methods in finite samples and report the results. In Section 5, we illustrate the method by applying it to the SOLVD(Studies of Left Ventricular Dysfunction, SOLVD Investigators, 1991) Treatment Trial data. We conclude with some discussion in Section 6. Proofs of the theorems can be found in Section 7.

3.2 Model and Estimation

Let $N^*(t)$ be the cumulative number of recurrent events that occur over the time interval $[0, t]$ and D be the survival time. Intuitively, $N^*(.)$ does not jump beyond D as subjects who die can not experience further recurrent events. Let τ be the study duration, C be the censoring time and $\boldsymbol{Z}$ be a $p \times 1$-vector of covariates. Throughout the paper, we assume that C may depend on $\boldsymbol{Z}$ but is independent of $\{N^*(.)\}$ and D given $\boldsymbol{Z}$. On the other hand, D_i is allowed to depend on $\{N^*(.)\}$, even conditionally on $\boldsymbol{Z}$. In general only the minimum of C and D is known and $\{N^*(.)\}$ can only be observed up to $\min(C, D)$. Further, we define $X = D \wedge C = \min(D, C)$, $N(t) = N^*(t \wedge C) = I(C \geq t)\, N^*(t)$, $Y(t) = I(C \geq t)$ and $\delta = I(D \leq C)$. For a random sample of n subjects, the observed data consist of $\{N_i(.), X_i, \delta_i, \boldsymbol{Z}_i\}$, $i = 1, \cdots, n$. We also define $E[dN^*(t)|\boldsymbol{Z}]$ as the marginal rate of recurrent events up to t associated with $\boldsymbol{Z}$. According to our setup, $E[dN^*(t)|\boldsymbol{Z}] = 0$ if $t > D$. We wish to formulate the effect of $\boldsymbol{Z}$ on $E[dN^*(t)|\boldsymbol{Z}]$ through a semiparametric model without specifying the dependence structure between recurrent events and terminal events, or among recurrent events.

We consider the following accelerated rate model,

$$E[dN^*(t)|\boldsymbol{Z}] = d\mu_0(te^{\boldsymbol{\beta}_0^T\boldsymbol{Z}}), \tag{3.1}$$

where $\boldsymbol{Z}$ is a $p \times 1$ vector of covariates, $\boldsymbol{\beta}_0$ is a $p \times 1$ vector of unknown regression parameters, and $d\mu_0(.)$ is an unspecified continuous function. Under this model, the effect of covariates is to accelerate or decelerate the rate function. In the absence of death, this model has been studied by Ghosh(2004). For convenience, we work with data on the transformed scale. Let $Y(t;\boldsymbol{\beta}) = Y(te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}) = I(C \geq te^{-\boldsymbol{\beta}^T\boldsymbol{Z}})$, $N^*(t;\boldsymbol{\beta}) = N^*(te^{-\boldsymbol{\beta}^T\boldsymbol{Z}})$ and $dN(t;\boldsymbol{\beta}) = Y(t;\boldsymbol{\beta})dN^*(t;\boldsymbol{\beta})$. Notice that

$$E[dN^*(t;\boldsymbol{\beta})|\boldsymbol{Z}] = E[dN^*(te^{-\boldsymbol{\beta}^T\boldsymbol{Z}})|\boldsymbol{Z}] = d\mu_0(te^{(\boldsymbol{\beta}_0-\boldsymbol{\beta})^T\boldsymbol{Z}}).$$

When $\boldsymbol{\beta} = \boldsymbol{\beta}_0$, model (3.1) is equivalent to

$$E[dN^*(t;\boldsymbol{\beta}_0)|\boldsymbol{Z}] = E[dN^*(te^{-\boldsymbol{\beta}_0^T\boldsymbol{Z}})|\boldsymbol{Z}] = d\mu_0(t). \tag{3.2}$$

When all the censoring times C are known, for fixed $\boldsymbol{\beta}$, a Brewlow-type estimator for $\mu_0(t)$ is given by

$$\widehat{\mu}_0^*(t;\boldsymbol{\beta}) = \sum_{i=1}^n \int_0^t \frac{dN_i(u;\boldsymbol{\beta})}{\sum_{i=1}^n Y_i(u;\boldsymbol{\beta})},\ t \in [0,\tau].$$

To estimate $\boldsymbol{\beta}$, we consider the following estimating function

$$\sum_{i=1}^n \int_0^\tau Y_i(t;\boldsymbol{\beta})\boldsymbol{Z}\left\{dN_i(u;\boldsymbol{\beta}) - d\widehat{\mu}_0^*(t;\boldsymbol{\beta})\right\}.$$

After substituting $\widehat{\mu}_0^*(t;\boldsymbol{\beta})$ into the above function and some algebraic manipulations , we obtain an unbiased estimating function for $\boldsymbol{\beta}_0$,

$$\boldsymbol{U}_n(\boldsymbol{\beta}) = \sum_{i=1}^{n} \int_0^{\tau} Y_i(t;\boldsymbol{\beta}) \left\{ \boldsymbol{Z}_i - \frac{\sum_{i=1}^{n} Y_i(t;\boldsymbol{\beta})\boldsymbol{Z}_i}{\sum_{i=1}^{n} Y_i(t;\boldsymbol{\beta})} \right\} dN_i(t;\boldsymbol{\beta}). \tag{3.3}$$

However, when a subject experiences a terminal event, the subject's censoring time is unobserved and unknown. Thus (3.3) cannot be evaluated. We want to replace $Y(t;\boldsymbol{\beta})$ in the above estimating function with a quantity of equal expectation. One approach is to use the inverse probability censoring weighting technique(IPCW) (Robins and Rotnitzky, 1992). Similar techniques can be seen in existing literatures(e.g., Lin and Ying(1993), Cheng, Wei and Ying(1995), Fine and Gray(1999), Ghosh and Lin(2002)) in different contexts. The IPCW function is defined as

$$W(t;\boldsymbol{\beta}) = \frac{I\left(C \geq D \wedge te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}\right) S(te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}|\boldsymbol{Z})}{S\left(X \wedge te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}|\boldsymbol{Z}\right)},$$

where $S(t|\boldsymbol{Z}) = P\left(C > t|\boldsymbol{Z}\right)$.

Since we allow C to depend on $\boldsymbol{Z}$, it is convenient to formulate $S(t|\boldsymbol{Z})$ through the Cox proportional hazards model,

$$\lambda(t|\boldsymbol{Z}) = \lambda_0(t)e^{\boldsymbol{\gamma}^T\boldsymbol{Z}}, \tag{3.4}$$

where $\lambda_0(.)$ is an unspecified baseline hazard function, $\boldsymbol{\gamma}$ is a $p \times 1$ vector of unknown regression parameter, and $\lambda(t|\boldsymbol{Z})$ is the hazard function corresponding to $S(t|\boldsymbol{Z})$.

We have the following estimator of $S(t|\boldsymbol{Z})$,

$$\widehat{S}(t|\boldsymbol{Z}) = \exp\{-\int_0^t e^{\widehat{\boldsymbol{\gamma}}^T\boldsymbol{Z}} d\widehat{\Lambda}_0(u)\},$$

where $\widehat{\boldsymbol{\gamma}}$ is the maximum partial likelihood estimator of $\boldsymbol{\gamma}$(Cox,1975) and $\widehat{\Lambda}_0(.)$ is the

Breslow estimator of $\Lambda_0(t) = \int_0^t \lambda_0(u)du$. Thus $W_i(t;\boldsymbol{\beta})$ can be estimated by

$$\widehat{W}(t;\boldsymbol{\beta}) = \frac{I\left(C \geq D \wedge te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}\right)\widehat{S}(te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}|\boldsymbol{Z})}{\widehat{S}\left(X \wedge te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}|\boldsymbol{Z}\right)}.$$

By the law of conditional expectation, it can be shown that (Appendix proof of Theorem 1),

$$E\left\{W(t;\boldsymbol{\beta})\right\} = E\left\{\widehat{W}(t;\boldsymbol{\beta})\right\} = E\left\{Y(t;\boldsymbol{\beta})\right\}.$$

By replacing $Y(t;\boldsymbol{\beta})$ in (3.3) with $\widehat{W}(t;\boldsymbol{\beta})$, we obtain a modified estimating function for $\boldsymbol{\beta}_0$:

$$\boldsymbol{U}_n(\boldsymbol{\beta}) = \sum_{i=1}^n \int_0^\tau \widehat{W}_i(t;\boldsymbol{\beta})\left\{\boldsymbol{Z}_i - \frac{\sum_{i=1}^n \widehat{W}_i(t;\boldsymbol{\beta})\boldsymbol{Z}_i}{\sum_{i=1}^n \widehat{W}_i(t;\boldsymbol{\beta})}\right\} dN_i(t;\boldsymbol{\beta}). \tag{3.5}$$

The estimating function is not continuous in $\boldsymbol{\beta}$. Therefore we estimate $\boldsymbol{\beta}_0$ by minimizing the norm $\|\boldsymbol{U}_n(\boldsymbol{\beta})\|$, where $\|\mathbf{a}\| = (\mathbf{a}^T\mathbf{a})^{1/2}$. Denote this estimator by $\widehat{\boldsymbol{\beta}}$. Since common derivative-based optimization methods will not work, we use the Nelder-Mead simplex method (1965) to find the minimum.

Given $\widehat{\boldsymbol{\beta}}$, we estimate $\mu_0(t)$ by the Nelson-Aalen-type estimator of $\widehat{\mu}_0(t)$, where

$$\widehat{\mu}_0(t) = \sum_{i=1}^n \int_0^t \frac{\widehat{W}_i(u;\widehat{\boldsymbol{\beta}})dN_i(u;\widehat{\boldsymbol{\beta}})}{\sum_{i=1}^n \widehat{W}_i(u;\widehat{\boldsymbol{\beta}})},\ t \in [0,\tau]. \tag{3.6}$$

3.3 Asymptotic properties

We consider the following assumptions,

(C1) $\{N_i, X_i, \boldsymbol{Z}_i\}$ $(i = 1, \cdots, n)$ are independent and identically distributed.

(C2) $P(C_i > \tau|\boldsymbol{Z}_i) > 0, i = 1, \cdots, n$, where τ is the study duration.

(C3) $E[N_i(\tau)] < \infty$, $i = 1, \cdots, n$,.

(C4) Covariates $\boldsymbol{Z}_i, i = 1, \cdots, n$, are bounded.

(C5) $\mu_0(t)$ is strictly increasing and has bounded second derivative in $[0, \tau]$.

(C6) $\lambda_0(t) > 0$, $t \in [0, \tau]$.

(C7) Matrix $\boldsymbol{A}$ is non-singular, where

$$\boldsymbol{A} \equiv \nabla_{\boldsymbol{\beta}} E\left\{\int_0^\tau W(t;\boldsymbol{\beta})\left[\boldsymbol{Z} - \frac{E\{W(t;\boldsymbol{\beta})\boldsymbol{Z}\}}{E\{W(t;\boldsymbol{\beta})\}}\right] dN(t;\boldsymbol{\beta})\right\}\Bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0},$$

and $\nabla_{\boldsymbol{\beta}}$ is the gradient of $\boldsymbol{\beta}$.

Condition (C2) implies the weight function $W(t; \boldsymbol{\beta})$ will be uniformly bounded away from zero. Conditions (C1), (C3)-(C7) are standard.

3.3.1 Asymptotic properties of $\widehat{\boldsymbol{\beta}}$

Theorem 3.3.1 *Under regularity conditions (C1)-(C7), the parameter estimate $\widehat{\boldsymbol{\beta}}$ is strongly consistent for $\boldsymbol{\beta}_0$, i.e. $\widehat{\boldsymbol{\beta}} \xrightarrow{a.s.} \boldsymbol{\beta}_0$. The random vector $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ converges in distribution to a zero-mean normal distribution with a covariance matrix $\boldsymbol{A}^{-1}\boldsymbol{\Sigma}(\boldsymbol{A}^{-1})^T$, where*

$$\begin{aligned}
\boldsymbol{A} &\equiv \nabla_{\boldsymbol{\beta}} E\left\{\int_0^\tau W(t;\boldsymbol{\beta})\left[\boldsymbol{Z} - \frac{E\{W(t;\boldsymbol{\beta})\boldsymbol{Z}\}}{E\{W(t;\boldsymbol{\beta})\}}\right] dN(t;\boldsymbol{\beta})\right\}\Bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}, \\
\boldsymbol{\Sigma} &= E[\boldsymbol{J}\boldsymbol{J}^T], \\
\boldsymbol{J} &= \int_0^\tau W(t;\boldsymbol{\beta}_0)\left[\boldsymbol{Z} - \frac{E\{W(t;\boldsymbol{\beta}_0)\boldsymbol{Z}\}}{E\{W(t;\boldsymbol{\beta}_0)\}}\right] dM(t;\boldsymbol{\beta}_0), \\
M(t;\boldsymbol{\beta}_0) &= N(t;\boldsymbol{\beta}_0) - \int_0^t d\mu_0(s). \qquad (3.7)
\end{aligned}$$

A consistent estimator for $\boldsymbol{A}^{-1}\boldsymbol{\Sigma}(\boldsymbol{A}^{-1})^T$ *is* $\widehat{\boldsymbol{A}}^{-1}\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{A}}^{-1})^T$*, where*

$$
\begin{aligned}
\widehat{\boldsymbol{\Sigma}} &= n^{-1}\sum_{i=1}^{n}\widehat{\boldsymbol{J}}_i\widehat{\boldsymbol{J}}_i^T, \\
\widehat{\boldsymbol{J}}_i &= \int_0^{\tau}\widehat{W}_i(t;\widehat{\boldsymbol{\beta}})\left[\boldsymbol{Z}_i - \frac{\boldsymbol{Q}^{(1)}(t;\widehat{\boldsymbol{\beta}})}{Q^{(0)}(t;\widehat{\boldsymbol{\beta}})}\right]d\widehat{M}_i(t;\widehat{\boldsymbol{\beta}}), \\
\boldsymbol{Q}^{(k)}(t;\boldsymbol{\beta}) &= n^{-1}\sum_{i=1}^{n}\widehat{W}_i(t;\boldsymbol{\beta})\boldsymbol{Z}_i^{\otimes k}, k=0,1, \\
\widehat{M}_i(t;\boldsymbol{\beta}) &= N_i(t;\boldsymbol{\beta}) - \int_0^{te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}} d\widehat{\mu}_0(s),
\end{aligned}
$$

$$
\widehat{\boldsymbol{A}} = \left(\frac{\boldsymbol{U}_n(\widehat{\boldsymbol{\beta}}+\boldsymbol{e}_1h_n)-\boldsymbol{U}_n(\widehat{\boldsymbol{\beta}})}{nh_n},\cdots,\frac{\boldsymbol{U}_n(\widehat{\boldsymbol{\beta}}+\boldsymbol{e}_ph_n)-\boldsymbol{U}_n(\widehat{\boldsymbol{\beta}})}{nh_n}\right). \tag{3.8}
$$

For a column vector $\boldsymbol{a}$*,* $\boldsymbol{a}^{\otimes 0}=1$*,* $\boldsymbol{a}^{\otimes 1}=\boldsymbol{a}$*. The canonical vector* $\mathbf{e}_i$ *is* $p\times 1$ *and takes value 1 at the* j*th element and 0 elsewhere,* $j=1,\cdots,p$*. The perturbation size is* $h_n = O(n^{-1/2})$

.

3.3.2 Inference on the mean function

The following theorem describes the asymptotic properties of $\widehat{\mu}_0(t)$.

Theorem 3.3.2 *Under the regularity conditions (C1)-(C7),* $\widehat{\mu}_0(t)$ *is strongly consistent for* $\mu_0(t)$ *uniformly in* t*, i.e.* $\widehat{\mu}_0(t)\xrightarrow{a.s.}\mu_0(t)$*,* $t\in[0,\tau]$*. The process* $n^{1/2}\left\{\widehat{\mu}_0(t)-\mu_0(t)\right\}$*,* $t\in[0,\tau]$*, converges weakly to a zero-mean Gaussian process with a covariance function* $E[\xi(s)\xi(t)]$*, where*

$$
\xi(t) = \int_0^t\frac{W(u;\boldsymbol{\beta}_0)dM(u;\boldsymbol{\beta}_0)}{E\{W(u;\boldsymbol{\beta}_0)\}} - \boldsymbol{b}(t)^T\boldsymbol{A}^{-1}\boldsymbol{J},\ t\in[0,\tau],
$$

$$\boldsymbol{b}(t) \;=\; -\int_0^t \frac{E\{W(u;\boldsymbol{\beta}_0)\boldsymbol{Z}\}}{E\{W(u;\boldsymbol{\beta}_0)\}} d\{\dot{\mu}_0(u)\,u\}\,, \tag{3.9}$$

where $\boldsymbol{A}$,$\boldsymbol{J}$ and $M(t;\boldsymbol{\beta}_0)$ are defined as in (3.7). A consistent estimator for the asymptotic covariance is

$$\widehat{\phi}(s,t) = n^{-1}\sum_{i=1}^{n}\widehat{\xi}_i(s)\widehat{\xi}_i(t),$$

where

$$\widehat{\xi}_i(t) \;=\; \int_0^t \frac{\widehat{W_i}(u;\widehat{\boldsymbol{\beta}})d\widehat{M_i}(u;\widehat{\boldsymbol{\beta}})}{Q^{(0)}(t;\widehat{\boldsymbol{\beta}})} - \widehat{\boldsymbol{b}}(t)^T\widehat{\boldsymbol{A}}^{-1}\widehat{\boldsymbol{J}}_i,$$

$$\widehat{\boldsymbol{b}}(t) \;=\; \left(\frac{\widehat{\mu}_0(t;\widehat{\boldsymbol{\beta}}+\boldsymbol{e}_1h_n)-\widehat{\mu}_0(t;\widehat{\boldsymbol{\beta}})}{h_n},\cdots,\frac{\widehat{\mu}_0(t;\widehat{\boldsymbol{\beta}}+\boldsymbol{e}_ph_n)-\widehat{\mu}_0(t;\widehat{\boldsymbol{\beta}})}{h_n}\right)^T, \tag{3.10}$$

where $\widehat{\boldsymbol{A}}$, $\widehat{\boldsymbol{J}}_i$, $\widehat{M_i}(t;\boldsymbol{\beta})$ and $\boldsymbol{Q}^{(0)}(t;\boldsymbol{\beta})$ are defined as in (3.8). The canonical vector $\boldsymbol{e}_i$ is $p\times 1$ and takes value 1 at the jth element and 0 elsewhere, $j=1,\cdots,p$. The perturbation size is $h_n = O(n^{-1/2})$.

Based on the asymptotic results in Theorem 3.3.2, we can construct confidence intervals for $\mu_0(t)$. We consider the transformed process $n^{1/2}\{\log\{\widehat{\mu}_0(t)\}-\log\{\mu_0(t)\}\}$ since $\mu_0(t)$ is non-negative. By the delta method, the process $n^{1/2}\{\log\{\widehat{\mu}_0(t)\}-\log\{\mu_0(t)\}\}$ converges weakly to a zero-mean Gaussian process with a covariance function $\phi(s,t)/\mu_0^2(t)$, $s,t\in[0,\tau]$. Therefore the 95% pointwise confidence interval for $\log\{\mu_0(t)\}$ is

$$\log\{\widehat{\mu}_0(t)\} \pm 1.96\,n^{-1/2}\,\frac{\widehat{\phi}(t,t)^{1/2}}{\widehat{\mu}_0(t)}.$$

Exponentiating the above interval, we obtain the 95% pointwise confidence interval for $\mu_0(t)$

$$\widehat{\mu}_0(t) \exp\left\{\pm 1.96\, n^{-1/2} \frac{\widehat{\phi}(t,t)^{1/2}}{\widehat{\mu}_0(t)}\right\}. \tag{3.11}$$

To construct simultaneous confidence bands for $\mu_0(t)$ over a time interval $[\tau_1, \tau_2]$, $0 < \tau_1 < \tau_2 \leq \tau$, we use a Monte-Carlo method. Specifically, let

$$V_G(t) = n^{-1/2} \sum_{i=1}^{n} \widehat{\xi}_i(t) G_i,$$

where $(G_1, \cdots, G_n)$ are independent standard normal variables. The following theorem states that $V(t) = n^{1/2}\{\widehat{\mu}_0(t) - \mu_0(t)\}$ and $V_G(t)$ have the same limiting distribution.

Theorem 3.3.3 *Under the regularity conditions (C1)-(C7), conditional on the observed data, the process $V_G(t)$, converges weakly to a zero-mean Gaussian process with a covariance function $\phi(s,t) = E[\xi(s)\xi(t)]$, where $\phi(s,t)$ is defined as in (3.9).*

Given a fixed set of observed data, $\widehat{\xi}_i(t), i = 1, \cdots, n$ are also fixed. We generate $(G_1, \cdots, G_n)$ independently from the standard normal distribution and calculate $V_G(t)$. Repeat this process B times, where B is sufficiently large, we obtain a collection of $V_G(t)$ which simulates the limiting distribution of $V(t)$. The 95% simultaneous confidence band for $\mu_0(t)$ over a time interval $[\tau_1, \tau_2]$, $0 < \tau_1 < \tau_2 \leq \tau$, is

$$\widehat{\mu}_0(t) \exp\left\{\pm c_\alpha\, n^{-1/2} \frac{\widehat{\phi}(t,t)^{1/2}}{\widehat{\mu}_0(t)}\right\}, \tag{3.12}$$

where c_α is the estimated 95th percentile of $\sup_{\tau_1 \leq t \leq \tau_2} \left| \frac{V_G(t)}{\widehat{\phi}(t,t)^{1/2}} \right|$.

3.4 Simulation Studies

A series of numerical simulation studies are conducted to evaluate the performance of the proposed estimator in the finite sample situation.

A frailty term ν is used to induce dependence among recurrent events and deaths. Let ν be a gamma variable with mean 1 and variance σ^2, we generate the failure times from the following model

$$\begin{aligned} \lambda^D(t|\nu, \boldsymbol{Z}) &= \nu\lambda_0^D e^{\gamma_D^T \boldsymbol{Z}} \qquad (3.13) \\ \Rightarrow P(D \geq t|\nu, \boldsymbol{Z}) &= \exp\{-\nu\lambda_0^D\, t\, e^{\gamma_D^T \boldsymbol{Z}}\}. \end{aligned}$$

We generate the recurrent events from the following model

$$E\{d\widetilde{N}(t)|\nu, \boldsymbol{Z}\} = c\, e^{\boldsymbol{\beta}^T \boldsymbol{Z}} \exp\{\nu\lambda_0^D\, t\, e^{\gamma_D^T \boldsymbol{Z}}\}, \qquad (3.14)$$

where $\widetilde{N}(.)$ is a pseudo recurrent process. Given ν, we assume the pseudo recurrent process is independent of death.

Here we show the recurrent events generated satisfy the proposed model. Since we assume $N^*(.)$ does not jump beyond D, we have $dN^*(t) = I(D \geq t)d\widetilde{N}(t)$. Therefore

$$\begin{aligned} E\{dN^*(t)|\boldsymbol{Z}\} &= E\{d\widetilde{N}(t)I(D \geq t)|\boldsymbol{Z}\} \\ &= E_\nu\left\{E\{d\widetilde{N}(t)I(D \geq t)|\nu, \boldsymbol{Z}\}\right\} \\ &= E_\nu\left\{E\{d\widetilde{N}(t)|\nu, \boldsymbol{Z}\}E\{I(D \geq t)|\nu, \boldsymbol{Z}\}\right\}, \ d\widetilde{N}(.)\perp D \text{ given } \nu \\ &= E_\nu\{E\{d\widetilde{N}(t)|\nu, \boldsymbol{Z}\}P(D \geq t|\nu, \boldsymbol{Z})\} \\ &= E_\nu\{c\, e^{\boldsymbol{\beta}^T \boldsymbol{Z}} \exp\{\nu\lambda_0^D\, t\, e^{\gamma_D^T \boldsymbol{Z}}\} \exp\{-\nu\lambda_0^D\, t\, e^{\gamma_D^T \boldsymbol{Z}}\}\} \\ &= E_\nu\{c\, e^{\boldsymbol{\beta}^T \boldsymbol{Z}}\} = c\, e^{\boldsymbol{\beta}^T \boldsymbol{Z}}, \end{aligned}$$

which is the proposed model (3.1) when $\mu_0(te^{\boldsymbol{\beta}_0^T \boldsymbol{Z}}) = c\, te^{\boldsymbol{\beta}_0^T \boldsymbol{Z}}$.

Independent censoring times are generated from the following model

$$\lambda^C(t|\boldsymbol{Z}) \;\; = \;\; \lambda_0^C e^{\gamma_C^T \boldsymbol{Z}} \tag{3.15}$$

We first consider Z as a single dichotomous covariate from the $Bernoulli(0.5)$ distribution. The following combinations of simulation parameters are chosen: $\beta_0 = 0, 0.2, 0.5$, $\sigma^2 = 0, 0.25, 0.5, 0.75, 1$, $\lambda_0^D = 0.25$, $\lambda_0^C = 1$, $\gamma_D = 0.3$, $\gamma_C = 0.2$, $c = 2$. The value of σ^2 controls the correlation among recurrent events and death. The correlation decreases as σ^2 becomes smaller, with $\sigma^2 = 0$ implying zero correlation. The average observed numbers of recurrent events range from about 1.8 to 2.4 events per subject. We consider sample sizes $n = 100, 200, 400$. For each setting, 500 simulation samples are generated. The simulations are programmed in MATLAB(version 7.7.0).

To obtain $\widehat{\boldsymbol{\beta}}$, we minimize the Euclidean norm $\|\boldsymbol{U}_n(\boldsymbol{\beta})\|$ using Nelder-Mead simplex method. In our simulation, the minimization procedure gives about $0.6\% - 2.8\%$ extreme values for $\boldsymbol{\beta}$ with $n = 100$ due to numerical instability. The fraction of extreme values decreases to about $0.2\% - 1.8\%$ when $n = 400$. Therefore we report the summary statistics after excluding outliers which are 1.5 inter-quartile range above the third quartile or below the first quartile. To estimate the asymptotic covariance of $\widehat{\boldsymbol{\beta}}$, we use the numerical derivative method as suggested in Theorem 1. In particular, we consider h_n be $n^{-1/2}$, $3n^{-1/2}$, and $5n^{-1/2}$. Our numerical experience shows that the estimates are robust to the choice of h_n as long as the sample size is over 100. Hence, we only report the results associated with $h_n = 3n^{-1/2}$. The results are presented in Table 3.1, Table 3.2 and Table 3.3.

Based on the simulation results, the coefficient estimator $\widehat{\boldsymbol{\beta}}_D$ appears to be approximately unbiased for all combinations of sample sizes, correlations and the true values of $\boldsymbol{\beta}_0$. The proposed standard error estimator provides a good estimate of the true variation of $\widehat{\boldsymbol{\beta}}$. As we increase the sample size, we see improvement in the coverage rate as they fall in the $0.94 - 0.96$ range. The accuracy of the asymptotic approximation appears to be unaffected by the amount of correlation between recurrent events and death.

We also examine the finite-sample properties of the proposed estimator for continuous covariates by considering $Z \sim Uniform(0, 2)$ and $Z \sim N(0, 1)$, under the same combination set of parameters, with sample n=100, 200. Table 3.4 and Table 3.5 summarize the results when $\beta_0 = 0.2$. The results are similar to those for the dichotomous covariate.

3.5 Application to the SOLVD data

We apply the proposed method to the SOLVD(Studies of Left Ventricular Dysfunction, SOLVD Investigators, 1991) Treatment Trial data. The SOLVD treatment trial was randomized, double-blind, and placebo-controlled. From June 1986 to March 1989, a total of 2569 patients were enrolled in the treatment trial. Eligible participants were of age between 21 and 80 years, inclusive, with overt symptoms of congestive heart failure, and left ventricular ejection fraction less than 35 percent. In addition to the conventional treatment, participants were randomly assigned to receive either placebo(n=1284) or enalapril(n=1285) at doses of 2.5 to 20 mg per day. During the 2-year follow-up period, detailed information for hospitalizations and mortality was recorded. It is of interest to know if enalapril reduced the repeated hospitalizations for patients while adjusting for the baseline ejection fraction.

In this data analysis, we only focus on the women subjects in the Treatment Trial. A total of 503 women subjects are used in the analysis, with 259 patients with a total of 755 recorded hospitalizations and 244 patients with a total of 564 recorded hospitalizations in the placebo and enalapril treatment, respectively. The Kaplan-Meier estimators of the survival functions are plotted in Figure 3.1. The value of log-rank test statistic is 2.253 with p-value=0.263, indicating there is no significant difference in survival between the two treatment groups. Table 3.6 summarizes the hospitalization and survival experiences for the women subjects in the two treatment groups. We are interested in assessing the average treatment effect (TRT) on repeated hospitalizations after adjusting for the baseline ejection fraction(EF).We center EF on the mean. TRT is one for enalapril and zero for placebo.

The results are presented in Table 3.7. Based on the proposed methods, we find the effect of enalapril treatment is significant. On average, the times to hospitalizations for women on Enalapril treatment are $e^{0.498} = 1.645$ times those on placebo. However, the baseline ejection fraction appears to be nonsignificant. In summary, enalapril effectively reduced the frequency of hospitalizations but not mortality among women.

Figure 3.2 displays the estimation of the mean frequency of hospitalizations from day 0 to day 1682 for women with 24.53% baseline ejection fraction who received enalapril treatment versus those who did not. The simultaneous confidence bands are calculated from (3.12) by 1000 realizations of $V_G^D(t)$. From the plots, we can see those treated with enalapril have fewer number of hospitalizations than those in the placebo group.

Since our proposed method involves modelling the censoring distribution, we use the Schoenfeld residuals(Schoenfeld, 1982) plots (Figure 3.3) to assess the validity of proportional hazard assumption. They suggest that we cannot reject the proportional hazard assumption. Thus it is reasonable to model the censoring distribution with a Cox proportional hazard model.

3.6 Concluding remarks

In this chapter, we consider a marginal mean accelerated failure time model for the recurrent event (Lin, Wei and Ying, 1998). In the presence of terminal event, we propose a modified estimating function by applying the inverse probability of censoring weighting technique. Our approach does not require to specify the dependence structure between the recurrent events and the terminal events. Simulation studies suggest the proposed methods work well on data with moderate size. The proposed methods provide parameter estimates with easy interpretations as well as the mean frequency estimate of recurrent event, which is a quantity of clinical interest.

3.7 Proofs of the theorems

Proof of Theorem 3.3.1:

In the empirical process notation, we can rewrite (3.5) as

$$\frac{1}{n}\boldsymbol{U}_n(\boldsymbol{\beta}) \quad = \quad \mathbb{P}_n\left\{\int_0^\tau \widehat{W}(t;\boldsymbol{\beta})\left[\boldsymbol{Z}-\frac{\mathbb{P}_n\{\widehat{W}(t;\boldsymbol{\beta})\boldsymbol{Z}\}}{\mathbb{P}_n\{\widehat{W}(t;\boldsymbol{\beta})\}}\right]dN(t;\boldsymbol{\beta})\right\}.$$

Let

$$\frac{1}{n}\boldsymbol{U}^*_n(\boldsymbol{\beta}) = \mathbb{P}_n\left\{\int_0^\tau W(t;\boldsymbol{\beta})\left[\boldsymbol{Z}-\frac{\mathbb{P}_n\{W(t;\boldsymbol{\beta})\boldsymbol{Z}\}}{\mathbb{P}_n\{W(t;\boldsymbol{\beta})\}}\right]dN(t;\boldsymbol{\beta})\right\}.$$

Notice that $\widehat{W}_i(t;\boldsymbol{\beta}) = \frac{I\left(C_i \geq D_i \wedge te^{-\boldsymbol{\beta}^T\boldsymbol{Z}_i}\right)\widehat{S}(te^{-\boldsymbol{\beta}^T\boldsymbol{Z}_i}|\boldsymbol{Z}_i)}{\widehat{S}\left(X\wedge te^{-\boldsymbol{\beta}^T\boldsymbol{Z}_i}|\boldsymbol{Z}_i\right)}, W_i(t;\boldsymbol{\beta}) = \frac{I\left(C_i \geq D_i \wedge te^{-\boldsymbol{\beta}^T\boldsymbol{Z}_i}\right)S(te^{-\boldsymbol{\beta}^T\boldsymbol{Z}_i}|\boldsymbol{Z}_i)}{S\left(X_i\wedge te^{-\boldsymbol{\beta}^T\boldsymbol{Z}_i}|\boldsymbol{Z}_i\right)}$, where $\widehat{S}(t|\boldsymbol{Z}_i)$ is the estimated survival function at time t given $\boldsymbol{Z}_i$ from the Cox proportional hazard model (3.4), and $S(t|\boldsymbol{Z}_i)$ is the true survival function at time t. We know that $\widehat{\boldsymbol{\gamma}} \xrightarrow{a.s.} \boldsymbol{\gamma}$ and $\widehat{\Lambda}(\cdot) \xrightarrow{a.s.} \Lambda_0(\cdot)$, then $\widehat{S}(te^{-\boldsymbol{\beta}^T\boldsymbol{Z}_i}|\boldsymbol{Z}_i) \xrightarrow{a.s.} S(te^{-\boldsymbol{\beta}^T\boldsymbol{Z}_i}|\boldsymbol{Z}_i)$ and $\widehat{S}(X_i \wedge te^{-\boldsymbol{\beta}^T\boldsymbol{Z}_i}|\boldsymbol{Z}_i) \xrightarrow{a.s.} S(X_i \wedge te^{-\boldsymbol{\beta}^T\boldsymbol{Z}_i}|\boldsymbol{Z}_i)$. Thus

$$\sup_{t\in[0,\tau]}\max_i\left|\widehat{W}_i(t;\boldsymbol{\beta}) - W_i(t;\boldsymbol{\beta})\right| \xrightarrow{a.s.} 0.$$

Since we assume $\boldsymbol{Z}$ is bounded, there exist constants C_1 , such that

$$\begin{aligned}
&\sup_{t\in[0,\tau]}\max_i\left|\frac{1}{n}\sum_{i=1}^n\widehat{W}_i(t;\boldsymbol{\beta})\boldsymbol{Z}_i - \frac{1}{n}\sum_{i=1}^n W_i(t;\boldsymbol{\beta})\boldsymbol{Z}_i\right| \\
\leq\ &\sup_{t\in[0,\tau]}\max_i\left|\widehat{W}_i(t;\boldsymbol{\beta}) - W_i(t;\boldsymbol{\beta})\right|\cdot C_1 \xrightarrow{a.s.} 0.
\end{aligned}$$

Thus we can show

$$\sup_{\boldsymbol{\beta}\in\mathcal{B}}\left|\frac{1}{n}\boldsymbol{U}_n(\boldsymbol{\beta})-\frac{1}{n}\boldsymbol{U}_n^*(\boldsymbol{\beta})\right|\xrightarrow{a.s.}0.$$

In $\boldsymbol{U}_n^*(\boldsymbol{\beta})$, the classes of function $\{\log t-\boldsymbol{\beta}^T\boldsymbol{Z}\}$ and $\{\log C\geq logD\wedge(\log t-\boldsymbol{\beta}^T\boldsymbol{Z})\}$ belong to VC-class. Since exponentiation is monotone, $\{te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}\}=\{e^{\log t-\boldsymbol{\beta}^T\boldsymbol{Z}}\}$ and $\{C\geq D\wedge te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}\}=\{e^{\log C\geq D\wedge(\log t-\boldsymbol{\beta}^T\boldsymbol{Z})}\}$ are also VC-classes and thus bounded P-Donsker. By the results of exercise 2.6.9 and 2.6.14 in Van der Vaart (1996), $I(C\geq D\wedge te^{-\boldsymbol{\beta}^T\boldsymbol{Z}})$ is a VC-class and bounded P-Donsker. Since the survival function $S(\cdot)$ is monotone on $[0,1]$, by the permanence property of VC-class, $S(te^{-\boldsymbol{\beta}^T\boldsymbol{Z}})$ is a VC-class and bounded P-Donsker. Similarly, we can show $S(X\wedge te^{-\boldsymbol{\beta}^T\boldsymbol{Z}})$ is a VC-class and bounded P-Donsker. Therefore, $W(t;\boldsymbol{\beta})$ is P-Donsker and P-Glivenko-Cantelli as the product of bounded Donsker classes is Donsker. There exists a partition $0\leq s_1\leq\cdots\leq s_k\leq\cdots\leq\tau$, such that $N(t;\boldsymbol{\beta})=N(te^{-\boldsymbol{\beta}^T\boldsymbol{Z}})=\sum_{k=1}^{\infty}N(s_k)I(s_k<te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}\leq s_{k+1})$. Notice that $\sum_{k=1}^{M}N(s_k)I(s_k<te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}\leq s_{k+1})$ is a VC-hull class since the class of function $\{N(s_k)\}$ is a VC-class, and $\sum_{k=1}^{M}I(s_k<te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}\leq s_{k+1})\leq 1$. Thus $\sum_{k}^{M}N(s_k)I(s_k<te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}\leq s_{k+1})$ is P-Donsker. By the permanence property of Donsker class, its closure, which is $N(t;\boldsymbol{\beta})$, is also Donsker. Trivially, $\boldsymbol{Z}$ is P-Donsker. Again since all products of bounded Donsker classes are Donsker, $\{W(t;\boldsymbol{\beta})N(t;\boldsymbol{\beta}):\boldsymbol{\beta}\in\mathcal{B},t\in[0,\tau]\}$, $\{W(t;\boldsymbol{\beta})\boldsymbol{Z}N(t;\boldsymbol{\beta}):\boldsymbol{\beta}\in\mathcal{B},t\in[0,\tau]\}$ and $\{W(t;\boldsymbol{\beta})\boldsymbol{Z}:\boldsymbol{\beta}\in\mathcal{B},t\in[0,\tau]\}$ are P-Donsker and thus P-Glivenko-Cantelli. Therefore, we have

$$\sup_{\boldsymbol{\beta}\in\mathcal{B}}\left|\frac{1}{n}\boldsymbol{U}_n^*(\boldsymbol{\beta})-\boldsymbol{U}(\boldsymbol{\beta})\right|\xrightarrow{a.s.}0,$$

where

$$\boldsymbol{U}(\boldsymbol{\beta})\equiv P\left\{\int_0^{\tau}W(t;\boldsymbol{\beta})\left[\boldsymbol{Z}-\frac{P\{W(t;\boldsymbol{\beta})\boldsymbol{Z}\}}{P\{W(t;\boldsymbol{\beta})\}}\right]dN(t;\boldsymbol{\beta})\right\}.$$

Since we have shown $\sup_{\boldsymbol{\beta}\in\mathcal{B}} \left|\frac{1}{n}\boldsymbol{U}_n(\boldsymbol{\beta}) - \frac{1}{n}\boldsymbol{U}^*_n(\boldsymbol{\beta})\right| \xrightarrow{a.s.} 0$, we have

$$\sup_{\boldsymbol{\beta}\in\mathcal{B}} \left|\frac{1}{n}\boldsymbol{U}_n(\boldsymbol{\beta}) - \boldsymbol{U}(\boldsymbol{\beta})\right| \xrightarrow{a.s.} 0.$$

Next, we show $\boldsymbol{U}(\boldsymbol{\beta}_0) = 0$ under the proposed model (3.2).

$$\begin{aligned}
&\boldsymbol{U}(\boldsymbol{\beta}_0) \\
= \;& P\left\{\int_0^\tau W(t;\boldsymbol{\beta}_0)\left[\boldsymbol{Z} - \frac{P\{W(t;\boldsymbol{\beta}_0)\boldsymbol{Z}\}}{P\{W(t;\boldsymbol{\beta}_0)\}}\right] dN(t;\boldsymbol{\beta}_0)\right\} \\
= \;& \int_0^\tau E\left\{W(t;\boldsymbol{\beta}_0)\left[\boldsymbol{Z} - \frac{E\{W(t;\boldsymbol{\beta}_0)\boldsymbol{Z}\}}{E\{W(t;\boldsymbol{\beta}_0)\}}\right] dN(t;\boldsymbol{\beta}_0)\right\} \\
= \;& \int_0^\tau E\left\{W(t;\boldsymbol{\beta}_0)\boldsymbol{Z} dN(t;\boldsymbol{\beta}_0)\right\} - \int_0^\tau \frac{E\{W(t;\boldsymbol{\beta}_0)\boldsymbol{Z}\}}{E\{W(t;\boldsymbol{\beta}_0)\}} E\left\{W(t;\boldsymbol{\beta}_0) dN(t;\boldsymbol{\beta}_0)\right\}
\end{aligned} \tag{3.16}$$

Now look at the first term of (3.16), since $I\left(C \geq D \wedge te^{-\boldsymbol{\beta}_0^T\boldsymbol{Z}}\right) \cdot I\left(C \geq te^{-\boldsymbol{\beta}_0^T\boldsymbol{Z}}\right) = I\left(C \geq te^{-\boldsymbol{\beta}_0^T\boldsymbol{Z}}\right)$, $S\left(X \wedge te^{-\boldsymbol{\beta}_0^T\boldsymbol{Z}}|\boldsymbol{Z}\right) = S\left(te^{-\boldsymbol{\beta}_0^T\boldsymbol{Z}}|\boldsymbol{Z}\right)$ when $I\left(C \geq te^{-\boldsymbol{\beta}_0^T\boldsymbol{Z}}\right) \cdot I\left(D \geq te^{-\boldsymbol{\beta}_0^T\boldsymbol{Z}}\right) = 1$. We have

$$\begin{aligned}
&E\left\{W(t;\boldsymbol{\beta}_0)\boldsymbol{Z} dN(t;\boldsymbol{\beta}_0)\right\} \\
= \;& E\left\{\frac{I\left(C \geq D \wedge te^{-\boldsymbol{\beta}_0^T\boldsymbol{Z}}\right) S(te^{-\boldsymbol{\beta}_0^T\boldsymbol{Z}}|\boldsymbol{Z})}{S\left(X \wedge te^{-\boldsymbol{\beta}_0^T\boldsymbol{Z}}|\boldsymbol{Z}\right)}\,\boldsymbol{Z}\, I\left(C \geq te^{-\boldsymbol{\beta}_0^T\boldsymbol{Z}}\right) dN^*(t;\boldsymbol{\beta}_0)\right\} \\
= \;& E\left\{I\left(C \geq te^{-\boldsymbol{\beta}_0^T\boldsymbol{Z}}\right)\boldsymbol{Z}\, dN^*(t;\boldsymbol{\beta}_0)\right\} \\
= \;& E\left\{I\left(C \geq te^{-\boldsymbol{\beta}_0^T\boldsymbol{Z}}\right)\boldsymbol{Z}\, E\left[dN^*(t;\boldsymbol{\beta}_0)|\boldsymbol{Z}\right]\right\} \\
= \;& E\left\{I\left(C \geq te^{-\boldsymbol{\beta}_0^T\boldsymbol{Z}}\right)\boldsymbol{Z}\, d\mu_0(t)\right\} = \int_0^\tau E\left\{Y(t;\boldsymbol{\beta}_0)\boldsymbol{Z}\right\} d\mu_0(t).
\end{aligned}$$

Using similar arguments, we can show $E\{W(t;\boldsymbol{\beta}_0)dN(t;\boldsymbol{\beta}_0)\} = E\{Y(t;\boldsymbol{\beta}_0)\}\,d\mu_0(t)$. As for the second term of (3.16),

$$
\begin{aligned}
& E\{W(t;\boldsymbol{\beta}_0)\} \qquad (3.17)\\
= \;& E\left\{\frac{I\left(C \geq D \wedge te^{-\boldsymbol{\beta}_0^T\boldsymbol{Z}}\right) S(te^{-\boldsymbol{\beta}_0^T\boldsymbol{Z}}|\boldsymbol{Z})}{S\left(X \wedge te^{-\boldsymbol{\beta}_0^T\boldsymbol{Z}}|\boldsymbol{Z}\right)}\right\}\\
= \;& E\left\{\frac{I\left(D \geq te^{-\boldsymbol{\beta}_0^T\boldsymbol{Z}}\right) I\left(C \geq D \wedge te^{-\boldsymbol{\beta}_0^T\boldsymbol{Z}}\right) S(te^{-\boldsymbol{\beta}_0^T\boldsymbol{Z}}|\boldsymbol{Z})}{S\left(X \wedge te^{-\boldsymbol{\beta}_0^T\boldsymbol{Z}}|\boldsymbol{Z}\right)}\right.\\
& \left. + \frac{I\left(D < te^{-\boldsymbol{\beta}_0^T\boldsymbol{Z}}\right) I\left(C \geq D \wedge te^{-\boldsymbol{\beta}_0^T\boldsymbol{Z}}\right) S(te^{-\boldsymbol{\beta}_0^T\boldsymbol{Z}}|\boldsymbol{Z})}{S\left(X \wedge te^{-\boldsymbol{\beta}_0^T\boldsymbol{Z}}|\boldsymbol{Z}\right)}\right\}\\
= \;& E\left\{I\left(D \geq te^{-\boldsymbol{\beta}_0^T\boldsymbol{Z}}\right) I\left(C \geq te^{-\boldsymbol{\beta}_0^T\boldsymbol{Z}}\right) + \frac{I\left(D < te^{-\boldsymbol{\beta}_0^T\boldsymbol{Z}}\right) I\left(C \geq D\right) S(te^{-\boldsymbol{\beta}_0^T\boldsymbol{Z}}|\boldsymbol{Z})}{S\left(D|\boldsymbol{Z}\right)}\right\}.
\end{aligned}
$$

The second term of (3.17) equals to

$$
E\left\{E\left\{\left.\frac{I\left(D < te^{-\boldsymbol{\beta}_0^T\boldsymbol{Z}}\right) I\left(C \geq D\right) S(te^{-\boldsymbol{\beta}_0^T\boldsymbol{Z}}|\boldsymbol{Z})}{S\left(D|\boldsymbol{Z}\right)}\right| \boldsymbol{Z}, D\right\}\right\}
$$

$$
\begin{aligned}
= \;& E\left\{\frac{I\left(D < te^{-\boldsymbol{\beta}_0^T\boldsymbol{Z}}\right) S(te^{-\boldsymbol{\beta}_0^T\boldsymbol{Z}}|\boldsymbol{Z})}{S\left(D|\boldsymbol{Z}\right)} E\left\{\left. I\left(C \geq D\right)\right| \boldsymbol{Z}, D\right\}\right\}\\
= \;& E\left\{\frac{I\left(D < te^{-\boldsymbol{\beta}_0^T\boldsymbol{Z}}\right) S(te^{-\boldsymbol{\beta}_0^T\boldsymbol{Z}}|\boldsymbol{Z})}{S\left(D|\boldsymbol{Z}\right)} S\left(D|\boldsymbol{Z}\right)\right\}\\
= \;& E\left\{I\left(D < te^{-\boldsymbol{\beta}_0^T\boldsymbol{Z}}\right) S(te^{-\boldsymbol{\beta}_0^T\boldsymbol{Z}}|\boldsymbol{Z})\right\} = E\left\{I\left(D < te^{-\boldsymbol{\beta}_0^T\boldsymbol{Z}}\right) I\left(C \geq te^{-\boldsymbol{\beta}_0^T\boldsymbol{Z}}\right)\right\}
\end{aligned}
$$

Thus (3.17) becomes

$$\begin{aligned} E\{W(t;\boldsymbol{\beta}_0)\} &= E\left\{I\left(D \geq te^{-\boldsymbol{\beta}_0^T\boldsymbol{Z}}\right)I\left(C \geq te^{-\boldsymbol{\beta}_0^T\boldsymbol{Z}}\right) + I\left(D < te^{-\boldsymbol{\beta}_0^T\boldsymbol{Z}}\right)I\left(C \geq te^{-\boldsymbol{\beta}_0^T\boldsymbol{Z}}\right)\right\} \\ &= E\left\{I\left(C \geq te^{-\boldsymbol{\beta}_0^T\boldsymbol{Z}}\right)\right\} = E\left\{Y(t;\boldsymbol{\beta}_0)\right\}. \end{aligned}$$

Similarly, we can show

$$E\{W(t;\boldsymbol{\beta}_0)\boldsymbol{Z}\} = E\left\{Y(t;\boldsymbol{\beta}_0)\boldsymbol{Z}\right\}.$$

Plugging the above results into (3.16) , we show $\boldsymbol{U}(\boldsymbol{\beta}_0) = 0$.

