SEMIPARAMETRIC TIME-VARYING COEFFICIENT REGRESSION MODEL FOR LONGITUDINAL DATA WITH CENSORED TIME ORIGIN

by

Qiong Shou

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Applied Mathematics

Charlotte

2012

Approved by:

____
Dr. Yanqing Sun

____
Dr. Zongwu Cai

____
Dr. Jiancheng Jiang

____
Dr. Ron Sass

ABSTRACT

QIONG SHOU.Semiparametric time-varying coefficient regression model for longitudinal data with censored time origin. (Under the direction of DR. YANQING SUN)

In preventive HIV vaccine efficacy trials, thousands of HIV-negative volunteers are randomized to receive vaccine or placebo, and are monitored for HIV infection. The primary objective is to assess vaccine efficacy to prevent HIV infection. An important aspect of vaccine efficacy trials is to assess whether vaccine decreases secondary transmission of HIV and ameliorates HIV disease progression in vaccine recipients who become infected.

This thesis investigates the vaccine effect on the post HIV longitudinal biomarkers (e.g., viral loads and CD4 counts) over time since the actual HIV acquisition. The method applies to the situation when the time of the actual HIV acquisition may be missing or censored.

The problem is investigated under the semiparametric additive time-varying coefficient model where the influences of some covariates vary nonparametrically with time while the effects of the other covariates remain constant. The weighted profile least squares estimators are developed for the unknown parameters as well as for the nonparametric coefficient functions. The method uses the expectation maximization approach to deal with the censored time origin. The asymptotic properties of both the parametric and nonparametric estimators are derived and the consistent estimates of the asymptotic variances are given. The numerical simulations are conducted to examine finite sample properties of the proposed estimators.

ACKNOWLEDGMENTS

Upon the completion of this thesis I would sincerely gratefully express my thanks to many people. First of all I would like to show my respect and gratitude to my supervisor, Dr. Yanqing Sun who was greatly helpful and offered invaluable guidance to me on both academic performance and personal life during my study at University of North Carolina at Charlotte. Her attitude to research work and her attitude to life deeply engraved in my heart and memory. Special thanks also go to the members of the supervisory committee, Dr. Zongwu Cai, Dr. Jiancheng Jiang and Dr. Ron Sass without whose solid wisdom and abundant assistance I would not have accomplished my doctoral study. Also I would never forget the generous support from Dr. Xiyuan Qian at East China University of Science and Technology and Dr. Peter B. Gilbert at University of Washington and Fred Hutchinson Cancer Research Center. Without their professional knowledge and endless patience the application would not have been realized so successfully. Not forgetting to my honorable professors in the Department of Mathematics and Statistics department who supported me on such an unforgettable and unique study experience for five years.

Here allow me to express my full love and gratitude to my beloved families. Speechless thanks for their understanding, their support and their love during so many years from my birth. Finally, deep gratitude also due to my graduate friends through the duration of my study in this university. Thank all of them for sharing their time with me and priceless assistance.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

1.1 A motivating example

In preventive HIV vaccine efficacy trials, thousands of HIV-negative volunteers are randomized to receive vaccine or placebo, and are monitored for HIV infection. The primary objective is to assess vaccine efficacy to prevent HIV infection. An important aspect of vaccine efficacy trials is to assess whether vaccine decreases secondary transmission of HIV and ameliorates HIV disease progression in vaccine recipients who become infected (cf., Clemens *et al.*, 1997; Halloran *et al.*, 1997; Clements-Mann, 1998; Nabel, 2001; Shiver *et al.*, 2002; HVTN, 2004; IAVI, 2004).

We propose to investigate the vaccine effect on the post HIV longitudinal biomarkers (e.g., viral loads and CD4 counts). Viral load and CD4 counts have been found to be highly prognostic for both secondary transmission and progression to clinical disease in observational studies (cf., Mellors *et al.*, 1997; HIV Surrogate Marker Collaborative Group, 2000; Quinn *et al.*, 2000; Gray *et al.*, 2001). All previous analyses of HIV vaccine efficacy trials assessed these biomarkers based on the time from HIV positive diagnosis. However, it is biologically meaningful to assess whether vaccination modifies or accelerates the development of these biomarkers over time since the actual HIV acquisition. This assessment can be challenging since exact times of actual HIV acquisition are often unobtainable for trial participants. A brief description of HIV vaccine efficacy trial's diagnosis algorithm is given in the following.

HIV vaccine trials test volunteers for anti-HIV antibodies at periodic intervals (e.g., every 3 or 6 months); these antibody-based tests have near-perfect sensitivity to detect infections that occurred at least four weeks ago but otherwise may miss the infection. For all subjects with an HIV antibody positive (Ab+) test, a "look-back" procedure is applied wherein earlier available blood samples are tested for HIV infection using a more sensitive antigen-based HIV-specific PCR assay, which has near-perfect sensitivity if the infection occurred at least one week ago. Therefore, each infected subject is classified into one of two groups, defined by whether the *earliest* HIV positive sample is Ab- and PCR+ or is Ab+ and PCR+. The actual HIV acquisition time is approximated well by the time at Ab- and PCR+, while actual infection time occur approximately between the first Ab+ and earlier Ab- test times in the case of Ab+ and PCR+. The Ab+ and PCR+ cases occur in between 20% and 70% of infected subjects, with the rate depending on the frequency of HIV testing.

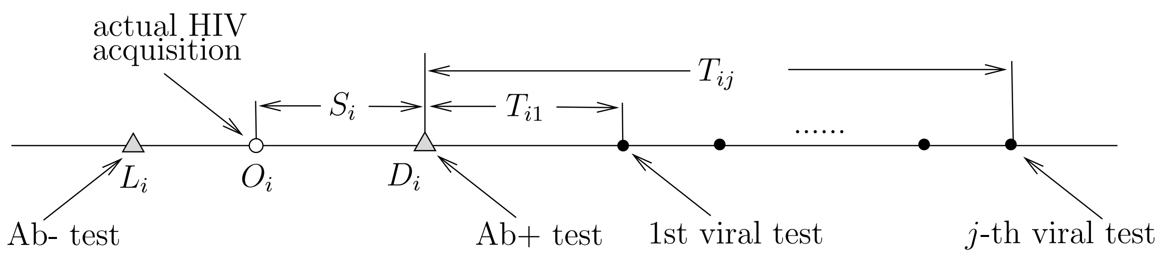


Figure 1.1: Time since actual HIV acqusition in case of Ab+ and PCR+.

Consider the $i = 1, \dots, n$ subjects who become HIV infected during the HIV vaccine efficacy trial. Let O_i be the time of actual HIV acquisition, D_i the HIV positive diagnosis time based on the trial's diagnosis algorithm (first Ab+ test time) and L_i the last Ab- test time. Post-infection biomarkers are measured at times $T_{i1}, \dots, T_{in_i}$, where T_{ij} is the time between the first Ab+ and the time at which the jth measurement is taken. Let S_i be the gap between HIV acquisition and the diagnosis, $S_i = D_i - O_i$. If subject i has an acute sample (Ab- and PCR+), the actual infection time can be well approximated by L_i and in this case let $S_i = D_i - L_i$. Otherwise, S_i is less than $D_i - L_i$. The S_i (time origin) is left censored by $D_i - L_i$ with censoring indicator R_i: $R_i = 1$ if S_i is observed and $R_i = 0$ if S_i is less than $D_i - L_i$. The time from actual HIV acquisition to the jth sampling time is then $T^o_{ij} = S_i + T_{ij}$. Figure 1.1 illustrates the set-up.

1.2 Existing works

The sampling times $T^o_{ij} = S_i + T_{ij}$ from the actual HIV acquisition are known when S_i is completely observed. In this case many existing statistical methods can be used to analyze model (2.1). Among others, recent works in this area include semiparametric methods by Moyeed & Diggle (1994), Zeger & Diggle (1994), and Liang, Wu & Carroll (2003), nonparametric methods by Hoover, Rice, Wu & Yang (1998), Wu, Chiang & Hoover (1998), Scheike & Zhang (1998), Wu & Zhang (2002), Wu & Liang (2004) and Sun & Wu (2003). Martinussen & Scheike (1999, 2000, 2001) and Lin & Ying (2001) considered time-varying coefficients regression models for longitudinal data and successfully integrated counting process techniques into the analysis of longitudinal data, providing further bridging between survival analysis, recurrent events, and time-dependent observations. Sun and Wu (2005) developed weighted least squares estimation procedure which avoids modeling of the sampling times is asymptotically more efficient than a single nearest neighbor smoothing which depends on estimation of the sampling model.

CHAPTER 2: ESTIMATION APPROACH THROUGH EM ALGORITHM

2.1 Preliminaries

Suppose that there is a random sample of n subjects. For subject i, let $Y_i(t)$ be the response process and let $X_i(t)$ and $Z_i(t)$ be the possibly time-dependent covariates of dimensions $(p+1)\times 1$ and $q\times 1$, respectively, where t is time since actual acquisition. The proposed general semiparametric time-varying coefficients regression model assumes that

$$Y_i(t) = \beta^T(t)X_i(t) + \gamma^T Z_i(t) + \epsilon_i(t), \qquad i = 1, \ldots, n \tag{2.1}$$

where $\beta(t)$ is an unspecified $(p+1)\times 1$ vector of smooth regression functions, γ is a $q\times 1$ dimensional vector of parameters, and $\epsilon_i(t)$ is a mean-zero process. The notation x^T represents transpose of a vector or matrix x. The first component of $X(t)$ is specified as 1 in general, giving to a model with a nonparametric baseline. The effect of $X(t)$ is modeled nonparametrically while the effect of $Z(t)$ follows a given parameter.

The observations of $Y_i(t)$ are taken at time points $T^o_{i1} < T^o_{i2} < \cdots < T^o_{in_i}$, where n_i is the total number of observations on the ith subject. The number of observations taken on the ith subject by time t is $N^o_i(t) = \sum_{j=1}^{n_i} I(T^o_{ij} \le t)$, where $I(\cdot)$ is the indicator function. Let C_i be the end of follow-up time or censoring time for the ith subject starting at HIV positive diagnosis (Ab+ test time). The censoring time C_i will be allowed to depend on the covariates $X_i(\cdot)$ and $Z_i(\cdot)$. The responses for the ith subject can only be observed at the time points before C_i. The censoring time since the actual time origin (HIV acquisition) is $S_i + C_i$.

Let

$$E\{dN_i^o(t)\,|\,X_i(t), Z_i(t)\} = \alpha(t, U_i(t))\,dt \equiv \alpha_i(t)\,dt, \qquad i = 1, \ldots, n, \tag{2.2}$$

where $U_i(t)$, a $m \times 1$ vector, is the part of the covariates $(X_i(t), Z_i(t))$ that affects the potential sampling times. The function $\alpha(t, \mathbf{u})$ is an unspecified nonnegative smooth function.

The time S_i from actual HIV acquisition to HIV positive diagnosis may be left censored. Let $R_i = I(S_i \geq V_i)$ be the censoring indicator. For the application concerned here, the censoring time V_i (e.g. $D_i - L_i$) is always observed. Let $\mathcal{D}_i = \{V_i, C_i, A_i, T_{ij}, X_i(T_{ij}^o), Z_i(T_{ij}^o), Y_i(T_{ij}^o), j = 1, \ldots, n_i\}$, where A_i is a collection of possible auxiliary variables that are not of interest in the modelling of $Y_i(t)$ but may be useful in predicting the distribution of S_i. The observed data for subject i can be expressed as $\mathcal{X}_i = \{R_i S_i, R_i, \mathcal{D}_i\}$. The observation is $\{S_i, \mathcal{D}_i\}$ if $R_i = 1$ and $\mathcal{D}_i$ if $R_i = 0$. Although exact times T_{ij}^o may be unobtainable, the values $X_{ij} = X_i(T_{ij}^o)$, $Z_{ij} = Z_i(T_{ij}^o)$ and $Y_{ij} = Y_i(T_{ij}^o)$ at T_{ij}^o are known. Denote the observed data by $\mathcal{X} = \{\mathcal{X}_i, i = 1, 2, \ldots, n\}$.

Assume that the censoring time C_i is noninformative in the sense that $E\{dN_i^o(t)$ $|X_i(t), Z_i(t), S_i + C_i \geq t\} = E\{dN_i^o(t)|X_i(t), Z_i(t)\}$ and $E\{Y_i(t)|X_i(t), Z_i(t), S_i + C_i \geq t\} = E\{Y_i(t)|X_i(t),\, Z_i(t)\}$. Assume also that $Y_i(t)$ and $N_i^o(t)$ are independent conditional on $X_i(t)$, $Z_i(t)$ and $S_i + C_i \geq t$. This assumption implies that, conditional on covariate processes, sampling times are noninformative for the response. Note that dependence between response and sampling times as well dependence between sampling times and the censoring time C_i is often induced by ignoring certain covariates (cf., Miloslavsky *et al.,* 2004 and Zeng, 2005). The stated conditional independence assumptions make the proposed methods applicable to situations where dependence may exist among response process, sampling times and censoring time C_i but becoming independent by including appropriate additional covariates. A recent work by Sun and Lee (2011) on testing independent censoring for longitudinal data provides needed procedures for checking such assumptions. Let $N_i(t) = \sum_{j=1}^{n_i} I(T_{ij} \le t)$. Assume $E\{Y_i(t)|X_i(\cdot), Z_i(\cdot), N_i(\cdot), S_i, V_i, C_i\} = E\{Y_i(t)|X_i(\cdot), Z_i(\cdot)\}$.

When all S_i's are observed, the existing statistical methods cited in Section 1.2 can be used to analyze model (2.1). However, none of these methods address the problem in which the time origin may be censored. We propose to extend the investigation of model (2.1) to accommodate this situation.

2.2 Estimation Procedures

It is important to note that if the unobserved or censored S_i is treated as missing, then S_i is not missing at random in the sense of Robin (1976). The inverse probability weighting of complete-cases method of Horvitz and Thompson (1952) and the augmented inverse probability weighted complete-case method of Robins, Rotnitzky and Zhao (1994), which have been successfully adapted in Sun and Gilbert (2011), Sun, Wang and Gilbert (2011) and by many other authors, will not work in this situation. We propose an estimation procedure based on the missing-data principle using the EM-algorithm. The EM-algorithm has been applied by Scheike and Sun (2007) to develop maximum likelihood estimation for tied survival data under Cox regression model.

Let $F_S(s|\mathcal{D}_i)$ be the conditional distribution of S_i given $\mathcal{D}_i$. The conditional distribution of S_i given $\mathcal{D}_i$ and $R_i = 0$, $F_S(s|\mathcal{D}_i, R_i = 0)$, equals $F_S(s|\mathcal{D}_i)/F_S(V_i|\mathcal{D}_i)$ for $s \le V_i$ and 1 for $s > V_i$. Assume that $\max\{S_i, V_i\}$ is bounded by a predetermined constant c. This is reasonable since for the application concerned here $\max\{S_i, V_i\}$ is less than the time interval between two consecutive testing times which is usually between 3 and 6 months. The distribution of S_i based on the left censored data can be estimated by using the right censored data through the transformation $\{\min\{c - S_i, c - V_i\}, R_i = I(c - S_i \le c - V_i)\}$. Thus, the Kaplan-Meier estimator can be used to estimate the distribution of S_i when S_i is independent of $\mathcal{D}_i$. Otherwise, a failure time regression model such as the Cox model (Cox, 1972) can be used to estimate the conditional distribution $F_S(s|\mathcal{D}_i)$. Observing the censoring time V_i for all subjects is a key factor in the estimation of $F_S(s|\mathcal{D}_i, R_i = 0)$. Otherwise $F_S(s|\mathcal{D}_i, R_i = 0)$ is not identifiable.