By Taylor expansion, we have

$$\boldsymbol{U}(\boldsymbol{\beta}) = \boldsymbol{U}(\boldsymbol{\beta}_0) + \boldsymbol{A}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + o(|\boldsymbol{\beta} - \boldsymbol{\beta}_0|) = \boldsymbol{A}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + o(|\boldsymbol{\beta} - \boldsymbol{\beta}_0|),$$

By assumption (C7), $\boldsymbol{A}$ is non-singular, for sufficiently large n and small enough ϵ, we have

$$\sup_{|\boldsymbol{\beta}-\boldsymbol{\beta}_0|=\epsilon}\left|\frac{1}{n}\boldsymbol{U}_n(\boldsymbol{\beta}) - \boldsymbol{A}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\right| \leq \inf_{|\boldsymbol{\beta}-\boldsymbol{\beta}_0|=\epsilon}\left|\boldsymbol{A}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\right|.$$

Notice that $\boldsymbol{A}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$ has a unique solution within $|\boldsymbol{\beta} - \boldsymbol{\beta}_0| \leq \epsilon$. By the degree theory(Deming 1985, Chapter 1), the above inequality implies $\frac{1}{n}\boldsymbol{U}_n(\boldsymbol{\beta})$ has the same number of non-zero solution as $\boldsymbol{A}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$. Therefore there exists $\widehat{\boldsymbol{\beta}}$ which is the solution to $\boldsymbol{U}_n(\boldsymbol{\beta}) = 0$ and $|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0| \leq \epsilon$ for sufficiently large n. Since ϵ can be chosen arbitrarily small, $\widehat{\boldsymbol{\beta}} \xrightarrow{a.s.} \boldsymbol{\beta}_0$. This concludes the proof of consistency of $\widehat{\boldsymbol{\beta}}$.

Let $M(t;\boldsymbol{\beta}) = N(t;\boldsymbol{\beta}) - \int_0^{te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}} d\mu_0(se^{\boldsymbol{\beta}_0^T\boldsymbol{Z}})$. By addition and subtraction,

$$\begin{aligned}
\frac{1}{n}\boldsymbol{U}_n(\widehat{\boldsymbol{\beta}}) &= \mathbb{P}_n\int_0^\tau \widehat{W}(t;\widehat{\boldsymbol{\beta}})\left[\boldsymbol{Z} - \frac{\mathbb{P}_n\{\widehat{W}(t;\widehat{\boldsymbol{\beta}})\boldsymbol{Z}\}}{\mathbb{P}_n\{\widehat{W}(t;\widehat{\boldsymbol{\beta}})\}}\right]\left\{dN(t;\widehat{\boldsymbol{\beta}}) - d\mu_0(t)\right\} \\
&= \mathbb{P}_n\int_0^\tau \widehat{W}(t;\widehat{\boldsymbol{\beta}})\left[\boldsymbol{Z} - \frac{\mathbb{P}_n\{\widehat{W}(t;\widehat{\boldsymbol{\beta}})\boldsymbol{Z}\}}{\mathbb{P}_n\{\widehat{W}(t;\widehat{\boldsymbol{\beta}})\}}\right]\left\{dN(t;\widehat{\boldsymbol{\beta}}) - d\mu_0(te^{(\boldsymbol{\beta}_0-\widehat{\boldsymbol{\beta}})^T\boldsymbol{Z}})\right\} \\
&\quad +\mathbb{P}_n\int_0^\tau \widehat{W}(t;\widehat{\boldsymbol{\beta}})\left[\boldsymbol{Z} - \frac{\mathbb{P}_n\{\widehat{W}(t;\widehat{\boldsymbol{\beta}})\boldsymbol{Z}\}}{\mathbb{P}_n\{\widehat{W}(t;\widehat{\boldsymbol{\beta}})\}}\right]d\left\{\mu_0(te^{(\boldsymbol{\beta}_0-\widehat{\boldsymbol{\beta}})^T\boldsymbol{Z}}) - \mu_0(t)\right\} \\
&= \mathbb{P}_n\int_0^\tau \widehat{W}(t;\widehat{\boldsymbol{\beta}})\left[\boldsymbol{Z} - \frac{\mathbb{P}_n\{\widehat{W}(t;\widehat{\boldsymbol{\beta}})\boldsymbol{Z}\}}{\mathbb{P}_n\{\widehat{W}(t;\widehat{\boldsymbol{\beta}})\}}\right]dM(t;\widehat{\boldsymbol{\beta}}) \\
&\quad +\mathbb{P}_n\int_0^\tau \widehat{W}(t;\widehat{\boldsymbol{\beta}})\left[\boldsymbol{Z} - \frac{\mathbb{P}_n\{\widehat{W}(t;\widehat{\boldsymbol{\beta}})\boldsymbol{Z}\}}{\mathbb{P}_n\{\widehat{W}(t;\widehat{\boldsymbol{\beta}})\}}\right]d\left\{\mu_0(te^{(\boldsymbol{\beta}_0-\widehat{\boldsymbol{\beta}})^T\boldsymbol{Z}}) - \mu_0(t)\right\}
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
\boldsymbol{U}(\widehat{\boldsymbol{\beta}}) &= P\int_0^\tau W(t;\widehat{\boldsymbol{\beta}})\left[\boldsymbol{Z} - \frac{P\{W(t;\widehat{\boldsymbol{\beta}})\boldsymbol{Z}\}}{P\{W(t;\widehat{\boldsymbol{\beta}})\}}\right]dM(t;\widehat{\boldsymbol{\beta}}) \\
&\quad +P\int_0^\tau W(t;\widehat{\boldsymbol{\beta}})\left[\boldsymbol{Z} - \frac{P\{W(t;\widehat{\boldsymbol{\beta}})\boldsymbol{Z}\}}{P\{W(t;\widehat{\boldsymbol{\beta}})\}}\right]d\left\{\mu_0(te^{(\boldsymbol{\beta}_0-\widehat{\boldsymbol{\beta}})^T\boldsymbol{Z}}) - \mu_0(t)\right\}
\end{aligned}$$

Thus

$$\begin{aligned}
&\sqrt{n}\left\{\frac{1}{n}\boldsymbol{U}_n(\widehat{\boldsymbol{\beta}}) - \boldsymbol{U}(\widehat{\boldsymbol{\beta}})\right\} \\
=\ &\sqrt{n}\mathbb{P}_n\int_0^\tau \widehat{W}(t;\widehat{\boldsymbol{\beta}})\left[\boldsymbol{Z} - \frac{\mathbb{P}_n\{\widehat{W}(t;\widehat{\boldsymbol{\beta}})\boldsymbol{Z}\}}{\mathbb{P}_n\{\widehat{W}(t;\widehat{\boldsymbol{\beta}})\}}\right]dM(t;\widehat{\boldsymbol{\beta}}) \\
&-\sqrt{n}P\int_0^\tau W(t;\widehat{\boldsymbol{\beta}})\left[\boldsymbol{Z} - \frac{P\{W(t;\widehat{\boldsymbol{\beta}})\boldsymbol{Z}\}}{P\{W(t;\widehat{\boldsymbol{\beta}})\}}\right]dM(t;\widehat{\boldsymbol{\beta}}) \\
&+\sqrt{n}\mathbb{P}_n\int_0^\tau \widehat{W}(t;\widehat{\boldsymbol{\beta}})\left[\boldsymbol{Z} - \frac{\mathbb{P}_n\{\widehat{W}(t;\widehat{\boldsymbol{\beta}})\boldsymbol{Z}\}}{\mathbb{P}_n\{\widehat{W}(t;\widehat{\boldsymbol{\beta}})\}}\right]d\left\{\mu_0(te^{(\boldsymbol{\beta}_0-\widehat{\boldsymbol{\beta}})^T\boldsymbol{Z}}) - \mu_0(t)\right\} \\
&-\sqrt{n}P\int_0^\tau W(t;\widehat{\boldsymbol{\beta}})\left[\boldsymbol{Z} - \frac{P\{W(t;\widehat{\boldsymbol{\beta}})\boldsymbol{Z}\}}{P\{W(t;\widehat{\boldsymbol{\beta}})\}}\right]d\left\{\mu_0(te^{(\boldsymbol{\beta}_0-\widehat{\boldsymbol{\beta}})^T\boldsymbol{Z}}) - \mu_0(t)\right\}
\end{aligned}$$

$$\begin{aligned}
&= \sqrt{n}(\mathbb{P}_n - P)\int_0^\tau \widehat{W}(t;\widehat{\boldsymbol{\beta}})\left[\boldsymbol{Z} - \frac{\mathbb{P}_n\{\widehat{W}(t;\widehat{\boldsymbol{\beta}})\boldsymbol{Z}\}}{\mathbb{P}_n\{\widehat{W}(t;\widehat{\boldsymbol{\beta}})\}}\right] dM(t;\widehat{\boldsymbol{\beta}}) \\
&\quad + \sqrt{n}P\int_0^\tau \widehat{W}(t;\widehat{\boldsymbol{\beta}})\left[\boldsymbol{Z} - \frac{\mathbb{P}_n\{\widehat{W}(t;\widehat{\boldsymbol{\beta}})\boldsymbol{Z}\}}{\mathbb{P}_n\{\widehat{W}(t;\widehat{\boldsymbol{\beta}})\}}\right] dM(t;\widehat{\boldsymbol{\beta}}) \\
&\quad - \sqrt{n}P\int_0^\tau W(t;\widehat{\boldsymbol{\beta}})\left[\boldsymbol{Z} - \frac{P\{W(t;\widehat{\boldsymbol{\beta}})\boldsymbol{Z}\}}{P\{W(t;\widehat{\boldsymbol{\beta}})\}}\right] dM(t;\widehat{\boldsymbol{\beta}}) \\
&\quad + \sqrt{n}(\mathbb{P}_n - P)\int_0^\tau \widehat{W}(t;\widehat{\boldsymbol{\beta}})\left[\boldsymbol{Z} - \frac{\mathbb{P}_n\{\widehat{W}(t;\widehat{\boldsymbol{\beta}})\boldsymbol{Z}\}}{\mathbb{P}_n\{\widehat{W}(t;\widehat{\boldsymbol{\beta}})\}}\right] d\left\{\mu_0(te^{(\boldsymbol{\beta}_0-\widehat{\boldsymbol{\beta}})^T\boldsymbol{Z}}) - \mu_0(t)\right\} \\
&\quad + \sqrt{n}P\int_0^\tau \left\{\widehat{W}(t;\widehat{\boldsymbol{\beta}})\left[\boldsymbol{Z} - \frac{\mathbb{P}_n\{\widehat{W}(t;\widehat{\boldsymbol{\beta}})\boldsymbol{Z}\}}{\mathbb{P}_n\{\widehat{W}(t;\widehat{\boldsymbol{\beta}})\}}\right]\right. \\
&\quad \left. - W(t;\widehat{\boldsymbol{\beta}})\left[\boldsymbol{Z} - \frac{P\{W(t;\widehat{\boldsymbol{\beta}})\boldsymbol{Z}\}}{P\{W(t;\widehat{\boldsymbol{\beta}})\}}\right]\right\} d\left\{\mu_0(te^{(\boldsymbol{\beta}_0-\widehat{\boldsymbol{\beta}})^T\boldsymbol{Z}}) - \mu_0(t)\right\} \\
&\equiv (i) + (ii) - (iii) + (iv) + (v). \qquad (3.18)
\end{aligned}$$

For any given $\boldsymbol{\beta}$,

$$\begin{aligned}
& P\left\{W(t;\boldsymbol{\beta})\boldsymbol{Z}dM(t;\boldsymbol{\beta})\right\} \\
&= E\left\{W(t;\boldsymbol{\beta})\boldsymbol{Z}\left\{I\left(C \ge te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}\right)dN^*(t;\boldsymbol{\beta}) - d\mu_0(te^{(\boldsymbol{\beta}_0-\boldsymbol{\beta})^T\boldsymbol{Z}})\right\}\right\} \\
&= E\left\{\frac{I\left(C \ge D \wedge te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}\right)S(te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}|\boldsymbol{Z})}{S\left(X \wedge te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}|\boldsymbol{Z}\right)}\boldsymbol{Z}I\left(C \ge te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}\right)dN^*(t;\boldsymbol{\beta})\right\} \\
&\quad - E\left\{W(t;\boldsymbol{\beta})\boldsymbol{Z}d\mu_0(te^{(\boldsymbol{\beta}_0-\boldsymbol{\beta})^T\boldsymbol{Z}})\right\} \qquad (3.19)
\end{aligned}$$

Since $I\left(C \ge D \wedge te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}\right) \cdot I\left(C \ge te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}\right) = I\left(C \ge te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}\right)$, $S\left(X \wedge te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}|\boldsymbol{Z}\right) = S\left(te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}|\boldsymbol{Z}\right)$ when $I\left(C \ge te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}\right) \cdot I\left(D \ge te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}\right) = 1$.

For the first term of (3.19) we have

$$E\left\{\frac{I\left(C \ge D \wedge te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}\right)S(te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}|\boldsymbol{Z})}{S\left(X \wedge te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}|\boldsymbol{Z}\right)}\boldsymbol{Z}I\left(C \ge te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}\right)dN^*(t;\boldsymbol{\beta})\right\}$$

$$
\begin{aligned}
&= E\left\{I\left(C \geq te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}\right)\boldsymbol{Z}dN^*(t;\boldsymbol{\beta})\right\} \\
&= E\left\{I\left(C \geq te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}\right)\boldsymbol{Z}E\left[dN^*(t;\boldsymbol{\beta})|\boldsymbol{Z}\right]\right\} \\
&= E\left\{I\left(C \geq te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}\right)\boldsymbol{Z}d\mu_0(te^{(\boldsymbol{\beta}_0-\boldsymbol{\beta})^T\boldsymbol{Z}})\right\} = E\left\{Y(t;\boldsymbol{\beta})\boldsymbol{Z}d\mu_0(te^{(\boldsymbol{\beta}_0-\boldsymbol{\beta})^T\boldsymbol{Z}})\right\}.
\end{aligned}
$$

Using similar arguments in the proof of (3.17), we can show the second term of (3.19),

$$
E\left\{W(t;\boldsymbol{\beta})\boldsymbol{Z}d\mu_0(te^{(\boldsymbol{\beta}_0-\boldsymbol{\beta})^T\boldsymbol{Z}})\right\} = E\left\{Y(t;\boldsymbol{\beta})\boldsymbol{Z}d\mu_0(te^{(\boldsymbol{\beta}_0-\boldsymbol{\beta})^T\boldsymbol{Z}})\right\}.
$$

Thus (3.19) $= P\left\{W(t;\boldsymbol{\beta})\boldsymbol{Z}dM(t;\boldsymbol{\beta})\right\} = 0$.

Similarly, we can show

$$
P\left\{\widehat{W}(t;\boldsymbol{\beta})\boldsymbol{Z}dM(t;\boldsymbol{\beta})\right\} = 0,
$$

$$
P\left\{W(t;\boldsymbol{\beta})\frac{P\{W(t;\boldsymbol{\beta})\boldsymbol{Z}\}}{P\{W(t;\boldsymbol{\beta})\}}dM(t;\boldsymbol{\beta})\right\} = 0,
$$

$$
P\left\{\widehat{W}(t;\boldsymbol{\beta})\frac{\mathbb{P}_n\{\widehat{W}(t;\boldsymbol{\beta})\boldsymbol{Z}\}}{\mathbb{P}_n\{\widehat{W}(t;\boldsymbol{\beta})\}}dM(t;\boldsymbol{\beta})\right\} = 0,
$$

Therefore, given $\widehat{\boldsymbol{\beta}}$, (ii) and (iii) of (3.18) are zero.

By Taylor expansion, for any given $\boldsymbol{\beta}$,

$$
\begin{aligned}
&\mu_0(te^{(\boldsymbol{\beta}_0-\boldsymbol{\beta})^T\boldsymbol{Z}_i}) - \mu_0(t) \qquad (3.20)\\
&= \mu_0(te^{\boldsymbol{Z}_i^T(\boldsymbol{\beta}_0-\boldsymbol{\beta})}) - \mu_0(t) \\
&= \mu_0\left\{t[1+\boldsymbol{Z}_i^T(\boldsymbol{\beta}_0-\boldsymbol{\beta})+o(\boldsymbol{Z}_i^T|\boldsymbol{\beta}_0-\boldsymbol{\beta}|)]\right\} - \mu_0(t) \\
&= \dot{\mu}_0(t)\left\{t[1+\boldsymbol{Z}_i^T(\boldsymbol{\beta}_0-\boldsymbol{\beta})+o(\boldsymbol{Z}_i^T|\boldsymbol{\beta}_0-\boldsymbol{\beta}|)]-t\right\} + \frac{\ddot{\mu}_0(t)}{2}\,t\,o(\boldsymbol{Z}_i^T|\boldsymbol{\beta}_0-\boldsymbol{\beta}|) \\
&= \dot{\mu}_0(t)\left\{t\boldsymbol{Z}_i^T(\boldsymbol{\beta}_0-\boldsymbol{\beta})\right\} + t\,o(\boldsymbol{Z}_i^T|\boldsymbol{\beta}_0-\boldsymbol{\beta}|),\ \ddot{\mu}_0(t) \text{ is bounded by Condition (C5).}
\end{aligned}
$$

By the consistency of $\widehat{\boldsymbol{\beta}}$, $\mu_0(te^{(\boldsymbol{\beta}_0-\boldsymbol{\beta})^T\boldsymbol{Z}_i})-\mu_0(t)=o_P(1)$. Since $\widehat{W}(t;\widehat{\boldsymbol{\beta}})\left[\boldsymbol{Z}-\frac{\mathbb{P}_n\{\widehat{W}(t;\widehat{\boldsymbol{\beta}})\boldsymbol{Z}\}}{\mathbb{P}_n\{\widehat{W}(t;\widehat{\boldsymbol{\beta}})\}}\right]$ is bounded and $\sqrt{n}(\mathbb{P}_n-P)=O(1)$, we have

$$(iv) \text{ of } (3.18) \quad = \quad P\int_0^\tau O(1)\cdot o_P(1)=o_P(1).$$

Notice that $\widehat{W}(t;\widehat{\boldsymbol{\beta}})-W(t;\widehat{\boldsymbol{\beta}})=O(n^{-1/2})$ since $\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}=O(n^{-1/2})$ and $\widehat{\Lambda}(\cdot)-\Lambda(\cdot)=O(n^{-1/2})$. We also have $\mathbb{P}_n-P=O(n^{-1/2})$, therefore

$$\widehat{W}(t;\widehat{\boldsymbol{\beta}})\left[\boldsymbol{Z}-\frac{\mathbb{P}_n\{\widehat{W}(t;\widehat{\boldsymbol{\beta}})\boldsymbol{Z}\}}{\mathbb{P}_n\{\widehat{W}(t;\widehat{\boldsymbol{\beta}})\}}\right]-W(t;\widehat{\boldsymbol{\beta}})\left[\boldsymbol{Z}-\frac{P\{W(t;\widehat{\boldsymbol{\beta}})\boldsymbol{Z}\}}{P\{W(t;\widehat{\boldsymbol{\beta}})\}}\right]=O(n^{-1/2}),$$

$$(v) \text{ of } (3.18)=\sqrt{n}P\int_0^\tau O(n^{-1/2})\cdot o_P(1)=O(1)\cdot o_P(1)=o_P(1).$$

Thus (3.18) becomes

$$\begin{aligned}
&\sqrt{n}\left\{\frac{1}{n}\boldsymbol{U}_n(\widehat{\boldsymbol{\beta}})-\boldsymbol{U}(\widehat{\boldsymbol{\beta}})\right\}\\
=\ &\mathbb{G}_n\int_0^\tau \widehat{W}(t;\widehat{\boldsymbol{\beta}})\left\{\boldsymbol{Z}-\frac{\mathbb{P}_n\{\widehat{W}(t;\widehat{\boldsymbol{\beta}})\boldsymbol{Z}\}}{\mathbb{P}_n\{\widehat{W}(t;\widehat{\boldsymbol{\beta}})\}}\right\}dM(t;\widehat{\boldsymbol{\beta}})+o_P(1). \qquad (3.21)
\end{aligned}$$

Using similar arguments as in proofing the consistency of $\widehat{\boldsymbol{\beta}}$, we can show $\{\widehat{W}(t;\widehat{\boldsymbol{\beta}}):\widehat{\boldsymbol{\beta}}\in\mathcal{B},t\in[0,\tau]\}$, $\{\widehat{W}(t;\widehat{\boldsymbol{\beta}})\boldsymbol{Z}:\widehat{\boldsymbol{\beta}}\in\mathcal{B},t\in[0,\tau]\}$, $\{\widehat{W}(t;\widehat{\boldsymbol{\beta}})\boldsymbol{Z}dM(t;\widehat{\boldsymbol{\beta}}):\widehat{\boldsymbol{\beta}}\in\mathcal{B},t\in[0,\tau]\}$, and $\{\widehat{W}(t;\widehat{\boldsymbol{\beta}})dM(t;\widehat{\boldsymbol{\beta}}):\widehat{\boldsymbol{\beta}}\in\mathcal{B},t\in[0,\tau]\}$ are P-Donsker and thus P-Glivenko-Cantelli. We can replace the operator $\mathbb{P}_n$ in (3.21) with P without altering the limiting distribution. By the strong consistency of $\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\gamma}}$, $\widehat{\Lambda}(\cdot)$ and Lemma 19.24 of Van der Vaart(1998), (3.21) becomes

$$\begin{aligned}
\sqrt{n}\left\{\frac{1}{n}\boldsymbol{U}_n(\widehat{\boldsymbol{\beta}})-\boldsymbol{U}(\widehat{\boldsymbol{\beta}})\right\} &= \mathbb{G}_n\int_0^\tau W(t;\boldsymbol{\beta}_0)\left\{\boldsymbol{Z}-\frac{P\{W(t;\boldsymbol{\beta}_0)\boldsymbol{Z}\}}{P\{W(t;\boldsymbol{\beta}_0)\}}\right\}dM(t;\boldsymbol{\beta}_0)+o_P(1)\\
&= \mathbb{G}_n\boldsymbol{J}+o_P(1). \qquad (3.22)
\end{aligned}$$

On the other hand, by Taylor expansion and the fact that $\boldsymbol{U}(\boldsymbol{\beta}_0) = 0$, we have

$$\begin{aligned} \boldsymbol{U}(\widehat{\boldsymbol{\beta}}) &= \boldsymbol{U}(\boldsymbol{\beta}_0) + \boldsymbol{A}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + o(|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0|) \\ &= \boldsymbol{A}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + o(|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0|). \end{aligned} \tag{3.23}$$

Combining (3.22), (3.23) and $\boldsymbol{U}_n(\widehat{\boldsymbol{\beta}}) = 0$, we have

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = -\boldsymbol{A}^{-1}\mathbb{G}_n \boldsymbol{J} + o_P(1). \tag{3.24}$$

Therefore $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ converges in distribution to a normal distribution with mean zero and covariance matrix $\boldsymbol{A}^{-1}\boldsymbol{\Sigma}(\boldsymbol{A}^{-1})^T$,$\boldsymbol{\Sigma} = E[\boldsymbol{J}\boldsymbol{J}^T]$, $\boldsymbol{A}$ and $\boldsymbol{J}$ as defined before. By replacing $\boldsymbol{\beta}_0$ with $\widehat{\boldsymbol{\beta}}$ and expectations with empirical means in the expressions of $\boldsymbol{J}$, we obtain an estimator $\widehat{\boldsymbol{\Sigma}}$, where $\widehat{\boldsymbol{\Sigma}}$ are defined in (3.8). Using similar arguments as in proofing the consistency and normality of $\widehat{\boldsymbol{\beta}}$, we can show that $\widehat{\boldsymbol{\Sigma}}$ is a consistent estimator of $\boldsymbol{\Sigma}$.

Next we show $\widehat{\boldsymbol{A}}$, as defined in (3.8), is a consistent estimator of A. From (3.22) and (3.23), for any canonical vector $\boldsymbol{e}_i$ and $h_n \to 0$, uniformly in a $\sqrt{n}$-neighborhood of $\boldsymbol{\beta}_0$,

$$\begin{aligned} \frac{1}{\sqrt{n}}\boldsymbol{U}_n(\widehat{\boldsymbol{\beta}}) &= \sqrt{n}\boldsymbol{A}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + \mathbb{G}_n\boldsymbol{J} + o_P(1) \\ \frac{1}{\sqrt{n}}\boldsymbol{U}_n(\widehat{\boldsymbol{\beta}} + \boldsymbol{e}_i h_n) &= \sqrt{n}\boldsymbol{A}(\widehat{\boldsymbol{\beta}} + \boldsymbol{e}_i h_n - \boldsymbol{\beta}_0) + \mathbb{G}_n\boldsymbol{J} + o_P(1) \\ \Rightarrow \frac{1}{\sqrt{n}}\left\{\boldsymbol{U}_n(\widehat{\boldsymbol{\beta}}) - \boldsymbol{U}_n(\widehat{\boldsymbol{\beta}} + \boldsymbol{e}_i h_n)\right\} &= \sqrt{n}\boldsymbol{A}\boldsymbol{e}_i h_n + o_P(1) \\ \Rightarrow \frac{\boldsymbol{U}_n(\widehat{\boldsymbol{\beta}} + \boldsymbol{e}_i h_n) - \boldsymbol{U}_n(\widehat{\boldsymbol{\beta}})}{n h_n} &= \boldsymbol{A}\,\boldsymbol{e}_i + o_P(\frac{1}{\sqrt{n}h_n}) \end{aligned}$$

If we take $h_n = O(n^{-1/2})$, then $o_P(\frac{1}{\sqrt{n}h_n}) = o_P(1)$, $\frac{\boldsymbol{U}_n(\widehat{\boldsymbol{\beta}}+\boldsymbol{e}_i h_n) - \boldsymbol{U}_n(\widehat{\boldsymbol{\beta}})}{n h_n} = \boldsymbol{A}\,e_i + o_P(1)$, $i = 1, \cdots, p$. Thus $\widehat{\boldsymbol{A}} = \left(\frac{\boldsymbol{U}_n(\widehat{\boldsymbol{\beta}}+\boldsymbol{e}_1 h_n) - \boldsymbol{U}_n(\widehat{\boldsymbol{\beta}})}{n h_n}, \cdots, \frac{\boldsymbol{U}_n(\widehat{\boldsymbol{\beta}}+\boldsymbol{e}_p h_n) - \boldsymbol{U}_n(\widehat{\boldsymbol{\beta}})}{n h_n}\right)$ converges in probability to $\boldsymbol{A}$.

<u>Proof of Theorem 3.3.2</u>:

Using empirical process notation, we write

$$\widehat{\mu}_0(t) = \mathbb{P}_n \int_0^t \frac{\widehat{W}(u;\widehat{\boldsymbol{\beta}})dN(u;\widehat{\boldsymbol{\beta}})}{\mathbb{P}_n\widehat{W}(u;\widehat{\boldsymbol{\beta}})},\ t \in [0,\tau].$$

By addition, subtraction and triangle inequality,

$$\begin{aligned}
& \sup_{t\in[0,\tau]} |\widehat{\mu}_0(t) - \mu_0(t)| \\
= & \sup_{t\in[0,\tau]} \left| \mathbb{P}_n \int_0^t \frac{\widehat{W}(u;\widehat{\boldsymbol{\beta}}) \left\{ dN(u;\widehat{\boldsymbol{\beta}}) - d\mu_0(u) \right\}}{\mathbb{P}_n\{\widehat{W}(u;\widehat{\boldsymbol{\beta}})\}} \right| \\
= & \sup_{t\in[0,\tau]} \left| \mathbb{P}_n \int_0^t \frac{\widehat{W}(u;\widehat{\boldsymbol{\beta}})dM(u;\widehat{\boldsymbol{\beta}})}{\mathbb{P}_n\{\widehat{W}(u;\widehat{\boldsymbol{\beta}})\}} + \mathbb{P}_n \int_0^t \frac{\widehat{W}(u;\widehat{\boldsymbol{\beta}})d\left\{ \mu_0(ue^{(\boldsymbol{\beta}_0-\widehat{\boldsymbol{\beta}})^T\boldsymbol{Z}}) - \mu_0(u)) \right\}}{\mathbb{P}_n\{\widehat{W}(u;\widehat{\boldsymbol{\beta}})\}} \right| \\
\leq & \sup_{t\in[0,\tau]} \left| (\mathbb{P}_n - P) \int_0^t \frac{\widehat{W}(u;\widehat{\boldsymbol{\beta}})dM(u;\widehat{\boldsymbol{\beta}})}{\mathbb{P}_n\{\widehat{W}(u;\widehat{\boldsymbol{\beta}})\}} \right| + \sup_{t\in[0,\tau]} \left| P \int_0^t \frac{\widehat{W}(u;\widehat{\boldsymbol{\beta}})dM(u;\widehat{\boldsymbol{\beta}})}{\mathbb{P}_n\{\widehat{W}(u;\widehat{\boldsymbol{\beta}})\}} \right| \\
& + \sup_{t\in[0,\tau]} \left| \mathbb{P}_n \int_0^t \frac{\widehat{W}(u;\widehat{\boldsymbol{\beta}})d\left\{ \mu_0(ue^{(\boldsymbol{\beta}_0-\widehat{\boldsymbol{\beta}})^T\boldsymbol{Z}}) - \mu_0(u)) \right\}}{\mathbb{P}_n\{\widehat{W}(u;\widehat{\boldsymbol{\beta}})\}} \right|, \qquad (3.25)
\end{aligned}$$

where $M(t;\boldsymbol{\beta}) = N(t;\boldsymbol{\beta}) - \int_0^{te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}} d\mu_0(se^{\boldsymbol{\beta}_0^T\boldsymbol{Z}})$.

By Taylor expansion, for any given $\boldsymbol{\beta}$,

$$\begin{aligned}
& \mu_0(te^{(\boldsymbol{\beta}_0-\boldsymbol{\beta})^T\boldsymbol{Z}_i}) - \mu_0(t) \qquad (3.26) \\
= & \mu_0(te^{\boldsymbol{Z}_i^T(\boldsymbol{\beta}_0-\boldsymbol{\beta})}) - \mu_0(t) \\
= & \mu_0\left\{ t[1 + \boldsymbol{Z}_i^T(\boldsymbol{\beta}_0 - \boldsymbol{\beta}) + o(\boldsymbol{Z}_i^T|\boldsymbol{\beta}_0 - \boldsymbol{\beta}|)] \right\} - \mu_0(t) \\
= & \dot{\mu}_0(t)\left\{ t[1 + \boldsymbol{Z}_i^T(\boldsymbol{\beta}_0 - \boldsymbol{\beta}) + o(\boldsymbol{Z}_i^T|\boldsymbol{\beta}_0 - \boldsymbol{\beta}|)] - t \right\} + \frac{\ddot{\mu}_0(t)}{2}\, t\, o(\boldsymbol{Z}_i^T|\boldsymbol{\beta}_0 - \boldsymbol{\beta}|) \\
= & \dot{\mu}_0(t)\left\{ t\boldsymbol{Z}_i^T(\boldsymbol{\beta}_0 - \boldsymbol{\beta}) \right\} + t\, o(\boldsymbol{Z}_i^T|\boldsymbol{\beta}_0 - \boldsymbol{\beta}|),\ \ddot{\mu}_0(t) \text{ is bounded by condition (C5).}
\end{aligned}$$

Since $\widehat{\boldsymbol{\beta}} \xrightarrow{a.s.} \boldsymbol{\beta}$, the third term of (3.25) converges almost surely to 0.

For any given $\boldsymbol{\beta}$,

$$
\begin{aligned}
& P\frac{\widehat{W}(u;\boldsymbol{\beta})dM(u;\boldsymbol{\beta})}{\mathbb{P}_n\{\widehat{W}(u;\boldsymbol{\beta})\}} \\
= \; & E\left\{\frac{\widehat{W}(u;\boldsymbol{\beta})}{\mathbb{P}_n\{\widehat{W}(u;\boldsymbol{\beta})\}}\left\{I\left(C \geq te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}\right)dN^*(t;\boldsymbol{\beta}) - d\mu_0(te^{(\boldsymbol{\beta}_0-\boldsymbol{\beta})^T\boldsymbol{Z}})\right\}\right\} \\
= \; & E\left\{\frac{I\left(C \geq D \wedge te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}\right)S(te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}|\boldsymbol{Z})}{\mathbb{P}_n\{\widehat{W}(u;\boldsymbol{\beta})\}S\left(X \wedge te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}|\boldsymbol{Z}\right)}I\left(C \geq te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}\right)dN^*(t;\boldsymbol{\beta})\right\} \\
& -E\left\{\frac{\widehat{W}(u;\boldsymbol{\beta})}{\mathbb{P}_n\{\widehat{W}(u;\boldsymbol{\beta})\}}d\mu_0(te^{(\boldsymbol{\beta}_0-\boldsymbol{\beta})^T\boldsymbol{Z}})\right\}
\end{aligned}
\tag{3.27}
$$

Since $I\left(C \geq D \wedge te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}\right) \cdot I\left(C \geq te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}\right) = I\left(C \geq te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}\right)$, $S\left(X \wedge te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}|\boldsymbol{Z}\right) = S\left(te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}|\boldsymbol{Z}\right)$ when $I\left(C \geq te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}\right) \cdot I\left(D \geq te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}\right) = 1$.

For the first term of (3.27) we have

$$
\begin{aligned}
& E\left\{\frac{I\left(C \geq D \wedge te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}\right)S(te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}|\boldsymbol{Z})}{\mathbb{P}_n\{\widehat{W}(u;\boldsymbol{\beta})\}S\left(X \wedge te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}|\boldsymbol{Z}\right)}I\left(C \geq te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}\right)dN^*(t;\boldsymbol{\beta})\right\} \\
= \; & E\left\{\frac{I\left(C \geq te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}\right)}{\mathbb{P}_n\{\widehat{W}(u;\boldsymbol{\beta})\}}dN^*(t;\boldsymbol{\beta})\right\} \\
= \; & E\left\{\frac{I\left(C \geq te^{-\boldsymbol{\beta}^T\boldsymbol{Z}}\right)}{\mathbb{P}_n\{\widehat{W}(u;\boldsymbol{\beta})\}}E\left[dN^*(t;\boldsymbol{\beta})|\boldsymbol{Z}\right]\right\} = E\left\{\frac{Y(t;\boldsymbol{\beta})}{\mathbb{P}_n\{\widehat{W}(u;\boldsymbol{\beta})\}}d\mu_0(te^{(\boldsymbol{\beta}_0-\boldsymbol{\beta})^T\boldsymbol{Z}})\right\}.
\end{aligned}
$$

Using similar arguments in the proof of (3.17), we can show the second term of (3.27),

$$
E\left\{\frac{\widehat{W}(u;\boldsymbol{\beta})}{\mathbb{P}_n\{\widehat{W}(u;\boldsymbol{\beta})\}}d\mu_0(te^{(\boldsymbol{\beta}_0-\boldsymbol{\beta})^T\boldsymbol{Z}})\right\} = E\left\{\frac{Y(t;\boldsymbol{\beta})}{\mathbb{P}_n\{\widehat{W}(u;\boldsymbol{\beta})\}}d\mu_0(te^{(\boldsymbol{\beta}_0-\boldsymbol{\beta})^T\boldsymbol{Z}})\right\}.
$$

We show $P\frac{\widehat{W}(u;\boldsymbol{\beta})dM(u;\boldsymbol{\beta})}{\mathbb{P}_n\{\widehat{W}(u;\boldsymbol{\beta})\}} = 0$. Thus the second term of (3.25) is zero.

Using similar arguments as in proofing the consistency of $\widehat{\boldsymbol{\beta}}$, we can show $\{\widehat{W}(u;\widehat{\boldsymbol{\beta}})dM(t;\widehat{\boldsymbol{\beta}}) : \widehat{\boldsymbol{\beta}} \in \mathcal{B}, t \in [0,\tau]\}$, $\{\widehat{W}(t;\widehat{\boldsymbol{\beta}}) : \widehat{\boldsymbol{\beta}} \in \mathcal{B}, t \in [0,\tau]\}$ are P-Donsker and thus P-Glivenko-Cantelli. The first term of (3.25) converges uniformly to zero. Thus $\widehat{\mu}_0(t) \xrightarrow{a.s.} \mu_0(t)$. We have proven the strong consistency of $\widehat{\mu}_0(t)$, $t \in [0,\tau]$.

By addition, subtraction and Taylor expansion (3.26), we have

$$
\begin{aligned}
& \sqrt{n}\left\{\widehat{\mu}_0(t) - \mu_0(t)\right\} \\
= \; & \sqrt{n}(\mathbb{P}_n - P)\int_0^t \frac{\widehat{W}(u;\widehat{\boldsymbol{\beta}})dM(u;\widehat{\boldsymbol{\beta}})}{\mathbb{P}_n\{\widehat{W}(u;\widehat{\boldsymbol{\beta}})\}} + \sqrt{n}P\int_0^t \frac{\widehat{W}(u;\widehat{\boldsymbol{\beta}})dM(u;\widehat{\boldsymbol{\beta}})}{\mathbb{P}_n\{\widehat{W}(u;\widehat{\boldsymbol{\beta}})\}} \\
& - \left\{\mathbb{P}_n\int_0^t \frac{\widehat{W}(u;\widehat{\boldsymbol{\beta}})\boldsymbol{Z}}{\mathbb{P}_n\{\widehat{W}(u;\widehat{\boldsymbol{\beta}})\}} d\left\{\dot{\mu}_0(u)\, u\right\}\right\}^T \sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + o_P(1) \qquad (3.28)
\end{aligned}
$$

Again since $P\left\{\widehat{W}(t;\boldsymbol{\beta})dM(t;\boldsymbol{\beta})\right\} = 0$, the second term of (3.28) is zero. Now look at the last term of (3.28), by similar arguments as in the proof of the consistency of $\widehat{\boldsymbol{\beta}}$, we can replace $\widehat{W}(u;\widehat{\boldsymbol{\beta}})$ with $W(t;\boldsymbol{\beta}_0)$ due to the strong consistency of $\widehat{\boldsymbol{\gamma}}$, $\widehat{\Lambda}(\cdot)$ and $\widehat{\boldsymbol{\beta}}$. Furthermore, $\{W(t;\boldsymbol{\beta}_0) : \boldsymbol{\beta}_0 \in \mathcal{B}, t \in [0,\tau]\}$, $\{W(t;\boldsymbol{\beta}_0)\boldsymbol{Z} : \boldsymbol{\beta} \in \mathcal{B}, t \in [0,\tau]\}$ are P-Donsker and thus P-Glivenko-Cantelli, we can replace the operator $\mathbb{P}_n$ with P in the last term. For the first term of (3.28), $\{\widehat{W}(u;\widehat{\boldsymbol{\beta}})dM(t;\widehat{\boldsymbol{\beta}}) : \widehat{\boldsymbol{\beta}} \in \mathcal{B}, t \in [0,\tau]\}$, $\{\widehat{W}(t;\widehat{\boldsymbol{\beta}}) : \widehat{\boldsymbol{\beta}} \in \mathcal{B}, t \in [0,\tau]\}$ are P-Donsker and thus P-Glivenko-Cantelli. We can replace the operator $\mathbb{P}_n$ with P in the denominator in the first term without altering the limiting distribution. By applying the strong consistency of $\widehat{\boldsymbol{\beta}}$,$\widehat{\boldsymbol{\gamma}}$, $\widehat{\Lambda}(\cdot)$ and Lemma 19.24 of Van der Vaart(1998) to the first term of (3.28) and the result in (3.24) to the last term of (3.28), we have

$$
\begin{aligned}
& \sqrt{n}\left\{\widehat{\mu}_0(t) - \mu_0(t)\right\} \\
= \; & \mathbb{G}_n\int_0^t \frac{W(u;\boldsymbol{\beta}_0)dM(u;\boldsymbol{\beta}_0)}{P\{W(u;\boldsymbol{\beta}_0)\}} + \left\{-\int_0^t \frac{P\left\{W(u;\boldsymbol{\beta}_0)\boldsymbol{Z}\right\}}{P\{W(u;\boldsymbol{\beta}_0)\}} d\left\{\dot{\mu}_0(u)\, u\right\}\right\}^T \sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + o_P(1)
\end{aligned}
$$

$$
\begin{aligned}
&= \mathbb{G}_n \int_0^t \frac{W(u;\boldsymbol{\beta}_0)dM(u;\boldsymbol{\beta}_0)}{P\{W(u;\boldsymbol{\beta}_0)\}} - \left\{-\int_0^t \frac{P\{W(u;\boldsymbol{\beta}_0)\boldsymbol{Z}\}}{P\{W(u;\boldsymbol{\beta}_0)\}} d\{\dot{\mu}_0(u)\,u\}\right\}^T \boldsymbol{A}^{-1}\mathbb{G}_n\boldsymbol{J} + o_P(1) \\
&= \mathbb{G}_n a(t) - \boldsymbol{b}(t)^T\boldsymbol{A}^{-1}\mathbb{G}_n\boldsymbol{J} + o_P(1) \\
&\equiv \mathbb{G}_n\xi(t) + o_P(1),\ t \in [0,\tau]. \qquad (3.29)
\end{aligned}
$$

Therefore the process $\sqrt{n}\,\{\widehat{\mu}_0(t) - \mu_0(t)\}$ converges weakly to a zero-mean Gaussian process with a covariance function $\phi(s,t) = E[\xi(s)\xi(t)]$, $s,t \in [0,\tau]$, where $\xi(t)$ is defined as in (3.9).

Write $\widehat{\mu}_0(t) = \widehat{\mu}_0(t;\widehat{\boldsymbol{\beta}})$, we now show $\widehat{\boldsymbol{b}}(t)$ is a consistent estimator of $\boldsymbol{b}(t)$, as defined in (3.9) and (3.10). From (3.29), for any canonical vector e_i and $h_n \to 0$, uniformly in a $\sqrt{n}$-neighborhood of $\boldsymbol{\beta}_0$,

$$
\begin{aligned}
\sqrt{n}\left\{\widehat{\mu}_0(t;\widehat{\boldsymbol{\beta}}) - \mu_0(t)\right\} &= \mathbb{G}_n a(t) + \boldsymbol{b}(t)^T\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + o_P(1) \\
\sqrt{n}\left\{\widehat{\mu}_0(t;\widehat{\boldsymbol{\beta}} + \boldsymbol{e}_i h_n) - \mu_0(t)\right\} &= \mathbb{G}_n a(t) + \boldsymbol{b}(t)^T\sqrt{n}(\widehat{\boldsymbol{\beta}} + \boldsymbol{e}_i h_n - \boldsymbol{\beta}_0) + o_P(1) \\
\Rightarrow \sqrt{n}\left\{\widehat{\mu}_0(t;\widehat{\boldsymbol{\beta}} + \boldsymbol{e}_i h_n) - \widehat{\mu}_0(t;\widehat{\boldsymbol{\beta}})\right\} &= \sqrt{n}h_n\boldsymbol{b}(t)^T\boldsymbol{e}_i + o_P(1) \\
\Rightarrow \frac{\widehat{\mu}_0(t;\widehat{\boldsymbol{\beta}} + \boldsymbol{e}_i h_n) - \widehat{\mu}_0(t;\widehat{\boldsymbol{\beta}})}{h_n} &= \boldsymbol{b}(t)\,\boldsymbol{e}_i + o_P(\frac{1}{\sqrt{n}h_n}).
\end{aligned}
$$

If we take $h_n = O(n^{-1/2})$, then $o_P(\frac{1}{\sqrt{n}h_n}) = o_P(1)$,

$$
\frac{\widehat{\mu}_0(t;\widehat{\boldsymbol{\beta}} + \boldsymbol{e}_i h_n) - \widehat{\mu}_0(t;\widehat{\boldsymbol{\beta}})}{h_n} = \boldsymbol{b}(t)\,\boldsymbol{e}_i + o_P(1),\ i = 1,\cdots,p.
$$

Thus $\widehat{\boldsymbol{b}}(t) = \left(\frac{\widehat{\mu}_0(t;\widehat{\boldsymbol{\beta}}+\boldsymbol{e}_1h_n)-\widehat{\mu}_0(t;\widehat{\boldsymbol{\beta}})}{h_n},\cdots,\frac{\widehat{\mu}_0(t;\widehat{\boldsymbol{\beta}}+\boldsymbol{e}_ph_n)-\widehat{\mu}_0(t;\widehat{\boldsymbol{\beta}})}{h_n}\right)^T$ converges in probability to $\boldsymbol{b}(t)$.

By replacing $\boldsymbol{b}(t)$ with $\widehat{\boldsymbol{b}}(t)$, $\boldsymbol{\beta}_0$ with $\widehat{\boldsymbol{\beta}}$, $\mu_0(\cdot)$ with $\widehat{\mu}_0(\cdot)$ and expectations with empirical means in the expression of $\phi(s,t)$, we obtain an estimator for the asymptotic covariance. Using similar argument as in proofing the consistency and normality of $\widehat{\boldsymbol{\beta}}$, we can show that $\widehat{\phi}(s,t)$ is a consistent estimator of $\phi(s,t)$.

<u>Proof of Theorem 3.3.3</u>:

$$V_G(t) = n^{-1/2} \sum_{i=1}^{n} \widehat{\xi}_i(t) G_i,$$

where

$$\widehat{\xi}_i(t) \quad = \quad \int_0^t \frac{\widehat{W_i}(u;\widehat{\boldsymbol{\beta}}) d\widehat{M_i}(u;\widehat{\boldsymbol{\beta}})}{Q^{(0)}(t;\widehat{\boldsymbol{\beta}})} - \widehat{\boldsymbol{b}}(t)^T \widehat{\boldsymbol{A}}^{-1} \widehat{\boldsymbol{J}}_i,$$

where $\widehat{\boldsymbol{b}}(t)$, $\widehat{\boldsymbol{A}}$, $\widehat{\boldsymbol{J}}_i$, $\widehat{M_i}(t;\boldsymbol{\beta})$ and $\boldsymbol{Q}^{(0)}(t;\boldsymbol{\beta})$ are defined as in (3.10).

By the Donsker property of $\{\widehat{W}(t;\widehat{\boldsymbol{\beta}}) : \widehat{\boldsymbol{\beta}} \in \mathcal{B}, t \in [0,\tau]\}$, $\{\widehat{W}(t;\widehat{\boldsymbol{\beta}})\boldsymbol{Z} : \widehat{\boldsymbol{\beta}} \in \mathcal{B}, t \in [0,\tau]\}$ $\{\widehat{W}(u;\widehat{\boldsymbol{\beta}}) dM(t;\widehat{\boldsymbol{\beta}}) : \widehat{\boldsymbol{\beta}} \in \mathcal{B}, t \in [0,\tau]\}$, and $\{\widehat{W}(u;\widehat{\boldsymbol{\beta}})\boldsymbol{Z} dM(t;\widehat{\boldsymbol{\beta}}) : \widehat{\boldsymbol{\beta}} \in \mathcal{B}, t \in [0,\tau]\}$, we can show $\widehat{\xi}_i(t)$ belongs to a Donsker class. $Var(G_i) = 1$ since $(G_1, \cdots, G_n)$ are independent standard normal variables. By Theorem 3.6.13 of Van der Vaart and Wellner(1996), conditionally on the observed data,

$$V_G(t) = \mathbb{G}_n \xi(t) + o_P(1).$$

Therefore, the process $V_G(t)$, converges weakly to a zero-mean Gaussian process with a covariance function $\phi(s,t) = E[\xi(s)\xi(t)]$, where $\phi(s,t)$ is defined as in (3.9).

Table 3.1: Summary of simulation results, $Z \sim Bernoulli(0.5)$, $\beta_0 = 0$, $h_n = 3n^{-1/2}$.

		$\beta_0 = 0$			
n	σ^2	Bias	SE	SEE	CP
100	0	0.00	0.214	0.207	0.933
	0.25	-0.02	0.196	0.203	0.945
	0.5	-0.01	0.217	0.204	0.919
	0.75	-0.01	0.211	0.203	0.938
	1	-0.01	0.208	0.200	0.945
200	0	-0.00	0.137	0.143	0.959
	0.25	-0.01	0.145	0.145	0.933
	0.5	-0.02	0.143	0.145	0.947
	0.75	-0.01	0.151	0.143	0.924
	1	-0.00	0.151	0.143	0.934
400	0	-0.00	0.096	0.102	0.960
	0.25	-0.01	0.102	0.102	0.951
	0.5	-0.00	0.104	0.103	0.950
	0.75	-0.01	0.100	0.103	0.946
	1	-0.00	0.100	0.104	0.949

Note: Bias is the sample average of the estimator of β minus β;
SE is the sample standard deviation of the estimator of β ;
SEE is the sample average of the standard error estimator;
CP is the coverage rate of the Wald 95% confidence interval.

Table 3.2: Summary of simulation results, $Z \sim Bernoulli(0.5)$, $\beta_0 = 0.2$, $h_n = 3n^{-1/2}$.

		$\beta_0 = 0.2$			
n	σ^2	Bias	SE	SEE	CP
100	0	0.01	0.204	0.212	0.949
	0.25	0.01	0.214	0.210	0.941
	0.5	-0.01	0.212	0.206	0.931
	0.75	-0.00	0.197	0.204	0.948
	1	0.01	0.200	0.207	0.948
200	0	-0.02	0.143	0.144	0.948
	0.25	-0.00	0.147	0.146	0.952
	0.5	0.01	0.145	0.145	0.940
	0.75	-0.01	0.149	0.144	0.942
	1	0.00	0.143	0.144	0.941
400	0	0.00	0.098	0.102	0.946
	0.25	-0.00	0.101	0.102	0.956
	0.5	-0.00	0.096	0.102	0.958
	0.75	-0.00	0.103	0.102	0.937
	1	-0.00	0.103	0.103	0.959

Note: Bias is the sample average of the estimator of β minus β;
SE is the sample standard deviation of the estimator of β ;
SEE is the sample average of the standard error estimator;
CP is the coverage rate of the Wald 95% confidence interval.

Table 3.3: Summary of simulation results, $Z \sim Bernoulli(0.5)$, $\beta_0 = 0.5$, $h_n = 3n^{-1/2}$.

		$\beta_0 = 0.5$			
n	σ^2	Bias	SE	SEE	CP
100	0	0.03	0.200	0.217	0.982
	0.25	0.04	0.192	0.214	0.973
	0.5	0.01	0.202	0.212	0.971
	0.75	0.03	0.207	0.212	0.951
	1	0.03	0.210	0.210	0.960
200	0	0.04	0.150	0.149	0.947
	0.25	0.02	0.142	0.147	0.960
	0.5	0.02	0.142	0.147	0.947
	0.75	0.01	0.141	0.143	0.947
	1	0.01	0.147	0.145	0.935
400	0	0.01	0.098	0.102	0.964
	0.25	0.01	0.104	0.103	0.950
	0.5	0.00	0.098	0.102	0.942
	0.75	0.01	0.105	0.102	0.934
	1	0.01	0.102	0.102	0.948

Note: Bias is the sample average of the estimator of β minus β;
SE is the sample standard deviation of the estimator of β ;
SEE is the sample average of the standard error estimator;
CP is the coverage rate of the Wald 95% confidence interval.

Table 3.4: Summary of simulation results, $Z \sim Uniform(0, 2)$, $\beta_0 = 0.2$, $h_n = 3n^{-1/2}$.

		$\beta_0 = 0.2$			
n	σ^2	Bias	SE	SEE	CP
100	0	-0.01	0.182	0.187	0.951
	0.25	0.00	0.192	0.188	0.939
	0.5	0.00	0.173	0.181	0.955
	0.75	0.01	0.183	0.178	0.939
	1	-0.01	0.200	0.180	0.939
200	0	0.00	0.121	0.129	0.949
	0.25	-0.00	0.120	0.127	0.955
	0.5	0.00	0.125	0.128	0.940
	0.75	-0.00	0.122	0.129	0.956
	1	-0.00	0.123	0.126	0.951

Note: Bias is the sample average of the estimator of β minus β;
SE is the sample standard deviation of the estimator of β ;
SEE is the sample average of the standard error estimator;
CP is the coverage rate of the Wald 95% confidence interval.

Table 3.5: Summary of simulation results, $Z \sim N(0,1)$, $\beta_0 = 0.2$, $h_n = 3n^{-1/2}$.

		$\beta_0 = 0.2$			
n	σ^2	Bias	SE	SEE	CP
100	0	-0.00	0.106	0.109	0.937
	0.25	0.01	0.104	0.111	0.953
	0.5	0.01	0.096	0.107	0.953
	0.75	-0.00	0.102	0.106	0.945
	1	-0.00	0.094	0.107	0.976
200	0	-0.00	0.071	0.074	0.956
	0.25	0.00	0.075	0.073	0.934
	0.5	0.00	0.070	0.073	0.963
	0.75	0.00	0.069	0.073	0.938
	1	-0.00	0.067	0.072	0.957

Note: Bias is the sample average of the estimator of β minus β;
SE is the sample standard deviation of the estimator of β ;
SEE is the sample average of the standard error estimator;
CP is the coverage rate of the Wald 95% confidence interval.

Table 3.6: SOLVD Treatment Trial: hospitalizations and survival experiences

		Number of hospitalizations							
Treatment	Number of subjects	0	1	2	3	4	5	≥ 6	Number of Deaths
placebo	259	88	70	36	30	10	11	14	96
enalapril	244	97	68	38	19	9	8	5	78

Table 3.7: SOLVD Treatment Trial: Regression analysis for the effects of enalapril on hospitalizations

	Hospitalizations		
Covariate	Estimate	SE	95% Wald Confidence Interval
Treatment	-0.498	0.187	$(-0.864, -0.132)$
Baseline EF	-0.006	0.020	$(-0.033, 0.044)$

Note: Treatment is coded as 1 for enalapril, 0 for placebo;
Estimate is the estimated regression coefficient;
SE is the estimated standard error;
P-value is the two-sided p-value.

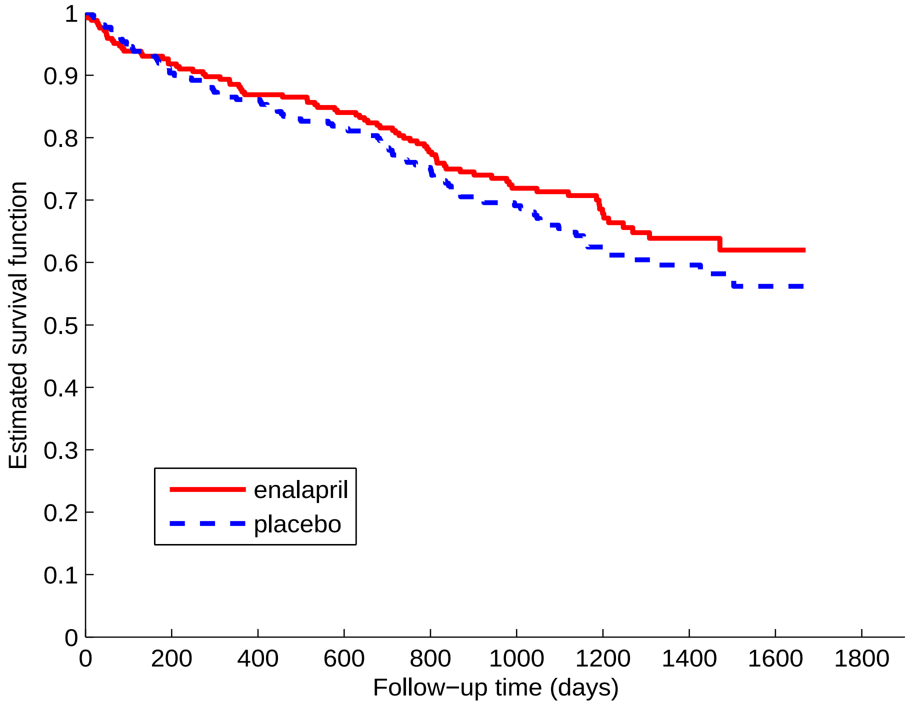

Figure 3.1: SOLVD Treatment Trial Data: Kaplan-Meier survival curves for the placebo group(shown by dashed lines) and the enalapril group (shown by solid lines).

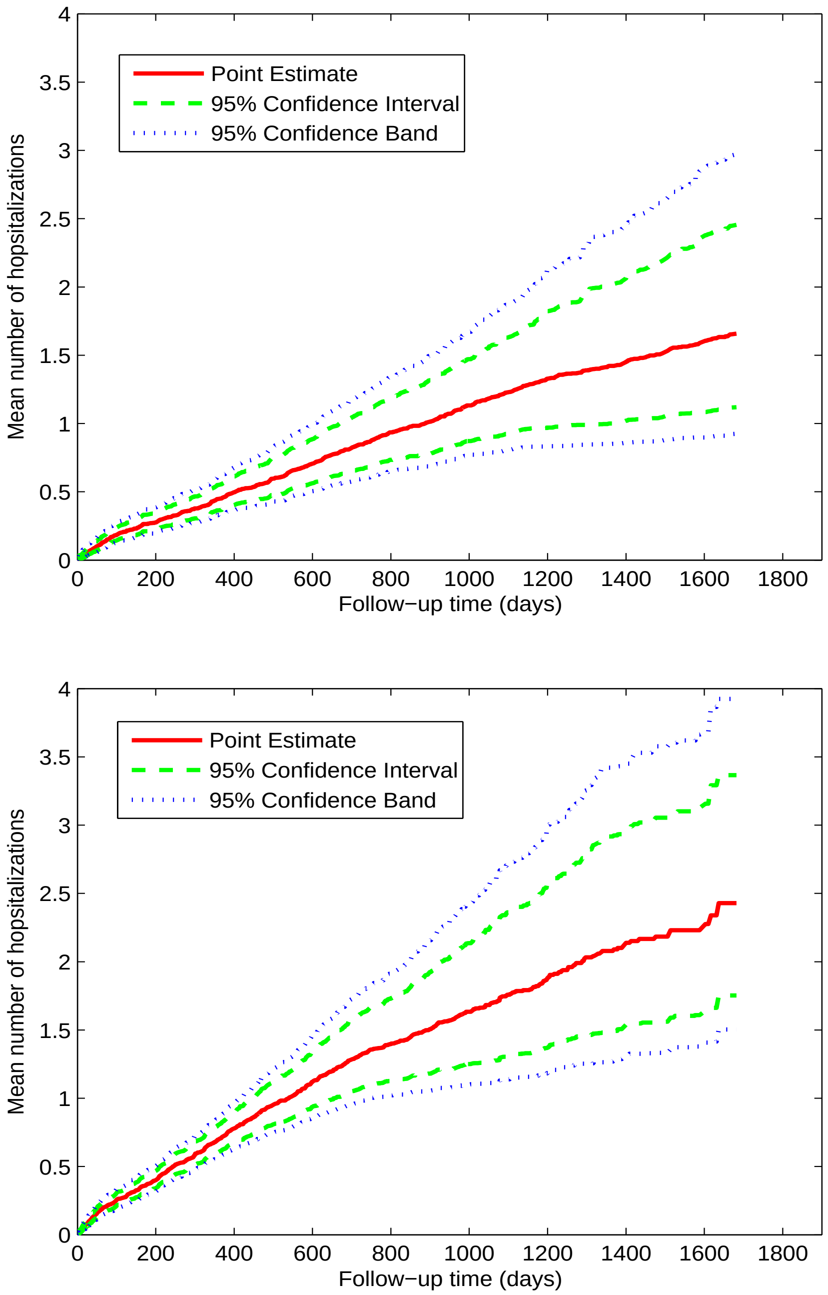

Figure 3.2: SOLVD Treatment Trial Data: Estimated mean frequency of hospitalizations for women with 24.53% baseline ejection fraction (a) receiving enalapril (b) receiving placebo. The confidence bands are based on 1000 simulations.

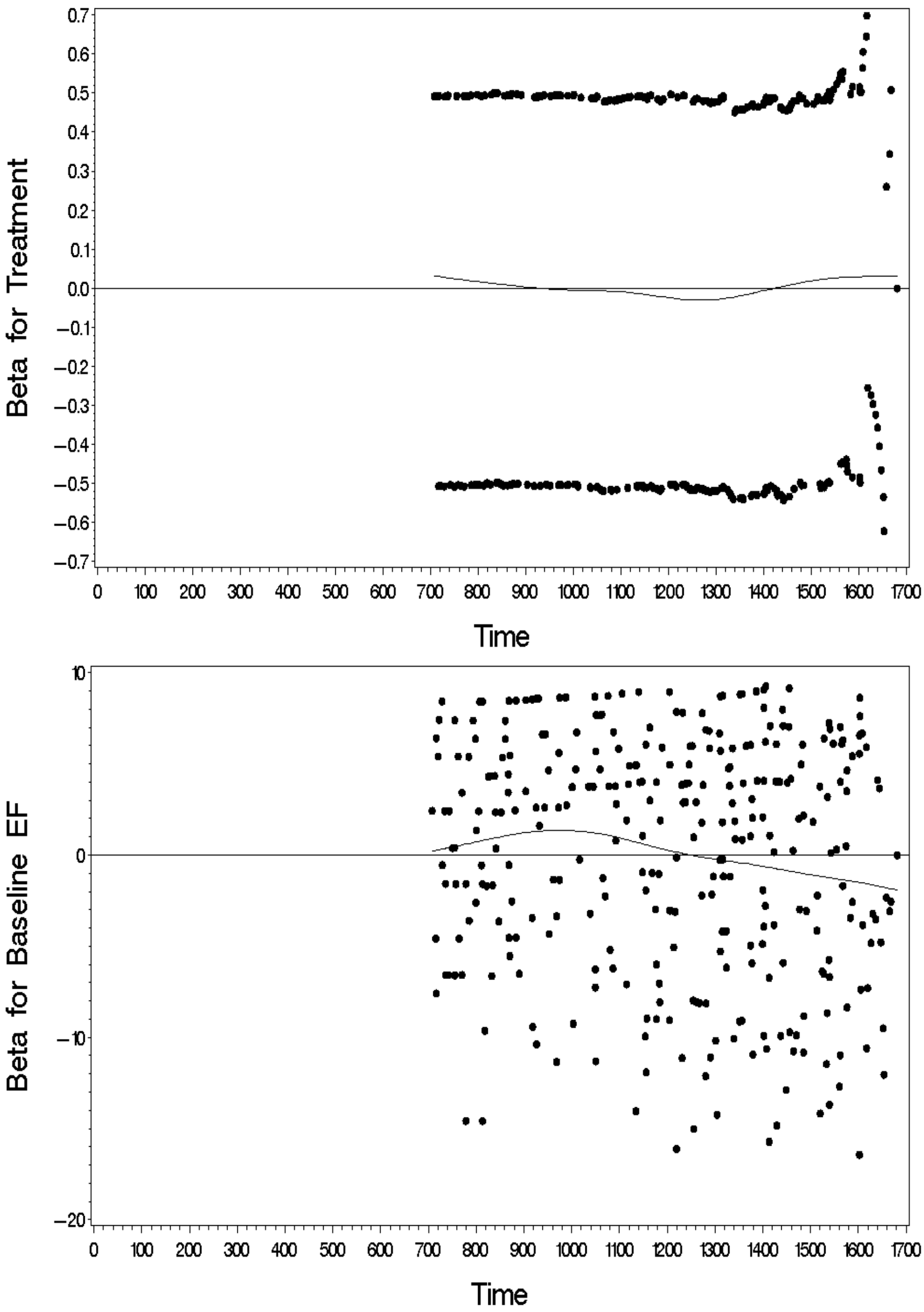

Figure 3.3: Schoenfeld residuals from censoring PH model versus covariates.

Chapter 4

Semiparametric Conditional Rate Model for Recurrent Events with Informative Terminal Event

4.1 Introduction

In many longitudinal follow-up studies, subjects often experience recurrent events. Examples include occurrence of new tumors in bladder cancer patients (Byar(1980)), hospitalizations for transplant candidates with kidney disease (Merion(2003)) and AIDS associated opportunistic diseases in HIV-infected patients (Abrams et al.(1994)). The recurrence of serious events often leads to a terminating event such as death, which precludes occurrence of further recurrent events. In such studies it is frequently of interest to evaluate the effects of a covariate such as treatment, on the basis of both recurrent event process and a terminal event.

In the absence of a terminal event, many authors have proposed methods for analyzing recurrent events under the Cox(1972) proportional hazard model framework, including methods based on inter-event times (Prentice, Williams and Peterson(1981), Chang and Hsiung (1994), Chang and Wang (1999)), methods based on the marginal hazards for individual recurrences (Wei, Lin and Weissfeld(1989)) and methods based on the intensity/rate function of the recurrent event process (Andersen and Gill(1982), Pepe and Cai(1993), Lawless and Nadeau(1995), Lawless, Nadeau and Cook(1997), Lin, Wei, Yang and Ying(2000)). All these methods take on the assumption that the terminal event is non-informative.

Methodologies to address analysis of data involving recurrent events with a terminal event have been studied in recent years. There are complete intensity approaches which involve jointly modelling both recurrent events and terminal event. Wang, Qing and Chiang(2001), Huang and Wang(2004), Liu, Wolfe and Huang(2004) and Ye, Kalbfleisch and Schaubel(2007) studied shared frailty models that assumed the proportional intensity model and proportional hazards model for the recurrent events and the terminal event, respectively. In their approaches, a common latent variable was used to model the association between the intensity of the recurrent event process and the hazard of the failure time.

Compared to the intensity models, rate models are attractive alternatives due to the easy interpretation of regression coefficients. Cook and Lawless(1997) studied mean and rate models for recurrent process conditional on survival at specific time points. Ghosh and Lin(2000) proposed a nonparametric estimator for the rate function of the recurrent events. Ghosh and Lin(2002) further considered a proportional rate model and obtained consistent estimators for the regression coefficients from an inverse probability weighted estimating equation. Miloslavsky et al(2004) also studied the proportional rate model and adopted an modified version of the estimating equation of Ghosh and Lin(2002). In these models, the dependence between the recurrent event process and the terminal event is unspecified.

In this chapter, we propose a proportional rate model for the recurrent event given the subjects are still alive. Our approach does not require specification of the dependence structure between the recurrent events and the terminal event. We consider two estimating procedures for the regression coefficients and the mean function of recurrent events. The remainder of this article is organized as follows. In Section 2, we introduce notations and present the proposed methods. In Section 3, we describe the asymptotic properties of the resulting estimators and discuss the inference procedures on the mean function of recurrent events. In Section 4, we conduct simulation studies to evaluate the proposed methods in finite samples and report the results. In Section 5, we illustrate the method by applying it to the SOLVD(Studies of Left Ventricular Dysfunction, SOLVD Investigators, 1991) Prevention Trial data. We conclude with some discussion in Section 6. Proofs of the theorems can be found in Section 7.

4.2 Model and estimation

Let $N^*(t)$ be the cumulative number of recurrent events that occur over the time interval $[0, t]$ in the absence of any censoring. In most applications, the follow-up time is subject to censoring. Let C and D denote, respectively, the censoring time and the survival time, and let $\boldsymbol{Z}(.)$ be a $p \times 1$-vector of covariates. Throughout the chapter, we assume that C may depend on $\boldsymbol{Z}$ but is independent of $\{N^*(.)\}$ and D given $\boldsymbol{Z}$. The survival time D is allowed to depend on $\{N^*(.)\}$, even conditionally on $\boldsymbol{Z}$. In general only the minimum of C and D is known and $\{N^*(.)\}$ can only be observed up to $\min(C, D)$. Further, suppose the study duration is τ and define $X = D \wedge C = \min(D, C)$, $N(t) = N^*(t \wedge D \wedge C) = I(X \geq t)\, N^*(t)$, $Y(t) = I(X \geq t)$ and $\delta = I(D \leq C)$. For a random sample of n subjects, the observed data consist of $\{N_i(.), X_i, \delta_i, \boldsymbol{Z}_i\}$, $i = 1, \cdots, n$. We wish to formulate the effect of $\boldsymbol{Z}$ on $\{N^*(.)\}$ without specifying the dependency structure between recurrent events and terminal events, or among recurrent events.

For the underlying recurrent process, we propose the following multiplicative conditional rate model when terminating events are present,

$$E\{dN^*(t)|D \geq t, \boldsymbol{Z}\} = I(D \geq t)e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}(t)} d\mu_0(t), \tag{4.1}$$

where $\boldsymbol{Z}$ is a $p \times 1$ vector of covariates, $\boldsymbol{\beta}_0$ is a $p \times 1$ vector of unknown regression parameters, and $d\mu_0(.)$ is an unspecified continuous function.

The corresponding model for the observed recurrent process is

$$E\{dN(t)|D \geq t, C, \boldsymbol{Z}\} = Y(t)e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}(t)} d\mu_0(t). \tag{4.2}$$

To estimate $\boldsymbol{\beta}$, we consider the following estimating equation (Method A),

$$\boldsymbol{U}_n(\boldsymbol{\beta}) = \sum_{i=1}^{n} \int_0^{\tau} Y_i(t) \left\{ \boldsymbol{Z}_i(t) - \frac{\sum_{i=1}^{n} Y_i(t) \boldsymbol{Z}_i(t) e^{\boldsymbol{\beta}^T \boldsymbol{Z}_i(t)}}{\sum_{i=1}^{n} Y_i(t) e^{\boldsymbol{\beta}^T \boldsymbol{Z}_i(t)}} \right\} dN_i(t). \tag{4.3}$$

The estimator of $\boldsymbol{\beta}_0$ is defined to be $\widehat{\boldsymbol{\beta}}$, the solution to $\boldsymbol{U}_n(\boldsymbol{\beta}) = \boldsymbol{0}$. It can be obtained using the Newton-Raphson method.

Given $\widehat{\boldsymbol{\beta}}$, we estimate $\mu_0(t)$ by the Nelson-Aalen-type estimator of $\widehat{\mu}_0(t)$, where

$$\widehat{\mu}_0(t) = \sum_{i=1}^{n} \int_0^{t} \frac{dN_i(u)}{\sum_{i=1}^{n} Y_i(u) e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}_i(u)}}, \ t \in [0, \tau]. \tag{4.4}$$

In most clinical studies, however, the survival time is of key interest, while the recurrent event process is affected by the survival distribution. We want to characterize subjects' entire clinical experience by making inference from both the survival distribution as well as the recurrent process. We consider the the inverse probability of survival weighting(IPSW) approach(Ghosh and Lin, 2002). This weight function takes the form

$$W_i^D(t) = \frac{Y_i(t)}{S_D(t|\boldsymbol{Z}_i)},$$

where $S_D(t|\boldsymbol{Z}) = P(D > t|\boldsymbol{Z})$.

As we allow D depend on $\boldsymbol{Z}$, it is convenient to formulate $S_D(t|\boldsymbol{Z})$ through the Cox proportional hazards model,

$$\lambda^D(t|\boldsymbol{Z}) = \lambda_0^D(t)e^{\gamma_D^T \boldsymbol{Z}(t)}, \tag{4.5}$$

where $\lambda_0^D(.)$ is an unspecified baseline hazard function, γ_D is a $p \times 1$ vector of unknown regression parameter, and $\lambda^D(t|\boldsymbol{Z})$ is the hazard function corresponding to $S_D(t|\boldsymbol{Z})$. Thus $S_D(t|\boldsymbol{Z})$ can be estimated by

$$\widehat{S}_D(t|\boldsymbol{Z}) = \exp\{-\int_0^t e^{\widehat{\gamma}_D^T \boldsymbol{Z}(u)} d\widehat{\Lambda}_0^D(u)\},$$

where $\widehat{\gamma}_D$ and $\widehat{\Lambda}_0^D(.)$ are the maximum partial likelihood (Cox, 1975) and Breslow estimators of γ_D and $\Lambda_0^D(t) = \int_0^t \lambda_0^D(u)du$, respectively.

Thus $W_i^D(t)$ can be estimated by

$$\widehat{W}_i^D(t) = \frac{Y_i(t)}{\widehat{S}_D(t|\boldsymbol{Z}_i)}.$$

Based on model (4.1), we propose the following estimating function for $\boldsymbol{\beta}_0$ (Method B),

$$\boldsymbol{U}_n^D(\boldsymbol{\beta}) = \sum_{i=1}^n \int_0^\tau \widehat{W}_i^D(t) \left\{ \boldsymbol{Z}_i(t) - \frac{\sum_{i=1}^n \widehat{W}_i^D(t)\boldsymbol{Z}_i(t)e^{\boldsymbol{\beta}^T \boldsymbol{Z}_i(t)}}{\sum_{i=1}^n \widehat{W}_i^D(t)e^{\boldsymbol{\beta}^T \boldsymbol{Z}_i(t)}} \right\} dN_i(t). \tag{4.6}$$

The estimator of $\boldsymbol{\beta}_0$ is defined to be $\widehat{\boldsymbol{\beta}}_D$, the solution to $\boldsymbol{U}_n^D(\boldsymbol{\beta}) = \mathbf{0}$. It can be obtained using the Newton-Raphson method.

Given $\widehat{\boldsymbol{\beta}}_D$, we estimate $\mu_0(t)$ by the Nelson-Aalen-type estimator of $\widehat{\mu}_0^D(t)$, where

$$\widehat{\mu}_0^D(t) = \sum_{i=1}^n \int_0^t \frac{\widehat{W}_i^D(u)dN_i(u)}{\sum_{i=1}^n \widehat{W}_i^D(u)e^{\widehat{\boldsymbol{\beta}}_D^T \boldsymbol{Z}_i(u)}},\ t \in [0, \tau]. \tag{4.7}$$

4.3 Asymptotic properties

We consider the following assumptions,

(C1) $\{N_i, X_i, \boldsymbol{Z}_i\}$ $(i = 1, \cdots, n)$ are independent and identically distributed.

(C2) $P(C_i > \tau | \boldsymbol{Z}_i) > 0, i = 1, \cdots, n$, where τ is the study duration.

(C3) $E[N_i(\tau)] < \infty$, $i = 1, \cdots, n,$.

(C4) Covariate $\boldsymbol{Z}_i(t), i = 1, \cdots, n$, are bounded and has finite total variations in $[0, \tau]$.

(C5) Matrices $\boldsymbol{A}$ and $\boldsymbol{A}_D$ are non-singular, where

$$\begin{aligned}
\boldsymbol{A} &= E\left\{\int_0^\tau Y(t)\left[\left\{\frac{E\{Y(t)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}{E\{Y(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}\right\}^{\otimes 2} - \frac{E\{Y(t)\boldsymbol{Z}(t)^{\otimes 2}e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}{E\{Y(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}\right]dN(t)\right\}, \\
\boldsymbol{A}_D &= E\left\{\int_0^\tau W^D(t)\left[\left\{\frac{E\{W^D(t)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}{E\{W^D(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}\right\}^{\otimes 2} - \frac{E\{W^D(t)\boldsymbol{Z}(t)^{\otimes 2}e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}{E\{W^D(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}\right]dN(t)\right\},
\end{aligned}$$

$\boldsymbol{a}^{\otimes 0} = 1$, $\boldsymbol{a}^{\otimes 1} = \boldsymbol{a}$, $\boldsymbol{a}^{\otimes 2} = \boldsymbol{a}\boldsymbol{a}^T$, where $\boldsymbol{a}$ is a column vector.

Condition (C2) implies the weight function $W^D(t)$ will be uniformly bounded away from zero. Conditions (C1), (C3), (C4) and (C5) are standard.

4.3.1 Asymptotic properties of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\beta}}_D$

Theorem 4.3.1 *Under regularity conditions (C1)-(C5), the parameter estimate $\widehat{\boldsymbol{\beta}}$ is strongly consistent for $\boldsymbol{\beta}_0$, i.e. $\widehat{\boldsymbol{\beta}} \xrightarrow{a.s.} \boldsymbol{\beta}_0$. The random vector $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ converges in distribution to a zero-mean normal distribution with a covariance matrix $\boldsymbol{A}^{-1}\boldsymbol{\Sigma}(\boldsymbol{A}^{-1})^T$, where*

$$\boldsymbol{A} \equiv \left.\frac{\partial \boldsymbol{U}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} = E\left\{\int_0^\tau Y(t)\left[\left\{\frac{E\{Y(t)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}{E\{Y(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}\right\}^{\otimes 2} - \frac{E\{Y(t)\boldsymbol{Z}(t)^{\otimes 2}e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}{E\{Y(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}\right]dN(t)\right\},$$

$$\begin{aligned}
\boldsymbol{\Sigma} &= E[\boldsymbol{J}\boldsymbol{J}^T], \\
\boldsymbol{J} &= \int_0^\tau Y(t)\left[\boldsymbol{Z}(t) - \frac{E\{Y(t)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}{E\{Y(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}\right] dM_{\boldsymbol{\beta}_0}(t), \\
M_{\boldsymbol{\beta}_0}(t) &= N(t) - \int_0^t Y(s)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(s)}d\mu_0(s). \qquad (4.8)
\end{aligned}$$

A consistent estimator for $\boldsymbol{A}^{-1}\boldsymbol{\Sigma}(\boldsymbol{A}^{-1})^T$ *is* $\widehat{\boldsymbol{A}}^{-1}\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{A}}^{-1})^T$*, where*

$$\begin{aligned}
\widehat{\boldsymbol{A}} &= n^{-1}\sum_{i=1}^n \left\{\int_0^\tau Y_i(t)\left[\left\{\frac{\boldsymbol{Q}^{(1)}(t;\widehat{\boldsymbol{\beta}})}{Q^{(0)}(t;\widehat{\boldsymbol{\beta}})}\right\}^{\otimes 2} - \frac{\boldsymbol{Q}^{(2)}(t;\widehat{\boldsymbol{\beta}})}{Q^{(0)}(t;\widehat{\boldsymbol{\beta}})}\right] dN_i(t)\right\}, \\
\widehat{\boldsymbol{\Sigma}} &= n^{-1}\sum_{i=1}^n \widehat{\boldsymbol{J}}_i\widehat{\boldsymbol{J}}_i^T, \\
\widehat{\boldsymbol{J}}_i &= \int_0^\tau Y_i(t)\left[\boldsymbol{Z}_i(t) - \frac{\boldsymbol{Q}^{(1)}(t;\widehat{\boldsymbol{\beta}})}{Q^{(0)}(t;\widehat{\boldsymbol{\beta}})}\right] d\widehat{M}_i(t), \\
\boldsymbol{Q}^{(k)}(t;\boldsymbol{\beta}) &= n^{-1}\sum_{i=1}^n Y_i(t)\boldsymbol{Z}_i(t)^{\otimes k}e^{\boldsymbol{\beta}^T\boldsymbol{Z}_i(t)}, k = 0, 1, 2, \\
\widehat{M}_i(t) &= N_i(t) - \int_0^t Y_i(s)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}_i(s)}d\widehat{\mu}_0(s). \qquad (4.9)
\end{aligned}$$

Theorem 4.3.2 *Under regularity conditions (C1)-(C5), the parameter estimate* $\widehat{\boldsymbol{\beta}}_D$ *is strongly consistent for* $\boldsymbol{\beta}_0$*, i.e.* $\widehat{\boldsymbol{\beta}}_D \xrightarrow{a.s.} \boldsymbol{\beta}_0$*. The random vector* $\sqrt{n}(\widehat{\boldsymbol{\beta}}_D - \boldsymbol{\beta}_0)$ *converges in distribution to a zero-mean normal distribution with a covariance matrix* $\boldsymbol{A}_D^{-1}\boldsymbol{\Sigma}_D(\boldsymbol{A}_D^{-1})^T$*, where*

$$\begin{aligned}
\boldsymbol{A}_D &\equiv \left.\frac{\partial \boldsymbol{U}^D(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \\
&= E\left\{\int_0^\tau W^D(t)\left[\left\{\frac{E\{W^D(t)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}{E\{W^D(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}\right\}^{\otimes 2} - \frac{E\{W^D(t)\boldsymbol{Z}(t)^{\otimes 2}e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}{E\{W^D(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}\right] dN(t)\right\}, \\
\boldsymbol{\Sigma}_D &= E[\boldsymbol{J}_D\boldsymbol{J}_D^T], \\
\boldsymbol{J}_D &= \int_0^\tau W^D(t)\left[\boldsymbol{Z}(t) - \frac{E\{W^D(t)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}{E\{W^D(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}\right] dM_{\boldsymbol{\beta}_0}(t), \\
M_{\boldsymbol{\beta}_0}(t) &= N(t) - \int_0^t Y(s)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(s)}d\mu_0(s). \qquad (4.10)
\end{aligned}$$

A consistent estimator for $\boldsymbol{A}_D^{-1}\boldsymbol{\Sigma}_D(\boldsymbol{A}_D^{-1})^T$ *is* $\widehat{\boldsymbol{A}}_D^{-1}\widehat{\boldsymbol{\Sigma}}_D(\widehat{\boldsymbol{A}}_D^{-1})^T$*, where*

$$\widehat{\boldsymbol{A}}_D = n^{-1}\sum_{i=1}^{n}\left\{\int_0^\tau \widehat{W}_i^D(t)\left[\left\{\frac{\boldsymbol{Q}_D^{(1)}(t;\widehat{\boldsymbol{\beta}}_D)}{Q_D^{(0)}(t;\widehat{\boldsymbol{\beta}}_D)}\right\}^{\otimes 2} - \frac{\boldsymbol{Q}_D^{(2)}(t;\widehat{\boldsymbol{\beta}}_D)}{Q_D^{(0)}(t;\widehat{\boldsymbol{\beta}}_D)}\right] dN_i(t)\right\},$$

$$\begin{aligned}
\widehat{\boldsymbol{\Sigma}}_D &= n^{-1}\sum_{i=1}^{n}\widehat{\boldsymbol{J}}_{D,i}\widehat{\boldsymbol{J}}_{D,i}^T,\\
\widehat{\boldsymbol{J}}_{D,i} &= \int_0^\tau \widehat{W}_i^D(t)\left[\boldsymbol{Z}_i(t) - \frac{\boldsymbol{Q}_D^{(1)}(t;\widehat{\boldsymbol{\beta}}_D)}{Q_D^{(0)}(t;\widehat{\boldsymbol{\beta}}_D)}\right] d\widehat{M}_i^D(t),\\
\boldsymbol{Q}_D^{(k)}(t;\boldsymbol{\beta}) &= n^{-1}\sum_{i=1}^{n}\widehat{W}_i^D(t)\boldsymbol{Z}_i(t)^{\otimes k}e^{\boldsymbol{\beta}^T\boldsymbol{Z}_i(t)}, k=0,1,2,\\
\widehat{M}_i^D(t) &= N_i(t) - \int_0^t Y_i(s)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}_i(s)}d\widehat{\mu}_0^D(s). \qquad (4.11)
\end{aligned}$$

4.3.2 Inference on the mean function

The following theorems describe the asymptotic properties of $\widehat{\mu}_0(t)$ and $\widehat{\mu}_0^D(t)$.