Let $\widehat{F}_S(s|\mathcal{D}_i)$ be the estimated conditional distribution of $F_S(s|\mathcal{D}_i)$. The probability $\pi_i = P(R_i = 1|\mathcal{D}_i) = P(S_i \geq V_i|\mathcal{D}_i)$ can be estimated by $\widehat{\pi}_i = 1 - \widehat{F}_S(V_i|\mathcal{D}_i)$. Let $dN_i^c(t) = I(S_i + C_i \geq t)dN_i^o(t)$. The estimation of model (2.1) will be based on targeting to minimize the following objective function:

$$
\begin{aligned}
l_t(\beta,\gamma) \;=\; & \sum_{i=1}^{n} R_i \int_0^{\tau} W_i(u)\{Y_i(u) - \beta^T(u)X_i(u) - \gamma^T Z_i(u)\}^2 \, dN_i^c(u) \\
& + \sum_{i=1}^{n} (1-R_i)\widehat{E}_S\Big\{ \int_0^{\tau} W_i(u)\{Y_i(u) - \beta^T(u)X_i(u) \\
& -\gamma^T Z_i(u)\}^2 \, dN_i^c(u)|\mathcal{X} \Big\}, \qquad (2.3)
\end{aligned}
$$

where $W_i(\cdot)$ is a nonnegative weight function, and $\widehat{E}_S\{\cdot|\mathcal{X}\}$ is the estimate of the conditional expectation, $E_S\{\cdot|\mathcal{X}\}$, of a function of S_i given $\mathcal{X}$. For a random function $G_n(t, X_i(t), Z_i(t), Y_i(t))$, $\widehat{E}_S\left\{\int_0^{\tau} G_n(u, X_i(u), Z_i(u), Y_i(u))dN_i^c(u)|\mathcal{X}\right\}$ equals

$$
\begin{aligned}
& \sum_{j=1}^{n_i} \widehat{E}_S\{G_n(S_i + T_{ij}, X_i(T_{ij}^o), Z_i(T_{ij}^o), Y_i(T_{ij}^o))I(C_i \geq T_{ij})I(S_i + T_{ij} \leq \tau)|\mathcal{X}\} \\
= & \sum_{j=1}^{n_i} \widehat{E}_S\{G_n(S_i + T_{ij}, X_{ij}, Z_{ij}, Y_{ij})I(C_i \geq T_{ij})I(S_i + T_{ij} \leq \tau)|\mathcal{X}\} \\
= & \sum_{j=1}^{n_i} \int_0^{\infty} G_n(s + T_{ij}, X_{ij}, Z_{ij}, Y_{ij})I(C_i \geq T_{ij})I(s + T_{ij} \leq \tau)\, d\widehat{F}_S(s|\mathcal{X})
\end{aligned}
$$

Since $F_s(s|\mathcal{X})$ is the conditional distribution of S_i given $\mathcal{X}$ for ith subject with $R_i = 0$ and $\mathcal{X}_i = \{S_i, R_i = 1, \mathcal{D}_i\} \cup \{\mathcal{D}_i, R_i = 0\}$, $F_s(s|\mathcal{X}) = F_s(s|\mathcal{D}_i, R_i = 0)$ by independence. This also holds for its estimator $\widehat{F}_s(s|\mathcal{X})$. Hence the above term equals

$$
\sum_{j=1}^{n_i} \int_0^{\infty} G_n(s + T_{ij}, X_{ij}, Z_{ij}, Y_{ij})I(C_i \geq T_{ij})I(s + T_{ij} \leq \tau)\, d\widehat{F}_S(s|\mathcal{D}_i, R_i = 0)
$$

$$= \sum_{j=1}^{n_i} \int_0^\infty G_n(s+T_{ij}, X_{ij}, Z_{ij}, Y_{ij}) I(C_i \ge T_{ij}) I(T_{ij} \le \tau - s)\, d\widehat{F}_S(s|\mathcal{D}_i)/\widehat{F}_S(V_i|\mathcal{D}_i).$$

Note that the function $G_n(u, X_i(u), Y_i(u), Z_i(u))$ maybe depend on the observed data which makes it measurable with respect to $\mathcal{X}$ for each fixed $(u, X_i(u), Y_i(u), Z_i(u))$.

Taking derivative of $l_t(\beta,\gamma)$ with respect to β for a fixed γ yields

$$\frac{\partial l_t(\beta,\gamma)}{\partial\beta} = -2\sum_{i=1}^{n} \ll W_i(t)X_i(t)\{Y_i(t) - \beta^T(t)X_i(t) - \gamma^T Z_i(t)\}\, dN_i^c(t) \gg_R, \quad (2.4)$$

where and hereafter, the notation $\ll H_i(t) \gg_R = R_i H_i(t) + (1-R_i)\widehat{E}_S\{H_i(t)|\mathcal{X}\}$ is used for a random function $H_i(t)$. This leads to the following estimating function

$$U_t(\beta,\gamma) = \sum_{i=1}^{n} \ll W_i(t)X_i(t)\{Y_i(t) - \beta^T(t)X_i(t) - \gamma^T Z_i(t)\}\, dN_i^c(t) \gg_R . \quad (2.5)$$

The root of the equation $U_t(\beta,\gamma) = 0$ is denoted by $\tilde{\beta}(t,\gamma)$. However, from $U_t(\beta,\gamma) = 0$ we obtain $\sum_{i=1}^n \ll W_i(t)X_i(t)X_i^T(t)\beta(t)\, dN_i^c(t) \gg_R = \sum_{i=1}^n \ll W_i(t)X_i(t)\{Y_i(t) - Z_i^T(t)\gamma\}\, dN_i^c(t) \gg_R$. The equation has no solution for $\beta(t)$ for fixed γ because of sparsity of the data at time t. However, the solution exists by gathering the data around a neighborhood of t. Let $\tilde{E}_{zx}(t) = n^{-1}\sum_{i=1}^n \ll \int_0^\tau K_h(u-t)Z_i(u)X_i^T(u)\, dN_i^c(u) \gg_R$, where $K_h(t) = h^{-1}K(t/h)$, $K(t)$ is a symmetric kernel function with a compact support and h is the bandwidth depending on n. The $\tilde{E}_{yx}(t)$ and $\tilde{E}_{xx}(t)$ are similarly defined by replacing Z_i with Y_i and X_i respectively. The local least squares estimator for $\beta(t)$ for fixed γ is then given by

$$\tilde{\beta}(t;\gamma) = \tilde{Y}_x^T(t) - \tilde{Z}_x^T(t)\gamma, \quad (2.6)$$

where $\tilde{Y}_x(t) = \tilde{E}_{yx}(t)(\tilde{E}_{xx}(t))^{-1}$ and $\tilde{Z}_x(t) = \tilde{E}_{zx}(t)(\tilde{E}_{xx}(t))^{-1}$. Replacing $\tilde{\beta}(t;\gamma)$ for $\beta(t)$ in (2.3) and taking derivative with respect to γ, we obtain the profile estimating equation for γ:

$$\sum_{i=1}^{n} \int_{t_1}^{t_2} \ll W_i(t)\{Z_i(t) - \tilde{Z}_x(t)X_i(t)\}\{Y_i(t) - X_i^T(t)\tilde{\beta}(t;\gamma)$$

$$-Z_i^T(t)\gamma\}\, dN_i^c(t) \gg_R = 0. \tag{2.7}$$

From (2.7), we solve for γ to get $\widehat{\gamma}$ which equals $\widehat{D}^{-1}\widehat{W}$ where

$$\begin{aligned}
\widehat{D} &= n^{-1}\sum_{i=1}^{n}\int_{t_1}^{t_2} \ll W_i(t)\{Z_i(t)-\tilde{Z}_x(t)X_i(t)\}^{\otimes 2} dN_i^c(t) \gg_R \\
\widehat{W} &= n^{-1}\sum_{i=1}^{n}\int_{t_1}^{t_2} \ll W_i(t)\{Z_i(t)-\tilde{Z}_x(t)X_i(t)\}\{Y_i(t)-X_i^T(t)\tilde{Y}_x^T(t)\} dN_i^c(t) \gg_R .
\end{aligned}$$

An estimator of $\beta(t)$ is given by $\widehat{\beta}(t) = \tilde{\beta}(t;\widehat{\gamma})$.

When S_i is observed for all subjects, $R_i = 1$. The estimators for $\beta(t)$ and γ are reduced to those under Sun and Wu (2005). However, when S_i is censored, the estimating equations (2.4) and (2.7) are weighted according to the conditional distribution of S_i so that the estimated covariate effects correspond to those at the time since the actual time origin. A key factor for this procedure to work is that the censoring time V_i is observed for all subjects so that the estimation of $F_S(s \mid \mathcal{D}_i, R_i = 0)$ is possible.

2.3 Computational algorithm

The boundary effects on the estimation of $\beta(t)$ and the covariance matrix of its estimator can be reduced by applying the equivalent kernel for the local linear approach; see Fan & Gijbels (1996).

Suppose the binary data $(T_1, B_1), (T_2, B_2), \cdots, (T_n, B_n)$ which are n independent and identically distributed copies from (T, B). To estimate $m(t_0) = E(B|T = t_0)$ is of interest. Suppose we use symmetric kernel $K(x)$ in local constant method. Then the local constant estimator of $m(t)$ at point t_0 will be

$$\widehat{m}_C = \frac{n^{-1}\sum_{i=1}^n K_h(T_i - t_0)B_i}{n^{-1}\sum_{i=1}^n K_h(T_i - t_0)}.$$

To get the equivalent kernel, we will mimic some notations in Fan & Gijbels

(1996).

$$S_{n,j}(t_0) = \sum_{i=1}^{n} K_h(T_i - t_0)(T_i - t_0)^j, \;\; j = 0, 1, 2.$$

Then

$$S_n(t_0) = \begin{pmatrix} S_{n,0}(t_0) & S_{n,1}(t_0) \\ S_{n,1}(t_0) & S_{n,2}(t_0) \end{pmatrix}.$$

Meanwhile the inverse can be written as

$$S_n^{-1}(t_0) = \frac{1}{S_{n,0}(t_0)S_{n,2}(t_0) - S_{n,1}^2(t_0)} \begin{pmatrix} S_{n,2}(t_0) & -S_{n,1}(t_0) \\ -S_{n,1}(t_0) & S_{n,0}(t_0) \end{pmatrix}.$$

As stated in the Section 3.2.2 of Fan & Gijbels (1996), the equivalent kernel for local linear approach is

$$K_h^*(t - t_0) = e_1^T S_n^{-1}(t_0)(1 \;\; t - t_0)^T K_h(t - t_0),$$

where $e_1 = (1 \;\; 0)^T$. Thus we can simplify the equivalent kernel as follows.

$$\begin{aligned} K_h^*(t - t_0) &= e_1^T S_n^{-1}(t_0)(1 \;\; t - t_0)^T K_h(t - t_0) \\ &= \frac{K_h(t - t_0)(\; 1 \;\; 0 \;)}{S_{n,0}(t_0)S_{n,2}(t_0) - S_{n,1}^2(t_0)} \begin{pmatrix} S_{n,2}(t_0) & -S_{n,1}(t_0) \\ -S_{n,1}(t_0) & S_{n,0}(t_0) \end{pmatrix} \begin{pmatrix} 1 \\ t - t_0 \end{pmatrix} \\ &= \frac{\{S_{n,2}(t_0) - S_{n,1}(t_0)(t - t_0)\}K_h(t - t_0)}{S_{n,0}(t_0)S_{n,2}(t_0) - S_{n,1}^2(t_0)}. \end{aligned}$$

Therefore, the local linear estimator $\widehat{m}_L$ at point t_0 under the model $B = m(T) + \epsilon$ is

$$\frac{n^{-1}\sum_{i=1}^n K_h^*(T_i - t_0)B_i}{n^{-1}\sum_{i=1}^n K_h^*(T_i - t_0)} = \frac{\sum_{i=1}^n \{S_{n,2}(t_0) - S_{n,1}(t_0)(T_i - t_0)\}K_h(T_i - t_0)B_i}{\sum_{i=1}^n \{S_{n,2}(t_0) - S_{n,1}(t_0)(T_i - t_0)\}K_h(T_i - t_0)}.$$

Compared to the local constant estimator above, if we use the following kernel

$$W_h(T_i - t_0) = \{S_{n,2}(t_0) - S_{n,1}(t_0)(T_i - t_0)\}K_h(T_i - t_0) \tag{2.8}$$

instead of $K_h(T_i - t_0)$, we simply obtain the local linear estimator.

Let $f(t)$ be the density function of T. For a interior point t_0, the local linear estimator is asymptotically equivalent to the local constant estimator as $h \to 0$ and $nh^5 = O(1)$ since for a symmetric kernel, $\int K(x)x\,dx = 0$. Then

$$\begin{aligned} n^{-1}ES_{n,j}(t_0) &= EK_h(T_i - t_0)(T_i - t_0)^j = \int K_h(t - t_0)(t - t_0)^j f(t)\,dt \\ &= \int K(x)h^j x^j f(t_0 + hx)\,dx = h^j(f(t_0) + o(h)) \int K(x)x^j\,dx = o(h). \end{aligned}$$

Especially note that $n^{-1}ES_{n,1}(t_0) = 0$. Hence

$$\begin{aligned} \widehat{m}_L &= \frac{\sum_{i=1}^n \{S_{n,2}(t_0) - S_{n,1}(t_0)(T_i - t_0)\}K_h(T_i - t_0)B_i}{\sum_{i=1}^n \{S_{n,2}(t_0) - S_{n,1}(t_0)(T_i - t_0)\}K_h(T_i - t_0)} \\ &= \frac{n^{-1}\sum_{i=1}^n \{n^{-1}S_{n,2}(t_0) - n^{-1}S_{n,1}(t_0)h\frac{T_i - t_0}{h}\}K_h(T_i - t_0)B_i}{n^{-1}\sum_{i=1}^n \{n^{-1}S_{n,2}(t_0) - n^{-1}S_{n,1}(t_0)h\frac{T_i - t_0}{h}\}K_h(T_i - t_0)} \\ &\approx \frac{n^{-1}\sum_{i=1}^n K_h(T_i - t_0)B_i}{n^{-1}\sum_{i=1}^n K_h(T_i - t_0)} + o_p(h^2) \\ &= \widehat{m}_C + o_p(h^2). \end{aligned}$$

Thus $(nh)^{1/2}(\widehat{m}_L - \widehat{m}_C) = o_p((nh^5)^{1/2})$, which means the asymptotic distributions for the local linear estimator and the local constant estimator are the same for an interior point t_0 as $h \to 0$ and $nh^5 = O(1)$. This enables using the equivalent kernel for the boundary time points while using the kernel in local constant approach for the interior time points.