Theorem 4.3.3 *Under the regularity conditions (C1)-(C5),* $\widehat{\mu}_0(t)$ *is strongly consistent for* $\mu_0(t)$ *uniformly in* t*, i.e.* $\widehat{\mu}_0(t) \xrightarrow{a.s.} \mu_0(t)$*,* $t \in [0,\tau]$*. The process* $n^{1/2}\{\widehat{\mu}_0(t) - \mu_0(t)\}$*,* $t \in [0,\tau]$*, converges weakly to a zero-mean Gaussian process with a covariance function* $\phi(s,t) = E[K(s)K(t)]$*, where*

$$\begin{aligned}
&K(t) \qquad (4.12)\\
&= \int_0^t \frac{dM_{\boldsymbol{\beta}_0}(u)}{E\{Y(u)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(u)}\}} + \left\{\int_0^t \frac{E\left\{Y(u)\boldsymbol{Z}(u)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(u)}\right\}}{E\{Y(u)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(u)}\}}d\mu_0(u)\right\}^T \boldsymbol{A}^{-1}\boldsymbol{J},\ t \in [0,\tau],
\end{aligned}$$

where $\boldsymbol{A}$*,*$\boldsymbol{J}$ *and* $M_{\boldsymbol{\beta}_0}(t)$ *are defined as in (4.8). A consistent estimator for* $\phi(s,t)$ *is*

$$\widehat{\phi}(s,t) = n^{-1}\sum_{i=1}^{n}\widehat{K}_i(s)\widehat{K}_i(t),$$

where

$$\widehat{K}_i(t) = \int_0^t \frac{d\widehat{M}_i(u)}{Q^{(0)}(t;\widehat{\boldsymbol{\beta}})} + \left\{ \int_0^t \frac{\boldsymbol{Q}^{(1)}(u;\widehat{\boldsymbol{\beta}})}{Q^{(0)}(u;\widehat{\boldsymbol{\beta}})} d\widehat{\mu}_0(u) \right\}^T \widehat{\boldsymbol{A}}^{-1} \widehat{\boldsymbol{J}}_i, \qquad (4.13)$$

where $\widehat{\boldsymbol{A}}$*,* $\widehat{\boldsymbol{J}}_i$*,* $\widehat{M}_i(t)$ *and* $\boldsymbol{Q}^{(k)}(t;\boldsymbol{\beta})$*,* $k = 0, 1$ *are defined as in (4.9).*

Theorem 4.3.4 *Under the regularity conditions (C1)-(C5),* $\widehat{\mu}_0^D(t)$ *is strongly consistent for* $\mu_0(t)$ *uniformly in* t*, i.e.* $\widehat{\mu}_0^D(t) \xrightarrow{a.s.} \mu_0(t)$*,* $t \in [0, \tau]$*. The process* $n^{1/2}\left\{\widehat{\mu}_0^D(t) - \mu_0(t)\right\}$*,* $t \in [0, \tau]$*, converges weakly to a zero-mean Gaussian process with a covariance function* $\phi_D(s,t) = E[K_D(s)K_D(t)]$*, where*

$$\begin{aligned} &K_D(t) &(4.14)\\ &= \int_0^t \frac{W^D(u)dM_{\boldsymbol{\beta}_0}(u)}{E\{W^D(u)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(u)}\}} + \left\{ \int_0^t \frac{E\left\{W^D(u)\boldsymbol{Z}(u)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(u)}\right\}}{E\{W^D(u)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(u)}\}} d\mu_0(u) \right\}^T \boldsymbol{A}_D^{-1}\boldsymbol{J}_D, \ t \in [0,\tau], \end{aligned}$$

where $\boldsymbol{A}_D$*,*$\boldsymbol{J}_D$ *and* $M_{\boldsymbol{\beta}_0}(t)$ *are defined as in (4.10). A consistent estimator for* $\phi_D(s,t)$ *is*

$$\widehat{\phi}_D(s,t) = n^{-1}\sum_{i=1}^n \widehat{K}_{D,i}(s)\widehat{K}_{D,i}(t),$$

where

$$\widehat{K}_{D,i}(t) = \int_0^t \frac{\widehat{W}_i^D(u)d\widehat{M}_i^D(u)}{Q_D^{(0)}(t;\widehat{\boldsymbol{\beta}}_D)} + \left\{ \int_0^t \frac{\boldsymbol{Q}_D^{(1)}(u;\widehat{\boldsymbol{\beta}}_D)}{Q_D^{(0)}(u;\widehat{\boldsymbol{\beta}}_D)} d\widehat{\mu}_0^D(u) \right\}^T \widehat{\boldsymbol{A}}_D^{-1}\widehat{\boldsymbol{J}}_{D,i}, \quad (4.15)$$

where $\widehat{\boldsymbol{A}}_D$*,* $\widehat{\boldsymbol{J}}_{D,i}$*,* $\widehat{M}_i^D(t)$ *and* $\boldsymbol{Q}_D^{(k)}(t;\boldsymbol{\beta})$*,* $k = 0, 1$ *are defined as in (4.11).*

Based on the asymptotic results in Theorem 4.3.3, we can construct confidence intervals for $\mu_0(t)$. We consider the transformed process $n^{1/2}\{\log\{\widehat{\mu}_0(t)\} - \log\{\mu_0(t)\}\}$ since $\mu_0(t)$ is non-negative. By the delta method, the process $n^{1/2}\{\log\{\widehat{\mu}_0(t)\} - \log\{\mu_0(t)\}\}$ converges weakly to a zero-mean Gaussian process with a covariance function $\phi(s,t)/\mu_0^2(t)$,

$s, t \in [0, \tau]$. Therefore the 95% pointwise confidence interval for $\log\{\mu_0(t)\}$ is

$$\log\{\widehat{\mu}_0(t)\} \pm 1.96\, n^{-1/2} \frac{\widehat{\phi}(t,t)^{1/2}}{\widehat{\mu}_0(t)}.$$

Exponentiating the above interval, we obtain the 95% pointwise confidence interval for $\mu_0(t)$

$$\widehat{\mu}_0(t) \exp\left\{ \pm 1.96\, n^{-1/2} \frac{\widehat{\phi}(t,t)^{1/2}}{\widehat{\mu}_0(t)} \right\}. \tag{4.16}$$

To construct simultaneous confidence bands for $\mu_0(t)$ over a time interval $[\tau_1, \tau_2]$, $0 < \tau_1 < \tau_2 \leq \tau$, we use a Monte-Carlo method. Specifically, let

$$V_G(t) = n^{-1/2} \sum_{i=1}^{n} \widehat{K}_i(t) G_i,$$

where $(G_1, \cdots, G_n)$ are independent standard normal variables. The following theorem states that $V(t)$ and $V_G(t)$ have the same limiting distribution.

Theorem 4.3.5 *Under the regularity conditions (C1)-(C5), conditional on the observed data, the process $V_G(t)$, converges weakly to a zero-mean Gaussian process with a covariance function $\phi(s,t) = E[K(s)K(t)]$, where $\phi(s,t)$ is defined as in (4.12).*

Given a fixed set of observed data, $\widehat{K}_i(t), i = 1, \cdots, n$ are also fixed. We generate $(G_1, \cdots, G_n)$ independently from the standard normal distribution and calculate $V_G(t)$. Repeat this process B times, where B is sufficiently large, we obtain a collection of $V_G(t)$ which simulates the limiting distribution of $V(t)$. The 95% simultaneous confidence band for $\mu_0(t)$ over a time interval $[\tau_1, \tau_2]$, $0 < \tau_1 < \tau_2 \leq \tau$, is

$$\widehat{\mu}_0(t) \exp\left\{ \pm c_\alpha\, n^{-1/2} \frac{\widehat{\phi}(t,t)^{1/2}}{\widehat{\mu}_0(t)} \right\}, \tag{4.17}$$

where c_α is the estimated 95th percentile of $\sup_{\tau 1\le t\le\tau_2}\left|\frac{V_G(t)}{\widehat{\phi}(t,t)^{1/2}}\right|$.

Similarly, we can construct confidence intervals/bands for $\mu_0(t)$ by the asymptotic normality of $\widehat{\mu}_0^D(t)$ and the consistency of variance estimator $\widehat{\phi}_D(s,t) = n^{-1}\sum_{i=1}^n \widehat{K}_{D,i}(s)\widehat{K}_{D,i}(t)$ as defined in (4.14). With the log-transformation, the 95% pointwise confidence interval for $\mu_0(t)$ is

$$\widehat{\mu}_0^D(t)\exp\left\{\pm 1.96\, n^{-1/2}\,\frac{\widehat{\phi}_D(t,t)^{1/2}}{\widehat{\mu}_0^D(t)}\right\}. \tag{4.18}$$

Let

$$V_G^D(t) = n^{-1/2}\sum_{i=1}^n \widehat{K}_{D,i}(t)G_i,$$

where $(G_1,\cdots,G_n)$ are independent standard normal variables. The following theorem states that $V(t)$ and $V_G^D(t)$ have the same limiting distribution.

Theorem 4.3.6 *Under the regularity conditions (C1)-(C5), conditional on the observed data, the process $V_G^D(t)$, converges weakly to a zero-mean Gaussian process with a covariance function $\phi_D(s,t) = E[K_D(s)K_D(t)]$, where $\phi_D(s,t)$ is defined as in (4.14).*

To approximate the distribution of $V(t)$, we obtain a collection of $V_G^D(t)$ by repeatedly generating $(G_1,\cdots,G_n)$ while fixing the observed data for a large number of times. The 95% simultaneous confidence band for $\mu_0(t)$ over a time interval $[\tau_1,\tau_2]$, $0<\tau_1<\tau_2\le\tau$, is

$$\widehat{\mu}_0^D(t)\exp\left\{\pm c_\alpha^*\, n^{-1/2}\,\frac{\widehat{\phi}_D(t,t)^{1/2}}{\widehat{\mu}_0^D(t)}\right\}, \tag{4.19}$$

where c_α^* is the estimated 95th percentile of $\sup_{\tau 1\le t\le\tau_2}\left|\frac{V_G^D(t)}{\widehat{\phi}_D(t,t)^{1/2}}\right|$.

When the covariates are centered at a given $\boldsymbol{z}$, we have $E\{dN^*(t)|D\ge t,\boldsymbol{Z}\} = I(D\ge t)e^{\boldsymbol{\beta}_0^T(\boldsymbol{Z}-\boldsymbol{z})}d\mu_0(t)$. Therefore $E\{dN^*(t)|D\ge t,\boldsymbol{z}\} = I(D\ge t)e^{\boldsymbol{\beta}_0^T(\boldsymbol{z}-\boldsymbol{z})}d\mu_0(t) = d\mu_0(t)$. In applications, if we want to estimate the mean function for a given set of covariate $\boldsymbol{z}$, we can center the covariates at $\boldsymbol{z}$ in the data set and obtain point and interval estimates for $E\{N^*(t)|D \geq t, \boldsymbol{z}\}$ by using the formulae (4.4),(4.7), (4.16), (4.17), (4.18) and (4.19).

4.4 Simulation studies

A series of numerical simulation studies are conducted to evaluate the performance of the proposed estimator in the finite sample situation. We focus on the estimation of $\boldsymbol{\beta}_0$.

We first generate the death time D from the Cox proportional hazard model $\lambda(t|\boldsymbol{Z}) = \lambda_0(t)e^{\boldsymbol{\gamma}^T\boldsymbol{Z}}$. The density function of the death time is $f_D(t) = \lambda_0(t)e^{\boldsymbol{\gamma}^T\boldsymbol{Z}}e^{-\Lambda_0(t)e^{\boldsymbol{\gamma}^T\boldsymbol{Z}}}$, and $P(D \geq t|\boldsymbol{Z}) = e^{-\Lambda_0(t)e^{\boldsymbol{\gamma}^T\boldsymbol{Z}}}$.

Given D and $\boldsymbol{Z}$, we generate the recurrent events using the following intensity model:

$$E\{dN^*(t)|N^*(t-), D, \boldsymbol{Z}, \nu\} = \nu g(D, \boldsymbol{Z})h(t)dt,$$

where ν is a gamma variable with mean 1 and variance σ^2 independent of D and $\boldsymbol{Z}$. We can control the correlations among the recurrent events by varying the value of σ^2. The above intensity model also implies

$$E\{dN^*(t)|D, \boldsymbol{Z}\} = g(D, \boldsymbol{Z})h(t)dt.$$

Next we determine the form of $g(x, \boldsymbol{Z})$.

$$\begin{aligned} E\{dN^*(t)|D \geq t, \boldsymbol{Z}\} &= \frac{E\{dN^*(t)I(D \geq t)|\boldsymbol{Z}\}}{E\{I(D \geq t)|\boldsymbol{Z}\}} = \frac{E\{E[dN^*(t)I(D \geq t)|D, \boldsymbol{Z}]\}}{E\{I(D \geq t)|\boldsymbol{Z}\}} \\ &= \frac{E\{I(D \geq t)E[dN^*(t)|D, \boldsymbol{Z}]\}}{P(D \geq t|\boldsymbol{Z})} = \frac{E\{I(D \geq t)g(D, \boldsymbol{Z})h(t)dt\}}{e^{-\Lambda_0(t)e^{\boldsymbol{\gamma}^T\boldsymbol{Z}}}} \\ &= \frac{\int_t^\infty g(x, \boldsymbol{Z})h(t)f_D(x)dx}{e^{-\Lambda_0(t)e^{\boldsymbol{\gamma}^T\boldsymbol{Z}}}} \\ &= \frac{\int_t^\infty g(x, \boldsymbol{Z})h(t)\lambda_0(x)e^{\boldsymbol{\gamma}^T\boldsymbol{Z}}e^{-\Lambda_0(x)e^{\boldsymbol{\gamma}^T\boldsymbol{Z}}}dx}{e^{-\Lambda_0(t)e^{\boldsymbol{\gamma}^T\boldsymbol{Z}}}} \qquad (4.20) \end{aligned}$$

Based on our proposed model (4.1), we set (4.20) equal to $\mu'(t)e^{\boldsymbol{\beta}^T \boldsymbol{Z}}dt$. Therefore

$$\int_t^\infty g(x, \boldsymbol{Z})\lambda_0(x)e^{\boldsymbol{\gamma}^T \boldsymbol{Z}}e^{-\Lambda_0(x)e^{\gamma^T \boldsymbol{z}}}dx = \frac{\mu'(t)}{h(t)}e^{\boldsymbol{\beta}^T \boldsymbol{Z}}e^{-\Lambda_0(t)e^{\gamma^T \boldsymbol{z}}}.$$

Differentiate both sides of the above equation with respect to t and set $t = x$, we have

$$\begin{aligned} g(x, \boldsymbol{Z}) &= \frac{1}{\lambda_0(x)e^{\boldsymbol{\gamma}^T \boldsymbol{Z}}}\left[-\left(\frac{\mu'(x)}{h(x)}\right)' e^{\boldsymbol{\beta}^T \boldsymbol{Z}} + \frac{\mu'(x)}{h(x)}\lambda_0(x)e^{(\boldsymbol{\beta}+\boldsymbol{\gamma})^T \boldsymbol{Z}}\right] \\ &= -\frac{1}{\lambda_0(x)}\left(\frac{\mu'(x)}{h(x)}\right)' e^{(\boldsymbol{\beta}-\boldsymbol{\gamma})^T \boldsymbol{Z}} + \frac{\mu'(x)}{h(x)}e^{\boldsymbol{\beta}^T \boldsymbol{Z}}. \end{aligned}$$

In particular, we can let $h(x) = h_0$,$\lambda_0(x) = \lambda_0$ and $\mu'(x) = h_0 e^{-\rho_0 x}$, $\rho_0 \geq 0$. Then

$$g(x, \boldsymbol{Z}) = e^{-\rho_0 x}e^{\boldsymbol{\beta}^T \boldsymbol{Z}}\left[1 + \frac{\rho_0}{\lambda_0}e^{-\boldsymbol{\gamma}^T \boldsymbol{Z}}\right].$$

Notice that ρ_0 controls the correlation between the recurrent events and death, with correlation being 0 when $\rho_0 = 0$. As we increase ρ_0, the correlation becomes larger.

Independent censoring time C is generated from the uniform $(0, \tau)$ distribution.

We first consider Z as a single dichotomous covariate from the $Bernoulli(0.5)$ distribution. The following combinations of simulation parameters are chosen: $\beta_0 = 0, 0.2, 0.5$, $\sigma^2 = 0.5$, $\gamma = 0.3$, $\lambda_0 = 1$, $h_0 = 8$, $\tau = 6$ and $\rho_0 = 0, 1, 4, 8$. The values of ρ_0 range from none to high correlation between recurrences and death. The average observed numbers of recurrent events range from 1 to 6 event(s) per subject. We consider sample sizes $n = 200, 400, 600$. For each setting, 1000 simulation samples are generated. The simulations are programmed in MATLAB(version 7.7.0). The results based on Method A and Method B are presented side-by-side in Table 4.1, Table 4.2 and Table 4.3.

For Method A, the coefficient estimator $\widehat{\boldsymbol{\beta}}$ appears to be approximately unbiased for all combinations of sample sizes, correlations and the true values of $\boldsymbol{\beta}_0$. The proposed standard error estimator provides a good estimate of the true variation of $\widehat{\boldsymbol{\beta}}$. For all sample sizes considered, the coverage rate fall in the $0.93 - 0.96$ range. The accuracy of the asymptotic approximation appears to be unaffected by the amount of correlation between recurrent events and death.

For Method B, the coefficient estimator $\widehat{\boldsymbol{\beta}}_D$ appears to be approximately unbiased for all combinations of sample sizes, correlations and the true values of $\boldsymbol{\beta}_0$. The proposed standard error estimator tends to underestimate the true standard error when the recurrent events and death are independent. However, as the correlation between the recurrences and death increases, we see substantial improvement in the estimated standard errors. For all sample sizes considered, the coverage rate falls in the $0.94 - 0.96$ range for moderately/highly correlated death and recurrences.

We also examine the finite-sample properties of the proposed estimator for continuous covariates by considering $Z \sim Uniform(0, 2)$ and $Z \sim Normal(0, 1)$, under the same combination set of parameters. The results are presented in Table 4.4, Table 4.5, Table 4.6, Table 4.7, Table 4.8 and Table 4.9.

Similar to the case when Z is dichotomous, both Method A and Method B yield approximately unbiased point estimates for $\boldsymbol{\beta}$. In term of the standard error estimator, Method A performs consistently well regardless of the correlation between the recurrences and death, while Method B underestimates the true standard error when death the recurrent events are independent. However, we see improvement as the correlation increases. For moderately/highly correlated recurrences and death data, Method B performs slightly better Method A in term of relative efficiency.

4.5 Application to the SOLVD data

As a demonstration, we apply the proposed methods to the SOLVD(Studies of Left Ventricular Dysfunction, SOLVD Investigators, 1991) Prevention Trial data. The SOLVD prevention trial was randomized, double-blind, and placebo-controlled. The trial had a three-year recruitment and a two-year follow-up. The basic inclusion criteria for the prevention trial were: age between 21 and 80 years, inclusive, no overt symptoms of congestive heart failure, and left ventricular EF less than 35%. EF is a number between 0 and 100 that measures the efficiency of the heart in ejecting blood. A total of 4228 patients with asymptomatic left ventricular dysfunction were randomly assigned to receive either enalapril or placebo in addition to usual care at one of the 83 hospitals linked to 23 centers in the United States, Canada, and Belgium. During the 2-year follow-up period, detailed information for hospitalizations and mortality was recorded. It is of interest to know if enalapril reduced the repeated hospitalizations for patients while adjusting for some baseline covariates such as center, gender, age and baseline ejection fraction.

Two subjects with invalid baseline ejection fraction record are excluded in the analysis. A total of 4226 subjects are included, with 2116 patients with a total of 2043 recorded hospitalizations and 2110 patients with a total of 1736 recorded hospitalizations in the placebo and enalapril treatment groups, respectively. The Kaplan-Meier estimators of the survival functions are plotted in Figure 4.1. The value of the log-rank test statistic is 2.593 with p-value=0.107, indicating there is no significant difference in survival between the two groups. Table 4.10 summarizes the hospitalization and survival experiences by groups. We are interested in assessing the average treatment effect (TRT) on repeated hospitalizations after adjusting for center, gender, age and baseline ejection fraction(EF). Both AGE and EF are centered on the mean. TRT is one for enalapril and zero for placebo. GENDER is one for male and zero for female. CENTER k ($k = 2, \cdots, 23$) is 1 for subjects in the kth center and 0 otherwise.

We conduct regression analysis by using the estimating equations (4.3) and (4.6). The two approaches yield similar results for point and interval estimates. In Table 4.11 we only present the results of the IPSW method based on (4.6). Enalapril treatment reduces the mean frequency of repeated hospitalizations by 16.8%(i.e., $1 - e^{-0.1834} = 0.168$) after adjusting for gender, centers, age and baseline ejection fraction. The reduction is statistically significant at the 0.05 level (p-value=0.001). Our analysis also indicates that the mean frequency of repeated hospitalization for heart failure increases by 0.9%(i.e., $e^{0.0085} - 1 = 0.009$) per year increase in age and decreases by 1.4%(i.e., $1 - e^{-0.0141} = 0.014$) with a 1% decrease in baseline ejection fraction. The effects of age and baseline ejection fraction are both statistically significant at the 0.05 level, with p-value=0.005 and p-value=0.006, respectively. Males have 13.5%(i.e., $1 - e^{-0.1453} = 0.135$) fewer repeated hospitalization than females, however, this reduction is not statistically significant at the 0.05 level (p-value=0.087). Effects of different centers on repeated hospitalizations appear to be nonsignificant except for Center 9 (p-value=0.005), Center 17 (p-value=0.017) and Center 22 (p-value=0.009). Center 9 is associated with larger number of hospitalizations, while patients in Center 17 and Center 22 had fewer hospitalizations. In terms of survival experience, older age(p-value$<$ 0.001) and smaller baseline ejection fraction(p-value$<$ 0.001) appear to be significantly associated with increased mortality. Although enalapril is seen to reduce mortality by 9.6%(i.e., $1 - e^{-0.1015} = 0.35$), the effect is not statistically significant(p-value=0.199). Males are 1.8%(i.e., $1 - e^{-0.0018} = 0.135$) less likely to die than females, however, the effect is again nonsignificant at the 0.05 level (p-value=0.989). Effects of different centers on mortality appear to be not statistically significant except for Center 7 (p-value=0.006), Center 10 (p-value=0.021) and Center 11 (p-value=0.018). Patients in Center 7 and Center 11 tend to have higher risk of death.

In summary, enalapril effectively reduced the frequency of hospitalizations but not mortality. Low baseline ejection fraction and old age are significantly associated with more frequent hospitalizations and higher risk of death. Gender appears to be not related to either hospitalizations or death.

Figure 4.2 displays the estimation of the mean frequency of hospitalizations from day

0 to day 1826 for a 59-year-old female patient with 28.3% baseline ejection fraction who received enalapril treatment versus one who did not. The simultaneous confidence bands are calculated from (4.19) by 1000 realizations of $V_G^D(t)$. From the plots, we can see those treated with enalapril have fewer hospitalizations than those in the placebo group.

Since our proposed method involves modelling the survival distribution, we plot the Schoenfeld residuals(Schoenfeld, 1982) to assess the validity of the proportional hazard assumption(Figure 4.3). They suggest that we cannot reject the proportional hazard assumption. Thus it is reasonable to model the survival distribution with a Cox proportional hazards model.

4.6 Concluding remarks

The conditional rate function of recurrent events given the subjects are still alive is of clinical interest as it contains information on recurrence up to death. We propose a multiplicative model for this rate function and develop two estimating procedures for the regression parameters as well as the mean function of recurrent events. Our approaches allow arbitrary dependence structure between the recurrent events and the terminal events. The proposed estimators are shown to be consistent and asymptotically normal. Simulation studies indicate that the first approach performs consistently well regardless of the correlation between the recurrences and death, while the second approach using the IPSW(Inverse Probability of Survival Weight) technique tends to underestimate the covariance of the regression parameter estimators when recurrences and death are independent or the correlation is small. However, the second approach works well if the correlation is moderate to high.

4.7 Proofs of the theorems

For convenience, we introduce the following notations: $\mathbb{P}_n f = \frac{1}{n}\sum_{i=1}^n f(U_i)$ as the expectation of f under the empirical measure, $Pf = \int f(u)dP(u)$ as the expectation of f under P, and $\mathbb{G}_n f = \sqrt{n}(\mathbb{P}_n f - Pf)$ as the empirical process evaluated at f, a Gaussian process $\mathbb{G}_P f$ as the limiting process of $\mathbb{G}_n f$.

Proof of Theorem 4.3.1:

In the empirical process notation, we can rewrite (4.3) as

$$\frac{1}{n}\boldsymbol{U}_n(\boldsymbol{\beta}) = \mathbb{P}_n\left\{\int_0^\tau Y(t)\left[\boldsymbol{Z}(t) - \frac{\mathbb{P}_n\{Y(t)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{Y(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}(t)}\}}\right]dN(t)\right\}.$$

Trivially, the class $\{\boldsymbol{\beta} \in \mathcal{B}\}$ and $\{\boldsymbol{Z}\}$ are both Donsker classes. Since products of bounded Donsker class are Donsker, $\{\boldsymbol{\beta}^T\boldsymbol{Z} : \boldsymbol{\beta} \in \mathcal{B}\}$ is a Donsker class. We know that exponentiation is Lipschitz continuous on compacts, $\{e^{\boldsymbol{\beta}^T\boldsymbol{Z}} : \boldsymbol{\beta} \in \mathcal{B}\}$ is also Donsker. By Lemma 4.1 of Kosorok(2008), N and Y are both Donsker as processes in $l^\infty([0,\tau])$. Again since all products of bounded Donsker classes are Donsker, $\{Y(t)\boldsymbol{Z}(t)dN(t), t \in [0,\tau]\}$, $\{Y(t)dN(t), t \in [0,\tau]\}$, $\{Y(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}(t)} : \boldsymbol{\beta} \in \mathcal{B}, t \in [0,\tau]\}$, $\{Y(t)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}(t)} : \boldsymbol{\beta} \in \mathcal{B}, t \in [0,\tau]\}$ are P-Donsker and thus P-Glivenko-Cantelli. Therefore, we have

$$\sup_{\boldsymbol{\beta}\in\mathcal{B}}\left|\frac{1}{n}\boldsymbol{U}_n(\boldsymbol{\beta}) - \boldsymbol{U}(\boldsymbol{\beta})\right| \xrightarrow{a.s.} 0,$$

where

$$\boldsymbol{U}(\boldsymbol{\beta}) \equiv P\left\{\int_0^\tau Y(t)\left[\boldsymbol{Z}(t) - \frac{P\{Y(t)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}(t)}\}}{P\{Y(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}(t)}\}}\right]dN(t)\right\}.$$

Next, we show $\boldsymbol{U}(\boldsymbol{\beta}_0) = 0$ under the proposed model (4.1).

$$\begin{aligned}
\boldsymbol{U}(\boldsymbol{\beta}_0) &= P\left\{\int_0^\tau Y(t)\left[\boldsymbol{Z}(t) - \frac{P\{Y(t)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}{P\{Y(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}\right]dN(t)\right\} \\
&= \int_0^\tau E\left\{Y(t)\left[\boldsymbol{Z}(t) - \frac{E\{Y(t)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}{E\{Y(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}\right]dN(t)\right\} \\
&= \int_0^\tau E\left\{Y(t)\boldsymbol{Z}(t)dN(t)\right\} - \int_0^\tau \frac{E\{Y(t)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}{E\{Y(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}E\left\{Y(t)dN(t)\right\} \\
&= \int_0^\tau E\left\{Y(t)\boldsymbol{Z}(t)E[dN(t)|D \geq t, \boldsymbol{Z}]\right\} \\
&\quad - \int_0^\tau \frac{E\{Y(t)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}{E\{Y(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}E\left\{Y(t)E[dN(t)|D \geq t, \boldsymbol{Z}]\right\} \\
&= \int_0^\tau E\left\{Y(t)\boldsymbol{Z}(t)I(D \geq t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}d\mu_0(t)\right\} \\
&\quad - \int_0^\tau \frac{E\{Y(t)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}{E\{Y(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}E\left\{Y(t)I(D \geq t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}d\mu_0(t)\right\} \\
&= \int_0^\tau E\left\{Y(t)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\right\}d\mu_0(t) \\
&\quad - \int_0^\tau \frac{E\{Y(t)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}{E\{Y(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}E\left\{Y(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\right\}d\mu_0(t) \\
&= 0
\end{aligned}$$

Using Taylor expansion, we have

$$\boldsymbol{U}(\boldsymbol{\beta}) = \boldsymbol{U}(\boldsymbol{\beta}_0) + \boldsymbol{A}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + o(|\boldsymbol{\beta} - \boldsymbol{\beta}_0|) = \boldsymbol{A}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + o(|\boldsymbol{\beta} - \boldsymbol{\beta}_0|).$$

By assumption (C5), $\boldsymbol{A}$ is non-singular, for sufficiently large n and small enough ϵ, we have

$$\sup_{|\boldsymbol{\beta}-\boldsymbol{\beta}_0|=\epsilon}\left|\frac{1}{n}\boldsymbol{U}_n(\boldsymbol{\beta}) - \boldsymbol{A}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\right| \leq \inf_{|\boldsymbol{\beta}-\boldsymbol{\beta}_0|=\epsilon}\left|\boldsymbol{A}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\right|.$$

Notice that $\boldsymbol{A}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$ has a unique solution within $|\boldsymbol{\beta} - \boldsymbol{\beta}_0| \leq \epsilon$. By the degree theory(Deming 1985, Chapter 1), the above inequality implies $\frac{1}{n}\boldsymbol{U}_n(\boldsymbol{\beta})$ has the same number of non-zero solution as $\boldsymbol{A}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)$. Therefore there exists $\widehat{\boldsymbol{\beta}}$ be the solution to $\boldsymbol{U}_n(\boldsymbol{\beta}) = 0$

and $|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0| \le \epsilon$ for sufficiently large n. Since ϵ can be chosen arbitrarily small, $\widehat{\boldsymbol{\beta}} \xrightarrow{a.s.} \boldsymbol{\beta}_0$. This concludes the proof of consistency of $\widehat{\boldsymbol{\beta}}$.

Let $M_{\boldsymbol{\beta}_0}(t) = N(t) - \int_0^t Y(s) e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}(s)} d\mu_0(s)$, and by addition and subtraction,

$$
\begin{aligned}
\frac{1}{n}\boldsymbol{U}_n(\widehat{\boldsymbol{\beta}}) &= \mathbb{P}_n \int_0^\tau Y(t)\left[\boldsymbol{Z}(t) - \frac{\mathbb{P}_n\{Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}\right]\left\{dN(t) - Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}d\mu_0(t)\right\} \\
&= \mathbb{P}_n \int_0^\tau Y(t)\left[\boldsymbol{Z}(t) - \frac{\mathbb{P}_n\{Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}\right]\left\{dN(t) - Y(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}d\mu_0(t)\right\} \\
&\quad -\mathbb{P}_n \int_0^\tau Y(t)\left[\boldsymbol{Z}(t) - \frac{\mathbb{P}_n\{Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}\right]\left\{e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)} - e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\right\}d\mu_0(t) \\
&= \mathbb{P}_n \int_0^\tau Y(t)\left[\boldsymbol{Z}(t) - \frac{\mathbb{P}_n\{Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}\right]dM_{\boldsymbol{\beta}_0}(t) \\
&\quad -\mathbb{P}_n \int_0^\tau Y(t)\left[\boldsymbol{Z}(t) - \frac{\mathbb{P}_n\{Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}\right]\left\{e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)} - e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\right\}d\mu_0(t).
\end{aligned}
$$

Similarly, we have

$$
\begin{aligned}
\boldsymbol{U}(\widehat{\boldsymbol{\beta}}) &= P \int_0^\tau Y(t)\left[\boldsymbol{Z}(t) - \frac{P\{Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}{P\{Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}\right]dM_{\boldsymbol{\beta}_0}(t) \\
&\quad -P \int_0^\tau Y(t)\left[\boldsymbol{Z}(t) - \frac{P\{Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}{P\{Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}\right]\left\{e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)} - e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\right\}d\mu_0(t).
\end{aligned}
$$

$$
\begin{aligned}
\sqrt{n}\left\{\frac{1}{n}\boldsymbol{U}_n(\widehat{\boldsymbol{\beta}}) - \boldsymbol{U}(\widehat{\boldsymbol{\beta}})\right\} &= \sqrt{n}\mathbb{P}_n \int_0^\tau Y(t)\left[\boldsymbol{Z}(t) - \frac{\mathbb{P}_n\{Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}\right]dM_{\boldsymbol{\beta}_0}(t) \\
&\quad -\sqrt{n}P \int_0^\tau Y(t)\left[\boldsymbol{Z}(t) - \frac{P\{Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}{P\{Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}\right]dM_{\boldsymbol{\beta}_0}(t)
\end{aligned}
$$

$$
\begin{aligned}
& -\sqrt{n}\mathbb{P}_n \int_0^\tau Y(t)\left[\boldsymbol{Z}(t) - \frac{\mathbb{P}_n\{Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}\right]\left\{e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)} - e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\right\} d\mu_0(t) \\
& +\sqrt{n}P \int_0^\tau Y(t)\left[\boldsymbol{Z}(t) - \frac{P\{Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}{P\{Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}\right]\left\{e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)} - e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\right\} d\mu_0(t) \\
= \; & \sqrt{n}(\mathbb{P}_n - P) \int_0^\tau Y(t)\left\{\boldsymbol{Z}(t) - \frac{\mathbb{P}_n\{Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}\right\} dM_{\boldsymbol{\beta}_0}(t) \\
& +\sqrt{n}P \int_0^\tau Y(t)\left\{\boldsymbol{Z}(t) - \frac{\mathbb{P}_n\{Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}\right\} dM_{\boldsymbol{\beta}_0}(t) \\
& -\sqrt{n}P \int_0^\tau Y(t)\left[\boldsymbol{Z}(t) - \frac{P\{Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}{P\{Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}\right] dM_{\boldsymbol{\beta}_0}(t) \\
& -\sqrt{n}(\mathbb{P}_n - P) \int_0^\tau Y(t)\left[\boldsymbol{Z}(t) - \frac{\mathbb{P}_n\{Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}\right]\left\{e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)} - e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\right\} d\mu_0(t) \\
& -\sqrt{n}P \int_0^\tau \left\{Y(t)\left[\boldsymbol{Z}(t) - \frac{\mathbb{P}_n\{Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}\right] - \right. \\
& \left. Y(t)\left[\boldsymbol{Z}(t) - \frac{P\{Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}{P\{Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}\right]\right\}\left\{e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)} - e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\right\} d\mu_0(t) \\
= \; & (i) + (ii) - (iii) - (iv) - (v) \qquad (4.21)
\end{aligned}
$$

Since $E\left\{dM_{\boldsymbol{\beta}_0}(t)|D \geq t, C, \boldsymbol{Z}\right\} = Y(t)E\left\{dN^*(t)|D \geq t, \boldsymbol{Z}\right\} - Y(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}d\mu_0(t) = 0$, we have

$$
\begin{aligned}
& P \int_0^\tau Y(t)\left\{\boldsymbol{Z}(t) - \frac{\mathbb{P}_n\{Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}\right\} dM_{\boldsymbol{\beta}_0}(t) \\
= \; & \int_0^\tau E\left\{Y(t)\left[\boldsymbol{Z}(t) - \frac{\mathbb{P}_n\{Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}\right] dM_{\boldsymbol{\beta}_0}(t)\right\} \\
= \; & \int_0^\tau E\left\{Y(t)\left[\boldsymbol{Z}(t) - \frac{\mathbb{P}_n\{Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}\right] E[dM_{\boldsymbol{\beta}_0}(t)|D \geq t, C, Z]\right\} \\
= \; & \int_0^\tau E\left\{Y(t)\left[\boldsymbol{Z}(t) - \frac{\mathbb{P}_n\{Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}\right] \cdot 0\right\} = 0.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
& P\int_0^\tau Y(t)\left\{\boldsymbol{Z}(t)-\frac{P\{Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}{P\{Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}\right\}dM_{\boldsymbol{\beta}_0}(t)\\
= & \int_0^\tau E\left\{Y(t)\left[\boldsymbol{Z}(t)-\frac{P\{Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}{P\{Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}\right]dM_{\boldsymbol{\beta}_0}(t)\right\}\\
= & \int_0^\tau E\left\{Y(t)\left[\boldsymbol{Z}(t)-\frac{P\{Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}{P\{Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}\right]E[dM_{\boldsymbol{\beta}_0}(t)|D\geq t,C,Z]\right\}\\
= & \int_0^\tau E\left\{Y(t)\left[\boldsymbol{Z}(t)-\frac{P\{Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}{P\{Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}\right]\cdot 0\right\}=0
\end{aligned}
$$

Therefore, (ii) and (iii) of (4.21) are zero.