In estimating $\beta(t)$, time points T may be unknown since S_i is left censored by V_i. Then we cannot simply use $S_{n,j}(t_0)$ defined above. Let

$$S_{n,j}(t) = \sum_{i=1}^n \ll \int_0^\tau K_h(u - t)(u - t)^j dN_i^c(u) \gg_R, \; j = 0, 1, 2.$$

Now under the new definition of $S_{n,j}(t_0)$, we still have the form of equivalent kernel in (2.8) for local linear approach of estimating $\beta(t)$.

2.4 Cross-validation bandwidth selection

The optimal theoretical bandwidth is difficult to achieve since it would involve estimating the second derivative $\beta''(t)$; see Fan and Gijbels (1996) and Cai and Sun (2002). In practice, the appropriate bandwidth selection can be based on a cross-validation method. This approach is widely used in nonparametric function estimation literature; see Rice and Silverman (1991) for leave-one-subject-out cross-validation approach and Tian, Zucker and Wei (2005) for K-fold cross-validation approach.

An analog of the K-fold cross-validation approach in the current setting is to divide the data into K equal-sized groups. Let D_k denote the kth subgroup of data, then the kth prediction error is given by

$$PE_k(h) = \sum_{i \in D_k} \ll \int_{t_1}^{t_2} \left[Y_i(t) - (\widehat{\beta}_{(-k)}(t))^T X_i(t) - \widehat{\gamma}_{(-k)}^T Z_i(t) \right]^2 dN_i^c(t) \gg_R, \quad (2.9)$$

for $k = 1, \ldots, K$, where $\widehat{\beta}_{(-k)}(t)$ and $\widehat{\gamma}_{(-k)}$ are the estimators of $\beta(t)$ and γ based on the data without the subgroup D_k. The data-driven bandwidth selection based on the K-fold cross-validation is to choose the bandwidth h that minimizes the total prediction error $PE(h) = \sum_{k=1}^{K} PE_k(h)$. This bandwidth selection procedure will be further studied and tested empirically through simulations.

CHAPTER 3: ASYMPTOTIC PROPERTIES

In this section we will explore the asymptotic properties of our estimators. First we will define some notations for the future use. Let

$$e_{zx}(t) = E(\xi_i(t)\alpha_i(t)Z_i(t)X_i^T(t)),$$

where $\xi_i(t) = I(S_i + C_i + c_1 \geq t)$. Similarly we can define $e_{xx}(t)$ and $e_{yx}(t)$. Also let $y_x(t) = e_{yx}(t)(e_{xx}(t))^{-1}$, $z_x(t) = e_{zx}(t)(e_{xx}(t))^{-1}$ and γ_0, $\beta_0(t)$ be the true value or true curve respectively. Then

$$\begin{aligned}
& y_x^T(t) - z_x^T(t)\gamma_0 \\
= \; & (e_{xx}(t))^{-1}e_{yx}^T(t) - (e_{xx}(t))^{-1}e_{zx}^T(t)\gamma_0 \\
= \; & (e_{xx}(t))^{-1}[E(\xi_i(t)\alpha_i(t)X_i(t)Y_i^T(t)) - E(\xi_i(t)\alpha_i(t)X_i(t)Z_i^T(t))\gamma_0] \\
= \; & (e_{xx}(t))^{-1}E(\xi_i(t)\alpha_i(t)X_i(t)[Y_i^T(t) - Z_i^T(t)\gamma_0]) \\
= \; & (e_{xx}(t))^{-1}E(\xi_i(t)\alpha_i(t)X_i(t)[X_i^T(t)\beta_0(t) + \epsilon^T(t)]) \\
= \; & (e_{xx}(t))^{-1}e_{xx}(t)\beta_0(t) + E(E[\xi_i(t)\alpha_i(t)X_i(t)\epsilon^T(t) \mid X_i(t), Z_i(t), S_i + C_i \geq t]) \\
= \; & \beta_0(t) + E(\xi_i(t)\alpha_i(t)X_i(t)E[\epsilon^T(t) \mid X_i(t), Z_i(t), S_i + C_i \geq t]) \\
= \; & \beta_0(t) + E(\xi_i(t)\alpha_i(t)X_i(t)E[\epsilon^T(t) \mid X_i(t), Z_i(t)]) = \beta_0(t).
\end{aligned}$$

Let $\beta^*(t) = \tilde{y}_x^T(t) - \tilde{z}_x^T(t)\gamma_0$ where $\tilde{y}_x(t) = \tilde{e}_{yx}(t)(\tilde{e}_{xx}(t))^{-1}$, $\tilde{z}_x(t) = \tilde{e}_{zx}(t)(\tilde{e}_{xx}(t))^{-1}$ and $\tilde{e}_{yx}(t) = \int_0^\tau K_h(u-t)e_{yx}(u)du$. We have the fact that $\tilde{e}_{yx}(t) = \int_0^\tau K_h(u-t)e_{yx}(u)du \xrightarrow{P} e_{yx}(u)$ as $h \to 0$. Similar definitions can de defined for $\tilde{e}_{zx}(t)$ and $\tilde{e}_{xx}(t)$. And similar facts hold too. Also we denote the transposes of the matrices by changing the order of the subscripts. Now let us state the following conditions.

Conditions (I). Assume that the $\{n_i\}$ are bounded; the $\{S_i\}$ are bounded by a large enough value L and independent of $\mathcal{D}_i$; the kernel function $K(\cdot)$ is symmetric with compact support on $[-1,1]$; the processes $X_i(t)$, $Z_i(t)$ and $\alpha_i(t)$, $0 \le t \le \tau$, are bounded by a constant, continuous and their total variations are bounded by a constant; the values of the jth measurement X_{ij} and Z_{ij} are also bounded; $(e_{xx}(t))^{-1}$ for $0 \le t \le \tau$ are bounded; the weight function $W_i(t)$ can be written as a difference of two monotone functions and each converges to a deterministic function so that $W_i(t)$ converges to $w(t)$ for all i.

Under the conditions above and by Lemma A.2.3 we can prove that $\tilde{E}_{zx}(t) \xrightarrow{P} e_{zx}(t)$ uniformly in $t \in [t_1, t_2] \subset [0, \tau]$. Similar asymptotic results hold for $\tilde{E}_{yx}(t)$ and $\tilde{E}_{xx}(t)$. By continuous mapping theorem, the above results lead to the conclusion that $\tilde{Y}_x(t)$ and $\tilde{Z}_x(t)$ converge to $y_x(t)$ and $z_x(t)$ uniformly in $t \in [t_1, t_2]$ respectively.

Both parametric and nonparametric estimators we proposed in the previous chapter are consistent. First we apply (2.6) to (2.3), we can get $n^{-1}\tilde{l}(\gamma)$ equals

$$n^{-1} \sum_{i=1}^{n} \ll \int_0^\tau W_i(s)\{Y_i(s) - \tilde{Y}_x(s)X_i(s) + \gamma^T(\tilde{Z}_x(s)X_i(s) - Z_i(s))\}^2 \, dN_i^c(s) \gg_R$$

which is a random function of γ. This function can be proved to uniformly converge to a deterministic function of γ. Then followed by Theorem 5.7 of van der Vaart (1998), we obtained the consistency of $\widehat{\gamma}$.

Theorem 3.1: (Consistency of $\widehat{\gamma}$) Under Condition (I), $\widehat{\gamma} = \widehat{D}^{-1}\widehat{W}$ converges to its true value γ_0 in probability as $n \to \infty$. $\square$

Then by the definition of $\widehat{\beta}(t)$, it is apparent to show

Theorem 3.2: (Consistency of $\widehat{\beta}(t)$) Under Condition (I), $\widehat{\beta}(t) = \tilde{\beta}(t; \widehat{\gamma})$ converges to $\beta_0(t)$ in probability uniformly on $[t_1, t_2]$ as $n \to \infty$, where $0 \le t_1 \le t_2 \le \tau$. $\square$

Also we can obtain the asymptotic distribution of $\widehat{\gamma}$ and $\widehat{\beta}(t)$ for a fix t. In

Section 2.2 $\widehat{\gamma}$ is the solution of (2.7). So denote $U(\gamma)$ as

$$\sum_{i=1}^{n} \int_{t_1}^{t_2} \ll W_i(t)\{Z_i(t) - \tilde{Z}_x(t)X_i(t)\}\{Y_i(t) - X_i^T(t)\tilde{\beta}(t;\gamma) - Z_i^T(t)\gamma\}\, dN_i^c(t) \gg_R$$

which is usually called the score function. Then the Taylor expansion of $U(\widehat{\gamma})$ at γ_0 is

$$n^{1/2}(\widehat{\gamma} - \gamma_0) = -\left(n^{-1}\frac{\partial U(\gamma^*)}{\partial \gamma^T}\right)^{-1} [n^{-1/2}U(\gamma_0)],$$

where γ^* is on the line segment between $\widehat{\gamma}$ and γ_0. To prove the asymptotic normality of $n^{1/2}(\widehat{\gamma}-\gamma_0)$, it is sufficient to prove the convergence in probability to a non-singular matrix of $n^{-1}\frac{\partial U(\gamma^*)}{\partial \gamma^T}$, and the weak convergence of $n^{-1/2}U(\gamma_0)$. The convergence in probability can be easily obtained by applying Lemma A.2.2. And $n^{-1/2}U(\gamma_0)$ can be derived to equal to

$$\begin{aligned} n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2} & w(t)\{Z_i(t) - z_x(t)X_i(t)\}\epsilon_i(t)[R_i dN_i^c(t) \\ & + E_s\{(1-R_i)dN_i^c(t) \mid \mathcal{D}_i, R_i = 0\}] + o_p(1). \end{aligned}$$

Then applying theorem 5.21 (van der Vaart, 1998) to the sore function, the asymptotic normality of $\widehat{\gamma}$ is presented in the following theorem.

Theorem 3.3: (Asymptotic Normality of $\widehat{\gamma}$) Under Condition (I), $n^{1/2}(\widehat{\gamma} - \gamma_0) \xrightarrow{D} \mathcal{N}(0, D^{-1}VD^{-1})$ as $n \to \infty$ where

$$\begin{aligned} D &= E\left(\int_{t_1}^{t_2} w(t)\{Z_i(t) - z_x(t)X_i(t)\}^{\otimes 2}\, dN_i^c(t)\right), \\ V &= E\left\{\int_{t_1}^{t_2} [R_i w(t)(Z_i(t) - z_x(t)X_i(t))\epsilon_i(t)dN_i^c(t) \right. \\ & \left. \quad + (1-R_i)E_s\{w(t)(Z_i(t) - z_x(t)X_i(t))\epsilon_i(t)dN_i^c(t) \mid \mathcal{D}_i, R_i = 0\}]\right\}^{\otimes 2}. \quad \square \end{aligned}$$

Based on the equations (A.9) and (A.11), the asymptotic variance above can be estimated by $\widehat{D}^{-1}\widehat{V}\widehat{D}^{-1}$ where

$$\widehat{D} = n^{-1}\sum_{i=1}^{n}\int_{t_1}^{t_2} \ll W_i(t)\{Z_i(t) - \tilde{Z}_x(t)X_i(t)\}^{\otimes 2}\, dN_i^c(t) \gg_R,$$

$$\widehat{V} = n^{-1}\sum_{i=1}^{n}\left\{\int_{t_1}^{t_2} \ll W_i(t)(Z_i(t) - \tilde{Z}_x(t)X_i(t))\widehat{\epsilon}_i(t)\,dN_i^c(t) \gg_R\right\}^{\otimes 2}$$

and $\widehat{\epsilon}_i(t) = Y_i(t) - \widehat{\beta}(t)^T X_i(t) - \widehat{\gamma}^T Z_i(t)$. This estimator is consistent estimator of the asymptotic variance by the consistency of $\widehat{D}$ and $\widehat{V}$.

Before demonstrating the asymptotic normality of $\widehat{\beta}(t)$ at each fixed time point t, we first introduce the following notations. We know that $N_i^c(t)$ is a counting process. Let the filtration $\mathcal{F}_t^c = \sigma\{N_i^c(s), R_i, X_i(\cdot), Z_i(\cdot), Y_i(\cdot), 0 \le s \le t\}$. By the Doob-Meyer decomposition theorem, under this filtration there is a unique pair of a martingale $M_i^c(t)$ and a compensator $A_i^c(t)$ which can be defined as $\int_0^t \sum_{j=1}^{n_i} I(T_{ij}^0 \ge s)\alpha_i^c(s)ds$ such that $N_i^c(t) = A_i^c(t) + M_i^c(t)$. Let $Y_i^c(t) = \sum_{j=1}^{n_i} I(T_{ij}^0 \ge t)$.

By definitions we can obtain that

$$\begin{aligned} & (nh)^{1/2}(\widehat{\beta}(t) - \beta^*(t)) \\ = \;& (nh)^{1/2}(\tilde{\beta}(t;\gamma_0) - \beta^*(t)) + (nh)^{1/2}(\widehat{\gamma} - \gamma_0)\frac{\partial\tilde{\beta}(t;\gamma_0)}{\partial\gamma} + O_p(n^{-1/2}h^{1/2}), \end{aligned}$$

and

$$\beta^*(t) = \beta_0(t) + (1/2)\mu_2 h^2 (e_{xx}(t))^{-1}[e''_{xy}(t) - e''_{xz}(t)\gamma_0 - e''_{xx}(t)\beta_0(t)] + o(h^2).$$

So it suffices to focus on the difference $(nh)^{1/2}(\tilde{\beta}(t;\gamma_0) - \beta^*(t))$ to get the following theorem.

Theorem 3.4: (Asymptotic Normality of $\widehat{\beta}(t)$) Under Condition (I), $((nh)^{1/2}(\widehat{\beta}(t) - \beta_0(t) - \beta_{Bias}(t)) \xrightarrow{D} \mathcal{N}(0, \mu_0\Sigma(t))$ for each fixed time point t as $n \to \infty$, $h \to 0$, $nh \to \infty$ and $nh^5 = O(1)$. Here $\mu_0 = \int_{-1}^{1} K^2(u)du$, $\mu_2 = \int_{-1}^{1} u^2 K^2(u)du$,

$$\begin{aligned} \beta_{Bias}(t) = \;& (1/2)\mu_2 h^2 (e_{xx}(t))^{-1}[e''_{xy}(t) - e''_{xz}(t)\gamma_0 - e''_{xx}(t)\beta_0(t) \\ & + 2e'_{xx}(t)\beta'_0(t) + e_{xx}(t)\beta''_0(t)], \end{aligned}$$

and $\Sigma(t)$ is a positive semidefinite matrix. $\square$

Based on the equation (A.14), the covariance matrix of $\widehat{\beta}(t)$ can be estimated by

$$n^{-2}(\tilde{E}_{xx}(t))^{-1}\left[\sum_{i=1}^{n}\left(\ll\int_0^\tau K_h(u-t)X_i(u)\widehat{\epsilon}_i(u)dN_i^c(u)\gg_R\right)^{\otimes 2}\right](\tilde{E}_{xx}(t))^{-1},$$

which is a consistent estimator base on the derivation in Appendix. And since

$$\begin{aligned}
& (nh)^{1/2}(\widehat{\beta}(t)-\beta_0(t)-\beta_{Bias}(t)) \\
= \; & (nh)^{1/2}(\tilde{\beta}(t;\gamma_0)-\beta_0(t)-\beta_{Bias}(t))+(nh)^{1/2}(\widehat{\gamma}-\gamma_0)\frac{\partial\tilde{\beta}(t;\gamma_0)}{\partial\gamma}+O_p(n^{-1/2}h^{1/2}) \\
= \; & n^{-1/2}\sum_{i=1}^{n}h^{1/2}\Bigg[(e_{xx}(t))^{-1}\Bigg(R_i\int_0^\tau K_h(u-t)X_i(u)\epsilon_i(u)dN_i^c(u) \\
& \qquad +(1-R_i)E_s\Big\{\int_0^\tau K_h(u-t)X_i(u)\epsilon_i(u)dN_i^c(u)\mid\mathcal{D}_i,R_i=0\Big\}\Bigg) \\
& \qquad -D^{-1}\Bigg(\int_{t_1}^{t_2}w(t)\{Z_i(t)-z_x(t)X_i(t)\}\epsilon_i(t)[R_idN_i^c(t) \\
& \qquad +E_s\{(1-R_i)dN_i^c(t)\mid\mathcal{D}_i,R_i=0\}]\Bigg)\tilde{z}_x(t)\Bigg] \\
& +O(h^{1/2})+o_p(h^{1/2})+O_p(n^{-1/2}h^{5/2})+O_p(n^{-1/2}h^{1/2}),
\end{aligned}$$

we can adjust the estimation of covariance matrix of $\widehat{\beta}(t)$ as follows

$$\begin{aligned}
& n^{-2}\sum_{i=1}^{n}\Bigg((\tilde{E}_{xx}(t))^{-1}\ll\int_0^\tau K_h(u-t)X_i(u)\widehat{\epsilon}_i(u)dN_i^c(u)\gg_R \\
& \qquad -\widehat{D}^{-1}\ll\int_{t_1}^{t_2}W(t)\{Z_i(t)-\tilde{Z}_x(t)X_i(t)\}\widehat{\epsilon}_i(t)dN_i^c(t)\gg_R\tilde{Z}_x(t)\Bigg)^{\otimes 2}.
\end{aligned}$$

CHAPTER 4: A SIMULATION STUDY

A numerical study is conducted to illustrate the feasibility and validity of the proposed methods. The performances of the estimator for γ are measured through the bias (Bias), the sample standard error of the estimates (SSE), the estimated standard error of $\widehat{\gamma}$(ESE) and the coverage probability of a 95% confidence interval for γ. The overall performance of the estimator for the jth component $\beta_j(\cdot)$ on the interval $[0, \tau]$ is evaluated through the square root of integrated average square error

$$RASE(\widehat{\beta}_j(\cdot)) = \left\{ \frac{1}{\tau} \int_0^\tau (\widehat{\beta}_j(t) - \beta_j(t))^2 \, dt) \right\}^{1/2},$$

where $\widehat{\beta}_j(t)$ is the estimate of $\beta_j(t)$. The simulation uses the unit weight function. The interval $[t_1, t_2] = [0.15, \pi]$ is taken to be $[0, \tau]$ in the estimating functions (2.7).

The performance of the proposed estimators are examined under the following selected setting of model (2.1). Let $Y_i(t)$ follow the semiparametric additive model:

$$Y_i(t) = \beta_0(t) + \beta_1(t)X_i + \gamma Z_i + \epsilon_i(t), \qquad i = 1, \ldots, n, \tag{4.1}$$

where $\beta_0(t) = 1 - t$, $\beta_1(t) = 5\sin(t)$, $\gamma = 8$, X_i is uniformly distributed on $[0, 1]$, and Z_i is a Bernoulli random variable with $P(Z_i = 1) = 0.5$. The error process $\epsilon_i(t)$ has a normal distribution with mean ϕ_i and variance 1 for subject i where ϕ_i follows a standard normal distribution.

For subject i, S_i is generated from the uniform distribution on $[0, 0.8]$. The first sampling point is set as $T_{i1} = 0$, and the rest T_{ij}'s are generated from a Poisson process $N_i(t)$ with the intensity rate of $\lambda_0 \exp(\eta_1 X_i + \eta_2 Z_i)$ where $\lambda_0 = 0.4$, $\eta_1 = 1$ and $\eta_2 = 0.3$. Let Y_{ij} be the responses $Y_i(t)$ at time points $T^0_{ij} = T_{ij} + S_i$ following model

(4.1). The censoring time C_i is exponentially distributed with the parameter adjusted to give an approximately 0% or 30% censoring in the time interval $[0, \tau] = [0, 4]$, which is the probability of $\max_{1 \leq j \leq n_i}\{T^0_{i,j} \wedge \tau\} > S_i + C_i$, denoted as c_R. The average number of observations in the interval $[0, \tau] = [0, 4]$ per subject is about 3.48.

The following four cases, including three different left censoring percentages for S_i,denoted as c_L, and the one that ignores S_i by mistreating T_{ij} as the measurement times since the actual time origin, are conducted to examine the behavior of both estimators: (1) $c_L = 0\%$ which means $\{S_i\}$ are observed for all the subjects; (2) $c_L = 20\%$; (3) $c_L = 50\%$; and (4) the last case treats T_{ij} as the time since the actual time origin and $Y_{ij} = Y_i(T^0_{ij})$ as the response at $t = T_{ij}$. The censoring time V_i is generated from an uniform distribution $[a, b]$ with the parameters a and b adjusted to yield desired percentages of left censoring for S_i.

The simulation presented in the following is carried out using local linear approach. As discussed in Section 2.3, to reduce the time consumption of simulations, the Epanechnikov kernel $K(u) = 0.75(1 - u^2)I(|u| \leq 1)$ is used for the inner points of time interval, i.e. $(3h, \tau - 3h)$ while the equivalent kernel in (2.8) is applied for the boundary points in $[0, 3h] \bigcup [\tau - 3h, \tau]$.

For sample sizes $n = 200$, 300 and 500, and bandwidths $h = 0.3$, 0.4 and 0.5, Table 4.1 shows the biases (Bias), the sample standard errors (SSE), the estimated standard errors (ESE) of $\widehat{\gamma}$, the coverage probabilities of a 95% confidence interval for γ and also the square root of integrated average square error (RASE) of both components of $\widehat{\beta}(t)$ for the first three cases based on 500 simulations when there is no right censoring. While Table 4.2 shows the same criterions for the first three cases based on 500 simulations when there is 30% of subjects right-censored during the time scale. The biases of $\widehat{\gamma}$ for the first three cases using the proposed method are small. The sample standard errors of $\widehat{\gamma}$ are close to its estimated standard errors. Both standard errors reduce as the sample size increases. When the left censoring percentage of S_i goes up, the standard errors rise a tiny bit since the increase of percentage means more unknown information of S_i. The coverage probabilities of $\widehat{\gamma}$ are slightly around 0.95 as expected. The square root of integrated average square error of $\widehat{\beta}_0(t)$ is smaller than that of $\widehat{\beta}_1(t)$ because $\beta_0(t)$ is a straight line while $\beta_1(t)$ is more curvy. Both RASE's increase together with the left censoring percentage of S_i.

Furthermore, as the bandwidth h changes from small values to big values, there are more data in the neighborhood. Then for the straight line $\beta_1(t)$, larger bandwidth makes the estimator fit better. As a result $RASE(\widehat{\beta}_0(\cdot))$ becomes smaller. However $\beta_2(t)$ is a curve. Larger bandwidth only leads to bigger value of $RASE(\widehat{\beta}_1(\cdot))$.

Table 4.3 present the biases, sample standard errors, estimated standard errors and the coverage probabilities related to γ in the case of mistreating T_{ij} as the measurement times since the actual time origin. Although both the standard errors of $\widehat{\gamma}$ increase compared to the third case with the same left censoring percentage, the biases are also small, the coverage probabilities are close to 0.95 and two types of standard errors are also close. This means even the time origin is mistreated, we can still get an unbiased estimator of γ since γ is time-independent.

Table 4.4 compare the RASE's in the two cases when the left censoring percentage of S_i is 50%. An obvious reduction of both RASE's is shown in the table.

Figure 4.1 shows the average estimates of $\beta(t) = (\beta_0(t), \beta_1(t))^T$ based on 500 simulations under four cases proposed above. Figure 4.1 (a), (b) and (c) are the plots of the average of the estimates based on the proposed method corresponding to 0%, 20% and 50% left censoring for S_i, and Figure 4.1 (d) corresponds to the fourth case. Figure 4.1 (a), (b) and (c) show that the estimated curves fit the true curve quite well. There is an obvious time shift for the covariate effect of X_i in Figure 4.1 (d).

Figure 4.2 shows both the standard errors of $\beta(t) = (\beta_0(t), \beta_1(t))^T$ based on 500 simulations under four cases proposed above. Figure 4.2 (a), (b) and (c) are the plots based on the proposed method corresponding to 0%, 20% and 50% left censoring for S_i, and Figure 4.2 (d) corresponds to the fourth case. In all four plots, the sample standard error curves are quite close to the estimated standard error curve. In the first three cases there are big variation at the beginning time while in the fourth case there are large variation at the end of the time scale. It is related to the amount of data. According to the generation of data, for each subject the first measure is taken at $T_ij = 0$. Then in fourth case there are most data at the beginning while least data at the end. On the other hand, in the first three cases the time point is $T_{ij}^o = S_i + T_{ij}$ which results in a time shift of length S_i. Then there are less data near the beginning and more data near $t = 4$ than in the fourth case.

Figure 4.3 shows the coverage probability of a pointwise 95% confidence interval for each component of $\beta(t) = (\beta_0(t), \beta_1(t))^T$ at each time point t based on 500 simulations under four cases proposed above. Figure 4.3 (a), (b) and (c) are the plots based on the proposed method corresponding to 0%, 20% and 50% left censoring for S_i, and Figure 4.3 (d) corresponds to the fourth case. The doted line in all four plots are the line when coverage probability is 95%. It is quite clear that all the coverage probabilities are close to 0.95.

Table 4.1: Summary statistics from the estimator $\widehat{\gamma}$ and $\widehat{\beta}(t)$ for no right censoring

c_L	n	h	Bias	SSE	ESE	CP	RASE($\widehat{\beta}_0(t)$)	RASE($\widehat{\beta}_1(t)$)
0%	200	0.3	−0.0090	0.1794	0.1780	0.958	0.0205	0.0479
		0.4	−0.0082	0.1794	0.1786	0.948	0.0172	0.0596
		0.5	−0.0078	0.1794	0.1790	0.954	0.0161	0.0858
	300	0.3	−0.0009	0.1386	0.1450	0.966	0.0182	0.0500
		0.4	0.0011	0.1385	0.1454	0.966	0.0163	0.0639
		0.5	0.0013	0.1384	0.1457	0.968	0.0160	0.0907
	500	0.3	−0.0083	0.1117	0.1134	0.950	0.0104	0.0323
		0.4	−0.0083	0.1116	0.1136	0.952	0.0064	0.0445
		0.5	−0.0081	0.1116	0.1137	0.950	0.0056	0.0724
20%	200	0.3	−0.0064	0.1809	0.1781	0.948	0.0279	0.0686
		0.4	−0.0064	0.1808	0.1788	0.946	0.0256	0.0758
		0.5	−0.0062	0.1810	0.1793	0.944	0.0241	0.0959
	300	0.3	0.0022	0.1426	0.1450	0.960	0.0314	0.0772
		0.4	0.0027	0.1427	0.1454	0.960	0.0310	0.0864
		0.5	0.0033	0.1426	0.1457	0.960	0.0289	0.1061
	500	0.3	−0.0059	0.1127	0.1135	0.942	0.0182	0.0704
		0.4	−0.0058	0.1127	0.1137	0.944	0.0154	0.0759
		0.5	−0.0057	0.1127	0.1139	0.944	0.0147	0.0914
50%	200	0.3	−0.0061	0.1821	0.1784	0.952	0.0905	0.2187
		0.4	−0.0055	0.1822	0.1795	0.952	0.0897	0.1960
		0.5	−0.0051	0.1822	0.1800	0.952	0.0547	0.1608
	300	0.3	0.0051	0.1418	0.1451	0.962	0.0725	0.1798
		0.4	0.0058	0.1417	0.1458	0.964	0.0672	0.1743
		0.5	0.0060	0.1417	0.1461	0.962	0.0585	0.1626
	500	0.3	−0.0050	0.1132	0.1138	0.942	0.0557	0.1824
		0.4	−0.0039	0.1147	0.1145	0.948	0.0544	0.1711
		0.5	−0.0041	0.1135	0.1143	0.942	0.0431	0.1615

Table 4.2: Summary statistics from the estimator $\widehat{\gamma}$ and $\widehat{\beta}(t)$ for 30% right censoring rate

c_L	n	h	Bias	SSE	ESE	CP	RASE($\widehat{\beta}_0(t)$)	RASE($\widehat{\beta}_1(t)$)
0%	200	0.3	−0.0131	0.1871	0.1836	0.946	0.0213	0.0479
		0.4	−0.0121	0.1873	0.1843	0.946	0.0179	0.0569
		0.5	−0.0113	0.1872	0.1848	0.950	0.0172	0.0823
	300	0.3	−0.0011	0.1436	0.1500	0.962	0.0243	0.0548
		0.4	−0.0009	0.1434	0.1504	0.968	0.0226	0.0667
		0.5	−0.0006	0.1432	0.1507	0.968	0.0223	0.0921
	500	0.3	−0.0092	0.1154	0.1173	0.948	0.0123	0.0334
		0.4	−0.0092	0.1152	0.1175	0.946	0.0076	0.0415
		0.5	−0.0089	0.1152	0.1177	0.944	0.0066	0.0677
20%	200	0.3	−0.0084	0.1874	0.1835	0.944	0.0330	0.0745
		0.4	−0.0085	0.1875	0.1844	0.950	0.0306	0.0784
		0.5	−0.0083	0.1879	0.1850	0.952	0.0290	0.0962
	300	0.3	0.0015	0.1468	0.1500	0.960	0.0376	0.0796
		0.4	0.0019	0.1469	0.1504	0.962	0.0381	0.0874
		0.5	0.0024	0.1470	0.1507	0.962	0.0362	0.1065
	500	0.3	−0.0066	0.1160	0.1174	0.942	0.0181	0.0773
		0.4	−0.0064	0.1158	0.1176	0.942	0.0148	0.0806
		0.5	−0.0063	0.1157	0.1178	0.942	0.0152	0.0924
50%	200	0.3	−0.0081	0.1897	0.1835	0.950	0.0921	0.2330
		0.4	−0.0077	0.1897	0.1847	0.952	0.0873	0.2065
		0.5	−0.0072	0.1898	0.1854	0.952	0.0565	0.1688
	300	0.3	0.0042	0.1467	0.1500	0.962	0.0773	0.1844
		0.4	0.0047	0.1468	0.1507	0.960	0.0736	0.1789
		0.5	0.0052	0.1467	0.1512	0.962	0.0653	0.1672
	500	0.3	−0.0057	0.1164	0.1175	0.942	0.0557	0.1935
		0.4	−0.0051	0.1163	0.1179	0.944	0.0491	0.1852
		0.5	−0.0049	0.1163	0.1182	0.942	0.0442	0.1683