Now look at (iv) of (4.21). By the consistency of $\widehat{\boldsymbol{\beta}}$, $e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}-e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}=o_P(1)$. Since $Y(t)\left[\boldsymbol{Z}(t)-\frac{\mathbb{P}_n\{Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}\right]$ is bounded and $\sqrt{n}(\mathbb{P}_n-P)=O(1)$, we have

$$
(iv) \text{ of } (4.21) \;=\; P\int_0^\tau O(1)\cdot o_P(1)=o_P(1).
$$

Again by using $\mathbb{P}_n-P=O(n^{-1/2})$, we have

$$
Y(t)\left[\boldsymbol{Z}(t)-\frac{\mathbb{P}_n\{Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}\right]-Y(t)\left[\boldsymbol{Z}(t)-\frac{P\{Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}{P\{Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}\right]=O(n^{-1/2}).
$$

Then

$$
(v) \text{ of } (4.21) \;=\; \sqrt{n}P\int_0^\tau O(n^{-1/2})\cdot o_P(1)=O(1)\cdot o_P(1)=o_P(1).
$$

Thus (4.21) becomes

$$
\begin{aligned}
& \sqrt{n}\left\{\frac{1}{n}\boldsymbol{U}_n(\widehat{\boldsymbol{\beta}})-\boldsymbol{U}(\widehat{\boldsymbol{\beta}})\right\}\\
= & \;\mathbb{G}_n\int_0^\tau Y(t)\left\{\boldsymbol{Z}(t)-\frac{\mathbb{P}_n\{Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}\right\}dM_{\boldsymbol{\beta}_0}(t)+o_P(1). \qquad (4.22)
\end{aligned}
$$

Using similar arguments as in proofing the consistency of $\widehat{\boldsymbol{\beta}}$, we can show $\{Y(t)\boldsymbol{Z}(t)dN(t), t \in [0,\tau]\}$, $\{Y(t)dN(t), t \in [0,\tau]\}$, $\{Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)} : \widehat{\boldsymbol{\beta}} \in \mathcal{B}, t \in [0,\tau]\}$, $\{Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)} : \widehat{\boldsymbol{\beta}} \in \mathcal{B}, t \in [0,\tau]\}$ are P-Donsker and thus P-Glivenko-Cantelli. We can replace the operator $\mathbb{P}_n$ in (4.22) with P without altering the limiting distribution. By the strong consistency of $\widehat{\boldsymbol{\beta}}$ and Lemma 19.24 of Van der Vaart(1998), (4.22) becomes

$$\begin{aligned} & \sqrt{n}\left\{\frac{1}{n}\boldsymbol{U}_n(\widehat{\boldsymbol{\beta}}) - \boldsymbol{U}(\widehat{\boldsymbol{\beta}})\right\} \\ = \;& \mathbb{G}_n \int_0^\tau Y(t)\left\{\boldsymbol{Z}(t) - \frac{P\{Y(t)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}{P\{Y(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}\right\} dM_{\boldsymbol{\beta}_0}(t) + o_P(1) \\ \equiv \;& \mathbb{G}_n \boldsymbol{J} + o_P(1). \end{aligned} \tag{4.23}$$

On the other hand, since $\boldsymbol{U}_n(\widehat{\boldsymbol{\beta}}) = 0$ and $\boldsymbol{U}(\boldsymbol{\beta}_0) = 0$, we have

$$\begin{aligned} \sqrt{n}\left\{\frac{1}{n}\boldsymbol{U}_n(\widehat{\boldsymbol{\beta}}) - \boldsymbol{U}(\widehat{\boldsymbol{\beta}})\right\} &= \sqrt{n}\left\{0 - (\boldsymbol{U}(\widehat{\boldsymbol{\beta}}) - \boldsymbol{U}(\boldsymbol{\beta}_0))\right\} \\ &= -\sqrt{n}\boldsymbol{A}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + o(|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0|) \end{aligned} \tag{4.24}$$

Combining (4.23) and (4.24), we have

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = -\boldsymbol{A}^{-1}\mathbb{G}_n \boldsymbol{J} + o_P(1). \tag{4.25}$$

Therefore $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ converges in distribution to a normal distribution with mean zero and covariance matrix $\boldsymbol{A}^{-1}\boldsymbol{\Sigma}(\boldsymbol{A}^{-1})^T$,$\boldsymbol{\Sigma} = E[\boldsymbol{J}\boldsymbol{J}^T]$, $\boldsymbol{A}$ and $\boldsymbol{J}$ as defined as in (4.8). By replacing $\boldsymbol{\beta}_0$ with $\widehat{\boldsymbol{\beta}}$ and expectations with empirical means in the expressions of $\boldsymbol{A}$ and $\boldsymbol{J}$, we obtain an estimator $\widehat{\boldsymbol{A}}^{-1}\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{A}}^{-1})^T$, where $\widehat{\boldsymbol{A}}$ and $\widehat{\boldsymbol{\Sigma}}$ are defined as in (4.9). Using similar arguments as in proofing the consistency and normality of $\widehat{\boldsymbol{\beta}}$, we can show that $\widehat{\boldsymbol{A}}^{-1}\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{A}}^{-1})^T$ is a consistent estimator of $\boldsymbol{A}^{-1}\boldsymbol{\Sigma}(\boldsymbol{A}^{-1})^T$.

<u>Proof of Theorem 4.3.2</u>:

In the empirical process notation, we can rewrite (4.6) as

$$\frac{1}{n}\boldsymbol{U}_n^D(\boldsymbol{\beta}) \;=\; \mathbb{P}_n\left\{\int_0^{\tau}\widehat{W}^D(t)\left[\boldsymbol{Z}(t)-\frac{\mathbb{P}_n\{\widehat{W}^D(t)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{\widehat{W}^D(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}(t)}\}}\right]dN(t)\right\},$$

and let

$$\frac{1}{n}\boldsymbol{U}_n^{D^*}(\boldsymbol{\beta})=\mathbb{P}_n\left\{\int_0^{\tau}W^D(t)\left[\boldsymbol{Z}(t)-\frac{\mathbb{P}_n\{W^D(t)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{W^D(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}(t)}\}}\right]dN(t)\right\}.$$

Notice that $\widehat{W}_i^D(t)=\frac{Y_i(t)}{\widehat{S}_D(t|\boldsymbol{Z}_i)}$, $W_i^D(t)=\frac{Y_i(t)}{S_D(t|\boldsymbol{Z}_i)}$, $\widehat{S}_D(t|\boldsymbol{Z}_i)$ is the estimated survival function given $\boldsymbol{Z}$ based on the Cox proportional hazard model (4.5), and $S_D(t|\boldsymbol{Z}_i)$ is the true survival function. We know that $\widehat{\boldsymbol{\gamma}}_D\xrightarrow{a.s.}\boldsymbol{\gamma}_D$ and $\widehat{\Lambda}^D(\cdot)\xrightarrow{a.s.}\Lambda_0^D(\cdot)$, then

$$\sup_{t\in[0,\tau]}\max_i\left|\widehat{W}_i^D(t)-W_i^D(t)\right|\xrightarrow{a.s.}0.$$

Since we assume $\boldsymbol{Z}(.)$ is bounded, there exist constants C_1 and C_2, such that

$$\sup_{t\in[0,\tau]}\max_i\left|\frac{1}{n}\sum_{i=1}^n\widehat{W}_i^D(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_i(t)}-\frac{1}{n}\sum_{i=1}^n W_i^D(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_i(t)}\right|\le\sup_{t\in[0,\tau]}\max_i\left|\widehat{W}_i^D(t)-W_i^D(t)\right|\cdot C_1\xrightarrow{a.s.}0,$$

$$\begin{aligned}&\sup_{t\in[0,\tau]}\max_i\left|\frac{1}{n}\sum_{i=1}^n\widehat{W}_i^D(t)\boldsymbol{Z}_i(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_i(t)}-\frac{1}{n}\sum_{i=1}^n W_i^D(t)\boldsymbol{Z}_i(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_i(t)}\right|\\ \le\;&\sup_{t\in[0,\tau]}\max_i\left|\widehat{W}_i^D(t)-W_i^D(t)\right|\cdot C_2\xrightarrow{a.s.}0.\end{aligned}$$

Thus we have

$$\sup_{\boldsymbol{\beta}\in\mathcal{B}}\left|\frac{1}{n}\boldsymbol{U}_n^D(\boldsymbol{\beta})-\frac{1}{n}\boldsymbol{U}_n^{D^*}(\boldsymbol{\beta})\right|\xrightarrow{a.s.}0.$$

In $\boldsymbol{U}_n^{D^*}(\boldsymbol{\beta})$, the class $\{\boldsymbol{\beta} \in \mathcal{B}\}$, $\{\boldsymbol{\gamma}_D \in \mathcal{D}\}$ and $\{\boldsymbol{Z}\}$ are all bounded Donsker classes. Since products of bounded Donsker class are Donsker, $\{\boldsymbol{\beta}^T\boldsymbol{Z} : \boldsymbol{\beta} \in \mathcal{B}\}$ and $\{\boldsymbol{\gamma}_D^T\boldsymbol{Z} : \boldsymbol{\gamma}_D \in \mathcal{D}\}$ are Donsker classes. We know that exponentiation is Lipschitz continuous on compacts, $\{e^{\boldsymbol{\beta}^T\boldsymbol{Z}} : \boldsymbol{\beta} \in \mathcal{B}\}$ and $\{e^{\boldsymbol{\gamma}_D^T\boldsymbol{Z}} : \boldsymbol{\gamma}_D \in \mathcal{D}\}$ are also Donsker. By Lemma 4.1 of Kosorok(2008), $\{S_0^D(t) = P(D \geq t) : t \in [0,\tau]\}$, $\{N(t) : t \in [0,\tau]\}$ and $\{Y(t) : t \in [0,\tau]\}$ are Donsker. Let $\phi : \mathbb{R}^2 \mapsto \mathbb{R}$ be defined by $\phi(x,y) = 1/(x^y)$. ϕ is Lipschitz continuous on the set of $[0,1] \times [k,k]$. Therefore $\{1/S_D(t|Z) = 1/[S_0^D(t)^{e^{\boldsymbol{\gamma}_D^T\boldsymbol{Z}}}] = \phi(S_0^D(t), e^{\boldsymbol{\gamma}_D^T\boldsymbol{Z}}) : t \in [0,\tau], \boldsymbol{\gamma}_D \in \mathcal{D}\}$ is P-Donsker and so is $\{W^D(t) = Y(t)/S_D(t|Z), t \in [0,\tau], \boldsymbol{\gamma}_D \in \mathcal{D}\}$. Again since all products of bounded Donsker classes are Donsker, $\{W^D(t)N(t) : t \in [0,\tau]\}$,$\{W^D(t)\boldsymbol{Z}(t)N(t) : t \in [0,\tau]\}$, $\{W^D(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}(t)} : \boldsymbol{\beta} \in \mathcal{B}, t \in [0,\tau]\}$, $\{W^D(t)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}(t)} : \boldsymbol{\beta} \in \mathcal{B}, t \in [0,\tau]\}$ are P-Donsker and thus P-Glivenko-Cantelli. Therefore, we have

$$\sup_{\boldsymbol{\beta}\in\mathcal{B}} \left| \frac{1}{n}\boldsymbol{U}_n^{D^*}(\boldsymbol{\beta}) - \boldsymbol{U}^D(\boldsymbol{\beta}) \right| \xrightarrow{a.s.} 0,$$

where

$$\boldsymbol{U}^D(\boldsymbol{\beta}) \equiv P\left\{ \int_0^\tau W^D(t) \left[\boldsymbol{Z}(t) - \frac{P\{W^D(t)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}(t)}\}}{P\{W^D(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}(t)}\}} \right] dN(t) \right\}.$$

Since we have shown $\sup_{\boldsymbol{\beta}\in\mathcal{B}} \left| \frac{1}{n}\boldsymbol{U}_n^D(\boldsymbol{\beta}) - \frac{1}{n}\boldsymbol{U}_n^{D^*}(\boldsymbol{\beta}) \right| \xrightarrow{a.s.} 0$, we have

$$\sup_{\boldsymbol{\beta}\in\mathcal{B}} \left| \frac{1}{n}\boldsymbol{U}_n^D(\boldsymbol{\beta}) - \boldsymbol{U}^D(\boldsymbol{\beta}) \right| \xrightarrow{a.s.} 0.$$

Next, we show $\boldsymbol{U}^D(\boldsymbol{\beta}_0) = 0$ under the proposed model (4.1).

$$\begin{aligned}
\boldsymbol{U}^D(\boldsymbol{\beta}_0) &= P\left\{\int_0^\tau W^D(t)\left[\boldsymbol{Z}(t) - \frac{P\{W^D(t)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}{P\{W^D(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}\right]dN(t)\right\} \\
&= \int_0^\tau E\left\{W^D(t)\left[\boldsymbol{Z}(t) - \frac{E\{W^D(t)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}{E\{W^D(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}\right]dN(t)\right\} \\
&= \int_0^\tau E\left\{W^D(t)\boldsymbol{Z}(t)dN(t)\right\} - \int_0^\tau \frac{E\{W^D(t)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}{E\{W^D(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}E\left\{W^D(t)dN(t)\right\} \\
&= \int_0^\tau E\left\{W^D(t)\boldsymbol{Z}(t)E[dN(t)|D \geq t, \boldsymbol{Z}]\right\} \\
&\quad - \int_0^\tau \frac{E\{W^D(t)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}{E\{W^D(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}E\left\{W^D(t)E[dN(t)|D \geq t, \boldsymbol{Z}]\right\} \\
&= \int_0^\tau E\left\{W^D(t)\boldsymbol{Z}(t)I(D \geq t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}d\mu_0(t)\right\} \\
&\quad - \int_0^\tau \frac{E\{W^D(t)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}{E\{Y(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}E\left\{W^D(t)I(D \geq t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}d\mu_0(t)\right\} \\
&= \int_0^\tau E\left\{W^D(t)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\right\}d\mu_0(t) \\
&\quad - \int_0^\tau \frac{E\{W^D(t)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}{E\{W^D(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}E\left\{W^D(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\right\}d\mu_0(t) \\
&= 0
\end{aligned}$$

By Taylor expansion, we have

$$\boldsymbol{U}^D(\boldsymbol{\beta}) = \boldsymbol{U}^D(\boldsymbol{\beta}_0) + \boldsymbol{A}_D(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + o(|\boldsymbol{\beta} - \boldsymbol{\beta}_0|) = \boldsymbol{A}_D(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + o(|\boldsymbol{\beta} - \boldsymbol{\beta}_0|),$$

By assumption (C5), $\boldsymbol{A}_D$ is non-singular, for sufficiently large n and small enough ϵ, we have

$$\sup_{|\boldsymbol{\beta}-\boldsymbol{\beta}_0|=\epsilon}\left|\frac{1}{n}\boldsymbol{U}_n^D(\boldsymbol{\beta}) - \boldsymbol{A}_D(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\right| \leq \inf_{|\boldsymbol{\beta}-\boldsymbol{\beta}_0|=\epsilon}\left|\boldsymbol{A}_D(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\right|.$$

Notice that $\boldsymbol{A}_D(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$ has a unique solution within $|\boldsymbol{\beta} - \boldsymbol{\beta}_0| \leq \epsilon$. By the degree theory(Deming 1985, Chapter 1), the above inequality implies $\frac{1}{n}\boldsymbol{U}_n^D(\boldsymbol{\beta})$ has the same number of non-zero solution as $\boldsymbol{A}_D(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$. Therefore there exists $\widehat{\boldsymbol{\beta}}_D$ which is the solution to $\boldsymbol{U}_n^D(\boldsymbol{\beta})=0$ and $|\widehat{\boldsymbol{\beta}}_D-\boldsymbol{\beta}_0|\leq\epsilon$ for sufficiently large n. Since ϵ can be chosen arbitrarily small, $\widehat{\boldsymbol{\beta}}_D \xrightarrow{a.s.} \boldsymbol{\beta}_0$. This concludes the proof of consistency of $\widehat{\boldsymbol{\beta}}_D$.

Let $M_{\boldsymbol{\beta}_0}(t)=N(t)-\int_0^t Y(s)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(s)}d\mu_0(s)$, and by addition and subtraction,

$$\begin{aligned}
\frac{1}{n}\boldsymbol{U}_n^D(\widehat{\boldsymbol{\beta}}_D) &= \mathbb{P}_n\int_0^\tau \widehat{W}^D(t)\left[\boldsymbol{Z}(t)-\frac{\mathbb{P}_n\{\widehat{W}^D(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{\widehat{W}^D(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}}\right]\left\{dN(t)-Y(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}d\mu_0(t)\right\}\\
&= \mathbb{P}_n\int_0^\tau \widehat{W}^D(t)\left[\boldsymbol{Z}(t)-\frac{\mathbb{P}_n\{\widehat{W}^D(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{\widehat{W}^D(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}}\right]\left\{dN(t)-Y(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}d\mu_0(t)\right\}\\
&\quad -\mathbb{P}_n\int_0^\tau \widehat{W}^D(t)\left[\boldsymbol{Z}(t)-\frac{\mathbb{P}_n\{\widehat{W}^D(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{\widehat{W}^D(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}}\right]\left\{e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}-e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\right\}d\mu_0(t)\\
&= \mathbb{P}_n\int_0^\tau \widehat{W}^D(t)\left[\boldsymbol{Z}(t)-\frac{\mathbb{P}_n\{\widehat{W}^D(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{\widehat{W}^D(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}}\right]dM_{\boldsymbol{\beta}_0}(t)\\
&\quad -\mathbb{P}_n\int_0^\tau \widehat{W}^D(t)\left[\boldsymbol{Z}(t)-\frac{\mathbb{P}_n\{\widehat{W}^D(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{\widehat{W}^D(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}}\right]\left\{e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}-e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\right\}d\mu_0(t).
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
\boldsymbol{U}^D(\widehat{\boldsymbol{\beta}}_D) &= P\int_0^\tau W^D(t)\left[\boldsymbol{Z}(t)-\frac{P\{W^D(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}}{P\{W^D(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}}\right]dM_{\boldsymbol{\beta}_0}(t)\\
&\quad -P\int_0^\tau W^D(t)\left[\boldsymbol{Z}(t)-\frac{P\{W^D(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}}{P\{W^D(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}}\right]\left\{e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}-e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\right\}d\mu_0(t).
\end{aligned}$$

$$\begin{aligned}
\sqrt{n}\left\{\frac{1}{n}\boldsymbol{U}_n^D(\widehat{\boldsymbol{\beta}}_D)-\boldsymbol{U}^D(\widehat{\boldsymbol{\beta}}_D)\right\} &= \sqrt{n}\mathbb{P}_n\int_0^\tau \widehat{W}^D(t)\left[\boldsymbol{Z}(t)-\frac{\mathbb{P}_n\{\widehat{W}^D(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{\widehat{W}^D(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}}\right]dM_{\boldsymbol{\beta}_0}(t)\\
&\quad -\sqrt{n}P\int_0^\tau W^D(t)\left[\boldsymbol{Z}(t)-\frac{P\{W^D(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}}{P\{W^D(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}}\right]dM_{\boldsymbol{\beta}_0}(t)
\end{aligned}$$

$$
\begin{aligned}
& -\sqrt{n}\mathbb{P}_n \int_0^\tau \widehat{W}^D(t) \left[\boldsymbol{Z}(t) - \frac{\mathbb{P}_n\{\widehat{W}^D(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{\widehat{W}^D(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}} \right] \left\{ e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)} - e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)} \right\} d\mu_0(t) \\
& +\sqrt{n}P \int_0^\tau W^D(t) \left[\boldsymbol{Z}(t) - \frac{P\{W^D(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}}{P\{W^D(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}} \right] \left\{ e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)} - e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)} \right\} d\mu_0(t) \\
= \; & \sqrt{n}(\mathbb{P}_n - P) \int_0^\tau \widehat{W}^D(t) \left\{ \boldsymbol{Z}(t) - \frac{\mathbb{P}_n\{\widehat{W}^D(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{\widehat{W}^D(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}} \right\} dM_{\boldsymbol{\beta}_0}(t) \\
& +\sqrt{n}P \int_0^\tau \widehat{W}^D(t) \left\{ \boldsymbol{Z}(t) - \frac{\mathbb{P}_n\{\widehat{W}^D(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{\widehat{W}^D(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}} \right\} dM_{\boldsymbol{\beta}_0}(t) \\
& -\sqrt{n}P \int_0^\tau W^D(t) \left[\boldsymbol{Z}(t) - \frac{P\{W^D(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}}{P\{W^D(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}} \right] dM_{\boldsymbol{\beta}_0}(t) \\
& -\sqrt{n}(\mathbb{P}_n - P) \int_0^\tau \widehat{W}^D(t) \left[\boldsymbol{Z}(t) - \frac{\mathbb{P}_n\{\widehat{W}^D(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{\widehat{W}^D(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}} \right] \left\{ e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)} - e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)} \right\} d\mu_0(t) \\
& -\sqrt{n}P \int_0^\tau \left\{ \widehat{W}^D(t) \left[\boldsymbol{Z}(t) - \frac{\mathbb{P}_n\{\widehat{W}^D(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{\widehat{W}^D(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}} \right] - \right. \\
& \left. W^D(t) \left[\boldsymbol{Z}(t) - \frac{P\{W^D(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}}{P\{W^D(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}} \right] \right\} \left\{ e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)} - e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)} \right\} d\mu_0(t) \\
= \; & (i) + (ii) - (iii) - (iv) - (v). \qquad (4.26)
\end{aligned}
$$

Since $E\left\{dM_{\boldsymbol{\beta}_0}(t)|D \geq t, C, \boldsymbol{Z}\right\} = Y(t)E\left\{dN^*(t)|D \geq t, \boldsymbol{Z}\right\} - Y(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}d\mu_0(t) = 0$, we have

$$
\begin{aligned}
& P \int_0^\tau \widehat{W}^D(t) \left\{ \boldsymbol{Z}(t) - \frac{\mathbb{P}_n\{\widehat{W}^D(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{\widehat{W}^D(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}} \right\} dM_{\boldsymbol{\beta}_0}(t) \\
= \; & \int_0^\tau E \left\{ \widehat{W}^D(t) \left[\boldsymbol{Z}(t) - \frac{\mathbb{P}_n\{\widehat{W}^D(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{\widehat{W}^D(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}} \right] dM_{\boldsymbol{\beta}_0}(t) \right\} \\
= \; & \int_0^\tau E \left\{ \widehat{W}^D(t) \left[\boldsymbol{Z}(t) - \frac{\mathbb{P}_n\{\widehat{W}^D(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{\widehat{W}^D(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}} \right] E[dM_{\boldsymbol{\beta}_0}(t)|D \geq t, C, Z] \right\} \\
= \; & \int_0^\tau E \left\{ \widehat{W}^D(t) \left[\boldsymbol{Z}(t) - \frac{\mathbb{P}_n\{\widehat{W}^D(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{\widehat{W}^D(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}} \right] \cdot 0 \right\} = 0.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
& P\int_0^\tau W^D(t)\left\{\boldsymbol{Z}(t)-\frac{P\{W^D(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}}{P\{W^D(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}}\right\}dM_{\boldsymbol{\beta}_0}(t) \\
= & \int_0^\tau E\left\{W^D(t)\left[\boldsymbol{Z}(t)-\frac{P\{W^D(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}}{P\{W^D(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}}\right]dM_{\boldsymbol{\beta}_0}(t)\right\} \\
= & \int_0^\tau E\left\{W^D(t)\left[\boldsymbol{Z}(t)-\frac{P\{W^D(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}}{P\{W^D(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}}\right]E[dM_{\boldsymbol{\beta}_0}(t)|D\geq t,C,Z]\right\} \\
= & \int_0^\tau E\left\{W^D(t)\left[\boldsymbol{Z}(t)-\frac{P\{W^D(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}}{P\{W^D(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}}\right]\cdot 0\right\}=0
\end{aligned}
$$

Therefore, (ii) and (iii) of (4.26) are zero.

Now look at (iv) of (4.21). By the consistency of $\widehat{\boldsymbol{\beta}}_D$, $e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}-e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}=o_P(1)$. Since $\widehat{W}^D(t)\left[\boldsymbol{Z}(t)-\frac{\mathbb{P}_n\{\widehat{W}^D(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{\widehat{W}^D(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}}\right]$ is bounded and $\sqrt{n}(\mathbb{P}_n-P)=O(1)$, we have

$$
(iv)\text{ of (4.26)} \quad = \quad P\int_0^\tau O(1)\cdot o_P(1)=o_P(1).
$$

Notice that $\widehat{W}^D(t)-W^D(t)=O(n^{-1/2})$ since $\widehat{\boldsymbol{\gamma}}_D-\boldsymbol{\gamma}_D=O(n^{-1/2})$ and $\widehat{\Lambda}^D(t)-\Lambda^D(t)=O(n^{-1/2})$. Again by using $\mathbb{P}_n-P=O(n^{-1/2})$, we have

$$
\begin{aligned}
& \widehat{W}^D(t)\left[\boldsymbol{Z}(t)-\frac{\mathbb{P}_n\{\widehat{W}^D(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{\widehat{W}^D(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}}\right]-W^D(t)\left[\boldsymbol{Z}(t)-\frac{P\{W^D(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}}{P\{W^D(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}}\right] \\
= \; & O(n^{-1/2}).
\end{aligned}
$$

Then

$$
(v)\text{ of (4.26)} \quad = \quad \sqrt{n}P\int_0^\tau O(n^{-1/2})\cdot o_P(1)=O(1)\cdot o_P(1)=o_P(1)
$$

Thus (4.26) becomes

$$\begin{aligned}
& \sqrt{n}\left\{\frac{1}{n}\boldsymbol{U}_n^D(\widehat{\boldsymbol{\beta}}_D)-\boldsymbol{U}^D(\widehat{\boldsymbol{\beta}}_D)\right\} \\
= \;& \mathbb{G}_n\int_0^\tau \widehat{W}^D(t)\left\{\boldsymbol{Z}(t)-\frac{\mathbb{P}_n\{\widehat{W}^D(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{\widehat{W}^D(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)}\}}\right\}dM_{\boldsymbol{\beta}_0}(t)+o_P(1). \qquad (4.27)
\end{aligned}$$

Using similar arguments as in proofing the consistency of $\widehat{\boldsymbol{\beta}}_D$, we can show $\{\widehat{W}^D(t)\boldsymbol{Z}(t)dN(t),t\in[0,\tau]\}$, $\{\widehat{W}^D(t)dN(t),t\in[0,\tau]\}$, $\{\widehat{W}^D(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}(t)}:\boldsymbol{\beta}\in\mathcal{B},t\in[0,\tau]\}$, $\{\widehat{W}^D(t)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}(t)}:\boldsymbol{\beta}\in\mathcal{B},t\in[0,\tau]\}$ are P-Donsker and thus P-Glivenko-Cantelli. We can replace the operator $\mathbb{P}_n$ in (4.27) with P without altering the limiting distribution. By applying the strong consistency of $\widehat{\boldsymbol{\beta}}_D$, $\widehat{\boldsymbol{\gamma}}_D$, $\widehat{\Lambda}^D(\cdot)$ and Lemma 19.24 of Van der Vaart(1998), we have

$$\begin{aligned}
& \sqrt{n}\left\{\frac{1}{n}\boldsymbol{U}_n^D(\widehat{\boldsymbol{\beta}}_D)-\boldsymbol{U}^D(\widehat{\boldsymbol{\beta}}_D)\right\} \\
= \;& \mathbb{G}_n\int_0^\tau W^D(t)\left\{\boldsymbol{Z}(t)-\frac{P\{W^D(t)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}_0\boldsymbol{Z}(t)}\}}{P\{W^D(t)e^{\boldsymbol{\beta}_0\boldsymbol{Z}(t)}\}}\right\}dM_{\boldsymbol{\beta}_0}(t)+o_P(1) \\
= \;& \mathbb{G}_n\boldsymbol{J}_D+o_P(1). \qquad (4.28)
\end{aligned}$$

On the other hand, since $\boldsymbol{U}_n^D(\widehat{\boldsymbol{\beta}})=0$ and $\boldsymbol{U}^D(\boldsymbol{\beta}_0)=0$, we have

$$\begin{aligned}
\sqrt{n}\left\{\frac{1}{n}\boldsymbol{U}_n^D(\widehat{\boldsymbol{\beta}})-\boldsymbol{U}^D(\widehat{\boldsymbol{\beta}})\right\} &= \sqrt{n}\left\{0-(\boldsymbol{U}^D(\widehat{\boldsymbol{\beta}})-\boldsymbol{U}^D(\boldsymbol{\beta}_0))\right\} \\
&= -\sqrt{n}\boldsymbol{A}_D(\widehat{\boldsymbol{\beta}}_D-\boldsymbol{\beta}_0)+o(|\widehat{\boldsymbol{\beta}}_D-\boldsymbol{\beta}_0|) \qquad (4.29)
\end{aligned}$$

Combining (4.28) and (4.29), we have

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_D-\boldsymbol{\beta}_0) = -\boldsymbol{A}_D^{-1}\mathbb{G}_n\,\boldsymbol{J}_D+o_P(1). \qquad (4.30)$$

Therefore $\sqrt{n}(\widehat{\boldsymbol{\beta}}_D-\boldsymbol{\beta}_0)$ converges in distribution to a normal distribution with mean zero and covariance matrix $\boldsymbol{A}_D^{-1}\boldsymbol{\Sigma}_D(\boldsymbol{A}_D^{-1})^T$,$\boldsymbol{\Sigma}_D=E[\boldsymbol{J}_D\boldsymbol{J}_D^T]$, where $\boldsymbol{A}_D$ and $\boldsymbol{J}_D$ are defined in (4.10). By replacing $\boldsymbol{\beta}_0$ with $\widehat{\boldsymbol{\beta}}_D$ and expectations with empirical means in the expressions of $\boldsymbol{A}_D$ and $\boldsymbol{J}_D$, we obtain an estimator $\widehat{\boldsymbol{A}}_D^{-1}\widehat{\boldsymbol{\Sigma}}_D(\widehat{\boldsymbol{A}}_D^{-1})^T$, where $\widehat{\boldsymbol{A}}_D$ and $\widehat{\boldsymbol{\Sigma}}_D$ are defined in (4.11). Using similar arguments as in proofing the consistency and normality of $\widehat{\boldsymbol{\beta}}_D$, we can show that $\widehat{\boldsymbol{A}}_D^{-1}\widehat{\boldsymbol{\Sigma}}_D(\widehat{\boldsymbol{A}}_D^{-1})^T$ is a consistent estimator of $\boldsymbol{A}_D^{-1}\boldsymbol{\Sigma}_D(\boldsymbol{A}_D^{-1})^T$.

Proof of Theorem 4.3.3:

Using empirical process notation, we write

$$\widehat{\mu}_0(t) = \mathbb{P}_n \int_0^t \frac{dN(u)}{\mathbb{P}_n\{Y(u)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(u)}\}},\ t \in [0,\tau].$$

By addition, subtraction and triangle inequality,

$$\begin{aligned}
&\sup_{t\in[0,\tau]} |\widehat{\mu}_0(t) - \mu_0(t)| \\
=\ & \sup_{t\in[0,\tau]} \left| \mathbb{P}_n \int_0^t \frac{\left\{dN(u) - Y(u)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(u)}d\mu_0(u)\right\}}{\mathbb{P}_n\{Y(u)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(u)}\}} \right| \\
=\ & \sup_{t\in[0,\tau]} \left| \mathbb{P}_n \int_0^t \frac{dM_{\boldsymbol{\beta}_0}(u)}{\mathbb{P}_n\{Y(u)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(u)}\}} - \mathbb{P}_n \int_0^t \frac{Y(u)\left\{e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(u)} - e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(u)}\right\}d\mu_0(u)}{\mathbb{P}_n\{Y(u)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(u)}\}} \right| \\
\leq\ & \sup_{t\in[0,\tau]} \left| \mathbb{P}_n \int_0^t \frac{dM_{\boldsymbol{\beta}_0}(u)}{\mathbb{P}_n\{Y(u)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(u)}\}} \right| \\
& + \sup_{t\in[0,\tau]} \left| \mathbb{P}_n \int_0^t \frac{Y(u)\left\{e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(u)} - e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(u)}\right\}d\mu_0(u)}{\mathbb{P}_n\{Y(u)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(u)}\}} \right| \\
\leq\ & \sup_{t\in[0,\tau]} \left| (\mathbb{P}_n - P) \int_0^t \frac{dM_{\boldsymbol{\beta}_0}(u)}{\mathbb{P}_n\{Y(u)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(u)}\}} \right| + \sup_{t\in[0,\tau]} \left| P \int_0^t \frac{dM_{\boldsymbol{\beta}_0}(u)}{\mathbb{P}_n\{Y(u)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(u)}\}} \right| \\
& + \sup_{t\in[0,\tau]} \left| \mathbb{P}_n \int_0^t \frac{Y(u)\left\{e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(u)} - e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(u)}\right\}d\mu_0(u)}{\mathbb{P}_n\{Y(u)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(u)}\}} \right| \qquad (4.31)
\end{aligned}$$

By the strong consistency of $\widehat{\boldsymbol{\beta}}^T$, the third term of (4.31) converges almost surely to 0.

Since $E\left\{dM_{\boldsymbol{\beta}_0}(t)|D \geq t, C, \boldsymbol{Z}\right\} = Y(t)E\{dN^*(t)|D \geq t, \boldsymbol{Z}\} - Y(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}d\mu_0(t) = 0$, the second term of (4.31) becomes

$$
\begin{aligned}
\sup_{t\in[0,\tau]}\left|P\int_0^t \frac{dM_{\boldsymbol{\beta}_0}(u)}{\mathbb{P}_n\{Y(u)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(u)}\}}\right| &= \sup_{t\in[0,\tau]}\left|\int_0^t E\left\{\frac{dM_{\boldsymbol{\beta}_0}(u)}{\mathbb{P}_n\{Y(u)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(u)}\}}\right\}\right| \\
&= \sup_{t\in[0,\tau]}\left|\int_0^t E\left\{\frac{E\left\{dM_{\boldsymbol{\beta}_0}(u)|D \geq t, C, \boldsymbol{Z}\right\}}{\mathbb{P}_n\{Y(u)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(u)}\}}\right\}\right| \\
&= \sup_{t\in[0,\tau]}\left|\int_0^t E\left\{\frac{0}{\mathbb{P}_n\{Y(u)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(u)}\}}\right\}\right| = 0.
\end{aligned}
$$

Using similar arguments as in proofing the consistency of $\widehat{\boldsymbol{\beta}}$, we can show $\{Y(t)dN(t), t \in [0,\tau]\}$, $\{Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)} : \widehat{\boldsymbol{\beta}} \in \mathcal{B}, t \in [0,\tau]\}$ are P-Donsker and P-Glivenko-Cantelli. Thus the first term of (4.31) converges uniformly to zero. We have proven the strong consistency of $\widehat{\mu}_0(t)$, $t \in [0,\tau]$.

Based on the results in (4.31) and Taylor expansion, we have

$$
\begin{aligned}
& \sqrt{n}\left\{\widehat{\mu}_0(t) - \mu_0(t)\right\} \\
= & \sqrt{n}(\mathbb{P}_n - P)\int_0^t \frac{dM_{\boldsymbol{\beta}_0}(u)}{\mathbb{P}_n\{Y(u)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(u)}\}} + \sqrt{n}P\int_0^t \frac{dM_{\boldsymbol{\beta}_0}(u)}{\mathbb{P}_n\{Y(u)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(u)}\}} \\
& -\sqrt{n}\left\{\mathbb{P}_n\int_0^t \frac{Y(u)\boldsymbol{Z}(u)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(u)}}{\mathbb{P}_n\{Y(u)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(u)}\}}d\mu_0(u)\right\}^T(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + o_P(1). \qquad (4.32)
\end{aligned}
$$

Again since $E\left\{dM_{\boldsymbol{\beta}_0}(t)|D \geq t, C, \boldsymbol{Z}\right\} = 0$,

$$
\begin{aligned}
P\int_0^t \frac{dM_{\boldsymbol{\beta}_0}(u)}{\mathbb{P}_n\{Y(u)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(u)}\}} &= \int_0^t E\left\{\frac{dM_{\boldsymbol{\beta}_0}(u)}{\mathbb{P}_n\{Y(u)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(u)}\}}\right\} \\
&= \int_0^t E\left\{\frac{E\left\{dM_{\boldsymbol{\beta}_0}(u)|D \geq t, C, \boldsymbol{Z}\right\}}{\mathbb{P}_n\{Y(u)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(u)}\}}\right\} \\
&= \int_0^t E\left\{\frac{0}{\mathbb{P}_n\{Y(u)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(u)}\}}\right\} = 0.
\end{aligned}
$$

Thus the second term of (4.32) is zero.

Now look at the last term of (4.32), $\{Y(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}(t)} : \boldsymbol{\beta} \in \mathcal{B}, t \in [0,\tau]\}$, $\{Y(u)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}(t)} : \boldsymbol{\beta} \in \mathcal{B}, t \in [0,\tau]\}$ are P-Donsker and thus P-Glivenko-Cantelli, we can replace the operators $\mathbb{P}_n$ with P.

We have shown $\{dN(t), t \in [0,\tau]\}$, $\{Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)} : \widehat{\boldsymbol{\beta}} \in \mathcal{B}, t \in [0,\tau]\}$ are P-Donsker and thus P-Glivenko-Cantelli. We can replace the operator $\mathbb{P}_n$ with P in the denominator in the first term of (4.32) without altering the limiting distribution. By applying the strong consistency of $\widehat{\boldsymbol{\beta}}$ and Lemma 19.24 of Van der Vaart(1998) to the first two terms of (4.32) and the result in (4.25) to the last term of (4.32), we have

$$
\begin{aligned}
& \sqrt{n}\,\{\widehat{\mu}_0(t) - \mu_0(t)\} \\
= \;& \mathbb{G}_n \int_0^t \frac{dM_{\boldsymbol{\beta}_0}(u)}{P\{Y(u)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(u)}\}} + \mathbb{G}_n \left\{ \int_0^t \frac{P\left\{Y(u)\boldsymbol{Z}(u)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(u)}\right\}}{P\{Y(u)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(u)}\}} d\mu_0(u) \right\}^T \boldsymbol{A}^{-1}\boldsymbol{J} + o_P(1) \\
\equiv \;& \mathbb{G}_n K(t) + o_P(1),\; t \in [0,\tau] \qquad (4.33)
\end{aligned}
$$

Therefore the process $\sqrt{n}\,\{\widehat{\mu}_0(t) - \mu_0(t)\}$ converges weakly to a zero-mean Gaussian process with a covariance function $\phi(s,t) = E[K(s)K(t)^T]$, $s,t \in [0,\tau]$, where $K(t)$ is defined as in (4.12). By replacing $\boldsymbol{\beta}_0$ with $\widehat{\boldsymbol{\beta}}$, $\mu_0(\cdot)$ with $\widehat{\mu}_0(\cdot)$ and expectations with empirical means in the expression of $E[K(s)K(t)]$, we obtain an estimator for the asymptotic covariance

$$
\widehat{\phi}(s,t) = n^{-1}\sum_{i=1}^{n} \widehat{K}_i(s)\widehat{K}_i(t),
$$

where $\widehat{K}_i(t)$ is defined as in (4.13). Using similar arguments as in proofing the consistency and normality of $\widehat{\boldsymbol{\beta}}$, we can show that $\widehat{\phi}(s,t)$ is a consistent estimator of $\phi(s,t)$.

Proof of Theorem 4.3.4:

Using the empirical process notations, we write

$$\widehat{\mu}_0^D(t) = \mathbb{P}_n \int_0^t \frac{\widehat{W}^D(u)dN(u)}{\mathbb{P}_n\{\widehat{W}^D(u)e^{\widehat{\boldsymbol{\beta}}_D^T \boldsymbol{Z}(u)}\}},\ t \in [0,\tau].$$

By addition, subtraction and triangle inequality,

$$\begin{aligned}
& \sup_{t\in[0,\tau]} \left|\widehat{\mu}_0^D(t) - \mu_0(t)\right| \\
= & \sup_{t\in[0,\tau]} \left| \mathbb{P}_n \int_0^t \frac{\widehat{W}^D(u)\left\{dN(u) - Y(u)e^{\widehat{\boldsymbol{\beta}}_D^T \boldsymbol{Z}(u)} d\mu_0(u)\right\}}{\mathbb{P}_n\{\widehat{W}^D(u)e^{\widehat{\boldsymbol{\beta}}_D^T \boldsymbol{Z}(u)}\}} \right| \\
= & \sup_{t\in[0,\tau]} \left| \mathbb{P}_n \int_0^t \frac{\widehat{W}^D(u)dM_{\boldsymbol{\beta}_0}(u)}{\mathbb{P}_n\{\widehat{W}^D(u)e^{\widehat{\boldsymbol{\beta}}_D^T \boldsymbol{Z}(u)}\}} - \mathbb{P}_n \int_0^t \frac{\widehat{W}^D(u)\left\{e^{\widehat{\boldsymbol{\beta}}_D^T \boldsymbol{Z}(u)} - e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}(u)}\right\} d\mu_0(u)}{\mathbb{P}_n\{\widehat{W}^D(u)e^{\widehat{\boldsymbol{\beta}}_D^T \boldsymbol{Z}(u)}\}} \right| \\
\leq & \sup_{t\in[0,\tau]} \left| \mathbb{P}_n \int_0^t \frac{\widehat{W}^D(u)dM_{\boldsymbol{\beta}_0}(u)}{\mathbb{P}_n\{\widehat{W}^D(u)e^{\widehat{\boldsymbol{\beta}}_D^T \boldsymbol{Z}(u)}\}} \right| \\
& + \sup_{t\in[0,\tau]} \left| \mathbb{P}_n \int_0^t \frac{\widehat{W}^D(u)\left\{e^{\widehat{\boldsymbol{\beta}}_D^T \boldsymbol{Z}(u)} - e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}(u)}\right\} d\mu_0(u)}{\mathbb{P}_n\{\widehat{W}^D(u)e^{\widehat{\boldsymbol{\beta}}_D^T \boldsymbol{Z}(u)}\}} \right| \\
\leq & \sup_{t\in[0,\tau]} \left| (\mathbb{P}_n - P) \int_0^t \frac{\widehat{W}^D(u)dM_{\boldsymbol{\beta}_0}(u)}{\mathbb{P}_n\{\widehat{W}^D(u)e^{\widehat{\boldsymbol{\beta}}_D^T \boldsymbol{Z}(u)}\}} \right| + \sup_{t\in[0,\tau]} \left| P \int_0^t \frac{\widehat{W}^D(u)dM_{\boldsymbol{\beta}_0}(u)}{\mathbb{P}_n\{\widehat{W}^D(u)e^{\widehat{\boldsymbol{\beta}}_D^T \boldsymbol{Z}(u)}\}} \right| \\
& + \sup_{t\in[0,\tau]} \left| \mathbb{P}_n \int_0^t \frac{\widehat{W}^D(u)\left\{e^{\widehat{\boldsymbol{\beta}}_D^T \boldsymbol{Z}(u)} - e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}(u)}\right\} d\mu_0(u)}{\mathbb{P}_n\{\widehat{W}^D(u)e^{\widehat{\boldsymbol{\beta}}_D^T \boldsymbol{Z}(u)}\}} \right| \qquad (4.34)
\end{aligned}$$

By the strong consistency of $\widehat{\boldsymbol{\beta}}_D^T$, the third term of (4.34) converges almost surely to 0.

Since $E\left\{dM_{\boldsymbol{\beta}_0}(t)|D \geq t, C, \boldsymbol{Z}\right\} = Y(t)E\left\{dN^*(t)|D \geq t, \boldsymbol{Z}\right\} - Y(t)e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}(t)}d\mu_0(t) = 0$, the second term of (4.34) becomes

$$\begin{aligned}
\sup_{t\in[0,\tau]}\left|P\int_0^t \frac{\widehat{W}^D(u)dM_{\boldsymbol{\beta}_0}(u)}{\mathbb{P}_n\{\widehat{W}^D(u)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(u)}\}}\right| &= \sup_{t\in[0,\tau]}\left|\int_0^t E\left\{\frac{\widehat{W}^D(u)dM_{\boldsymbol{\beta}_0}(u)}{\mathbb{P}_n\{\widehat{W}^D(u)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(u)}\}}\right\}\right| \\
&= \sup_{t\in[0,\tau]}\left|\int_0^t E\left\{\frac{\widehat{W}^D(u)E\left\{dM_{\boldsymbol{\beta}_0}(u)|D \geq t, C, \boldsymbol{Z}\right\}}{\mathbb{P}_n\{\widehat{W}^D(u)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(u)}\}}\right\}\right| \\
&= \sup_{t\in[0,\tau]}\left|\int_0^t E\left\{\frac{\widehat{W}^D(u)\cdot 0}{\mathbb{P}_n\{\widehat{W}^D(u)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(u)}\}}\right\}\right| = 0.
\end{aligned}$$

Using similar arguments as in proofing the consistency of $\widehat{\boldsymbol{\beta}}_D$, we can show $\{\widehat{W}^D(t)dN(t), t \in [0,\tau]\}$, $\{\widehat{W}^D(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)} : \widehat{\boldsymbol{\beta}}_D \in \mathcal{B}, t \in [0,\tau]\}$ are P-Donsker and P-Glivenko-Cantelli. The first term of (4.34) converges uniformly to zero. Thus $\widehat{\mu}_0^D(t) \xrightarrow{a.s.} \mu_0(t)$. We have proven the strong consistency of $\widehat{\mu}_0^D(t)$, $t \in [0,\tau]$.