Table 4.3: Summary statistics from the estimator $\widehat{\gamma}$ for misplaced time origin with $c_L = 50\%$

c_R	n	h	Bias	SSE	ESE	CP
0%	200	0.3	-0.0016	0.2126	0.2119	0.946
		0.4	-0.0005	0.2122	0.2127	0.950
		0.5	0.0006	0.2121	0.2134	0.946
	300	0.3	0.0019	0.1746	0.1733	0.944
		0.4	0.0026	0.1746	0.1738	0.944
		0.5	0.0033	0.1745	0.1742	0.948
	500	0.3	-0.0066	0.1410	0.1349	0.932
		0.4	-0.0058	0.1407	0.1352	0.932
		0.5	-0.0052	0.1404	0.1354	0.932
30%	200	0.3	0.0003	0.2251	0.2262	0.946
		0.4	0.0014	0.2251	0.2272	0.946
		0.5	0.0027	0.2250	0.2281	0.946
	300	0.3	0.0019	0.1853	0.1865	0.938
		0.4	0.0029	0.1848	0.1871	0.940
		0.5	0.0040	0.1845	0.1876	0.946
	500	0.3	-0.0106	0.1487	0.1449	0.936
		0.4	-0.0096	0.1486	0.1452	0.936
		0.5	-0.0086	0.1480	0.1455	0.942

Table 4.4: Summary statistics from the estimator $\widehat{\beta}(t)$ for misplaced time origin with $c_L = 50\%$

			RASE($\widehat{\beta}_0(t)$)		RASE($\widehat{\beta}_1(t)$)	
c_R	n	h	Our method	Misplaced time origin	Our method	Misplaced time origin
0%	200	0.3	0.0905	0.3959	0.2187	1.4308
		0.4	0.0897	0.3979	0.1960	1.4260
		0.5	0.0547	0.3991	0.1608	1.4229
	300	0.3	0.0725	0.3860	0.1798	1.4345
		0.4	0.0672	0.3870	0.1743	1.4309
		0.5	0.0585	0.3871	0.1626	1.4284
	500	0.3	0.0557	0.4050	0.1824	1.4057
		0.4	0.0544	0.4063	0.1711	1.4024
		0.5	0.0431	0.4070	0.1615	1.3995
30%	200	0.3	0.0921	0.4009	0.2330	1.4222
		0.4	0.0873	0.4007	0.2065	1.4099
		0.5	0.0565	0.4044	0.1688	1.4035
	300	0.3	0.0773	0.3799	0.1844	1.4373
		0.4	0.0736	0.3805	0.1789	1.4329
		0.5	0.0653	0.3812	0.1672	1.4305
	500	0.3	0.0557	0.4061	0.1935	1.3985
		0.4	0.0491	0.4071	0.1852	1.3940
		0.5	0.0442	0.4077	0.1683	1.3917

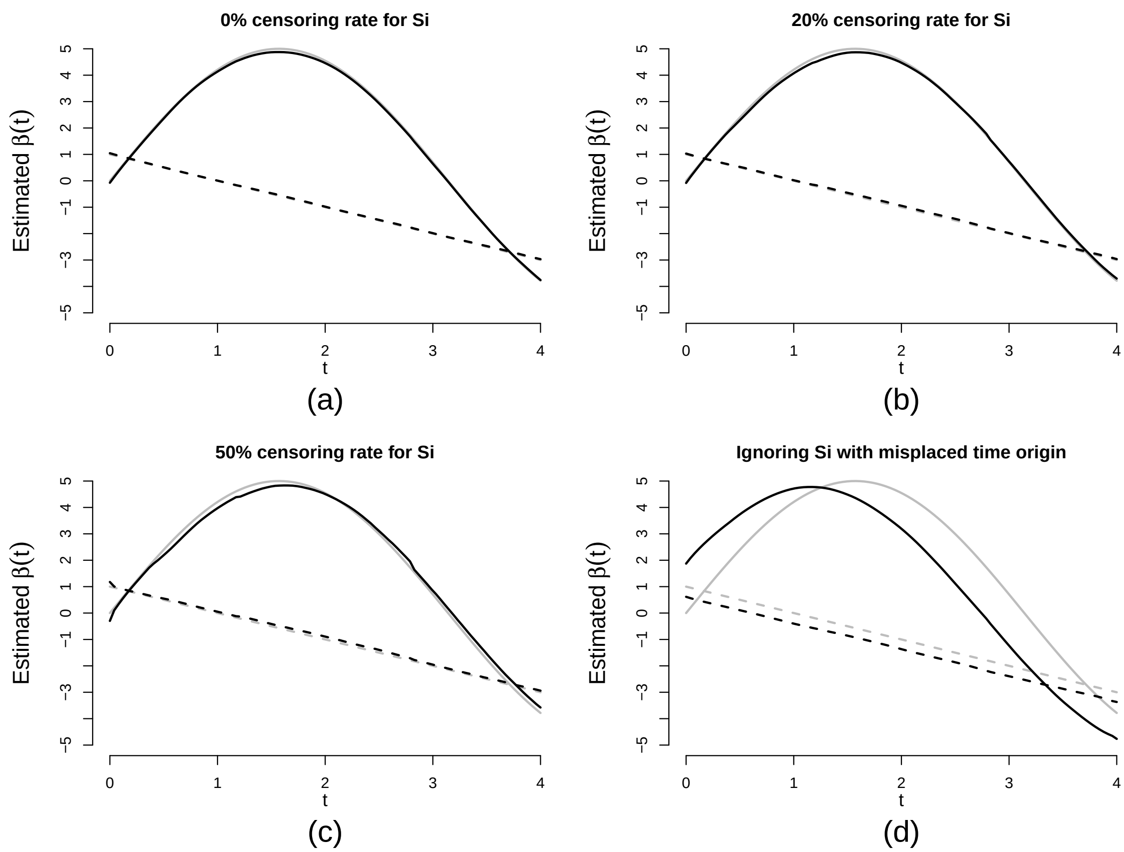


Figure 4.1: Averages in estimating $\beta(t)$ for $n = 300$ and $h = 0.4$. The solid lines are for $\beta_1(t)$ and the dashed lines are for $\beta_0(t)$. The grey lines are the true cures.

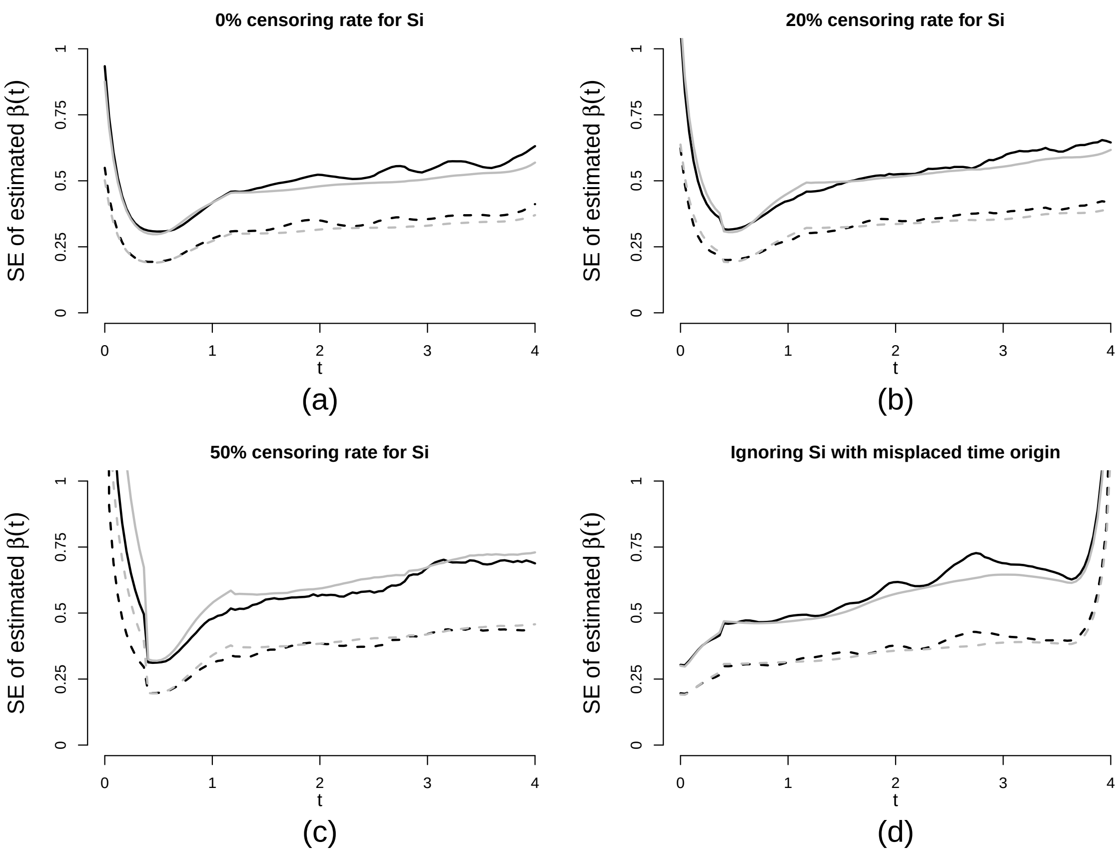


Figure 4.2: Sample and estimated standard errors in estimating $\beta(t)$ for $n = 300$ and $h = 0.4$. The solid lines are for $\beta_1(t)$ and the dashed lines are for $\beta_0(t)$. The grey lines are the estimated standard error and the black ones are the sample standard error.

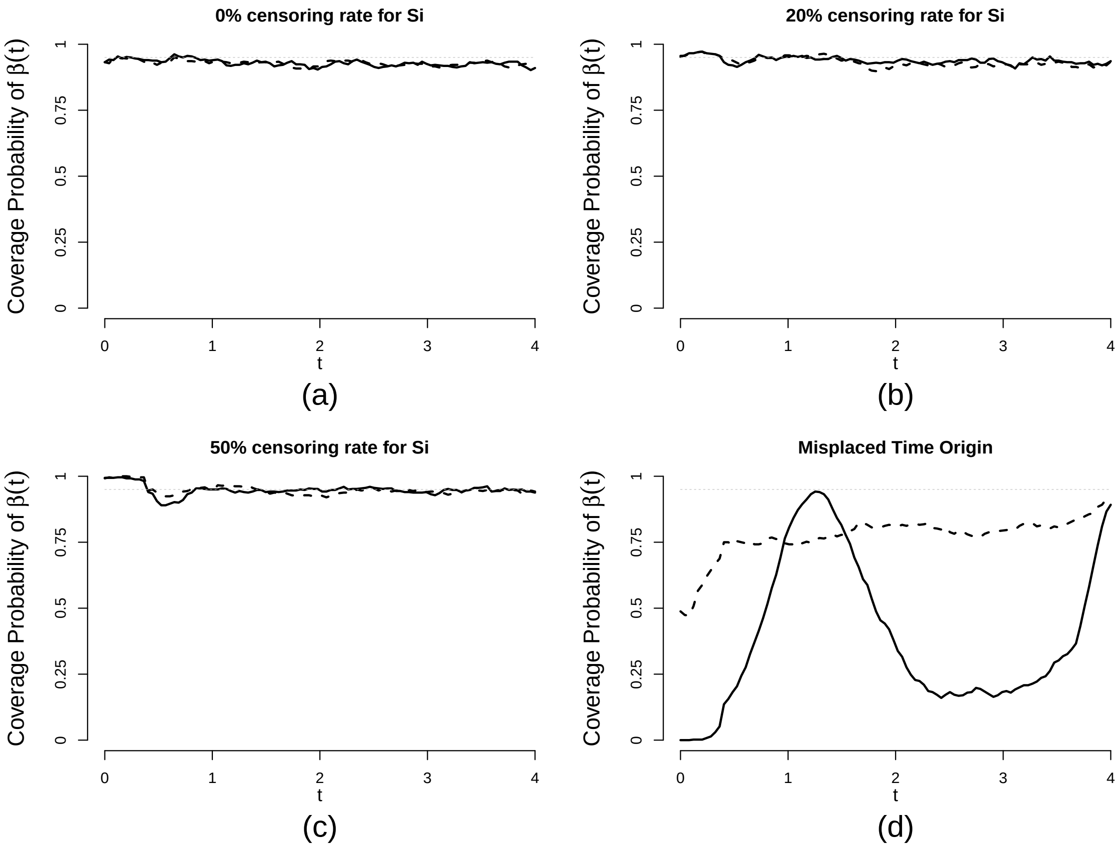


Figure 4.3: Coverage probability of a 95% confidence interval of $\beta(t)$ for $n = 300$ and $h = 0.4$. The solid lines are for $\beta_1(t)$ and the dashed lines are for $\beta_0(t)$.

CHAPTER 5: REAL DATA APPLICATION

In this chapter a real data from the step study (cf., Buchbinder et al., 2008; Fitzgerald et al., 2011) is analyzed by applying the methods discussed in previous chapters. The step study was a multicenter, double-blind, randomized, placebo-controlled, phase II test-of-concept study to determine whether the MRKAd5 HIV-1 gag/pol/nef vaccine, which elicits T cell immunity, is capable to result in controlling the replication of the Human immunodeficiency virus among the participants who got HIV-infected after vaccination. This study opened in December 2004 and was conducted at 34 sites in North America, the Caribbean, South America, and Australia. Three thousand HIV-1 negative participants aged from 18 to 45 who were at high risk of HIV-infection were enrolled and randomly assigned to receive vaccine ($X_2 = 1$) or placebo ($X_2 = 0$) in ratio 1:1, stratified by sex, study site ($Z_3 = 1$ if North America or Australia and 0 otherwise) and adenovirus type 5 (Ad5) antibody titer at baseline ($Z_1 = \ln Ad5$). Some of the participant were fully adherent to vaccinations ($Z_2 = 1$) while others not ($Z_2 = 0$).

The analysis in this chapter includes a subset of the 3000 participants which involves all 174 MITT cases as of September 22, 2009. It is recommended to study males only, for the entire analysis to avoid the effect of sex since there are only 15 females that are $< 10\%$ of the sample. All 159 males got HIV-infected at time 0 which may be not observed. However, each participants had the records of the dates of their first positive Elisa confirmed by Western Blot or RNA (D_i's in the above chapters), their first evidence of infection, and the estimated dates of infection which is considered as the midpoint between last RNA negative visit date (L_i's in the above chapters) which is not given in the data, and the date of first evidence of infection. Using the above dates, we can calculate our V_i in the above chapters by $V_i = D_i - L_i = D_i - 2\times$ estimated infection dates + the date of first evidence of infection. And R_i, the indicator of whether the actual acquisition of ith subject is observed or not, is 1 if the date of first evidence of infection is before the date of first positive Elisa. Otherwise $R_i = 0$. When $R_i = 1$, $V_i = S_i$. Otherwise S_i is left censored by V_i.

After the participant was infected, there were 18 scheduled post-infection visit per subject at weeks 0, 1, 2, 8, 12, 26, and every 26 weeks thereafter through week 338. However, the actual times and dates of visits may vary due to each individual. During jth visit, the ith subject received tests to have the measurements of HIV virus load ($Y_{ij} = \log_{10}$(virus load)) and CD4 cell counts (X_{1ij} = square root of CD4 counts) before the subject started the antiretroviral therapy (ART) or was censored. And the time from the first positive Elisa to the jth visit for ith subject is T_{ij} in the above chapters. The time between the first positive Elisa and ART initiation or censoring is the right censoring time. All the time in this chapter is in year. Our main interest is to see the effect of vaccine on the HIV virus load response.

In the data 159 males made a total of 791 pre-ART visits. Among them there are 156 missing in CD4 cell counts and 5 missing in HIV virus load. Since there are no missing in CD4 and virus load at the same time, we could use a simple imputation model to create a complete data set. At each time point separately, we use a linear regression model linking $\log_{10}$(viral load) to square root of CD4 count (for those with data on both), and use the viral load value for those with missing data to fill in the missing CD4 cell count or predict missing virus load data by CD4 values. However, at three time points there are no complete data for conducting the linear regression model fitting; at two other points there are only one complete data which is unable to complete the linear model fitting; at another time point one predicted value of virus load is relatively far beyond the range of other values of virus load and may affect the analysis results. Therefore, we delete these six visits to get the complete data for the entire analysis.

Now in this complete data set there are 159 subjects with 785 visits. 97 Of all the participants were in the vaccine group while 62 received the placebo. 122 subjects participate in the study in North America or Australia and the rest are residents in the other sites mentioned at the beginning of this chapter. The left censoring rate of S_i is 70.44% and the right censoring rate of T_{ij} is 69.81%. Figure 5.1 to Figure 5.3 are further exploration of the data. It is easy to figure out that there are few data after time point 2.5. Therefore, we will choose $t_1 = 0$ and $t_2 = 2.5$ to estimate γ, and also plot the estimators of $\beta(t)$'s for the time points in the interval [0,2.5]. Finally, Figure 5.4 shows the Kaplan Meier estimator of the distribution of S_i. Note that the smallest observed S_i is 0.14. Before that time we do not have enough information to get the estimator of the distribution. However, since time is always nonnegative, the probability of S_i reduce to 0 at $S_i = 0$.

After preliminary exploration of the data, we propose the following model for virus load response of the ith subject in this study:

$$Y_i(t) = \beta_0(t) + \beta_1(t)X_{1i}(t) + \beta_2(t)X_{2i} + \gamma_1 Z_{1i} + \gamma_2 Z_{2i} + \gamma_3 Z_{3i} + \epsilon_i(t). \qquad (5.1)$$

By the study of simulation and several tries of different bandwidths, a possible reasonable choice of the bandwidth for this data set is 0.5. And we still consider the unit weight for the analysis. The estimates of γ_1, γ_2 and γ_3 are 0.0302, -0.1467 and 0.1956, with the standard deviations 0.0389, 0.1492 and 0.1540, respectively. The p-values for testing $H_0 : \gamma_1 = 0$, $H_0 : \gamma_2 = 0$ and $H_0 : \gamma_3 = 0$ are equal to 0.4375, 0.3255 and 0.2042, respectively, which indicates that there are no significant effects of baseline Ad5 titer, study sites or the pre-protocol on the HIV viral load level. The estimates of time-dependent effects and their 95% pointwise confidence interval are shown in Figure 5.5. From the graph the effects of vaccine or CD4 cell count on the HIV viral load level are not significant, either. Further hypothesis test study will be done in the future. Finally Figure 5.6 shows the scatter plot of the residuals from fitting the model (5.1).

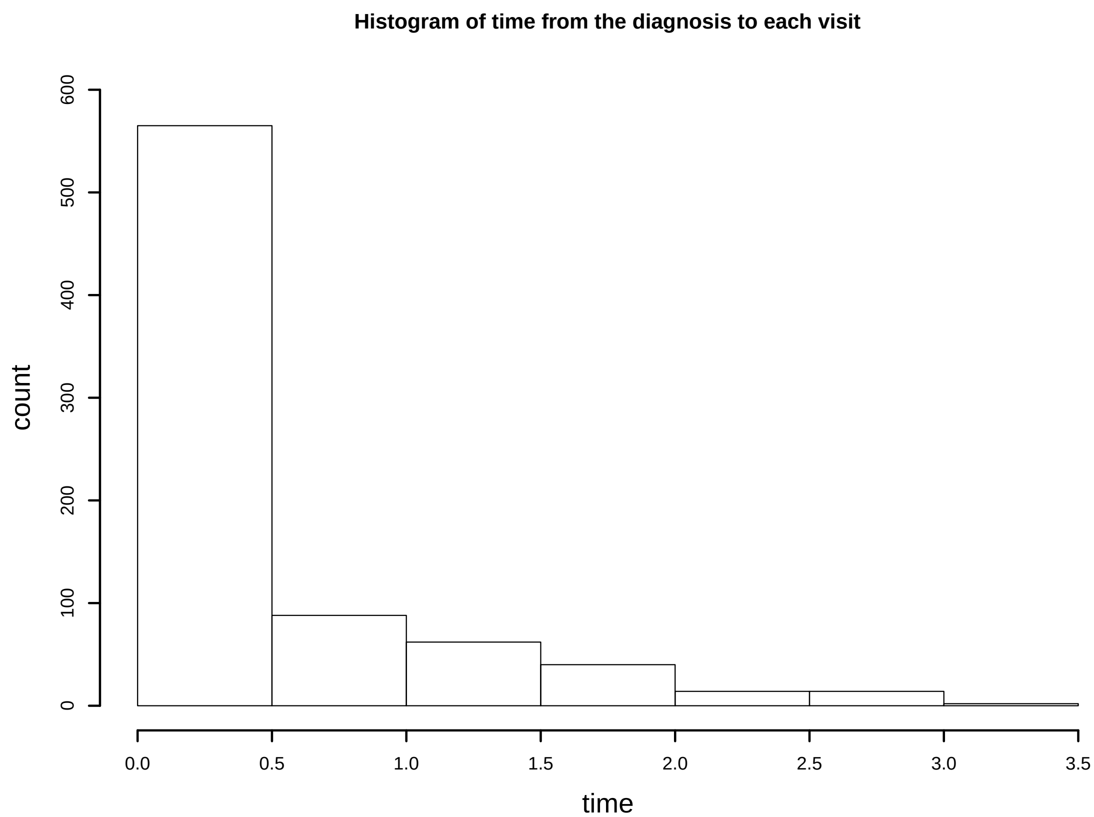


Figure 5.1: Histogram of the time from the first positive Elisa confirmed by Western Blot or RNA to each visit, denoted as T_{ij} in the paper.

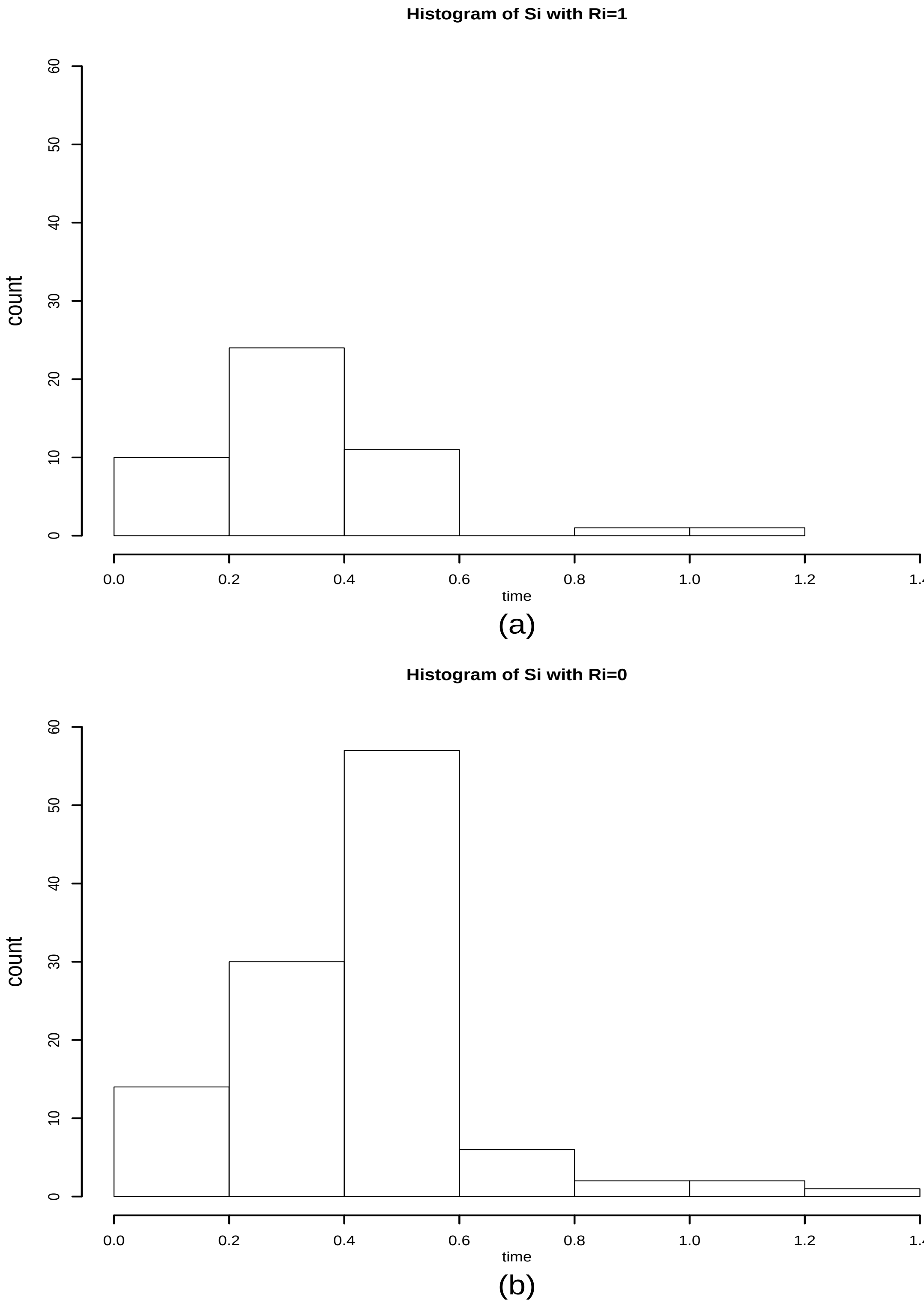


Figure 5.2: Histograms of the time from actual HIV acquisition to the first positive Elisa confirmed by Western Blot or RNA, denoted as S_i in the paper. Figure (a) shows the observed ones ($R_i = 1$) while figure (b) shows the counts of censored ones ($R_i = 0$).

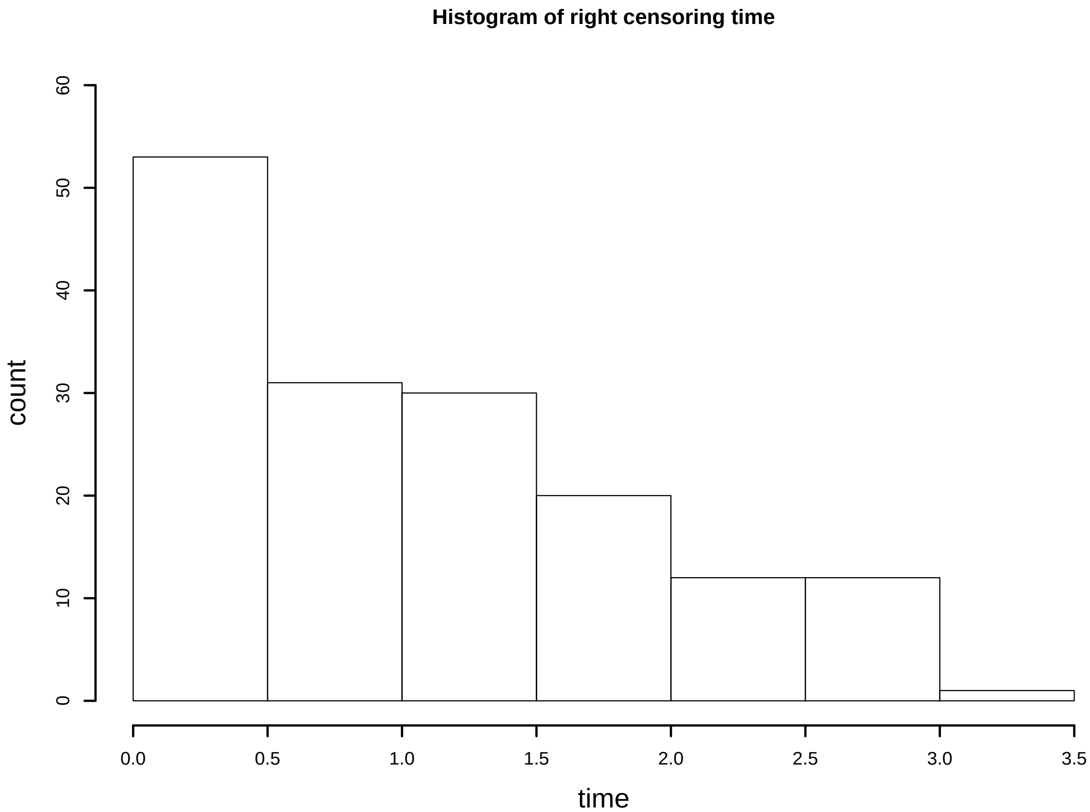


Figure 5.3: Histograms of the time from the first positive Elisa confirmed by Western Blot or RNA to ART initiation or censoring, denoted as C_i in the paper.