Based on the results in (4.34) and Taylor expansion, we have

$$\begin{aligned}
&\sqrt{n}\left\{\widehat{\mu}_0^D(t) - \mu_0(t)\right\} \\
= \ &\sqrt{n}(\mathbb{P}_n - P)\int_0^t \frac{\widehat{W}^D(u)dM_{\boldsymbol{\beta}_0}(u)}{\mathbb{P}_n\{\widehat{W}^D(u)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(u)}\}} + \sqrt{n}P\int_0^t \frac{\widehat{W}^D(u)dM_{\boldsymbol{\beta}_0}(u)}{\mathbb{P}_n\{\widehat{W}^D(u)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(u)}\}} \\
&-\sqrt{n}\left\{\mathbb{P}_n\int_0^t \frac{\widehat{W}^D(u)\boldsymbol{Z}(u)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(u)}}{\mathbb{P}_n\{\widehat{W}^D(u)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(u)}\}}d\mu_0(u)\right\}^T(\widehat{\boldsymbol{\beta}}_D - \boldsymbol{\beta}_0) + o_P(1) \qquad (4.35)
\end{aligned}$$

Since $E\left\{dM_{\boldsymbol{\beta}_0}(t)|D \geq t, C, \boldsymbol{Z}\right\} = 0$,

$$\begin{aligned}
P\int_0^t \frac{\widehat{W}^D(u)dM_{\boldsymbol{\beta}_0}(u)}{\mathbb{P}_n\{\widehat{W}^D(u)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(u)}\}} &= \int_0^t E\left\{\frac{\widehat{W}^D(u)dM_{\boldsymbol{\beta}_0}(u)}{\mathbb{P}_n\{\widehat{W}^D(u)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(u)}\}}\right\} \\
&= \int_0^t E\left\{\frac{\widehat{W}^D(u)E\left\{dM_{\boldsymbol{\beta}_0}(u)|D \geq t, C, \boldsymbol{Z}\right\}}{\mathbb{P}_n\{\widehat{W}^D(u)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(u)}\}}\right\} \\
&= \int_0^t E\left\{\frac{\widehat{W}^D(u)\cdot 0}{\mathbb{P}_n\{\widehat{W}^D(u)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(u)}\}}\right\} = 0.
\end{aligned}$$

Thus the second term of (4.35) is zero.

Now look at the last term of (4.35), $\{\widehat{W}^D(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}(t)} : \boldsymbol{\beta} \in \mathcal{B}, t \in [0,\tau]\}$ and $\{\widehat{W}^D(t)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}(t)} : \boldsymbol{\beta} \in \mathcal{B}, t \in [0,\tau]\}$ are P-Donsker and thus P-Glivenko-Cantelli, we can replace the operators $\mathbb{P}_n$ with P. We have shown $\{\widehat{W}^D(t)dN(t), t \in [0,\tau]\}$, $\{\widehat{W}^D(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)} : \widehat{\boldsymbol{\beta}}_D \in \mathcal{B}, t \in [0,\tau]\}$ are P-Donsker and thus P-Glivenko-Cantelli. We can replace the operator $\mathbb{P}_n$ with P in the denominator in the first term of (4.35) without altering the limiting distribution. By applying the strong consistency of $\widehat{\boldsymbol{\gamma}}_D$, $\widehat{\Lambda}^D(\cdot)$, $\widehat{\boldsymbol{\beta}}_D$ and Lemma 19.24 of Van der Vaart(1998) to the first term of (4.35) and the result in (4.30) to the last term of (4.35), we have

$$\begin{aligned}
\sqrt{n}\left\{\widehat{\mu}_0^D(t) - \mu_0(t)\right\} &= \mathbb{G}_n \int_0^t \frac{W^D(u)dM_{\boldsymbol{\beta}_0}(u)}{P\{W^D(u)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(u)}\}} \\
&\quad + \mathbb{G}_n \left\{ \int_0^t \frac{P\left\{W^D(u)\boldsymbol{Z}(u)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(u)}\right\}}{P\{W^D(u)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(u)}\}} d\mu_0(u) \right\}^T \boldsymbol{A}_D^{-1}\boldsymbol{J}_D + o_P(1) \\
&\equiv \mathbb{G}_n K_D(t) + o_P(1),\ t \in [0,\tau]. \qquad (4.36)
\end{aligned}$$

Therefore the process $\sqrt{n}\left\{\widehat{\mu}_0^D(t) - \mu_0(t)\right\}$ converges weakly to a zero-mean Gaussian process with a covariance function $\phi_D(s,t) = E[K_D(s)K_D(t)^T]$, $s,t \in [0,\tau]$, where $K_D(t)$ is defined as in (4.14). By replacing $\boldsymbol{\beta}_0$ with $\widehat{\boldsymbol{\beta}}_D$, $\mu_0(\cdot)$ with $\widehat{\mu}_0^D(\cdot)$ and expectations with empirical means in the expression of $E[K_D(s)K_D(t)]$, we obtain an estimator for the asymptotic covariance,

$$\widehat{\phi}_D(s,t) = n^{-1}\sum_{i=1}^n \widehat{K}_{D,i}(s)\widehat{K}_{D,i}(t),$$

where $\widehat{K}_{D,i}(t)$ is defined as in (4.15). Using similar argument as in proofing the consistency and normality of $\widehat{\boldsymbol{\beta}}_D$, we can show that $\widehat{\phi}_D(s,t)$ is a consistent estimator of $\phi_D(s,t)$.

Proof of Theorem 4.3.5:

$$V_G(t) = n^{-1/2} \sum_{i=1}^{n} \widehat{K}_i(t) G_i,$$

where

$$\widehat{K}_i(t) \quad = \quad \int_0^t \frac{d\widehat{M}_i(u)}{Q^{(0)}(t; \widehat{\boldsymbol{\beta}})} + \left\{ \int_0^t \frac{\boldsymbol{Q}^{(1)}(u; \widehat{\boldsymbol{\beta}})}{Q^{(0)}(u; \widehat{\boldsymbol{\beta}})} d\widehat{\mu}_0(u) \right\}^T \widehat{\boldsymbol{A}}^{-1} \widehat{\boldsymbol{J}}_i,$$

$\widehat{\boldsymbol{A}}$, $\widehat{\boldsymbol{J}}_i$, $\widehat{M}_i(t)$ and $\boldsymbol{Q}^{(k)}(t; \boldsymbol{\beta})$, $k = 0, 1$ are defined in (4.9).

By the Donsker property of $\{Y(t)\boldsymbol{Z}(t)dN(t), t \in [0, \tau]\}$, $\{Y(t)dN(t), t \in [0, \tau]\}$, $\{Y(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)} : \widehat{\boldsymbol{\beta}} \in \mathcal{B}, t \in [0, \tau]\}$, $\{Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)} : \widehat{\boldsymbol{\beta}} \in \mathcal{B}, t \in [0, \tau]\}$, we can show $\widehat{K}_i(t)$ belongs to a Donsker class. $Var(G_i) = 1$ since $(G_1, \cdots, G_n)$ are independent standard normal variables. By Theorem 3.6.13 of Van der Vaart and Wellner(1996), conditionally on the observed data,

$$V_G(t) = \mathbb{G}_n K(t) + o_P(1).$$

Therefore, the process $V_G(t)$, converges weakly to a zero-mean Gaussian process with a covariance function $\phi(s, t) = E[K(s)K(t)]$, where $\phi(s, t)$ is defined as (4.12).

<u>Proof of Theorem 4.3.6</u>:

$$V_G^D(t) = n^{-1/2} \sum_{i=1}^{n} \widehat{K}_{D,i}(t) G_i,$$

where

$$\widehat{K}_{D,i}(t) \quad = \quad \int_0^t \frac{\widehat{W}_i^D(u) d\widehat{M}_i^D(u)}{Q_D^{(0)}(t; \widehat{\boldsymbol{\beta}}_D)} + \left\{ \int_0^t \frac{\boldsymbol{Q}_D^{(1)}(u; \widehat{\boldsymbol{\beta}}_D)}{Q_D^{(0)}(u; \widehat{\boldsymbol{\beta}}_D)} d\widehat{\mu}_0^D(u) \right\}^T \widehat{\boldsymbol{A}}_D^{-1} \widehat{\boldsymbol{J}}_{D,i},$$

$\widehat{\boldsymbol{A}}_D$, $\widehat{\boldsymbol{J}}_{D,i}$, $\widehat{M}_i^D(t)$ and $\boldsymbol{Q}_D^{(k)}(t; \boldsymbol{\beta})$, $k = 0, 1$ are defined in (4.11).

By the Donsker property of $\{\widehat{W}^D(t)\boldsymbol{Z}(t)dN(t), t \in [0, \tau]\}$, $\{\widehat{W}^D(t)dN(t), t \in [0, \tau]\}$,

$\{\widehat{W}^D(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)} : \widehat{\boldsymbol{\beta}}_D \in \mathcal{B}, t \in [0,\tau]\}$, $\{\widehat{W}^D(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}_D^T\boldsymbol{Z}(t)} : \widehat{\boldsymbol{\beta}}_D \in \mathcal{B}, t \in [0,\tau]\}$, we can show $\widehat{K}_{D,i}(t)$ belongs to a Donsker class. $Var(G_i) = 1$ since $(G_1, \cdots, G_n)$ are independent standard normal variables. By Theorem 3.6.13 of Van der Vaart and Wellner(1996), conditionally on the observed data,

$$V_G^D(t) = \mathbb{G}_n K_D(t) + o_P(1).$$

Therefore, the process $V_G^D(t)$, converges weakly to a zero-mean Gaussian process with a covariance function $\phi_D(s,t) = E[K_D(s)K_D(t)]$, where $\phi_D(s,t)$ is defined as in (4.14).

Table 4.1: Summary of simulation results, $Z \sim Bernoulli(0.5)$, $\beta_0 = 0$

		$\beta_0 = 0$							
		Proposed Method A				Proposed Method B			
n	ρ_0	Bias	SE	SEE	CP	Bias	SE	SEE	CP
200	0	-0.00	0.150	0.143	0.936	-0.02	0.258	0.199	0.868
200	1	0.01	0.152	0.153	0.947	0.00	0.181	0.175	0.939
200	4	0.01	0.222	0.222	0.953	0.01	0.219	0.219	0.950
200	8	0.02	0.296	0.291	0.944	0.02	0.291	0.287	0.943
400	0	-0.00	0.105	0.103	0.950	-0.00	0.221	0.166	0.873
400	1	0.00	0.109	0.109	0.949	0.00	0.135	0.128	0.946
400	4	-0.00	0.161	0.157	0.947	0.00	0.159	0.155	0.944
400	8	-0.00	0.217	0.207	0.941	-0.00	0.213	0.204	0.944
600	0	0.00	0.084	0.085	0.963	-0.00	0.175	0.148	0.898
600	1	-0.00	0.092	0.089	0.940	-0.00	0.112	0.106	0.946
600	4	-0.00	0.129	0.129	0.952	-0.00	0.128	0.127	0.947
600	8	-0.00	0.171	0.168	0.943	-0.00	0.167	0.166	0.937

Note: Bias is the sample average of the estimator of β minus β;
SE is the sample standard deviation of the estimator of β ;
SEE is the sample average of the standard error estimator;
CP is the coverage rate of the Wald 95% confidence interval.

Table 4.2: Summary of simulation results, $Z \sim Bernoulli(0.5)$, $\beta_0 = 0.2$

		$\beta_0 = 0.2$							
		Proposed Method A				Proposed Method B			
n	ρ_0	Bias	SE	SEE	CP	Bias	SE	SEE	CP
200	0	-0.00	0.148	0.143	0.933	-0.01	0.251	0.198	0.888
200	1	0.01	0.156	0.152	0.943	0.01	0.184	0.171	0.934
200	4	0.01	0.217	0.219	0.957	0.01	0.214	0.216	0.957
200	8	0.01	0.294	0.287	0.952	0.01	0.288	0.283	0.958
400	0	0.00	0.108	0.102	0.934	0.00	0.214	0.163	0.875
400	1	0.00	0.108	0.107	0.942	0.00	0.129	0.125	0.940
400	4	-0.00	0.156	0.155	0.943	-0.00	0.153	0.153	0.950
400	8	-0.00	0.207	0.204	0.954	-0.00	0.204	0.201	0.952
600	0	0.00	0.084	0.085	0.959	-0.00	0.171	0.147	0.908
600	1	-0.00	0.091	0.088	0.942	-0.00	0.109	0.104	0.932
600	4	-0.00	0.126	0.127	0.955	-0.00	0.124	0.125	0.958
600	8	-0.01	0.167	0.166	0.937	-0.01	0.165	0.164	0.938

Note: Bias is the sample average of the estimator of β minus β;
SE is the sample standard deviation of the estimator of β ;
SEE is the sample average of the standard error estimator;
CP is the coverage rate of the Wald 95% confidence interval.

Table 4.3: Summary of simulation results, $Z \sim Bernoulli(0.5)$, $\beta_0 = 0.5$

		$\beta_0 = 0.5$							
		Proposed Method A				Proposed Method B			
n	ρ_0	Bias	SE	SEE	CP	Bias	SE	SEE	CP
200	0	-0.00	0.150	0.141	0.932	-0.01	0.246	0.193	0.882
200	1	0.00	0.152	0.150	0.946	0.01	0.176	0.168	0.939
200	4	0.02	0.220	0.216	0.946	0.02	0.218	0.212	0.944
200	8	0.01	0.293	0.282	0.935	0.01	0.287	0.278	0.937
400	0	0.00	0.104	0.102	0.941	-0.00	0.208	0.160	0.870
400	1	-0.00	0.105	0.106	0.953	-0.00	0.128	0.122	0.939
400	4	-0.00	0.153	0.153	0.954	-0.00	0.149	0.150	0.959
400	8	-0.00	0.208	0.201	0.934	-0.00	0.204	0.197	0.937
600	0	0.00	0.085	0.085	0.952	-0.01	0.174	0.144	0.884
600	1	-0.00	0.090	0.087	0.941	-0.00	0.104	0.101	0.953
600	4	-0.00	0.124	0.125	0.955	-0.00	0.122	0.123	0.951
600	8	-0.01	0.164	0.164	0.943	-0.01	0.160	0.161	0.941

Note: Bias is the sample average of the estimator of β minus β;
SE is the sample standard deviation of the estimator of β ;
SEE is the sample average of the standard error estimator;
CP is the coverage rate of the Wald 95% confidence interval.

Table 4.4: Summary of simulation results, $Z \sim Uniform(0, 2)$, $\beta_0 = 0$

		$\beta_0 = 0$							
		Proposed Method A				Proposed Method B			
n	ρ_0	Bias	SE	SEE	CP	Bias	SE	SEE	CP
200	0	0.00	0.133	0.124	0.933	-0.01	0.223	0.205	0.903
200	1	0.01	0.134	0.133	0.940	0.01	0.168	0.154	0.933
200	4	0.01	0.193	0.186	0.942	0.01	0.194	0.194	0.974
200	8	0.01	0.245	0.240	0.945	0.01	0.241	0.237	0.942
400	0	0.00	0.095	0.090	0.936	-0.01	0.193	0.145	0.877
400	1	0.00	0.095	0.095	0.952	0.00	0.124	0.117	0.935
400	4	-0.00	0.135	0.133	0.940	-0.00	0.133	0.135	0.957
400	8	-0.01	0.180	0.173	0.934	-0.01	0.178	0.170	0.934
600	0	0.00	0.076	0.075	0.959	-0.01	0.170	0.129	0.880
600	1	0.00	0.081	0.077	0.938	0.00	0.106	0.096	0.919
600	4	-0.00	0.105	0.108	0.965	-0.00	0.104	0.109	0.982
600	8	-0.00	0.140	0.143	0.945	-0.00	0.140	0.138	0.948

Note: Bias is the sample average of the estimator of β minus β;
SE is the sample standard deviation of the estimator of β ;
SEE is the sample average of the standard error estimator;
CP is the coverage rate of the Wald 95% confidence interval.

Table 4.5: Summary of simulation results, $Z \sim Uniform(0, 2)$, $\beta_0 = 0.2$

		$\beta_0 = 0.2$							
		Proposed Method A				Proposed Method B			
n	ρ_0	Bias	SE	SEE	CP	Bias	SE	SEE	CP
200	0	0.00	0.130	0.122	0.934	-0.01	0.220	0.195	0.939
200	1	0.01	0.134	0.129	0.936	0.01	0.164	0.150	0.916
200	4	0.01	0.186	0.180	0.943	0.01	0.185	0.209	0.971
200	8	0.02	0.244	0.233	0.942	0.02	0.240	0.229	0.943
400	0	0.00	0.092	0.089	0.951	-0.00	0.191	0.141	0.877
400	1	0.00	0.093	0.092	0.942	0.00	0.124	0.111	0.928
400	4	-0.00	0.131	0.129	0.945	-0.00	0.129	0.140	0.973
400	8	-0.00	0.174	0.167	0.931	-0.00	0.171	0.164	0.934
600	0	-0.00	0.077	0.074	0.935	-0.00	0.168	0.127	0.865
600	1	0.00	0.076	0.075	0.946	-0.00	0.101	0.093	0.929
600	4	-0.00	0.106	0.105	0.948	-0.00	0.105	0.113	0.977
600	8	-0.00	0.135	0.135	0.945	-0.00	0.133	0.133	0.943

Note: Bias is the sample average of the estimator of β minus β;
SE is the sample standard deviation of the estimator of β ;
SEE is the sample average of the standard error estimator;
CP is the coverage rate of the Wald 95% confidence interval.

Table 4.6: Summary of simulation results, $Z \sim Uniform(0, 2)$, $\beta_0 = 0.5$

		$\beta_0 = 0.5$							
		Proposed Method A				Proposed Method B			
n	ρ_0	Bias	SE	SEE	CP	Bias	SE	SEE	CP
200	0	0.00	0.128	0.121	0.929	-0.01	0.217	0.247	0.928
200	1	0.00	0.129	0.125	0.948	0.01	0.156	0.144	0.914
200	4	0.01	0.182	0.174	0.943	0.01	0.180	0.327	0.981
200	8	0.01	0.228	0.225	0.943	0.01	0.225	0.221	0.944
400	0	0.00	0.092	0.088	0.937	-0.00	0.195	0.139	0.850
400	1	0.00	0.089	0.090	0.949	0.00	0.115	0.108	0.932
400	4	-0.00	0.125	0.125	0.949	-0.00	0.124	0.208	0.983
400	8	-0.00	0.163	0.162	0.946	0.00	0.160	0.159	0.950
600	0	0.00	0.076	0.073	0.940	-0.01	0.167	0.125	0.869
600	1	-0.00	0.076	0.073	0.933	0.00	0.099	0.091	0.924
600	4	-0.00	0.099	0.101	0.951	-0.00	0.099	0.165	0.983
600	8	-0.00	0.128	0.131	0.952	-0.00	0.126	0.129	0.953

Note: Bias is the sample average of the estimator of β minus β;
SE is the sample standard deviation of the estimator of β ;
SEE is the sample average of the standard error estimator;
CP is the coverage rate of the Wald 95% confidence interval.

Table 4.7: Summary of simulation results, $Z \sim N(0,1)$, $\beta_0 = 0$

		$\beta_0 = 0$							
		Proposed Method A				Proposed Method B			
n	ρ_0	Bias	SE	SEE	CP	Bias	SE	SEE	CP
200	0	0.00	0.076	0.070	0.922	-0.01	0.122	0.097	0.877
200	1	0.01	0.084	0.081	0.938	0.00	0.093	0.087	0.927
200	4	0.01	0.127	0.119	0.930	0.01	0.123	0.117	0.933
200	8	0.00	0.166	0.153	0.926	0.00	0.162	0.150	0.929
400	0	-0.00	0.055	0.051	0.923	-0.00	0.098	0.079	0.894
400	1	-0.00	0.060	0.059	0.951	-0.00	0.071	0.065	0.929
400	4	-0.00	0.087	0.086	0.948	-0.00	0.085	0.083	0.943
400	8	-0.00	0.115	0.111	0.938	-0.01	0.113	0.108	0.941
600	0	0.00	0.043	0.043	0.956	0.00	0.086	0.070	0.896
600	1	0.00	0.049	0.048	0.938	-0.00	0.058	0.054	0.931
600	4	0.00	0.072	0.070	0.943	0.00	0.070	0.068	0.940
600	8	0.00	0.097	0.091	0.933	0.00	0.094	0.089	0.936

Note: Bias is the sample average of the estimator of β minus β;
SE is the sample standard deviation of the estimator of β ;
SEE is the sample average of the standard error estimator;
CP is the coverage rate of the Wald 95% confidence interval.

Table 4.8: Summary of simulation results, $Z \sim N(0,1)$, $\beta_0 = 0.2$

		$\beta_0 = 0.2$							
		Proposed Method A				Proposed Method B			
n	ρ_0	Bias	SE	SEE	CP	Bias	SE	SEE	CP
200	0	-0.00	0.073	0.070	0.934	-0.01	0.122	0.097	0.876
200	1	0.00	0.084	0.078	0.931	0.00	0.097	0.086	0.919
200	4	0.01	0.120	0.114	0.931	0.00	0.119	0.113	0.927
200	8	-0.00	0.158	0.148	0.928	-0.00	0.155	0.146	0.929
400	0	-0.00	0.053	0.051	0.939	-0.00	0.101	0.078	0.881
400	1	-0.00	0.058	0.056	0.938	-0.00	0.071	0.064	0.916
400	4	-0.00	0.081	0.082	0.948	-0.00	0.080	0.081	0.947
400	8	-0.00	0.108	0.106	0.940	-0.00	0.106	0.104	0.937
600	0	0.00	0.041	0.042	0.955	0.00	0.089	0.071	0.895
600	1	-0.00	0.046	0.046	0.950	-0.00	0.057	0.054	0.936
600	4	-0.00	0.068	0.067	0.941	-0.00	0.067	0.066	0.944
600	8	-0.00	0.090	0.087	0.937	-0.00	0.089	0.085	0.934

Note: Bias is the sample average of the estimator of β minus β;
SE is the sample standard deviation of the estimator of β ;
SEE is the sample average of the standard error estimator;
CP is the coverage rate of the Wald 95% confidence interval.

Table 4.9: Summary of simulation results, $Z \sim N(0,1)$, $\beta_0 = 0.5$

		$\beta_0 = 0.5$							
		Proposed Method A				Proposed Method B			
n	ρ_0	Bias	SE	SEE	CP	Bias	SE	SEE	CP
200	0	-0.00	0.075	0.072	0.929	-0.01	0.126	0.099	0.868
200	1	-0.00	0.085	0.080	0.925	-0.00	0.102	0.089	0.902
200	4	0.00	0.125	0.115	0.914	0.00	0.126	0.114	0.916
200	8	-0.01	0.155	0.148	0.932	-0.01	0.154	0.146	0.922
400	0	-0.00	0.055	0.053	0.943	-0.00	0.107	0.081	0.864
400	1	-0.00	0.059	0.058	0.932	-0.00	0.077	0.067	0.910
400	4	-0.00	0.085	0.083	0.938	-0.00	0.087	0.083	0.929
400	8	-0.01	0.111	0.107	0.935	-0.00	0.110	0.106	0.934
600	0	-0.00	0.044	0.045	0.943	-0.00	0.094	0.074	0.879
600	1	0.00	0.048	0.048	0.947	-0.00	0.062	0.057	0.927
600	4	-0.00	0.070	0.067	0.938	-0.00	0.071	0.067	0.940
600	8	-0.00	0.094	0.087	0.924	-0.00	0.094	0.087	0.921

Note: Bias is the sample average of the estimator of β minus β;
SE is the sample standard deviation of the estimator of β ;
SEE is the sample average of the standard error estimator;
CP is the coverage rate of the Wald 95% confidence interval.

Table 4.10: SOLVD Prevention Trial: Hospitalizations and survival experience

		Number of hospitalizations								
Treatment	Number of subjects	0	1	2	3	4	5	6	≥ 7	Number of Deaths
placebo	2116	1148	506	195	135	55	31	19	27	334
enalapril	2110	1226	487	209	79	50	18	18	23	312

Table 4.11: SOLVD Prevention Trial: Regression analysis for the effects of enalapril on hospitalizations and death of patients, using Method B

	Repeated Hospitalizations			Death		
Covariate	Estimate	SE	p-value	Estimate	SE	p-value
TRT	-0.183	0.055	0.001	-0.102	0.079	0.199
EF	-0.014	0.005	0.006	-0.038	0.007	< 0.001
AGE	0.009	0.003	0.005	0.031	0.004	< 0.001
GENDER	-0.145	0.085	0.087	-0.002	0.127	0.989
CENTER 2	0.135	0.162	0.406	-0.019	0.245	0.938
CENTER 3	0.252	0.163	0.122	0.091	0.242	0.708
CENTER 4	0.094	0.191	0.621	-0.107	0.272	0.695
CENTER 5	0.006	0.176	0.972	0.324	0.258	0.209
CENTER 6	-0.070	0.147	0.636	-0.077	0.218	0.723
CENTER 7	0.302	0.242	0.213	0.716	0.263	0.006
CENTER 8	0.336	0.165	0.041	0.454	0.230	0.049
CENTER 9	0.482	0.173	0.005	-0.030	0.269	0.913
CENTER 10	0.072	0.180	0.688	-0.708	0.306	0.021
CENTER 11	-0.006	0.180	0.975	0.552	0.232	0.018
CENTER 12	-0.345	0.192	0.072	-0.014	0.240	0.955
CENTER 13	0.224	0.172	0.191	-0.251	0.255	0.324
CENTER 14	-0.145	0.173	0.402	-0.324	0.255	0.204
CENTER 15	0.232	0.178	0.194	0.268	0.275	0.330
CENTER 16	-0.225	0.159	0.158	-0.412	0.251	0.101
CENTER 17	-0.712	0.299	0.017	-0.051	0.417	0.904
CENTER 18	-0.334	0.180	0.063	-0.413	0.280	0.140
CENTER 19	0.198	0.236	0.400	0.386	0.263	0.142
CENTER 20	0.021	0.182	0.909	-0.297	0.299	0.320
CENTER 21	-0.090	0.188	0.632	0.005	0.284	0.985
CENTER 22	-0.759	0.291	0.009	-0.162	0.328	0.622
CENTER 23	0.095	0.213	0.656	0.058	0.266	0.827

Note: TRT is coded as 1 for enalapril, 0 for placebo; GENDER is coded as 1 for male, 0 for female;
Estimate is the estimated regression coefficient;SE is the estimated standard error;
p-value is the two-sided p-value.

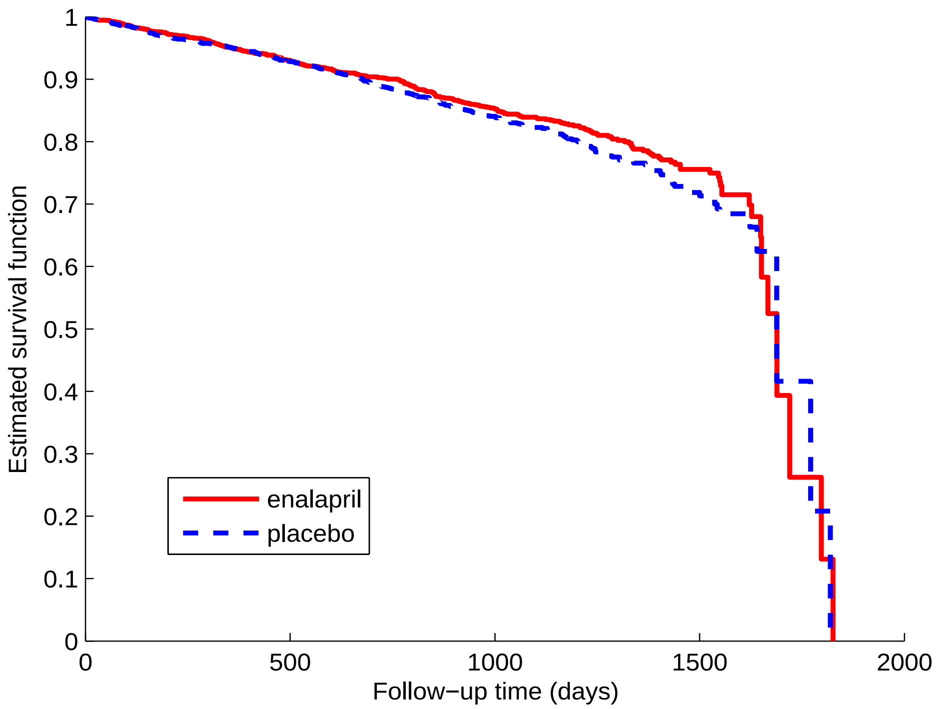

Figure 4.1: SOLVD Prevention Trial Data: Kaplan-Meier survival curves for the placebo group(shown by dashed lines) and the enalapril group (shown by solid lines)

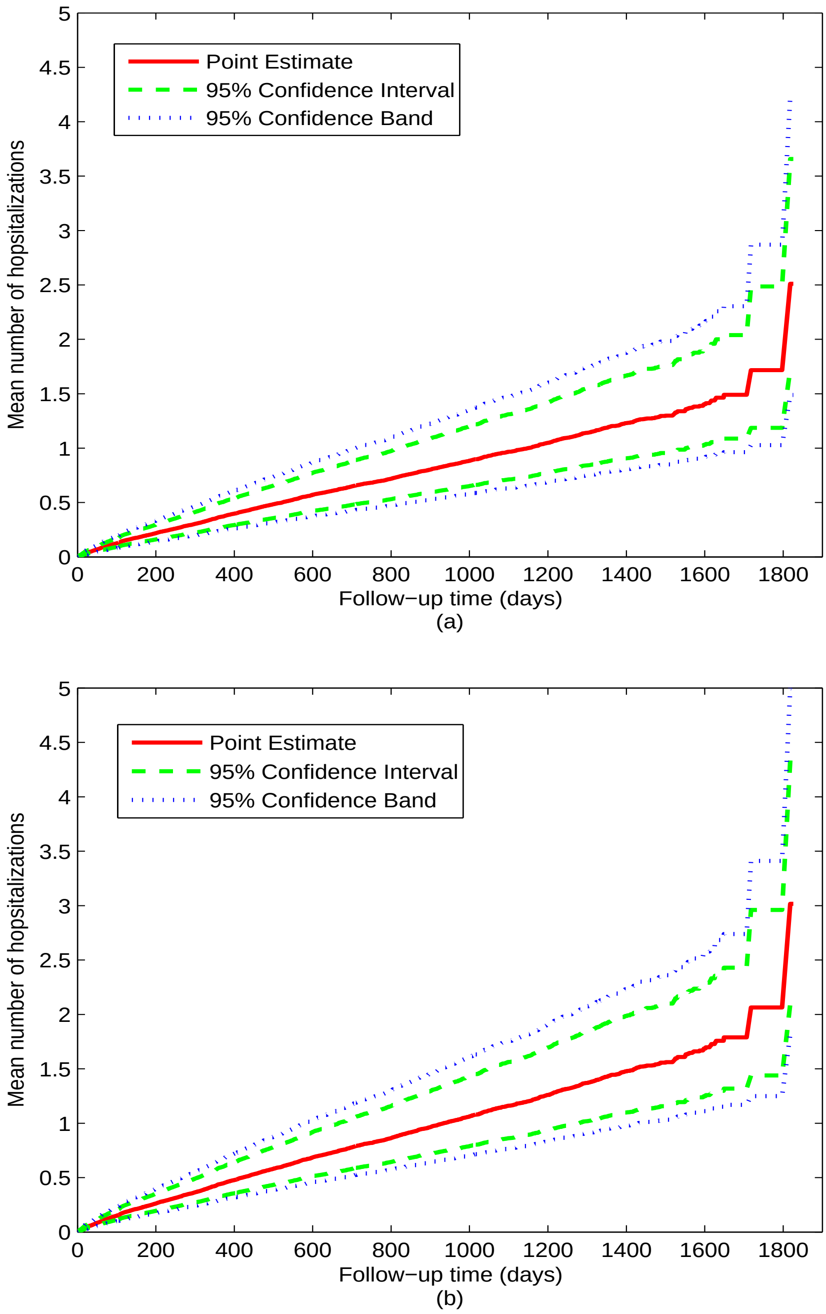

Figure 4.2: SOLVD Prevention Trial Data: Estimated mean frequency of hospitalizations for 59-year-old female patients with 28.3% baseline ejection fraction (a) receiving enalapril (b) receiving placebo. The confidence bands are based on 1000 simulations

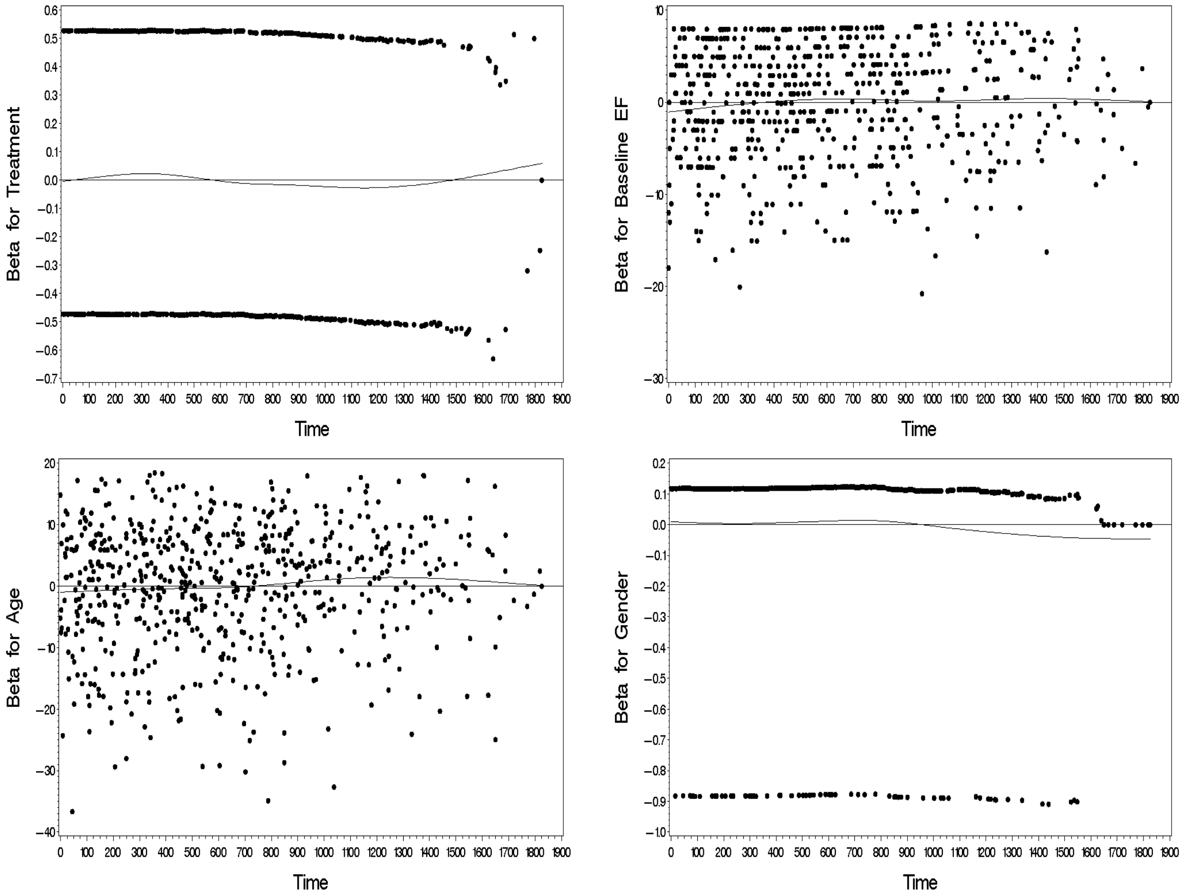

Figure 4.3: Schoenfeld residuals from censoring PH model versus covariates.

Chapter 5

Conditional Recurrent Event Rate Model with Incomplete Covariate Measurements

5.1 Introduction

In many longitudinal follow-up studies, certain components of the covariates may be incomplete for some subjects. For example, in the Studies of Left Ventricular Dysfunction(SOLVD, 1991) Prevention Trial, 605 patients have missing covariate measurements out of a total of 4228 patients. Baseline cardiothoracic ratio is missing for 604 patients, and baseline ejection fraction is missing for 2 patients. Interest focuses on assessing the average treatment effect on repeated hospitalizations after adjusting for certain covariates.

Several authors considered various methods to deal with missing covariate data under the framework of the Cox proportional hazard model(Cox, 1972). Under the assumptions of missing at random(MAR) and the missing pattern is monotonic, Prentice(1982) suggested using the partial likelihood based on the induced intensity process. Zhou and Pepe(1995) proposed a method based on the estimated induced partial likelihood. Their method yields more efficient estimates than those based on the maximum partial likelihood approach from the complete-case analysis. By assuming the data are missing completely at random(MCAR), Lin and Ying (1993) proposed approximate partial likelihood estimates that accommodate any pattern of missing data. Paik and Tsai (1997) proposed an imputation method that can be used when data are MAR. Chen and Little (1999) proposed a nonparametric maximum likelihood procedure when data are MAR. But their method works only when the missing covariates are all discrete or all normally distributed. Herring and Ibrahim (2001) developed a likelihood-based methodology for MAR covariates based on partial likelihood using an EM-type algorithm.

In Chapter 4, we studied a proportional rate model for recurrent events and applied it to the aforementioned SOLVD Prevention Trial data. In the presence of missing covariate measurements, we adopt the idea of Lin and Yin(1993) and propose an estimating procedure for model (4.1). The remainder of this article is organized as follows. In Section 2, we introduce notations and presents the proposed methods. In Section 3, we describe the asymptotic properties of the resulting estimators. In Section 4, we conduct simulation studies to evaluate the proposed methods in finite samples and report the results. In Section 5, we illustrate the method by applying it to the SOLVD(Studies of Left Ventricular Dysfunction, SOLVD Investigators, 1991) Prevention Trial data. We conclude with some discussion in Section 6. Proofs of the theorems can be found in Section 7.

5.2 Model and estimation

Let $N^*(t)$ be the cumulative number of recurrent events that occur over the time interval $[0, t]$ in the absence of any censoring. In most applications, the follow-up time is subject to censoring. Let C and D denote, respectively, the censoring time and the survival time, and let $\boldsymbol{Z}(.)$ be a $p \times 1$-vector of covariates. Throughout the chapter, we assume that C may depend on $\boldsymbol{Z}(.)$ but is independent of $\{N^*(.)\}$ and D given $\boldsymbol{Z}(.)$. The survival time D is allowed to depend on $\{N^*(.)\}$, even conditionally on $\boldsymbol{Z}$. In general only the minimum of C and D is known and $\{N^*(.)\}$ can only be observed up to $\min(C, D)$. Further, suppose the study duration is τ and define $X = D \wedge C = \min(D, C)$, $N(t) = N^*(t \wedge D \wedge C) = I(X \geq t)\, N^*(t)$, $Y(t) = I(X \geq t)$ and $\delta = I(D \leq C)$.

For $i = 1, \cdots, n$, suppose $\boldsymbol{Z}_i(.)$ may not be completely observed. Let $\boldsymbol{H}_i(.)$ be a $p \times p$ diagonal matrix with diagonal elements $\{H_{1i}(.), \cdots, H_{pi}(.)\}$, where $H_{ji}(t) = 1$ if $Z_{ji}(t)$ is available at time t and 0 otherwise, for $j = 1, \cdots, p$. Define $h_i(.) = I(H_{ji}(.) = 1, \text{ for all } j = 1, \cdots, p)$, an indicator function for whether $\boldsymbol{Z}_i(.)$ is completely observed or not. Conditional on $\{X_i \geq t\}$, we assume $H_{ji}(t)$, $j = 1, \cdots, p$ are independent of all other random variables. This corresponds to the missing completely at random(MCAR) assumption of Rubin(1976). For a random sample of n subjects, the observed data consist of $\{N_i(.), X_i, \delta_i, \boldsymbol{Z}_i, \boldsymbol{H}_i(.), h_i(.)\}$, $i = 1, \cdots, n$. We wish to formulate the effect of $\boldsymbol{Z}(.)$ on $\{N^*(.)\}$ without specifying the dependence structure between recurrent events and terminal events, or among recurrent events.

For the underlying recurrent process, we propose the following multiplicative conditional rate model when terminating events are present,

$$E\{dN^*(t)|D \geq t, \boldsymbol{Z}\} = I(D \geq t)e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}(t)} d\mu_0(t), \tag{5.1}$$

where $\boldsymbol{Z}(.)$ is a $p \times 1$ vector of covariates, $\boldsymbol{\beta}_0$ is a $p \times 1$ vector of unknown regression parameters, and $d\mu_0(.)$ is an unspecified continuous function. The corresponding model for the observed recurrent process is

$$E\{dN(t)|D \geq t, C, \boldsymbol{Z}\} = Y(t)e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}(t)} d\mu_0(t). \tag{5.2}$$

In Chapter 4, we propose an estimating equation to estimate $\boldsymbol{\beta}$,

$$\boldsymbol{U}_n^{full}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \int_0^{\tau} Y_i(t) \left\{ \boldsymbol{Z}_i(t) - \frac{\sum_{i=1}^{n} Y_i(t)\boldsymbol{Z}_i(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_i(t)}}{\sum_{i=1}^{n} Y_i(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_i(t)}} \right\} dN_i(t).$$

In the presence of missing covariates, we propose the following estimating equation

$$\boldsymbol{U}_n(\boldsymbol{\beta}) = \sum_{i=1}^{n} \int_0^{\tau} \boldsymbol{H}_i(t)Y_i(t) \left\{ \boldsymbol{Z}_i(t) - \frac{\sum_{i=1}^{n} h_i(t)Y_i(t)\boldsymbol{Z}_i(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_i(t)}}{\sum_{i=1}^{n} h_i(t)Y_i(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_i(t)}} \right\} dN_i(t), \qquad (5.3)$$

where $h_i(.)$ determines if the ith subject is included in estimating $\frac{\sum_{i=1}^{n} Y_i(t)\boldsymbol{Z}_i(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_i(t)}}{\sum_{i=1}^{n} Y_i(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_i(t)}}$, and $H_{ji}(.), j = 1, \cdots, p$, determine if the ith subject contributes to the jth component of the estimating function. The estimator of $\boldsymbol{\beta}_0$ is defined to be $\widehat{\boldsymbol{\beta}}$, the solution to $\boldsymbol{U}_n(\boldsymbol{\beta}) = \mathbf{0}$. It can be obtained by using the Newton-Raphson method. Given $\widehat{\boldsymbol{\beta}}$, we estimate $\mu_0(t)$ by the Nelson-Aalen-type estimator of $\widehat{\mu}_0(t)$, where

$$\widehat{\mu}_0(t) = \sum_{i=1}^{n} \int_0^{t} \frac{h_i(t)dN_i(u)}{\sum_{i=1}^{n} h_i(t)Y_i(u)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}_i(u)}}, \; t \in [0, \tau]. \qquad (5.4)$$

5.3 Asymptotic properties of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\mu}_0(t)$

We consider the following assumptions,

(C1) $\{N_i, X_i, \boldsymbol{Z}_i(.), \boldsymbol{H}_i(.), h_i(.)\}$ $(i = 1, \cdots, n)$ are independent and identically distributed.