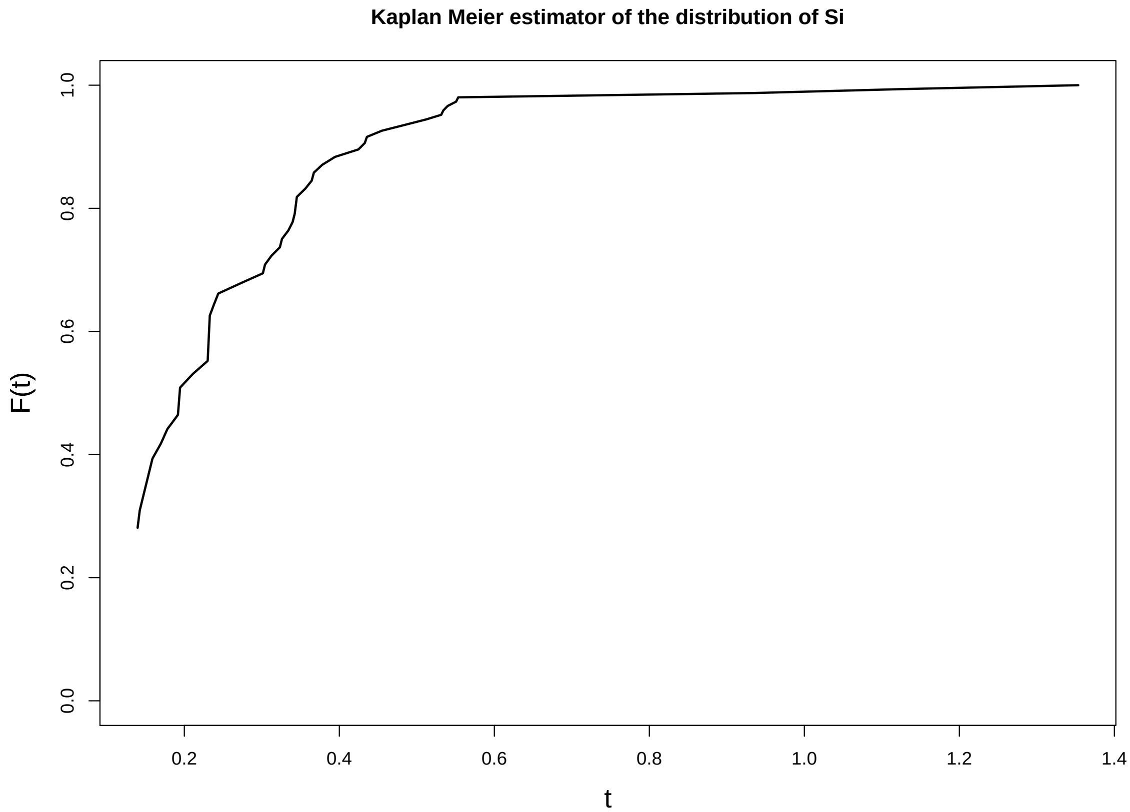
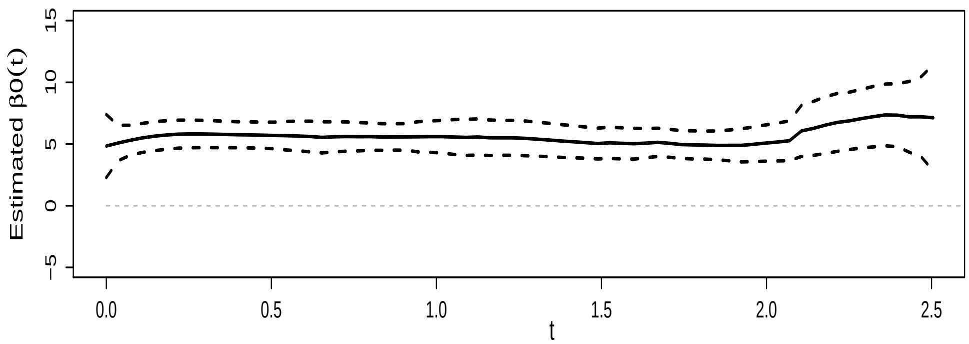
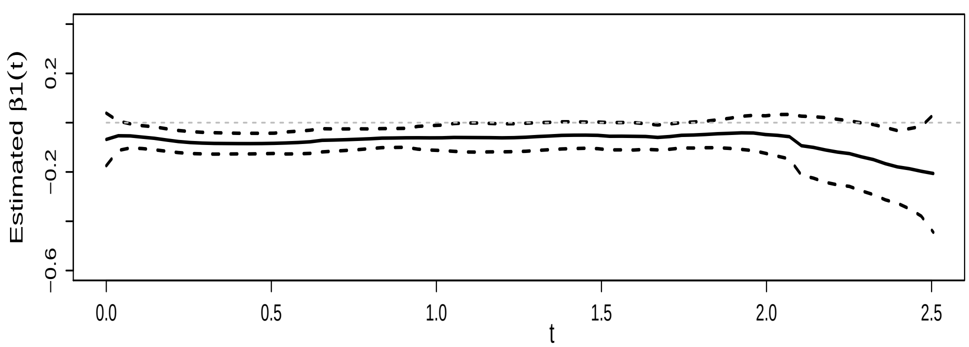


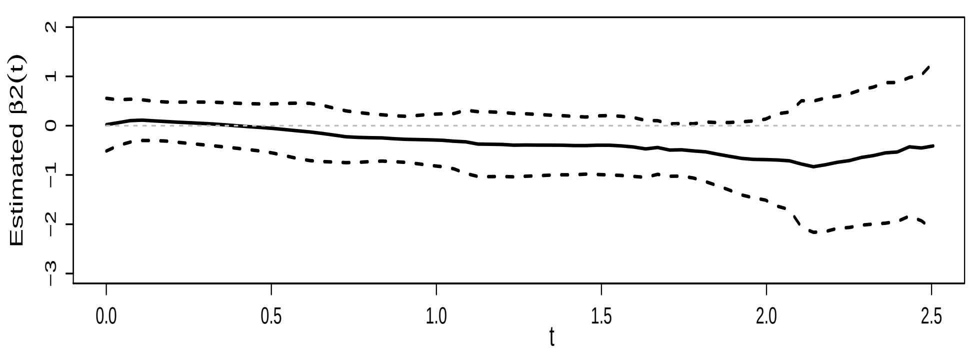
Figure 5.4: The Kaplan Meier estimator of the distribution function of the time from actual HIV acquisition to the first positive Elisa confirmed by Western Blot or RNA.



(a)



(b)



(c)

Figure 5.5: Figure (a) shows the estimated intercept effect, $\beta_0(t)$ curve and its 95% pointwise confidence intervals. Figure (b) shows the estimated squared CD4 effect, $\beta_1(t)$ curve and its 95% pointwise confidence intervals. Figure (c) shows the estimated treatment effect, $\beta_2(t)$ curve and its 95% pointwise confidence intervals. The solid curves are the estimated curves and the dashed curves are the confidence intervals.

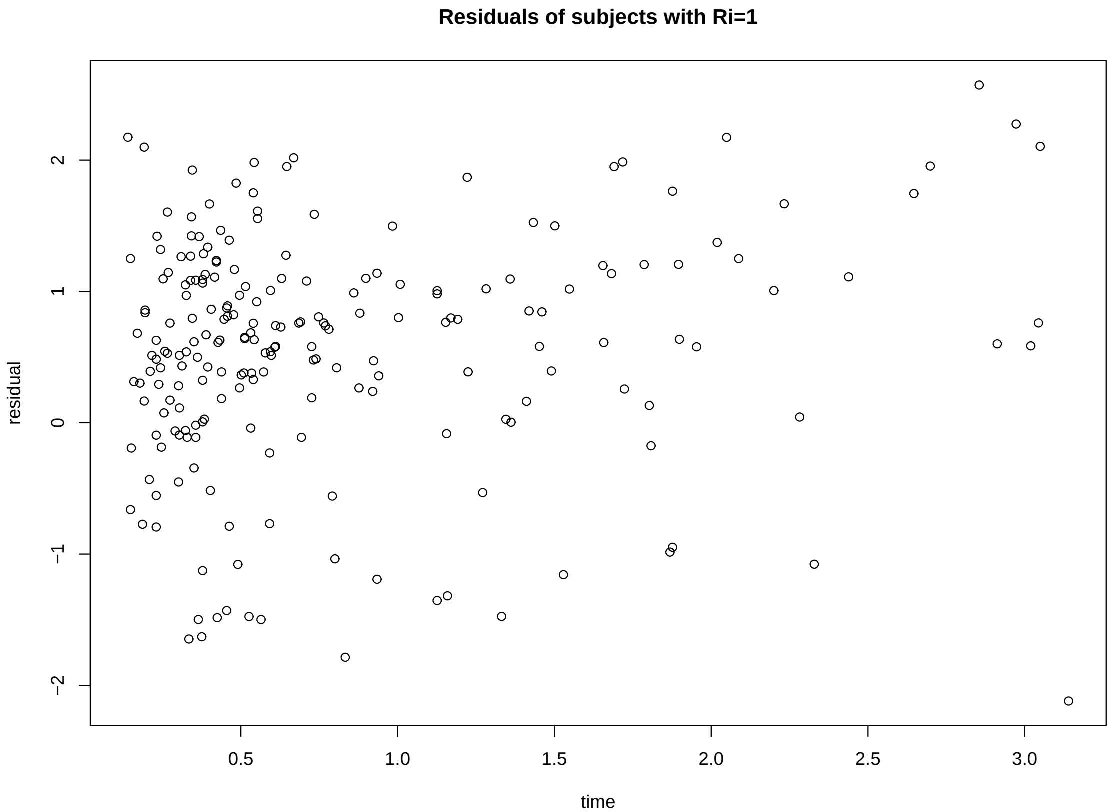


Figure 5.6: Scatter plot of residuals of the subjects with $R_i = 1$.

REFERENCES

ANDERSEN, P. K. AND GILL, R. D. 1982. Cox's regression model for counting processes: a large sample study. *The Annals of Statistics 10*, 1100–1120.

BUCHBINDER, S. P., MEHROTRA, D. V., DUERR, A., FITZGERALD, D. W., MOGG, R., LI, D., GILBERT, P. B., LAMA, J. R., MARMOR, M., DEL RIO, C., MCELRATH, M. J., CASIMIRO, D. R., GOTTESDIENER, K. M., CHODAKEWITZ, J. A., COREY, L., AND ROBERTSON, M. N. 2008. Efficacy assessment of a cell-mediated immunity hiv-1 vaccine (the step study): adouble-blind, randomised, placebo-controlled, test-of-concept trial. *Lancet 372 (9653)*, 1881–1893. PMCID: 2721012.

CAI, Z. AND SUN, Y. 2002. Local linear estimation for time-dependent coefficients in cox's regression models. *Scandinavian Journal of Statistics 30*, 93–111.

CLEMENS, J. D., NAFICY, A., AND RAO, M. R. 1997. Long-term evaluation of vaccine protection: Methodological issues for phase 3 trials and phase 4 studies. in: Levine mm, woodrow gc, kaper jb, cobon gs, eds. *New Generation Vaccines New York: Marcel Dekker, Inc.*, 47–67.

CLEMENTS-MANN, M. L. 1998. Lessons for aids vaccine development from non-aids vaccines. *AIDS Research and Human Retroviruses 14 Suppl 3*, S197–S203.

COX, D. R. 1972. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society B 34*, 187–220.

Fan, J. and Gijbels, I. 1996. *Local polynomial modeling and its applications.* Chapman & Hall/CRC Press LLC, Boca Raton, Florida.

Fitzgerald, D. W., Janes, H., Robertson, M., Coombs, R., Frank, I., Gilbert, P. B., Loufty, M., Mehrotra, D., and Duerr, A. 2011. An ad5-vectored hiv-1 vaccine elicitscell-mediated immunity but does not affect disease progression in hiv-1-infected male subjects: results from a randomized placebo-controlled trial (the step study). *The Journal of Infectious Diseases 203(6)*, 765–772.

Fleming, T. R. and Harrington, D. P. 1991. *Counting processes and survival analysis.* John Wiley & Sons, Inc., Hoboken, New Jersey.

Gray, R. H., Wawer, M. J., Brookmeyer, R., Sewankambo, N. K., Serwadda, D., Wabwire-Mangen, F., Lutalo, T., Li, X., vanCott, T., and Quinn, T. C. 2001. Probability of hiv-1 transmission per coital act in monogamous, heterosexual, hiv-1 discordant couples in rakai, uganda. *Lancet 357*, 1149–1153.

Halloran, M. E., Struchiner, C. J., and Longini, I. M. 1997. Study designs for evaluating different efficacy and effectiveness aspects of vaccines. *American Journal of Epidemiology 146*, 789–803.

HIVSMCGroup. 2000. Human immunodeficiency virus type 1 rna level and cd4 count as prognostic markers and surrogate endpoints: A meta-analysis. *AIDS Research and Human Retroviruses 16*, 1123–1133. HIV Surrogate Marker Collaborative Group.

Hoover, D., Rice, J. A., Wu, C. O., and Yang, L. P. 1998. Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika 85*, 809–822.

Horvitz, D. G. and Thompson, D. J. 1952. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association 47*, 663–685.

HVTN. 2004. The pipeline project. (HVTN stands for HIV Vaccine Trials Network.) Available at: http://www.hvtn.org/.

IAVI. 2004. State of current aids vaccine research. (IAVA stands for International AIDS Vaccine Initiative.) Available at: http://www.iavi.org/.

Liang, H.and Wu, H. and Carroll, R. J. 2003. The relationship between virologic and immunologic responses in aids clinical research using mixed-effects varying-coefficient semiparametric models with measurement error. *Biostatistics 4*, 297–312.

Lin, D. Y. and Ying, Z. 2001. Semiparametric and nonparametric regression analysis of longitudinal data (with discussion). *Journal of the American Statistical Association 96*, 103–113.

Martinussen, T. and Scheike, T. H. 1999. A semiparametric additive regression model for longitudinal data. *Biometrika 86*, 691–702.

Martinussen, T. and Scheike, T. H. 2000. A nonparametric dynamic additive regression model for longitudinal data. *The Annals of Statistics 28*, 1000–1025.

Martinussen, T. and Scheike, T. H. 2001. Sampling adjusted analysis of dynamic additive regression models for longitudinal data. *Scandinavian Journal of Statistics 28*, 303–323.

Mellors, J. W., Munoz, A., Giorgi, J. V., Margolick, J. B., Tassoni, C. J., Gupta, P., Kingsley, L. A., Todd, J. A., Saah, A. J., Detels, R., Phair,

J. P., and Rinaldo, C. R. J. 1997. Plasma viral load and cd4+ lymphocytes as prognostic markers of hiv-1 infection. *Annals of Internal Medicine 126*, 946–954.

Miloslavsky, M., Keles, S., van der Laan, M. J., and Butler, 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.

Moyeed, R. A. and Diggle, P. J. 1994. Rates of convergence in semiparametric modelling of longitudinal data. *Australian Journal of Statistics 36*, 75–93.

Nabel, G. J. 2001. Challenges and opportunities for development of an aids vaccine. *Nature 410*, 1002–1007.

Quinn, T. C., Wawer, M. J., Sewankambo, N., Serwadda, D., Li, C., Wabwire-Mangen, F., Meehan, M. O., and Lutalo, T. amd Gray, R. H. 2000. Viral load and heterosexual transmission of human immunodeficiency virus type 1. *New England Journal of Medicine 342*, 921–929.

Rice, J. A. and Silverman, B. W. 1991. Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society B 53*, 233–243.

Robin, D. B. 1976. Inference and missing data. *Biometrika 63*, 581–592.

Robins, J. M., Rotnitzky, A., and Zhao, L. P. 1994. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association 89*, 846–866.

Scheike, T. and Zhang, M. 1998. Cumulative regression function tests for longitudinal data. *The Annals of Statistics 26*, The Annals of Statistics.

SCHEIKE, T. H. AND SUN, Y. 2007. Maximum likelihood estimation for tied survival data under cox regression model via em-algorithm. *Lifetime Data Analysis 13*, 399–420.