(C2) $P(C_i > \tau | \boldsymbol{Z}_i) > 0, i = 1, \cdots, n$, where τ is the study duration.

(C3) $E[N_i(\tau)] < \infty$, $i = 1, \cdots, n,$.

(C4) Covariates $\boldsymbol{Z}_i(t), i = 1, \cdots, n$, are bounded and has finite total variations in $[0, \tau]$.

(C5) Matrices $\boldsymbol{A}$ is non-singular, where

$$\begin{aligned}
\boldsymbol{A} &= E\left\{\int_0^\tau \boldsymbol{H}(t)Y(t)\left[\left\{\frac{E\{h(t)Y(t)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}{E\{h(t)Y(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}\right\}^{\otimes 2} - \right.\right. \\
&\qquad \left.\left.\frac{E\{h(t)Y(t)\boldsymbol{Z}(t)^{\otimes 2}e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}{E\{h(t)Y(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}\right]dN(t)\right\},
\end{aligned}$$

$\boldsymbol{a}^{\otimes 0}=1$, $\boldsymbol{a}^{\otimes 1}=\boldsymbol{a}$, $\boldsymbol{a}^{\otimes 2}=\boldsymbol{a}\boldsymbol{a}^T$, where $\boldsymbol{a}$ is a column vector.

Theorem 5.3.1 *Under regularity conditions (C1)-(C5), the parameter estimate $\widehat{\boldsymbol{\beta}}$ is strongly consistent for $\boldsymbol{\beta}_0$, i.e. $\widehat{\boldsymbol{\beta}} \xrightarrow{a.s.} \boldsymbol{\beta}_0$. The random vector $\sqrt{n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)$ converges in distribution to a zero-mean normal distribution with a covariance matrix $\boldsymbol{A}^{-1}\boldsymbol{\Sigma}(\boldsymbol{A}^{-1})^T$, where*

$$\begin{aligned}
\boldsymbol{A} &\equiv \left.\frac{\partial \boldsymbol{U}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \\
&= E\left\{\int_0^\tau \boldsymbol{H}(t)Y(t)\left[\left\{\frac{E\{h(t)Y(t)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}{E\{h(t)Y(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}\right\}^{\otimes 2} - \frac{E\{h(t)Y(t)\boldsymbol{Z}(t)^{\otimes 2}e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}{E\{h(t)Y(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}\right]dN(t)\right\},
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{\Sigma} &= E[\boldsymbol{J}\boldsymbol{J}^T], \\
\boldsymbol{J} &= \int_0^\tau \boldsymbol{H}(t)Y(t)\left\{\boldsymbol{Z}(t) - \frac{E\{h(t)Y(t)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}{E\{h(t)Y(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}\right\}dM_{\boldsymbol{\beta}_0}(t) + \\
&\quad \int_0^\tau \left\{\boldsymbol{H}(t) - \frac{E\{\boldsymbol{H}(t)\}}{E\{h(t)\}}h(t)\right\}\left[\boldsymbol{Z}(t) - \frac{E\{h(t)Y(t)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}{E\{h(t)Y(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}\right]Y(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}d\mu_0(t), \\
M_{\boldsymbol{\beta}_0}(t) &= N(t) - \int_0^t Y(s)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(s)}d\mu_0(s). \qquad (5.5)
\end{aligned}$$

A consistent estimator for $\boldsymbol{A}^{-1}\boldsymbol{\Sigma}(\boldsymbol{A}^{-1})^T$ is $\widehat{\boldsymbol{A}}^{-1}\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{A}}^{-1})^T$, where

$$
\begin{aligned}
\widehat{\boldsymbol{A}} &= n^{-1}\sum_{i=1}^{n}\left\{\int_0^\tau \boldsymbol{H}_i(t)Y_i(t)\left[\left\{\frac{\boldsymbol{Q}^{(1)}(t;\widehat{\boldsymbol{\beta}})}{Q^{(0)}(t;\widehat{\boldsymbol{\beta}})}\right\}^{\otimes 2}-\frac{\boldsymbol{Q}^{(2)}(t;\widehat{\boldsymbol{\beta}})}{Q^{(0)}(t;\widehat{\boldsymbol{\beta}})}\right]dN_i(t)\right\},\\
\widehat{\boldsymbol{\Sigma}} &= n^{-1}\sum_{i=1}^{n}\widehat{\boldsymbol{J}}_i\widehat{\boldsymbol{J}}_i^T,\\
\widehat{\boldsymbol{J}}_i &= \int_0^\tau \boldsymbol{H}_i(t)Y_i(t)\left\{\boldsymbol{Z}_i(t)-\frac{\boldsymbol{Q}^{(1)}(t;\widehat{\boldsymbol{\beta}})}{Q^{(0)}(t;\widehat{\boldsymbol{\beta}})}\right\}d\widehat{M}_i(t)+\\
&\quad \int_0^\tau\left\{\boldsymbol{H}_i(t)-\frac{\sum_{i=1}^n\{\boldsymbol{H}_i(t)\}}{\sum_{i=1}^n\{h_i(t)\}}h_i(t)\right\}\left[\boldsymbol{Z}_i(t)-\frac{\boldsymbol{Q}^{(1)}(t;\widehat{\boldsymbol{\beta}})}{Q^{(0)}(t;\widehat{\boldsymbol{\beta}})}\right]Y_i(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}_i(t)}d\widehat{\mu}_0(t),\\
\boldsymbol{Q}^{(k)}(t;\boldsymbol{\beta}) &= n^{-1}\sum_{i=1}^{n}h_i(t)Y_i(t)\boldsymbol{Z}_i(t)^{\otimes k}e^{\boldsymbol{\beta}^T\boldsymbol{Z}_i(t)}, k=0,1,2,\\
\widehat{M}_i(t) &= N_i(t)-\int_0^t Y_i(s)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}_i(s)}d\widehat{\mu}_0(s). \qquad (5.6)
\end{aligned}
$$

Theorem 5.3.2 *Under the regularity conditions (C1)-(C5), $\widehat{\mu}_0(t)$ is strongly consistent for $\mu_0(t)$ uniformly in t, i.e. $\widehat{\mu}_0(t)\xrightarrow{a.s.}\mu_0(t)$, $t\in[0,\tau]$. The process $n^{1/2}\{\widehat{\mu}_0(t)-\mu_0(t)\}$, $t\in[0,\tau]$, converges weakly to a zero-mean Gaussian process with a covariance function $\phi(s,t)=E[K(s)K(t)]$, where*

$$
\begin{aligned}
&K(t) \qquad (5.7)\\
&= \int_0^t\frac{h(u)dM_{\boldsymbol{\beta}_0}(u)}{E\{h(u)Y(u)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(u)}\}}+\left\{\int_0^t\frac{E\left\{h(u)Y(u)\boldsymbol{Z}(u)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(u)}\right\}}{E\{h(u)Y(u)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(u)}\}}d\mu_0(u)\right\}^T\boldsymbol{A}^{-1}\boldsymbol{J},\ t\in[0,\tau],
\end{aligned}
$$

where $\boldsymbol{A}$,$\boldsymbol{J}$ and $M_{\boldsymbol{\beta}_0}(t)$ are defined as in (5.5). A consistent estimator for $\phi(s,t)$ is

$$\widehat{\phi}(s,t)=n^{-1}\sum_{i=1}^{n}\widehat{K}_i(s)\widehat{K}_i(t),$$

where

$$\widehat{K}_i(t) = \int_0^t \frac{h_i(u)d\widehat{M}_i(u)}{Q^{(0)}(t;\widehat{\boldsymbol{\beta}})} + \left\{\int_0^t \frac{\boldsymbol{Q}^{(1)}(u;\widehat{\boldsymbol{\beta}})}{Q^{(0)}(u;\widehat{\boldsymbol{\beta}})} d\widehat{\mu}_0(u)\right\}^T \widehat{\boldsymbol{A}}^{-1}\widehat{\boldsymbol{J}}_i, \quad (5.8)$$

where $\widehat{\boldsymbol{A}}$, $\widehat{\boldsymbol{J}}_i$, $\widehat{M}_i(t)$ *and* $\boldsymbol{Q}^{(k)}(t;\boldsymbol{\beta})$, $k = 0, 1$ *are defined as in (5.6).*

5.4 Simulation studies

A series of numerical simulation studies are conducted to evaluate the performance of the proposed estimator in the finite sample situation.

We first generate the death time D from the Cox proportional hazard model $\lambda(t|\boldsymbol{Z}) = \lambda_0 e^{\boldsymbol{\gamma}^T \boldsymbol{Z}}$. Given D and $\boldsymbol{Z}$, we generate the recurrent events using the following intensity model:

$$E\{dN^*(t)|N^*(t-), D, \boldsymbol{Z}, \nu\} = \nu g(D, \boldsymbol{Z}) h_0 dt,$$

where

$$g(D, \boldsymbol{Z}) = e^{-\rho_0 D} e^{\boldsymbol{\beta}^T \boldsymbol{Z}} \left[1 + \frac{\rho_0}{\lambda_0} e^{-\boldsymbol{\gamma}^T \boldsymbol{Z}}\right],$$

ν is a gamma variable with mean 1 and variance σ^2 independent of D and $\boldsymbol{Z}$, and h_0 is a positive constant. We can control the correlations among the recurrent events by varying the value of σ^2. Notice that ρ_0 controls the correlation between the recurrent events and death, with correlation being 0 when $\rho_0 = 0$. As we increase ρ_0, the correlation becomes larger. In Chapter 4 Section 4, we showed the above intensity model satisfies the conditional rate model (5.1). Independent censoring time C is generated from the uniform $(0, \tau)$ distribution.

We consider $\boldsymbol{Z} = (Z_1, Z_2)^T$, with $Z_1 \sim Bernoulli(0.5)$ as a 0/1 treatment indicator and $Z_2 \sim Uniform(0, 1)$. The corresponding parameter vector is $\boldsymbol{\beta} = (\beta_1, \beta_2)^T$. Our primary interest focuses on making inference on β_1 when measurements on Z_1 are complete but incomplete on Z_2. The following combinations of simulation parameters are chosen: $\boldsymbol{\beta}_0 = (0, 0.5)^T, (ln(2), 0.5)^T$, $\sigma^2 = 0.5$, $\gamma = (0.25, 0.3)^T$, $\lambda_0 = 0.4$, $h_0 = 6$, $\tau = 3$ and $\rho_0 = 0, 1, 4$. We consider $20\%, 50\%, 80\%$ missingness in Z_2, respectively. The values of ρ_0 range from none to moderate correlation between recurrences and death. The average observed numbers of recurrent events range from 2 to 11 events per subject. We consider sample sizes $n = 200, 400, 600$. For each setting, 1000 simulation samples are generated. The simulations are programmed in MATLAB(version 7.7.0).

Results for the complete covariate Z_1 are presented in Table 5.1, Table 5.2, Table 5.3, Table 5.4, Table 5.5 and Table 5.6. Results for the incomplete covariate Z_2 are presented in Table 5.7, Table 5.8, and Table 5.9. The coefficient estimator for $\boldsymbol{\beta} = (\beta_1, \beta_2)^T$ appears to be approximately unbiased for all combinations of sample size, missingness percentage, correlations and the true values of $\boldsymbol{\beta}$. The proposed standard error estimator provides a good estimate of the true variation of $\boldsymbol{\beta}$. The coverage rate falls in the $0.93 - 0.96$ range. The accuracy of the asymptotic approximation appears to be unaffected by the amount of correlation between recurrent events and death.

For comparison, we also evaluate the methods based on complete-cases only and full data analysis. The loss of efficiency for our proposed method relative to the full data analysis is smaller when the missing proportion becomes smaller and the recurrences and death become more correlated. For the estimation for the regression coefficient of the completely observed covariate, the proposed method generally appears to be more efficient than the complete-case analysis for situations with fewer missing measurements and more correlated recurrences and complete case analysis when the recurrent events are independent of death. For the estimation for the regression coefficient of the incomplete covariate, the proposed method has better coverage probability than the complete-case analysis, especially for smaller sample with high missing proportion.

5.5 Application to the SOLVD data

As a demonstration, we apply the proposed methods to the SOLVD(Studies of Left Ventricular Dysfunction, SOLVD Investigators, 1991) Prevention Trial data. The SOLVD prevention trial was randomized, double-blind, and placebo-controlled. The trial had a three-year recruitment and a two-year follow-up. The basic inclusion criteria for the prevention trial were: age between 21 and 80 years, inclusive, no overt symptoms of congestive heart failure, and left ventricular EF less than 35%. EF is a number between 0 and 100 that measures the efficiency of the heart in ejecting blood. A total of 4228 patients with asymptomatic left ventricular dysfunction were randomly assigned to receive either enalapril or placebo in addition to usual care at one of the 83 hospitals linked to 23 centers in the United States, Canada, and Belgium. During the 2-year follow-up period, detailed information for hospitalizations and mortality was recorded.

There are 2117 patients with a total of 2045 recorded hospitalizations and 2111 patients with a total of 1737 recorded hospitalizations in the placebo and enalapril treatment groups, respectively. The value of log-rank test statistic is 1.63 with p-value=0.202, indicating there is no significant difference in survival between the two groups. Table 5.10 summarizes the hospitalization and survival experiences by groups. We are interested in assessing the average treatment effect (TRT) on repeated hospitalizations after adjusting for center, gender, age, baseline ejection fraction(EF) and baseline cardiothoracic ratio(CTR). However, 605 patients have missing covariate measurements(603 patients have missing baseline cardiothoracic ratio measurement, 1 patient has missing baseline ejection fraction, and 1 patient is missing in both). We center AGE, EF and CTR on the mean. TRT is 1 for enalapril and 0 for placebo. GENDER is 1 for male and 0 for female. CENTER k ($k = 2, \cdots, 23$) is 1 for subjects in the kth center and 0 otherwise.

We conduct regression analysis by using the proposed estimating equation (5.3). The analysis is based on 4228 subjects. The results for TRT, EF, AGE, GENDER and CTR are presented in Table 5.11. Enalapril treatment reduces the mean frequency of repeated hospitalizations by 18.9%(i.e., $1-e^{-0.210} = 0.189$) after adjusting for gender, centers, age, baseline ejection fraction and baseline cardiothoracic ratio. The reduction is statistically significant at the 0.05 level (p-value< 0.001). Our analysis also indicates that the mean frequency of repeated hospitalization for heart failure increases by 1.01%(i.e., $e^{0.010} - 1 = 0.01$) per year increase in age and decreases by 1.4%(i.e., $1 - e^{-0.014} = 0.014$) with a 1% decrease in baseline ejection fraction. The effects of age and baseline ejection fraction are both statistically significant at the 0.05 level, with p-value=0.002 and p-value=0.006, respectively. Males have 11.2%(i.e., $1 - e^{-0.119} = 0.112$) fewer repeated hospitalization than females, however, this reduction is not statistically significant at the 0.05 level (p-value=0.184). Although the mean frequency of repeated hospitalization for heart failure increases by 86.1%(i.e., $e^{0.621} - 1 = 0.861$) with a 1% increase in baseline cardiothoracic ratio, the effect is not statistically significant at the 0.05 level(p-value=0.174). Table 5.4 also summarizes the results from two other procedures as comparison to the proposed method. One is analysis based on 3623 subjects with complete covariates. The other one is based on 4227 subjects without adjusting for baseline cardiothoracic ratio. These procedures yield similar results, while our proposed method provides stronger evidence for the effectiveness of enalapril treatment.

In summary, enalapril effectively reduced the frequency of hospitalizations but not mortality. Low baseline ejection fraction and old age are significantly associated with more frequent hospitalizations. Gender and baseline cardiothoracic ratio appears to be not related to repeated hospitalizations.

5.6 Concluding remarks

In this chapter, we propose an estimating procedure for model (4.1) in the presence of missing covariate measurements. We show the resulting estimates are unbiased and asymptotically normal. We compare the proposed method with complete-case analysis and the full data analysis via simulation studies as well as real data application. Our numerical results suggest that under appropriate conditions(large proportion of missing data, moderately/highly correlated recurrence and terminating censoring), the proposed method yields more efficient estimates than complete-case analysis.

5.7 Proofs of the theorems

For convenience, we introduce the following notations: $\mathbb{P}_n f = \frac{1}{n}\sum_{i=1}^{n} f(U_i)$ as the expectation of f under the empirical measure, $Pf = \int f(u)dP(u)$ as the expectation of f under P, and $\mathbb{G}_n f = \sqrt{n}(\mathbb{P}_n f - Pf)$ as the empirical process evaluated at f, a Gaussian process $\mathbb{G}_P f$ as the limiting process of $\mathbb{G}_n f$.

<u>Proof of Theorem 5.3.1</u>:

In the empirical process notation, we can rewrite (5.3) as

$$\frac{1}{n}\boldsymbol{U}_n(\boldsymbol{\beta}) = \mathbb{P}_n\left\{\int_0^\tau \boldsymbol{H}(t)Y(t)\left[\boldsymbol{Z}(t) - \frac{\mathbb{P}_n\{h(t)Y(t)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{h(t)Y(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}(t)}\}}\right]dN(t)\right\}.$$

Trivially, the class $\{\boldsymbol{\beta} \in \mathcal{B}\}$ and $\{\boldsymbol{Z}\}$ are both Donsker classes. Since products of bounded Donsker class are Donsker, $\{\boldsymbol{\beta}^T\boldsymbol{Z} : \boldsymbol{\beta} \in \mathcal{B}\}$ is a Donsker class. We know that exponentiation is Lipschitz continuous on compacts, $\{e^{\boldsymbol{\beta}^T\boldsymbol{Z}} : \boldsymbol{\beta} \in \mathcal{B}\}$ is also Donsker. By Lemma 4.1 of Kosorok(2008), N, Y, $\boldsymbol{H}$ and h are both Donsker as processes in $l^\infty([0,\tau])$. Again since all products of bounded Donsker classes are Donsker, $\{\boldsymbol{H}(t)Y(t)\boldsymbol{Z}(t)dN(t), t \in [0,\tau]\}$, $\{\boldsymbol{H}(t)Y(t)dN(t), t \in [0,\tau]\}$, $\{\boldsymbol{H}(t)Y(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}(t)} : \boldsymbol{\beta} \in \mathcal{B}, t \in [0,\tau]\}$, $\{\boldsymbol{H}(t)Y(t)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}(t)} : \boldsymbol{\beta} \in \mathcal{B}, t \in [0,\tau]\}$, $\{h(t)Y(t)\boldsymbol{Z}(t)dN(t), t \in [0,\tau]\}$, $\{h(t)Y(t)dN(t), t \in [0,\tau]\}$, $\{h(t)Y(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}(t)} :$

$\boldsymbol{\beta} \in \mathcal{B}, t \in [0,\tau]\}$, $\{h(t)Y(t)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}(t)} : \boldsymbol{\beta} \in \mathcal{B}, t \in [0,\tau]\}$ are P-Donsker and thus P-Glivenko-Cantelli. Therefore, we have

$$\sup_{\boldsymbol{\beta}\in\mathcal{B}} \left| \frac{1}{n}\boldsymbol{U}_n(\boldsymbol{\beta}) - \boldsymbol{U}(\boldsymbol{\beta}) \right| \xrightarrow{a.s.} 0,$$

where

$$\boldsymbol{U}(\boldsymbol{\beta}) \equiv P\left\{ \int_0^\tau \boldsymbol{H}(t)Y(t)\left[\boldsymbol{Z}(t) - \frac{P\{h(t)Y(t)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}(t)}\}}{P\{h(t)Y(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}(t)}\}}\right] dN(t) \right\}.$$

Next, we show $\boldsymbol{U}(\boldsymbol{\beta}_0) = 0$ under the proposed model (5.1). Since we assume missing completely at random (MCAR), $\boldsymbol{H}(t)$ and $h(t)$ are independent of all other random variables, conditional on $Y(t)$. We have

$$\begin{aligned}
\boldsymbol{U}(\boldsymbol{\beta}_0) &= P\left\{ \int_0^\tau \boldsymbol{H}(t)Y(t)\left[\boldsymbol{Z}(t) - \frac{P\{h(t)Y(t)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}{P\{h(t)Y(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}\right] dN(t) \right\} \\
&= \int_0^\tau E\left\{ \boldsymbol{H}(t)Y(t)\left[\boldsymbol{Z}(t) - \frac{E\{h(t)Y(t)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}{E\{h(t)Y(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}\right] dN(t) \right\} \\
&= \int_0^\tau E\left\{\boldsymbol{H}(t)Y(t)\boldsymbol{Z}(t)dN(t)\right\} - \int_0^\tau \frac{E\{h(t)Y(t)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}{E\{h(t)Y(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}} E\left\{\boldsymbol{H}(t)Y(t)dN(t)\right\} \\
&= \int_0^\tau E\left\{\boldsymbol{H}(t)\right\} E\left\{Y(t)\boldsymbol{Z}(t)E[dN(t)|D \geq t, C, \boldsymbol{Z}]\right\} \\
&\quad - \int_0^\tau \frac{E\left\{h(t)\right\} E\{Y(t)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}{E\left\{h(t)\right\} E\{Y(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}} E\left\{\boldsymbol{H}(t)\right\} E\left\{Y(t)E[dN(t)|D \geq t, C, \boldsymbol{Z}]\right\} \\
&= \int_0^\tau E\left\{\boldsymbol{H}(t)\right\} E\left\{Y(t)\boldsymbol{Z}(t)I(D \geq t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}d\mu_0(t)\right\} \\
&\quad - \int_0^\tau \frac{E\{Y(t)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}{E\{Y(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}} E\left\{\boldsymbol{H}(t)\right\} E\left\{Y(t)I(D \geq t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}d\mu_0(t)\right\} \\
&= \int_0^\tau E\left\{\boldsymbol{H}(t)\right\} E\left\{Y(t)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\right\} d\mu_0(t) \\
&\quad - \int_0^\tau \frac{E\{Y(t)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}{E\{Y(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}} E\left\{\boldsymbol{H}(t)\right\} E\left\{Y(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\right\} d\mu_0(t) \\
&= 0
\end{aligned}$$

Using Taylor expansion, we have

$$\boldsymbol{U}(\boldsymbol{\beta}) = \boldsymbol{U}(\boldsymbol{\beta}_0) + \boldsymbol{A}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + o(|\boldsymbol{\beta} - \boldsymbol{\beta}_0|) = \boldsymbol{A}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + o(|\boldsymbol{\beta} - \boldsymbol{\beta}_0|).$$

By assumption (C5), $\boldsymbol{A}$ is non-singular, for sufficiently large n and small enough ϵ, we have

$$\sup_{|\boldsymbol{\beta}-\boldsymbol{\beta}_0|=\epsilon} \left| \frac{1}{n}\boldsymbol{U}_n(\boldsymbol{\beta}) - (\boldsymbol{A}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)) \right| \leq \inf_{|\boldsymbol{\beta}-\boldsymbol{\beta}_0|=\epsilon} \left| \boldsymbol{A}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) \right|.$$

Notice that $\boldsymbol{A}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$ has a unique solution within $|\boldsymbol{\beta} - \boldsymbol{\beta}_0| \leq \epsilon$. By the degree theory(Deming 1985, Chapter 1), the above inequality implies $\frac{1}{n}\boldsymbol{U}_n(\boldsymbol{\beta})$ has the same number of non-zero solution as $\boldsymbol{A}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$. Therefore there exists $\widehat{\boldsymbol{\beta}}$ which is the solution to $\boldsymbol{U}_n(\boldsymbol{\beta}) = 0$ and $|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0| \leq \epsilon$ for sufficiently large n. Since ϵ can be chosen arbitrarily small, $\widehat{\boldsymbol{\beta}} \xrightarrow{a.s.} \boldsymbol{\beta}_0$. This concludes the proof of consistency of $\widehat{\boldsymbol{\beta}}$.

Let $M_{\boldsymbol{\beta}_0}(t) = N(t) - \int_0^t Y(s) e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}(s)} d\mu_0(s)$, and by addition and subtraction,

$$
\begin{aligned}
& \frac{1}{n}\boldsymbol{U}_n(\widehat{\boldsymbol{\beta}}) \\
= \; & \mathbb{P}_n \int_0^\tau \boldsymbol{H}(t) Y(t) \left[\boldsymbol{Z}(t) - \frac{\mathbb{P}_n\{h(t)Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{h(t)Y(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}} \right] \left\{ dN(t) - Y(t) e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)} d\mu_0(t) \right\} \\
& + \mathbb{P}_n \int_0^\tau \left\{ \boldsymbol{H}(t) - \frac{E\{\boldsymbol{H}(t)\}}{E\{h(t)\}} h(t) \right\} \left[\boldsymbol{Z}(t) - \frac{\mathbb{P}_n\{h(t)Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{h(t)Y(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}} \right] Y(t) e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)} d\mu_0(t) \\
= \; & \mathbb{P}_n \int_0^\tau \boldsymbol{H}(t) Y(t) \left[\boldsymbol{Z}(t) - \frac{\mathbb{P}_n\{h(t)Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{h(t)Y(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}} \right] \left\{ dN(t) - Y(t) e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}(t)} d\mu_0(t) \right\} \\
& - \mathbb{P}_n \int_0^\tau \boldsymbol{H}(t) Y(t) \left[\boldsymbol{Z}(t) - \frac{\mathbb{P}_n\{h(t)Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{h(t)Y(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}} \right] \left\{ e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)} - e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}(t)} \right\} d\mu_0(t) \\
& + \mathbb{P}_n \int_0^\tau \left\{ \boldsymbol{H}(t) - \frac{E\{\boldsymbol{H}(t)\}}{E\{h(t)\}} h(t) \right\} \left[\boldsymbol{Z}(t) - \frac{\mathbb{P}_n\{h(t)Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{h(t)Y(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}} \right] Y(t) e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)} d\mu_0(t) \\
= \; & \mathbb{P}_n \int_0^\tau \boldsymbol{H}(t) Y(t) \left[\boldsymbol{Z}(t) - \frac{\mathbb{P}_n\{h(t)Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{h(t)Y(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}} \right] dM_{\boldsymbol{\beta}_0}(t) \\
& - \mathbb{P}_n \int_0^\tau \boldsymbol{H}(t) Y(t) \left[\boldsymbol{Z}(t) - \frac{\mathbb{P}_n\{h(t)Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{h(t)Y(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}} \right] \left\{ e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)} - e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}(t)} \right\} d\mu_0(t) \\
& + \mathbb{P}_n \int_0^\tau \left\{ \boldsymbol{H}(t) - \frac{E\{\boldsymbol{H}(t)\}}{E\{h(t)\}} h(t) \right\} \left[\boldsymbol{Z}(t) - \frac{\mathbb{P}_n\{h(t)Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{h(t)Y(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}} \right] Y(t) e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)} d\mu_0(t).
\end{aligned}
$$

Similarly, we have

$$
\begin{aligned}
& \boldsymbol{U}(\widehat{\boldsymbol{\beta}}) \\
= \; & P \int_0^\tau \boldsymbol{H}(t) Y(t) \left[\boldsymbol{Z}(t) - \frac{P\{h(t)Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}}{P\{h(t)Y(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}} \right] dM_{\boldsymbol{\beta}_0}(t) \\
& - P \int_0^\tau \boldsymbol{H}(t) Y(t) \left[\boldsymbol{Z}(t) - \frac{P\{h(t)Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}}{P\{h(t)Y(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}} \right] \left\{ e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)} - e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}(t)} \right\} d\mu_0(t) \\
& + P \int_0^\tau \left\{ \boldsymbol{H}(t) - \frac{E\{\boldsymbol{H}(t)\}}{E\{h(t)\}} h(t) \right\} \left[\boldsymbol{Z}(t) - \frac{P\{h(t)Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}}{P\{h(t)Y(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}} \right] Y(t) e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)} d\mu_0(t). \\
& \sqrt{n} \left\{ \frac{1}{n} \boldsymbol{U}_n(\widehat{\boldsymbol{\beta}}) - \boldsymbol{U}(\widehat{\boldsymbol{\beta}}) \right\} \\
= \; & \sqrt{n} \mathbb{P}_n \int_0^\tau \boldsymbol{H}(t) Y(t) \left[\boldsymbol{Z}(t) - \frac{\mathbb{P}_n\{h(t)Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{h(t)Y(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}} \right] dM_{\boldsymbol{\beta}_0}(t) \\
& - \sqrt{n} P \int_0^\tau \boldsymbol{H}(t) Y(t) \left[\boldsymbol{Z}(t) - \frac{P\{h(t)Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}}{P\{h(t)Y(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}} \right] dM_{\boldsymbol{\beta}_0}(t) \\
& - \sqrt{n} \mathbb{P}_n \int_0^\tau \boldsymbol{H}(t) Y(t) \left[\boldsymbol{Z}(t) - \frac{\mathbb{P}_n\{h(t)Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{h(t)Y(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}} \right] \left\{ e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)} - e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}(t)} \right\} d\mu_0(t) \\
& + \sqrt{n} P \int_0^\tau \boldsymbol{H}(t) Y(t) \left[\boldsymbol{Z}(t) - \frac{P\{h(t)Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}}{P\{h(t)Y(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}} \right] \left\{ e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)} - e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}(t)} \right\} d\mu_0(t) \\
& + \sqrt{n} \mathbb{P}_n \int_0^\tau \left\{ \boldsymbol{H}(t) - \frac{E\{\boldsymbol{H}(t)\}}{E\{h(t)\}} h(t) \right\} \left[\boldsymbol{Z}(t) - \frac{\mathbb{P}_n\{h(t)Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{h(t)Y(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}} \right] Y(t) e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)} d\mu_0(t) \\
& - \sqrt{n} P \int_0^\tau \left\{ \boldsymbol{H}(t) - \frac{E\{\boldsymbol{H}(t)\}}{E\{h(t)\}} h(t) \right\} \left[\boldsymbol{Z}(t) - \frac{P\{h(t)Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}}{P\{h(t)Y(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}} \right] Y(t) e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)} d\mu_0(t) \\
= \; & \mathbb{G}_n \int_0^\tau \boldsymbol{H}(t) Y(t) \left\{ \boldsymbol{Z}(t) - \frac{\mathbb{P}_n\{h(t)Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{h(t)Y(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}} \right\} dM_{\boldsymbol{\beta}_0}(t) \\
& + \sqrt{n} P \int_0^\tau \boldsymbol{H}(t) Y(t) \left\{ \boldsymbol{Z}(t) - \frac{\mathbb{P}_n\{h(t)Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{h(t)Y(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}} \right\} dM_{\boldsymbol{\beta}_0}(t) \\
& - \sqrt{n} P \int_0^\tau \boldsymbol{H}(t) Y(t) \left[\boldsymbol{Z}(t) - \frac{P\{h(t)Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}}{P\{h(t)Y(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}} \right] dM_{\boldsymbol{\beta}_0}(t) \\
& - \mathbb{G}_n \int_0^\tau \boldsymbol{H}(t) Y(t) \left[\boldsymbol{Z}(t) - \frac{\mathbb{P}_n\{h(t)Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{h(t)Y(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}} \right] \left\{ e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)} - e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}(t)} \right\} d\mu_0(t)
\end{aligned}
$$

$$
\begin{aligned}
&-\sqrt{n}P\int_0^\tau \left\{\boldsymbol{H}(t)Y(t)\left[\boldsymbol{Z}(t)-\frac{\mathbb{P}_n\{h(t)Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{h(t)Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}\right]-\right.\\
&\left.\boldsymbol{H}(t)Y(t)\left[\boldsymbol{Z}(t)-\frac{P\{h(t)Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}{P\{h(t)Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}\right]\right\}\left\{e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}-e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\right\}d\mu_0(t)\\
&+\mathbb{G}_n\int_0^\tau\left\{\boldsymbol{H}(t)-\frac{E\{\boldsymbol{H}(t)\}}{E\{h(t)\}}h(t)\right\}\left[\boldsymbol{Z}(t)-\frac{\mathbb{P}_n\{h(t)Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{h(t)Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}\right]Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}d\mu_0(t)\\
&-\sqrt{n}P\int_0^\tau\left\{\boldsymbol{H}(t)-\frac{E\{\boldsymbol{H}(t)\}}{E\{h(t)\}}h(t)\right\}\left[\frac{\mathbb{P}_n\{h(t)Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{h(t)Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}-\right.\\
&\left.\frac{P\{h(t)Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}{P\{h(t)Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}\right]Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}d\mu_0(t)\\
=\;&(i)+(ii)-(iii)-(iv)-(v)+(vi)-(vii). \qquad (5.9)
\end{aligned}
$$

Since $E\left\{dM_{\boldsymbol{\beta}_0}(t)|D\geq t,C,\boldsymbol{Z}\right\}=Y(t)E\left\{dN^*(t)|D\geq t,\boldsymbol{Z}\right\}-Y(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}d\mu_0(t)=0$,

$$
\begin{aligned}
&P\int_0^\tau\boldsymbol{H}(t)Y(t)\left\{\boldsymbol{Z}(t)-\frac{\mathbb{P}_n\{h(t)Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{h(t)Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}\right\}dM_{\boldsymbol{\beta}_0}(t)\\
=\;&\int_0^\tau E\left\{\boldsymbol{H}(t)Y(t)\left[\boldsymbol{Z}(t)-\frac{\mathbb{P}_n\{h(t)Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{h(t)Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}\right]dM_{\boldsymbol{\beta}_0}(t)\right\}\\
=\;&\int_0^\tau E\left\{\boldsymbol{H}(t)Y(t)\left[\boldsymbol{Z}(t)-\frac{\mathbb{P}_n\{h(t)Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{h(t)Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}\right]E[dM_{\boldsymbol{\beta}_0}(t)|D\geq t,C,Z]\right\}\\
=\;&\int_0^\tau E\left\{\boldsymbol{H}(t)Y(t)\left[\boldsymbol{Z}(t)-\frac{\mathbb{P}_n\{h(t)Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{h(t)Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}\right]\cdot 0\right\}=0.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
&P\int_0^\tau\boldsymbol{H}(t)Y(t)\left\{\boldsymbol{Z}(t)-\frac{P\{h(t)Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}{P\{h(t)Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}\right\}dM_{\boldsymbol{\beta}_0}(t)\\
=\;&\int_0^\tau E\left\{\boldsymbol{H}(t)Y(t)\left[\boldsymbol{Z}(t)-\frac{P\{h(t)Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}{P\{h(t)Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}\right]dM_{\boldsymbol{\beta}_0}(t)\right\}\\
=\;&\int_0^\tau E\left\{\boldsymbol{H}(t)Y(t)\left[\boldsymbol{Z}(t)-\frac{P\{h(t)Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}{P\{h(t)Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}\right]E[dM_{\boldsymbol{\beta}_0}(t)|D\geq t,C,Z]\right\}\\
=\;&\int_0^\tau E\left\{\boldsymbol{H}(t)Y(t)\left[\boldsymbol{Z}(t)-\frac{P\{h(t)Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}{P\{h(t)Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}\right]\cdot 0\right\}=0
\end{aligned}
$$

Therefore, (ii) and (iii) of (5.9) are zero.

Now look at (iv) of (5.9). By the consistency of $\widehat{\boldsymbol{\beta}}$, $e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)} - e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}(t)} = o_P(1)$. Since

$$\boldsymbol{H}(t)Y(t)\left[\boldsymbol{Z}(t) - \frac{\mathbb{P}_n\{h(t)Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{h(t)Y(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}}\right]$$

is bounded and $\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - P) = O(1)$, we have

$$(iv) \text{ of } (5.9) \quad = \quad P\int_0^\tau O(1)\cdot o_P(1) = o_P(1).$$

Again by using $\mathbb{P}_n - P = O(n^{-1/2})$, we have

$$\boldsymbol{H}(t)Y(t)\left\{\left[\boldsymbol{Z}(t) - \frac{\mathbb{P}_n\{h(t)Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{h(t)Y(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}}\right] - \left[\boldsymbol{Z}(t) - \frac{P\{h(t)Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}}{P\{h(t)Y(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}}\right]\right\} = O(n^{-1/2}).$$

Then

$$(v) \text{ of } (5.9) \quad = \quad \sqrt{n}P\int_0^\tau O(n^{-1/2})\cdot o_P(1) = O(1)\cdot o_P(1) = o_P(1).$$

Similarly, by $\mathbb{P}_n - P = O(n^{-1/2})$,

$$\sqrt{n}\left\{\frac{\mathbb{P}_n\{h(t)Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{h(t)Y(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}} - \frac{P\{h(t)Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}}{P\{h(t)Y(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}\}}\right\}Y(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)}d\mu_0(t) = O(1).$$

For constant C, we have

$$\begin{aligned}
& P\int_0^\tau \left\{\boldsymbol{H}(t) - \frac{E\{\boldsymbol{H}(t)\}}{E\{h(t)\}}h(t)\right\}C \\
= & \int_0^\tau E\left\{\boldsymbol{H}(t) - \frac{E\{\boldsymbol{H}(t)\}}{E\{h(t)\}}h(t)\right\}C \\
= & \int_0^\tau \left\{E(\boldsymbol{H}(t)) - \frac{E\{\boldsymbol{H}(t)\}}{E\{h(t)\}}E(h(t))\right\}C = 0
\end{aligned}$$

Therefore the (vii) term of (5.9) is zero.

Thus (5.9) becomes

$$
\begin{aligned}
&\sqrt{n}\left\{\frac{1}{n}\boldsymbol{U}_n(\widehat{\boldsymbol{\beta}})-\boldsymbol{U}(\widehat{\boldsymbol{\beta}})\right\} && (5.10)\\
= \;& \mathbb{G}_n\int_0^\tau \boldsymbol{H}(t)Y(t)\left\{\boldsymbol{Z}(t)-\frac{\mathbb{P}_n\{h(t)Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{h(t)Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}\right\}dM_{\boldsymbol{\beta}_0}(t)\\
&+\mathbb{G}_n\int_0^\tau\left\{\boldsymbol{H}(t)-\frac{E\{\boldsymbol{H}(t)\}}{E\{h(t)\}}h(t)\right\}\left[\boldsymbol{Z}(t)-\frac{\mathbb{P}_n\{h(t)Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}{\mathbb{P}_n\{h(t)Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}\}}\right]Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}d\mu_0(t)\\
&+o_P(1).
\end{aligned}
$$

Using similar arguments as in proofing the consistency of $\widehat{\boldsymbol{\beta}}$, we can show $\{\boldsymbol{H}(t)Y(t)\boldsymbol{Z}(t)dN(t), t\in[0,\tau]\}$, $\{\boldsymbol{H}(t)Y(t)dN(t), t\in[0,\tau]\}$, $\{h(t)Y(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}:\widehat{\boldsymbol{\beta}}\in\mathcal{B}, t\in[0,\tau]\}$, $\{h(t)Y(t)\boldsymbol{Z}(t)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(t)}:\widehat{\boldsymbol{\beta}}\in\mathcal{B}, t\in[0,\tau]\}$ are P-Donsker and thus P-Glivenko-Cantelli. We can replace the operator $\mathbb{P}_n$ in (5.10) with P without altering the limiting distribution. By the strong consistency of $\widehat{\boldsymbol{\beta}}$ and Lemma 19.24 of Van der Vaart(1998), (5.10) becomes

$$
\begin{aligned}
&\sqrt{n}\left\{\frac{1}{n}\boldsymbol{U}_n(\widehat{\boldsymbol{\beta}})-\boldsymbol{U}(\widehat{\boldsymbol{\beta}})\right\} && (5.11)\\
= \;& \mathbb{G}_n\int_0^\tau \boldsymbol{H}(t)Y(t)\left\{\boldsymbol{Z}(t)-\frac{P\{h(t)Y(t)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}{P\{h(t)Y(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}\right\}dM_{\boldsymbol{\beta}_0}(t)\\
&+\mathbb{G}_n\int_0^\tau\left\{\boldsymbol{H}(t)-\frac{E\{\boldsymbol{H}(t)\}}{E\{h(t)\}}h(t)\right\}\left[\boldsymbol{Z}(t)-\frac{P\{h(t)Y(t)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}{P\{h(t)Y(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}\}}\right]Y(t)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(t)}d\mu_0(t)\\
&+o_P(1)\\
\equiv \;& \mathbb{G}_n\boldsymbol{J}+o_P(1). && (5.12)
\end{aligned}
$$

On the other hand, since $\boldsymbol{U}_n(\widehat{\boldsymbol{\beta}})=0$ and $\boldsymbol{U}(\boldsymbol{\beta}_0)=0$, we have

$$
\begin{aligned}
\sqrt{n}\left\{\frac{1}{n}\boldsymbol{U}_n(\widehat{\boldsymbol{\beta}})-\boldsymbol{U}(\widehat{\boldsymbol{\beta}})\right\} &= \sqrt{n}\left\{0-(\boldsymbol{U}(\widehat{\boldsymbol{\beta}})-\boldsymbol{U}(\boldsymbol{\beta}_0))\right\}\\
&= -\sqrt{n}\boldsymbol{A}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)+o(\sqrt{n}|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0|) \qquad (5.13)
\end{aligned}
$$

Combining (5.11) and (5.13), we have

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \quad = \quad -\boldsymbol{A}^{-1}\mathbb{G}_n \boldsymbol{J} + o_P(1). \tag{5.14}$$

Therefore $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ converges in distribution to a normal distribution with mean zero and covariance matrix $\boldsymbol{A}^{-1}\boldsymbol{\Sigma}(\boldsymbol{A}^{-1})^T$,$\boldsymbol{\Sigma} = E[\boldsymbol{J}\boldsymbol{J}^T]$, $\boldsymbol{A}$ and $\boldsymbol{J}$ are defined as in (5.5). By replacing $\boldsymbol{\beta}_0$ with $\widehat{\boldsymbol{\beta}}$ and expectations with empirical means in the expressions of $\boldsymbol{A}$ and $\boldsymbol{J}$, we obtain an estimator $\widehat{\boldsymbol{A}}^{-1}\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{A}}^{-1})^T$, where $\widehat{\boldsymbol{A}}$ and $\widehat{\boldsymbol{\Sigma}}$ are defined as in (5.6). Using similar arguments as in proofing the consistency and normality of $\widehat{\boldsymbol{\beta}}$, we can show that $\widehat{\boldsymbol{A}}^{-1}\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{A}}^{-1})^T$ is a consistent estimator of $\boldsymbol{A}^{-1}\boldsymbol{\Sigma}(\boldsymbol{A}^{-1})^T$.

Proof of Theorem 5.3.2:

Using empirical process notation, we write

$$\widehat{\mu}_0(t) = \mathbb{P}_n \int_0^t \frac{h(u)dN(u)}{\mathbb{P}_n\{h(u)Y(u)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(u)}\}},\ t \in [0,\tau].$$

By addition, subtraction and triangle inequality,

$$\begin{aligned}
& \sup_{t\in[0,\tau]} |\widehat{\mu}_0(t) - \mu_0(t)| \\
= & \sup_{t\in[0,\tau]} \left| \mathbb{P}_n \int_0^t \frac{h(u)\left\{dN(u) - Y(u)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(u)}d\mu_0(u)\right\}}{\mathbb{P}_n\{h(u)Y(u)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(u)}\}} \right| \\
= & \sup_{t\in[0,\tau]} \left| \mathbb{P}_n \int_0^t \frac{h(u)dM_{\boldsymbol{\beta}_0}(u)}{\mathbb{P}_n\{h(u)Y(u)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(u)}\}} - \mathbb{P}_n \int_0^t \frac{h(u)Y(u)\left\{e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(u)} - e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(u)}\right\}d\mu_0(u)}{\mathbb{P}_n\{h(u)Y(u)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(u)}\}} \right| \\
\leq & \sup_{t\in[0,\tau]} \left| \mathbb{P}_n \int_0^t \frac{h(u)dM_{\boldsymbol{\beta}_0}(u)}{\mathbb{P}_n\{h(u)Y(u)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(u)}\}} \right| \\
& + \sup_{t\in[0,\tau]} \left| \mathbb{P}_n \int_0^t \frac{h(u)Y(u)\left\{e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(u)} - e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(u)}\right\}d\mu_0(u)}{\mathbb{P}_n\{h(u)Y(u)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(u)}\}} \right|
\end{aligned}$$

$$\le \sup_{t\in[0,\tau]} \left| (\mathbb{P}_n - P) \int_0^t \frac{h(u)dM_{\boldsymbol{\beta}_0}(u)}{\mathbb{P}_n\{h(u)Y(u)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(u)}\}} \right| + \sup_{t\in[0,\tau]} \left| P \int_0^t \frac{h(u)dM_{\boldsymbol{\beta}_0}(u)}{\mathbb{P}_n\{h(u)Y(u)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(u)}\}} \right|$$
$$+ \sup_{t\in[0,\tau]} \left| \mathbb{P}_n \int_0^t \frac{h(u)Y(u)\left\{e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(u)} - e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}(u)}\right\} d\mu_0(u)}{\mathbb{P}_n\{h(u)Y(u)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(u)}\}} \right| \tag{5.15}$$

By the strong consistency of $\widehat{\boldsymbol{\beta}}^T$, the third term of (5.15) converges almost surely to 0.

Since $E\left\{dM_{\boldsymbol{\beta}_0}(t)|D \ge t, C, \boldsymbol{Z}\right\} = Y(t)E\{dN^*(t)|D \ge t, \boldsymbol{Z}\} - Y(t)e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}(t)}d\mu_0(t) = 0$, the second term of (5.15) becomes

$$\begin{aligned}
\sup_{t\in[0,\tau]} \left| P \int_0^t \frac{h(u)dM_{\boldsymbol{\beta}_0}(u)}{\mathbb{P}_n\{h(u)Y(u)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(u)}\}} \right| &= \sup_{t\in[0,\tau]} \left| \int_0^t E\left\{ \frac{h(u)dM_{\boldsymbol{\beta}_0}(u)}{\mathbb{P}_n\{h(u)Y(u)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(u)}\}} \right\} \right| \\
&= \sup_{t\in[0,\tau]} \left| \int_0^t E\left\{ \frac{E\left\{h(u)dM_{\boldsymbol{\beta}_0}(u)|D \ge t, C, \boldsymbol{Z}\right\}}{\mathbb{P}_n\{h(u)Y(u)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(u)}\}} \right\} \right| \\
&= \sup_{t\in[0,\tau]} \left| \int_0^t E\left\{ \frac{h(u)\cdot 0}{\mathbb{P}_n\{h(u)Y(u)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(u)}\}} \right\} \right| = 0.
\end{aligned}$$

Using similar arguments as in proofing the consistency of $\widehat{\boldsymbol{\beta}}$, we can show $\{h(t)Y(t)dN(t), t \in [0,\tau]\}$, $\{h(t)Y(t)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(t)} : \widehat{\boldsymbol{\beta}} \in \mathcal{B}, t \in [0,\tau]\}$ are P-Donsker and P-Glivenko-Cantelli. Thus the first term of (5.15) converges uniformly to zero. We have proven the strong consistency of $\widehat{\mu}_0(t)$, $t \in [0,\tau]$.

Based on the results in (5.15) and Taylor expansion, we have

$$\begin{aligned}
&\sqrt{n}\left\{\widehat{\mu}_0(t) - \mu_0(t)\right\} \\
&= \sqrt{n}(\mathbb{P}_n - P) \int_0^t \frac{h(u)dM_{\boldsymbol{\beta}_0}(u)}{\mathbb{P}_n\{h(u)Y(u)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(u)}\}} + \sqrt{n}P \int_0^t \frac{h(u)dM_{\boldsymbol{\beta}_0}(u)}{\mathbb{P}_n\{h(u)Y(u)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(u)}\}} \\
&\quad -\sqrt{n}\left\{ \mathbb{P}_n \int_0^t \frac{h(u)Y(u)\boldsymbol{Z}(u)e^{\boldsymbol{\beta}_0^T \boldsymbol{Z}(u)}}{\mathbb{P}_n\{h(u)Y(u)e^{\widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}(u)}\}} d\mu_0(u) \right\}^T (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + o_P(1).
\end{aligned} \tag{5.16}$$

Again since $E\left\{dM_{\boldsymbol{\beta}_0}(t)|D \geq t, C, \boldsymbol{Z}\right\} = 0$,

$$\begin{aligned}
P\int_0^t \frac{h(u)dM_{\boldsymbol{\beta}_0}(u)}{\mathbb{P}_n\{h(u)Y(u)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(u)}\}} &= \int_0^t E\left\{\frac{h(u)dM_{\boldsymbol{\beta}_0}(u)}{\mathbb{P}_n\{h(u)Y(u)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(u)}\}}\right\} \\
&= \int_0^t E\left\{\frac{E\left\{h(u)dM_{\boldsymbol{\beta}_0}(u)|D \geq t, C, \boldsymbol{Z}\right\}}{\mathbb{P}_n\{h(u)Y(u)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(u)}\}}\right\} \\
&= \int_0^t E\left\{\frac{h(u)\cdot 0}{\mathbb{P}_n\{h(u)Y(u)e^{\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}(u)}\}}\right\} = 0.
\end{aligned}$$

Thus the second term of (5.16) is zero.

Now look at the last term of (5.16), $\{h(t)Y(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}(t)} : \boldsymbol{\beta} \in \mathcal{B}, t \in [0,\tau]\}$, $\{h(t)Y(u)\boldsymbol{Z}(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}(t)} : \boldsymbol{\beta} \in \mathcal{B}, t \in [0,\tau]\}$ are P-Donsker and thus P-Glivenko-Cantelli, we can replace the operators $\mathbb{P}_n$ with P.

We have shown $\{h(t)dN(t), t \in [0,\tau]\}$, $\{h(t)Y(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}(t)} : \boldsymbol{\beta} \in \mathcal{B}, t \in [0,\tau]\}$ are P-Donsker and thus P-Glivenko-Cantelli. We can replace the operator $\mathbb{P}_n$ with P in the denominator in the first term of (5.16) without altering the limiting distribution. By applying the strong consistency of $\widehat{\boldsymbol{\beta}}$ and Lemma 19.24 of Van der Vaart(1998) to the first two terms of (5.16) and the result in (5.14) to the last term of (5.16), we have

$$\begin{aligned}
&\sqrt{n}\left\{\widehat{\mu}_0(t) - \mu_0(t)\right\} \\
= \;& \mathbb{G}_n \int_0^t \frac{h(u)dM_{\boldsymbol{\beta}_0}(u)}{P\{h(u)Y(u)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(u)}\}} + \mathbb{G}_n\left\{\int_0^t \frac{P\left\{h(u)Y(u)\boldsymbol{Z}(u)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(u)}\right\}}{P\{h(u)Y(u)e^{\boldsymbol{\beta}_0^T\boldsymbol{Z}(u)}\}}d\mu_0(u)\right\}^T \boldsymbol{A}^{-1}\boldsymbol{J} \\
&+o_P(1) \\
\equiv \;& \mathbb{G}_n K(t) + o_P(1),\ t \in [0,\tau] \qquad (5.17)
\end{aligned}$$

Therefore the process $\sqrt{n}\left\{\widehat{\mu}_0(t) - \mu_0(t)\right\}$ converges weakly to a zero-mean Gaussian process with a covariance function $\phi(s,t) = E[K(s)K(t)^T]$, $s, t \in [0,\tau]$, where $K(t)$ is defined as in (5.7). By replacing $\boldsymbol{\beta}_0$ with $\widehat{\boldsymbol{\beta}}$, $\mu_0(\cdot)$ with $\widehat{\mu}_0(\cdot)$ and expectations with empirical means in the expression of $E[K(s)K(t)]$, we obtain an estimator for the asymptotic covariance

$$\widehat{\phi}(s,t) = n^{-1}\sum_{i=1}^{n}\widehat{K}_i(s)\widehat{K}_i(t),$$

where $\widehat{K}_i(t)$ is defined as in (5.8). Using similar arguments as in proofing the consistency and normality of $\widehat{\boldsymbol{\beta}}$, we can show that $\widehat{\phi}(s,t)$ is a consistent estimator of $\phi(s,t)$.

Table 5.1: $\boldsymbol{Z} \sim (Bernoulli(0.5), Uniform(0,1))$, Z_1 is complete while Z_2 is incomplete, $\boldsymbol{\beta} = (0, 0.5)^T$.

		β_1=0, $n = 200$													
		Proposed Method				Complete Case Only					Full Data				
Missing %	ρ_0	Bias	SE	SEE	CP	Bias	SE	SEE	CP	RE1	Bias	SE	SEE	CP	RE2
20%	0	-.00	.168	.164	.952	.00	.152	.149	.949	.91	.00	.140	.135	.931	.82
	1	.01	.213	.200	.940	.01	.218	.203	.929	1.02	.01	.195	.182	.927	.91
	4	.01	.297	.298	.950	.02	.322	.321	.946	1.08	.02	.288	.286	.944	.96
50%	0	-.01	.232	.233	.950	.00	.192	.184	.935	.79	.00	.140	.135	.931	.58
	1	-.01	.246	.248	.954	.00	.261	.257	.951	1.04	.01	.195	.182	.927	.73
	4	.00	.349	.329	.937	.02	.438	.404	.924	1.23	.02	.288	.286	.944	.87
80%	0	-.01	.480	.427	.929	.01	.328	.270	.885	.67	.00	.140	.135	.931	.33
	1	-.01	.448	.406	.932	.00	.450	.392	.915	.99	.01	.195	.182	.927	.46
	4	.02	.476	.448	.937	.04	.823	.631	.900	1.43	.02	.288	.286	.944	.65

Note: Bias is the sample average of $\widehat{\beta}_1$ minus β_1; SE is the sample standard deviation of $\widehat{\beta}_1$;
SEE is the sample average of the standard error estimator; CP is the coverage rate of the Wald 95% confidence interval;
RE1 is relative efficiency of the proposed method vs. complete case only;
RE2 is relative efficiency of the proposed method vs. full data analysis.

Table 5.2: $\boldsymbol{Z} \sim (Bernoulli(0.5), Uniform(0,1))$, Z_1 is complete while Z_2 is incomplete, $\boldsymbol{\beta} = (0, 0.5)^T$.

β_1=0, $n = 400$															
		Proposed Method				Complete Case Only					Full Data				
Missing %	ρ_0	Bias	SE	SEE	CP	Bias	SE	SEE	CP	RE1	Bias	SE	SEE	CP	RE2
20%	0	.00	.121	.116	.946	.00	.109	.108	.949	.93	.00	.098	.096	.945	.83
	1	.01	.150	.142	.934	.01	.152	.144	.928	1.01	.00	.130	.129	.953	.91
	4	.01	.219	.211	.940	.02	.231	.228	.949	1.08	.01	.210	.205	.952	.97
50%	0	.00	.163	.163	.946	.01	.135	.134	.943	.82	.00	.098	.096	.945	.59
	1	.01	.175	.174	.943	.01	.187	.182	.947	1.05	.00	.130	.129	.953	.74
	4	.01	.242	.231	.946	.01	.307	.287	.925	1.24	.01	.210	.205	.952	.89
80%	0	.01	.303	.287	.936	.01	.223	.203	.925	.72	.00	.098	.096	.945	.34
	1	.01	.268	.273	.960	.02	.312	.285	.922	1.05	.00	.130	.129	.953	.48
	4	-.03	.308	.303	.947	-.00	.491	.452	.929	1.96	.01	.210	.205	.952	.88

Note: Bias is the sample average of $\widehat{\beta}_1$ minus β_1; SE is the sample standard deviation of $\widehat{\beta}_1$;

SEE is the sample average of the standard error estimator; CP is the coverage rate of the Wald 95% confidence interval;

RE1 is relative efficiency of the proposed method vs. complete case only;

RE2 is relative efficiency of the proposed method vs. full data analysis.

Table 5.3: $\boldsymbol{Z} \sim (Bernoulli(0.5), Uniform(0,1))$, Z_1 is complete while Z_2 is incomplete, $\boldsymbol{\beta} = (0, 0.5)^T$.

		β_1=0, $n = 600$													
		Proposed Method				Complete Case Only					Full Data				
Missing %	ρ_0	Bias	SE	SEE	CP	Bias	SE	SEE	CP	RE1	Bias	SE	SEE	CP	RE2
20%	0	.01	.093	.095	.954	.01	.088	.088	.954	.93	.00	.081	.079	.949	.83
	1	-.00	.115	.116	.955	-.01	.115	.119	.953	1.03	.00	.107	.106	.940	.91
	4	-.00	.171	.173	.954	.00	.183	.187	.956	1.08	.00	.171	.167	.939	.97
50%	0	-.00	.133	.133	.953	-.00	.115	.111	.937	.83	.00	.081	.079	.949	.59
	1	-.00	.141	.142	.949	.01	.151	.150	.949	1.06	.00	.107	.106	.940	.75
	4	.00	.184	.189	.953	.02	.244	.236	.938	1.25	.00	.171	.167	.939	.88
80%	0	.00	.238	.230	.936	-.01	.176	.170	.941	.75	.00	.081	.079	.949	.35
	1	.01	.217	.220	.955	.01	.244	.233	.944	1.06	.00	.107	.106	.940	.48
	4	.01	.247	.246	.950	.02	.398	.379	.937	2.01	.00	.171	.167	.939	.88

Note: Bias is the sample average of $\widehat{\beta}_1$ minus β_1; SE is the sample standard deviation of $\widehat{\beta}_1$;

SEE is the sample average of the standard error estimator; CP is the coverage rate of the Wald 95% confidence interval;

RE1 is relative efficiency of the proposed method vs. complete case only;

RE2 is relative efficiency of the proposed method vs. full data analysis.

Table 5.4: $\boldsymbol{Z} \sim (Bernoulli(0.5), Uniform(0,1))$, Z_1 is complete while Z_2 is incomplete, $\boldsymbol{\beta} = (ln(2), 0.5)^T$.

		$\beta_1 = ln(2)$, $n = 200$													
		Proposed Method				Complete Case Only					Full Data				
Missing %	ρ_0	Bias	SE	SEE	CP	Bias	SE	SEE	CP	RE1	Bias	SE	SEE	CP	RE2
20%	0	.01	.169	.162	.935	.00	.152	.146	.934	.91	.01	.141	.132	.927	.82
	1	.00	.202	.198	.938	.00	.201	.200	.940	1.01	.00	.185	.179	.939	.90
	4	.01	.301	.293	.937	.02	.330	.316	.936	1.08	.00	.291	.282	.947	.96
50%	0	.00	.242	.232	.946	.00	.190	.180	.938	.78	.01	.141	.132	.927	.57
	1	-.01	.257	.247	.942	.00	.258	.252	.936	1.02	.00	.185	.179	.939	.73
	4	.01	.337	.323	.934	.01	.429	.394	.927	1.22	.00	.291	.282	.947	.87
80%	0	.04	.486	.430	.920	.01	.325	.263	.876	.63	.01	.141	.132	.927	.32
	1	.03	.417	.399	.935	.02	.420	.385	.931	.98	.00	.185	.179	.939	.45
	4	.02	.461	.445	.937	.04	.960	.615	.898	1.38	.00	.291	.282	.947	.64

Note: Bias is the sample average of $\widehat{\beta}_1$ minus β_1; SE is the sample standard deviation of $\widehat{\beta}_1$;

SEE is the sample average of the standard error estimator; CP is the coverage rate of the Wald 95% confidence interval;

RE1 is relative efficiency of the proposed method vs. complete case only;

RE2 is relative efficiency of the proposed method vs. full data analysis.

Table 5.5: $\boldsymbol{Z} \sim (Bernoulli(0.5), Uniform(0,1))$, Z_1 is complete while Z_2 is incomplete, $\boldsymbol{\beta} = (ln(2), 0.5)^T$.

		$\beta_1 = ln(2)$, $n = 400$													
		Proposed Method				Complete Case Only					Full Data				
Missing %	ρ_0	Bias	SE	SEE	CP	Bias	SE	SEE	CP	RE1	Bias	SE	SEE	CP	RE2
20%	0	.00	.119	.115	.951	.00	.108	.106	.947	.92	.00	.100	.095	.935	.83
	1	.02	.144	.140	.947	.01	.147	.142	.948	1.01	.00	.129	.127	.946	.91
	4	.01	.210	.207	.957	.01	.230	.223	.951	1.08	.00	.202	.200	.951	.97
50%	0	-.00	.167	.162	.943	.00	.134	.131	.944	.81	.00	.100	.095	.935	.59
	1	.01	.176	.173	.955	.00	.188	.179	.936	1.03	.00	.129	.127	.946	.73
	4	.01	.228	.228	.943	.02	.284	.282	.948	1.24	.00	.202	.200	.951	.88
80%	0	.03	.291	.288	.942	.00	.218	.200	.928	.71	.00	.100	.095	.935	.34
	1	.01	.280	.271	.936	.01	.276	.281	.947	1.04	.00	.129	.127	.946	.47
	4	.01	.311	.301	.943	.03	.491	.440	.921	1.46	.00	.202	.200	.951	.66

Note: Bias is the sample average of $\widehat{\beta}_1$ minus β_1; SE is the sample standard deviation of $\widehat{\beta}_1$;

SEE is the sample average of the standard error estimator; CP is the coverage rate of the Wald 95% confidence interval;

RE1 is relative efficiency of the proposed method vs. complete case only;

RE2 is relative efficiency of the proposed method vs. full data analysis.

Table 5.6: $\boldsymbol{Z} \sim (Bernoulli(0.5), Uniform(0,1))$, Z_1 is complete while Z_2 is incomplete, $\boldsymbol{\beta} = (ln(2), 0.5)^T$.

		$\beta_1 = ln(2)$, $n = 600$													
		Proposed Method				Complete Case Only					Full Data				
Missing %	ρ_0	Bias	SE	SEE	CP	Bias	SE	SEE	CP	RE1	Bias	SE	SEE	CP	RE2
20%	0	-.00	.096	.094	.946	.00	.090	.087	.939	.93	.00	.079	.078	.949	.86
	1	.01	.119	.114	.936	.01	.117	.116	.949	1.02	.00	.105	.104	.941	.91
	4	.01	.169	.169	.950	.01	.183	.182	.941	1.08	.00	.166	.164	.947	.97
50%	0	.00	.136	.132	.947	.01	.110	.109	.943	.83	.00	.079	.078	.949	.59
	1	.01	.139	.141	.954	.00	.155	.146	.937	1.04	.00	.105	.104	.941	.74
	4	-.00	.196	.186	.930	-.01	.244	.231	.936	1.24	.00	.166	.164	.947	.88
80%	0	.01	.241	.230	.938	.00	.184	.167	.913	.73	.00	.079	.078	.949	.34
	1	.00	.224	.218	.941	.01	.240	.230	.934	1.06	.00	.105	.104	.941	.48
	4	.00	.247	.244	.941	.02	.366	.362	.932	1.48	.00	.166	.164	.947	.67

Note: Bias is the sample average of $\widehat{\beta}_1$ minus β_1; SE is the sample standard deviation of $\widehat{\beta}_1$;

SEE is the sample average of the standard error estimator; CP is the coverage rate of the Wald 95% confidence interval;

RE1 is relative efficiency of the proposed method vs. complete case only;

RE2 is relative efficiency of the proposed method vs. full data analysis.

Table 5.7: $\boldsymbol{Z} \sim (Bernoulli(0.5), Uniform(0,1))$, Z_1 is complete while Z_2 is incomplete, $\boldsymbol{\beta} = (0, 0.5)^T$.

		β_2=0.5, $n = 200$													
		Proposed Method				Complete Case Only					Full Data				
Missing %	ρ_0	Bias	SE	SEE	CP	Bias	SE	SEE	CP	RE1	Bias	SE	SEE	CP	RE2
20%	0	.01	.271	.258	.936	.00	.269	.255	.938	.98	-.00	.231	.231	.945	.90
	1	.02	.354	.352	.946	.02	.354	.351	.950	1.00	-.01	.322	.316	.942	.90
	4	.00	.602	.547	.917	-.00	.602	.547	.917	1.00	-.02	.519	.492	.944	.90
50%	0	.01	.353	.331	.933	.01	.347	.317	.935	.96	-.00	.231	.231	.945	.70
	1	.01	.466	.450	.943	.00	.462	.444	.942	.99	-.01	.322	.316	.942	.70
	4	-.00	.748	.690	.920	-.00	.744	.684	.923	.99	-.02	.519	.492	.944	.71
80%	0	.03	.605	.614	.933	.00	.563	.457	.884	.75	-.00	.231	.231	.945	.38
	1	.03	.788	.746	.918	.02	.758	.670	.894	.90	-.01	.322	.316	.942	.42
	4	.03	1.382	1.131	.920	.02	1.299	1.064	.902	.94	-.02	.519	.492	.944	.44

Note: Bias is the sample average of $\widehat{\beta}_2$ minus β_2; SE is the sample standard deviation of $\widehat{\beta}_2$;
SEE is the sample average of the standard error estimator; CP is the coverage rate of the Wald 95% confidence interval;
RE1 is relative efficiency of the proposed method vs. complete case only;
RE2 is relative efficiency of the proposed method vs. full data analysis.

Table 5.8: $\boldsymbol{Z} \sim (Bernoulli(0.5), Uniform(0,1))$, Z_1 is complete while Z_2 is incomplete, $\boldsymbol{\beta} = (0, 0.5)^T$.

		β_2=0.5, $n = 400$													
		Proposed Method				Complete Case Only					Full Data				
Missing %	ρ_0	Bias	SE	SEE	CP	Bias	SE	SEE	CP	RE1	Bias	SE	SEE	CP	RE2
20%	0	-.01	.190	.186	.944	-.01	.190	.185	.943	.99	.00	.169	.166	.944	.89
	1	-.00	.250	.250	.952	-.00	.250	.250	.950	1.00	.01	.230	.225	.950	.90
	4	.01	.406	.393	.941	.01	.406	.393	.940	1.00	-.02	.365	.353	.937	.90
50%	0	.01	.245	.235	.938	.00	.243	.231	.936	.98	.00	.169	.166	.944	.71
	1	-.00	.324	.319	.939	-.00	.323	.317	.939	.99	.01	.230	.225	.950	.71
	4	.01	.511	.497	.941	.01	.510	.496	.943	1.00	-.02	.365	.353	.937	.71
80%	0	-.01	.397	.396	.947	-.01	.378	.346	.926	.87	.00	.169	.166	.944	.42
	1	.02	.544	.511	.930	.02	.538	.491	.921	.96	.01	.230	.225	.950	.44
	4	.00	.859	.776	.921	-.00	.859	.763	.916	.98	-.02	.365	.353	.937	.45

Note: Bias is the sample average of $\widehat{\beta}_2$ minus β_2; SE is the sample standard deviation of $\widehat{\beta}_2$;
SEE is the sample average of the standard error estimator; CP is the coverage rate of the Wald 95% confidence interval;
RE1 is relative efficiency of the proposed method vs. complete case only;
RE2 is relative efficiency of the proposed method vs. full data analysis.

Table 5.9: $\boldsymbol{Z} \sim (Bernoulli(0.5), Uniform(0,1))$, Z_1 is complete while Z_2 is incomplete, $\boldsymbol{\beta} = (0, 0.5)^T$.

		β_2=0.5, $n = 600$													
		Proposed Method				Complete Case Only					Full Data				
Missing %	ρ_0	Bias	SE	SEE	CP	Bias	SE	SEE	CP	RE1	Bias	SE	SEE	CP	RE2
20%	0	.00	.158	.153	.944	.00	.157	.152	.939	.99	.00	.144	.136	.936	.89
	1	-.00	.202	.206	.952	-.00	.201	.206	.952	1.00	.00	.184	.184	.944	.89
	4	-.02	.347	.323	.927	-.02	.347	.323	.927	1.00	.01	.291	.289	.940	.89
50%	0	-.00	.203	.194	.934	-.00	.202	.192	.933	.99	.00	.144	.136	.936	.70
	1	.00	.275	.260	.936	.00	.277	.259	.935	1.00	.00	.184	.184	.944	.71
	4	.02	.410	.408	.952	.02	.409	.407	.950	1.00	.01	.291	.289	.940	.71
80%	0	.01	.323	.316	.946	.01	.316	.291	.928	.92	.00	.144	.136	.936	.43
	1	-.01	.418	.411	.947	-.01	.416	.401	.945	.98	.00	.184	.184	.944	.45
	4	-.02	.662	.635	.936	-.02	.659	.629	.935	.99	.01	.291	.289	.940	.46

Note: Bias is the sample average of $\widehat{\beta}_2$ minus β_2; SE is the sample standard deviation of $\widehat{\beta}_2$;
SEE is the sample average of the standard error estimator; CP is the coverage rate of the Wald 95% confidence interval;
RE1 is relative efficiency of the proposed method vs. complete case only;
RE2 is relative efficiency of the proposed method vs. full data analysis.

Table 5.10: SOLVD Prevention Trial: Hospitalizations and survival experience

		Number of hospitalizations								
Treatment	Number of subjects	0	1	2	3	4	5	6	≥ 7	Number of Deaths
placebo	2117	1148	506	196	135	55	31	19	27	334
enalapril	2111	1226	488	209	79	50	18	18	23	313

Table 5.11: SOLVD Prevention Trial: Regression analysis for the effects of enalapril on hospitalizations of patients

	Proposed Method			Complete Case Only			Without CTR		
Covariate	Estimate	SE	p-value	Estimate	SE	p-value	Estimate	SE	p-value
TRT	-.210	.056	$< .001$	-.162	.058	.005	-.182	.054	.001
EF	-.014	.005	.006	-.014	.006	.010	-.014	.005	.006
AGE	.010	.003	.002	.006	.003	.056	.009	.003	.003
GENDER	-.119	.090	.184	-.104	.093	.267	-.144	.084	.088
CTR	.621	.457	.174	.875	.499	.080	–	–	–

Note: TRT is coded as 1 for enalapril, 0 for placebo;

GENDER is coded as 1 for male, 0 for female;

Estimate is the estimated regression coefficient;

SE is the estimated standard error;

p-value is the two-sided p-value.

Chapter 6

Summary and Future Research

In this dissertation, we have studied statistical methods for recurrent events data with the presence of a terminal event. While most existing literatures on the analysis of recurrence data assume independent terminating censorship, the assumption is not realistic as the terminal event is likely to be informative about the recurrent events. In Chapter 3, we have considered an accelerated failure time marginal rate model for the cumulative number of the recurrent events over time, while taking the terminal event into account. Our marginal approach leaves the dependence between the recurrences and the terminal event unspecified, and the mean function incorporates the facts that subjects who die cannot experience any further recurrent events. We have applied the inverse probability of censoring weighting technique(Robins and Rotnitzky,1992) to construct an unbiased estimating equation. In Chapter 4, we have considered a proportional rate model for the recurrent event given the subjects are still alive. Again our approach is marginal as we do not need to specify the dependence structure between the recurrent events and the terminal event. We have proposed two estimating procedures. One involves using the inverse survival probability weight(Ghosh and Lin, 2002). Missing covariates is a frequent problem for the statistical analysis of recurrent event data. In Chapter 5, we have proposed a procedure to deal with missing covariate data for the proportional rate model considered in Chapter 4.

In all three chapters, we focus on the estimations of the covariate effects as well as the mean frequency of recurrence. Asymptotic properties of the proposed estimators are studied. Using modern empirical theory, we have shown the proposed estimators are consistent and asymptotically normal. Finite sample properties of the proposed estimators are examined via simulation studies. Numerical results under various setups confirm that the proposed methods work properly under reasonable finite sample sizes. For illustration, we have also applied these methods to the SOLVD(Studies of Left Ventricular Dysfunction, SOLVD Investigators, 1991) data.

Developing graphical and numerical methods for assessing the adequacy of the proposed models (3.1) and (4.1) will be our future work. In constructing the estimating equation (3.3), we consider the inverse probability censoring weight. We can explore other possible weight functions that incorporate the information of death and have the same expectation of $I(C_i \wedge D_i \geq t)$. The estimation of the inverse probability censoring weight involves fitting a Cox proportional hazard model to the censoring distribution which may not be true for some data. It would be worthwhile to explore the performance of our proposed method under model misspecification. Similarly, we can consider other weighting techniques for the proportional rate model (4.1) and compare the efficiency of the estimators based on different weights. In Chapter 3, we consider a marginal accelerated rate model. A natural extension would be to consider other types of models including the proportional odds model, the additive model or the semiparametric transformation model. We can also extend the proportional rate model proposed in Chapter 4 to other alternatives such as the accelerated failure time model, the additive model or the semiparametric transformation model.

Bibliography

Andersen, P.K., Borgan, O., Gill, R.D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer, New York.

Andersen, P.K. and Gill, R.D. (1982). Cox's regression model for counting processes: A Large-Sample Study. *The Annals of Statistics*, **10**, 1100-1120.

Breslow, N. (1974). Covariance analysis of censored survival data. *Biometrics*, **30**, 89-99. *Lecture Notes in Math*, **876**, 1-72.

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

Cai, J. and Prentice, R.L. (1995). Estimating equations for hazard ratio parameters based on correlated failure time data. *Biometrika*, **82**, 151-164.

Cai, J. and Prentice, R.L. (1997). Regression estimation using multivariate failure time data and a common baseline hazard function model. *Lifetime Data Analysis*, **3**, 197-213.

Cai, J. and Schaubel D.E. (2004). Analysis of recurrent event data. *Handbook of Statistics*, **23**, 603-623.

Carroll, R.J., Spiegelman, C.H., Lan, K.K.G., Gailey, K.T. and Abbott, R.D. (1984). On errors in variables for binary r egression models. *Biometrika*, **71**, 19-25.

Chambers, J.M., Mallows, C.L. and Stuck, B.W. (1976). A method for simulating stable random variables. *Journal of the American Statistical Association*, **71**, 340-344.

Chang, I.S. and Hsiung, C.A. (1994). Information and asymptotic efficiency in some generalized proportional hazards models for counting processes. *Annals of Statistics*, **22**, 1275-1298.

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

Chang, S.H. and Wang, M.C. (1999). Conditional regression analysis for recurrence time data. *Journal of the American Statistical Association*, **94**, 1221-1230.

Chen, H. Y. and Little, R.J.A. (1999). Proportional Hazards Regression with Missing Covariates. *Journal of the American Statistical Association*, **94**, 896-908.

Cheng, S.C., Wei, L.J. and Ying, Z. (1995). Analysis of transformation models with censored data. *Biometrika*, **82**, 835-845.

Clayton, D.G. and Cuzick, J. (1985). Multivariate generalizations of the proportional hazards model (with discussion). *Journal of the Royal Statistical Society, Series A*, **148**, 82-117.

Cook, R.J. and Lawless, J.F. (1997). Marginal analysis of recurrent events and a terminating event. *Statistics in Medicine*, **16**, 911-924.

Cox, D.R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B*, **34**, 182-220.

Cox, D.R. (1975). Partial likelihood. *Biometrika*, **62**, 269-276.

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

Deming,K.(1985). *Nonlinear functional analysis.* Springer, New York.

Fleming, T.R. and Harrington, D.P. (1991). *Counting Processes and Survival Analysis.* John Wiley and Sons, New York.

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.

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

Ghosh, D. and Lin, D.Y. (2002). Marginal regression models for recurrent and terminal events. *Statistica Sinica*, **12**, 663-688.

Ghosh, D. and Lin, D.Y. (2003). Semiparemetric analysis of recurrent events data in the presence of dependent censoring. *Biometrics*, **59**, 877-885.

Ghosh, D. (2004). Accelerated rates regression models for recurrent failure time data. *Lifetime Data Analysis*, **10**, 247-261.

Herring, A. H. and Ibrahim, J. G. (2001). Likelihood-based Methods for Missing Covariates in the Cox Proportional Hazards Model. *Journal of the American Statistical Association*, **96**, 292-302.

Hougaard P. (1986). Survival models for heterogeneous populations derived from positive stable distributions. *Biometrika*, **73**, 387-396.

Hougaard P. (1986). A class of multivariate failure time distributions. *Biometrika*, **73**, 671-678.

Huang, Y. (2000). Two-sample multistate accelerated sojourn times model. *Journal of the American Statistical Association*, **95**, 619-627.

Huang, Y. (2002). Calibration regression of censored lifetime medical cost. *Journal of the American Statistical Association*, **97**, 318-327.

Huang, Y. and Louis, T.A. (1998). Nonparametric estimation of the joint distribution of survival time and mark variables. *Biometrika*, **85**, 785-789.

Huang, Y. and Wang, M.C. (2003). Frequency of recurrent events at failure time: modelling and inference. *Journal of the American Statistical Association*, **98**, 663-670.

Huang, Y. and Wang, M.C. (2004). Joint modeling and estimation for recurrent event processes and failure time data. *Journal of the American Statistical Association*, **99**, 1153-1165.

Hughes, M.D. (1993), Regression dilution in the proportional hazards model. *Biometrika*, **49**, 1056-1066.

Kalbfleisch, J.D. and Prentice, R.L. (2002). *The statistical analysis of failure time data.* John Wiley and Sons, Hoboken, NJ.

Kanter, M. (1975). Stable density under change of scale and total variation inequalities. *The Annals of Probability*, **3**, 697-707.

Kirkpatrick, S., Jr., C. G., and Vecchi, M. (1983). Optimization by simulated annealing. *Science*, 220.

Kosorok, M.R. (2008). *Introduction to Empirical Processes and Semiparametric Inference.* Springer, New York.

Lee, E.W., Wei, L.J. and Ying, Z. (1993) Linear regression analysis for highly stratified failure time data. *Journal of the American Statistical Association*, **88**, 557-565.

Liang, K.Y., Self, S.G. and Chang, Y.C. (1993). Modelling marginal hazards in multivariate failure time data. *Journal of the Royal Statistical Society, Series B*, **55**, 441-453.

Lai, T.L. and Ying, Z. (1991a). Rank regression methods for left-truncated and right-censored data. *Annals of Statistics*, **19**, 531-556.

Lai, T.L. and Ying, Z. (1991b). Large sample theory of a modified Buckley-James estimator for regression analysis with censored data. *Annals of Statistics*, **19**, 1370-1402.

Li, Q. and Lagakos, S. (1997). Use of the Wei-Lin-Weissfeld method for the analysis of a recurring and a terminating event. *Statistics in Medicine*, **16**, 925-940.

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

Lee, E.W., Wei, L.J. and Amato, D.A. (1992). Cox-type regression analysis for large numbers of small groups of correlated failure time observations. In *Survival Analysis: State of the Arts* (Eds., J.P. Klein and P.K. Goel), Kluwer Academic Publishers, Dordrecht, 237-247

Lin, D.Y. and Geyer, C.J. (1992). Computational methods for semiparametric linear regression with censored data. *Journal of Computational and Graphical Statistics*, **1**, 77-90.

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

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

Lin, D.Y. and Ying, Z. (1993). Cox regression with incomplete covariate measurements. *Journal of the American Statistical Association*, **88**, 1341-1349.

Lin, J.S. and Wei, L.J. (1992). Linear regression analysis for multivariate failure time observations. *Journal of the American Statistical Association*, **87**, 1091-1097.

Liu, L., Wolfe, R.A. and Huang, X. (2004). Shared frailty models for recurrent events and a terminal event. *Biometrics*, **60**, 747-756.

McGilchrist, C.A. and Aisbett, C.W. (1991). Regression with frailty in survival analysis. *Biometrics*, **47**, 461-466.

Miloslavsky, M., Keles, M., var der Laan, M.J. and Butlers, S. (2004). Recurrent events analysis in the presence of time-dependent covariates and dependent censoring. *Journal of the Royal Statistical Society* **B**, **66**, 239-257.

Nadaraya, E.A. (1964). On estimating regression. *Theory of Probability and Its Applications*, **10**, 186-190.

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

Nelder, J.A., and Mead, R. (1965). A simplex algorithm for function minimization. *Computer Journal*, **7**, 308-313.

Paik, M. C. and Tsai, W.-Y. (1997). On Using the Cox Proportional Hazards Model with Missing Covariates. *Biometrika*, **84**, 579-593.

Pan, W. (2001). Using frailties in the accelerated failure time model. *Lifetime Data Analysis*, **7**, 55-64.

Parzen, M.I., Wei, L.J. and Ying, Z. (1994). A resampling methods based on pivotal estimating function. *Biometrika*, **81**, 341-350.

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

Pepe, M.S. and Fleming, T.R. (1991). A non-parametric method for dealing with mismeasured covariate data. *Journal of the American Statistical Association*, **86**, 108-113.

Pepe, M.S., Self, S.G. and Prentice, R.L. (1998). Further results on covariate measurement errors in cohort studies with time to response data. *Statistics in Medicine*, **8**, 1167-1178.

Peto, R. and Peto, J. (1972). Asymptotically efficient rank invariant procedures. *Journal of the Royal Statistical Society, Series A*, **135**, 185-207.

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

Prentice, R.L. (1982). Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika*, **69**, 331-342.

Prentice, R.L., Williams, B.J., and Peterson, A.V. (1981). On the regression analysis of multivariate failure time data. *Biometrika*, **68**, 373-389.

Pollard, D. (1984). *Convergence of Stochastic Processes.* Springer-Verlag, New York.

Reid, N. (1994). A conversation with Sir David Cox. *Statistical Science*, **9**, 439-455.

Ritov, Y. (1990). Estimation in a linear regression model with censored data. *Annals of Statistics*, **18**, 303-328.

Robins, J.M. and Rotnitzky, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. In: Jewell, N., Deitz, K., Farewell, V. (Eds.), AIDS *Epidemiology-Methodological Issues, Proceedings of the Biopharmaceutical Section.* American Statistical Association, pp.24-33.

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

Sen, P. and Singer, J.M. (1993). *Large Sample Methods in Statistics:An Introduction With Applications.* Chapman and Hall, 382 pp.

Shorack, G.R. and Wellner, J.A. (1986). *Empirical Processes with Applications to Statistics.* John Wiley and Sons, New York.

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

Strawderman, R.L. (2006). A regression model for dependent gap times. *The international Journal of Biostatistics*, **2**, No.1, Article 1.

SOLVD Investigators (1991). Effect of Enalapril on survival in patients with reduced left ventricular ejection fractions and congestive heart failure. *The New England Journal of Medicine*, **325**, 293-302.

Tsiatis, A.A. (1981). A large sample study of Cox's regression model. *Annals of Statistics*, **9**, 93-108.

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

Van der Vaart, A.W. (1998). *Asymptotic Statistics.* Cambridge University Press.

Van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics.* Springer-Verlag, New York.

Wang, M.C., Qin, J. and Chiang, C.T. (2001). Analyzing recurrent event data with informative censoring. *Journal of the American Statistical Association*, **96**, 1057-1065.

Watson, G.S. (1964) Smooth regression analysis. *Sankhya*, **26**, 359-372.

Wei, L.J. (1992). The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis. *Statistics in Medicine*, **11**, 1871-1879.

Wei, L.J. and Glidden, D.V. (1997). An overview of statistical methods for multiple failure time data in clinical trials(with discussion). *Statistics in Medicine*, **16**, 833-839.

Wei, L.J., Lin, E.Y. and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modelling marginal distributions. *Journal of the American Statistical Association*, **84**, 1065-1073.

Wei, L.J. (1992). The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis. *Statistics in Medicine*, **11**, 1871-1879.

Wei, L.J. and Glidden, D.V. (1997). An overview of statistical methods for multiple failure time data in clinical trials(with discussion). *Statistics in Medicine*, **16**, 833-839.

Wei, L.J., Lin, E.Y. and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modelling marginal distributions. *Journal of the American Statistical Association*, **84**, 1065-1073.

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

Ye,Y., Kalbfleisch, J., and Schaubel, D.E. (2007). Semiparametric analysis for correlated recurrent and terminal events. *Biometrics*, **63**, 78-87.

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

Zhou, H. and Wang, C.Y. (2000). Failure time regression with continuous covariates measured with error. *Journal of the Royal Statistical Society, Series B*, **62**, 657-665.

Zhou, H. and Pepe, M.S. (1995). Auxiliary covariate data in failure time regression. *Biometrika*, **82**, 139-149.