SHIVER, J. W., FU, T. M., CHEN, L., CASIMIRO, D. R., DAVIES, M. E., EVANS, R. K., ZHANG, Z. Q., SIMON, A. J., TRIGONA, W. L., DUBEY, S. A., HUANG, L., HARRIS, V. A., LONG, R. S., LIANG, X., HANDT, L., SCHLEIF, W. A., ZHU, L., FREED, D. C., PERSAUD, N. V., GUAN, L., PUNT, K. S., TANG, A., CHEN, M., WILSON, K. A., COLLINS, K. B., HEIDECKER, G. J., FERNANDEZ, V. R., PERRY, H. C., JOYCE, J. G., GRIMM, K. M., COOK, J. C., KELLER, P. M., KRESOCK, D. S., MACH, H., TROUTMAN, R. D., ISOPI, L. A., WILLIAMS, D. M., XU, Z., BOHANNON, K. E., VOLKIN, D. B., MONTEFIORI, D. C., MIURA, A., KRIVULKA, G. R., LIFTON, M. A., KURODA, M. J., SCHMITZ, J. E., LETVIN, N. L., CAULFIELD, M. J., BETT, A. J., YOUIL, R., KASLOW, D. C., AND EMINI, E. A. 2002. Replication-incompetent adenoviral vaccine vector elicits effective anti-immunodeficiency virus immunity. *Nature 415*, 331–335.

SUN, Y. AND GILBERT, P. B. 2012. Estimation of stratified mark-specific proportional hazards models with missing marks. *Scandinavian Journal of Statistics 39*, 34–52.

SUN, Y. AND LEE, J. 2011. Testing independent censoring for longitudinal data. *Statistica Sinica 21*, 1315–1339.

SUN, Y., WANG, J. H., AND GILBERT, P. B. Quantile regression for competing risks data with missing cause of failure. To appear in Statistica Sinica.

SUN, Y. AND WU, H. 2003. Auc-based tests for nonparametric functions with longitudinal data. *Statistica Sinica 13*, 593–612.

Sun, Y. and Wu, H. 2005. Semiparametric time-varying coefficients regression model for longitudinal data. *Scandinavian Journal of Statistics 32*, 21–47.

Tian, L., Zucker, D., and Wei, L. J. 2005. On the cox model with time-varying regression coefficients. *Journal of the American Statistical Association 100*, 172–183.

van der Vaart, A. W. 1998. *Asymptotic statistics*. Cambridge University Press, Cambridge.

Wu, C. O., Chiang, C. T., and Hoover, D. R. 1998. Asymptotic confidence regions for kernel smoothing of a time-varying coefficient model with longitudinal data. *Journal of the American Statistical Association 88*, 1388–1402.

Wu, H. and Liang, H. 2004. Backfitting random varying-coefficient models with time-dependent smoothing covariates. *Scandinavian Journal of Statistics 31*, 3–19.

Wu, H. and Zhang, J. T. 2002. Local polynomial mixed-effects models for longitudinal data. *Journal of the American Statistical Association 97*, 883–897.

Zeger, S. L. and Diggle, P. J. 1994. Semiparametric models for longitudinal data with application to cd4 cell numbers in hiv seroconverters. *Biometrics 50*, 689–699.

Zeng, D. 2005. Likelihood approach for martingale proportional hazards regression in the presence of dependent censoring. *The Annals of Statistics 33*, 501–521.

APPENDIX A: PROOFS OF LEMMA AND THEOREM

Now we will show the detailed proofs of five lemmas and four theorems we present in Chapter 3. In Section A.2, Lemma A.2.1 is used to prove Lemma A.2.2. The results of Lemma A.2.2 and Lemma A.2.3 states the consistent properties of our proposed notation $\ll\gg_R$. Lemma A.2.4 is the basis of getting Lemma A.2.5. We will repeatedly apply Lemmas A.2.2, A.2.3 and A.2.5 in proofs of theorems in Section A.3.

A.1 Preliminaries

Preparing for future application in this section, we first derive the martingale decomposition of the Kaplan-Meier estimator of the survival function for the left censored data.

In general, we have the i.i.d. data structure of the left censored data as follows,

$$\{T_i = max(S_i, C_i), \delta_i = I(S_i \geq C_i)\},$$

where S_i is the failure time censored by C_i, T_i is observed time and δ_i is the indicator of non-censorship for ith subject. Suppose L be a large enough number so that all $S_i < L$. Then

$$\{L - T_i = min(L - S_i, L - C_i), \delta_i = I(L - S_i \leq L - C_i)\}$$

is the corresponding right censored data structure. Let $S(t) = P(S_i > t)$ and $S^R(t) = P(L - S_i > t)$ be the survival functions of the failure time for the left and right censored data respectively. And $\widehat{S}(t)$, $\widehat{S}^R(t)$ are the Kaplan-Meyer estimators of the survival functions respectively. Now define the counting process $N_i^R(t) = I(L - T_i \leq t, \delta_i = 1)$.

By the Doob-Meyer decomposition, there is a compensator $\int_0^t Y_i^R(s)d\Lambda^R(s)$ and a martingale $M_i^R(t)$ so that $N_i^R(t) = \int_0^t Y_i^R(s)d\Lambda^R(s) + M_i^R(t)$. Here $Y_i^R(t) = I(L - T_i \geq t)$ is the at risk indicator and $\Lambda^R(t)$ is the cumulative hazard function. Let $N^R(t) = \sum_{i=1}^n N_i^R(t)$, $M^R(t) = \sum_{i=1}^n M_i^R(t)$ and $Y^R(t) = \sum_{i=1}^n Y_i^R(t) = \sum_{i=1}^n I(T_i \leq L - t)$. Assume that $Y^R(t)/n \xrightarrow{P} y^R(t)$. Hence according to Equation (2.11) in Chapter 3 on Page 98 of Fleming & Harrington (1991), we have the decomposition

$$n^{1/2}(\widehat{S}^R(t) - S^R(t)) = -n^{1/2}S^R(t)\int_0^t \frac{\widehat{S}^R(s-)}{S^R(s)}\frac{I(Y^R(s) > 0)}{Y^R(s)}dM^R(s) + o_p(1).$$

Since

$$S(t) = P(S_i > t) = P(L - S_i < L - t) = 1 - P(L - S_i \geq L - t) = 1 - S^R((L-t)-),$$

then for the left censored data

$$\begin{aligned}
& n^{1/2}(\widehat{S}(t) - S(t)) \\
= \; & -n^{1/2}[\widehat{S}^R((L-t)-) - S^R((L-t)-)] \\
= \; & n^{1/2}S^R((L-t)-)\int_0^{(L-t)-} \frac{\widehat{S}^R(s-)}{S^R(s)}\frac{I(Y^R(s) > 0)}{Y^R(s)}dM^R(s) + o_p(1) \\
= \; & n^{-1/2}(1 - S(t))\int_0^{(L-t)-} \frac{1 - \widehat{S}(L-s)}{1 - S((L-s)-)}\frac{I(Y^R(s) > 0)}{Y^R(s)/n}dM^R(s) + o_p(1) \\
= \; & n^{-1/2}(1 - S(t))\int_0^{(L-t)-} \frac{1}{y^R(s)}dM^R(s) + o_p(1). \qquad \text{(A.1)}
\end{aligned}$$

Now let us define the following notations for the future use.

$$\begin{aligned}
X_{zi}^I(u) &= \int_0^u [R_iZ_i(w)X_i^T(w)dN_i^c(w) - E(R_i\xi_i(w)\alpha_i(w)Z_i(w)X_i^T(w))dw], \\
X_{zi}^{II}(t) &= \int_0^\infty \int_0^L \int_{t_1}^t E\Big\{(1 - R_i)Z_i(u)X_i^T(u)\alpha_i^*(u-s)\frac{I(x \leq (L - V_i)-)}{F_s(V_i)}\Big\} \\
&\quad \cdot(e_{xx}(u))^{-1}dudF_s(s)\frac{dM_i^R(x)}{y^R(x)} \\
&\quad - \int_0^{L-}\int_0^{(L-x)-}\int_{t_1}^t E\Big\{(1 - R_i)\frac{Z_i(u)X_i^T(u)\alpha_i^*(u-s)}{F_s(V_i)}\Big\} \\
&\quad \cdot(e_{xx}(u))^{-1}du\, dF_s(s)\frac{dM_i^R(x)}{y^R(x)}
\end{aligned}$$

$$
\begin{aligned}
& + \int_0^L \int_{t_1}^t E\Big\{(1-R_i)\frac{Z_i(u)X_i^T(u)\alpha_i^*(u-s)}{F_s(V_i)}\Big\}F_s(s) \\
& \quad \cdot (e_{xx}(u))^{-1}du\frac{dM_i^R((L-s)-)}{y^R((L-s)-)}, \\
X_{zi}^{III}(u) \quad = \quad & \int_0^u (E_s\{(1-R_i)Z_i(w)X_i^T(w)dN_i^c(w) \mid \mathcal{D}_i, R_i = 0\} \\
& -E\{(1-R_i)\xi_i(w)\alpha_i(w)Z_i(w)X_i^T(w)\}dw),
\end{aligned}
$$

and

$$
X_{zn}^I(u) = n^{-1/2}\sum_{i=1}^n X_{zi}^I(u), X_{zn}^{II}(t) = n^{-1/2}\sum_{i=1}^n X_{zi}^{II}(t), X_{zn}^{III}(u) = n^{-1/2}\sum_{i=1}^n X_{zi}^{III}(u).
$$

Similarly, we can define $X_{yi}^I(u)$, $X_{yi}^{II}(t)$, $X_{yi}^{III}(u)$, $X_{yn}^I(u)$, $X_{yn}^{II}(t)$, $X_{yn}^{III}(u)$, $X_{xi}^I(u)$, $X_{xi}^{III}(u)$, $X_{xn}^I(u)$, $X_{xn}^{III}(u)$ by replacing $Z_i(\cdot)$ above with $Y_i(\cdot)$ and $X_i(\cdot)$ respectively. However

$$
\begin{aligned}
X_{xi}^{II}(t) \quad = \quad & \int_0^\infty \int_0^L \int_{t_1}^t \beta^T(u)E\Big\{(1-R_i)X_i(u)X_i^T(u)\alpha_i^*(u-s)\frac{I(x \le (L-V_i)-)}{F_s(V_i)}\Big\} \\
& \quad \cdot (e_{xx}(u))^{-1}dudF_s(s)\frac{dM_i^R(x)}{y^R(x)} \\
& - \int_0^{L-} \int_0^{(L-x)-} \int_{t_1}^t \beta^T(u)E\Big\{(1-R_i)\frac{X_i(u)X_i^T(u)\alpha_i^*(u-s)}{F_s(V_i)}\Big\} \\
& \quad \cdot (e_{xx}(u))^{-1}du\, dF_s(s)\frac{dM_i^R(x)}{y^R(x)} \\
& + \int_0^L \int_{t_1}^t \beta^T(u)E\Big\{(1-R_i)\frac{X_i(u)X_i^T(u)\alpha_i^*(u-s)}{F_s(V_i)}\Big\}F_s(s) \\
& \quad \cdot (e_{xx}(u))^{-1}du\frac{dM_i^R((L-s)-)}{y^R((L-s)-)}.
\end{aligned}
$$

Then $X_{xn}^{II}(t) = n^{-1/2}\sum_{i=1}^n X_{zi}^{II}(t)$.

A.2 Some Lemmas

Lemma A.2.1: Let a random function $g_i(t) = g(t, X_i(t), Z_i(t), Y_i(t))$. Then under Conditions (I), for $t \in [t_1, t_2] \subset [0, \tau]$,

$$
n^{-1}\sum_{i=1}^n (1-R_i)\widehat{E}_s\Big\{\int_{t_1}^{t_2} g_i(u)dN_i^c(u) \mid \mathcal{D}_i, R_i = 0\Big\} \xrightarrow{P} E\Big\{(1-R_i)\int_{t_1}^{t_2} g_i(u)dN_i^c(u)\Big\}
$$

as $n \to \infty$.

Proof. As mentioned in Section 2.2,

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^{n}(1-R_i)\widehat{E}_s\Big\{\int_{t_1}^{t_2} g_i(u)dN_i^c(u) \mid \mathcal{D}_i, R_i=0\Big\} \\
= \;& n^{-1}\sum_{i=1}^{n}(1-R_i)\int_0^L \sum_{j=1}^{n_i} g_i(s+T_{ij})I(C_i \geq T_{ij})\frac{d\widehat{F}_s(s \mid \mathcal{D}_i)}{\widehat{F}_s(V_i \mid \mathcal{D}_i)} \\
= \;& n^{-1}\sum_{i=1}^{n}(1-R_i)\int_0^L \sum_{j=1}^{n_i} g_i(s+T_{ij})I(C_i \geq T_{ij})\frac{dF_s(s \mid \mathcal{D}_i)}{F_s(V_i \mid \mathcal{D}_i)} \qquad \text{(A.2)}\\
& +n^{-1}\sum_{i=1}^{n}(1-R_i)\int_0^L \sum_{j=1}^{n_i} g_i(s+T_{ij})I(C_i \geq T_{ij})\Big(\frac{dF_s(s \mid \mathcal{D}_i)}{\widehat{F}_s(V_i \mid \mathcal{D}_i)} - \frac{dF_s(s \mid \mathcal{D}_i)}{F_s(V_i \mid \mathcal{D}_i)}\Big) \\
& +n^{-1}\sum_{i=1}^{n}(1-R_i)\int_0^L \sum_{j=1}^{n_i} g_i(s+T_{ij})I(C_i \geq T_{ij})\frac{d\widehat{F}_s(s \mid \mathcal{D}_i) - dF_s(s \mid \mathcal{D}_i)}{\widehat{F}_s(V_i \mid \mathcal{D}_i)}
\end{aligned}
$$

If $\widehat{F}_s(s \mid \mathcal{D}_i)$ is the Kaplan-Meier estimator of conditional survival function, we still have $\widehat{F}_s(s \mid \mathcal{D}_i) \xrightarrow{P} F_s(s \mid \mathcal{D}_i)$, $\widehat{F}_s(V_i \mid \mathcal{D}_i) \xrightarrow{P} F_s(V_i \mid \mathcal{D}_i)$. Then by continuous theorem, $1/\widehat{F}_s(V_i \mid \mathcal{D}_i) \xrightarrow{P} 1/F_s(V_i \mid \mathcal{D}_i)$. So under the Conditions (I) the second term in (A.2) which is equal to

$$
n^{-1}\sum_{i=1}^{n}(1-R_i)\int_0^L \sum_{j=1}^{n_i} g_i(s+T_{ij})I(C_i \geq T_{ij})\Big(\frac{1}{\widehat{F}_s(V_i \mid \mathcal{D}_i)} - \frac{1}{F_s(V_i \mid \mathcal{D}_i)}\Big)dF_s(s \mid \mathcal{D}_i)
$$

converges to zero in probability. Since S_i is independent of $\mathcal{D}_i$ and remind that $N_i(t) = \sum_{j=1}^{n_i} I(T_{ij} \leq t)$, the third term in (A.2) is equal to

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^{n}(1-R_i)\int_0^L \sum_{j=1}^{n_i} g_i(s+T_{ij})I(C_i \geq T_{ij})\frac{d\widehat{F}_s(s) - dF_s(s)}{F_s(V_i \mid \mathcal{D}_i)} + o_p(1) \\
= \;& n^{-1}\sum_{i=1}^{n}(1-R_i)\int_0^L \Big(\int_{t_1-s}^{t_2-s} g_i(s+v)I(C_i \geq v)dN_i(v)\Big)\frac{d(\widehat{F}_s(s) - F_s(s))}{F_s(V_i \mid \mathcal{D}_i)} \\
& +o_p(1) \\
= \;& \int_0^L \Big[n^{-1}\sum_{i=1}^{n}(1-R_i)\int_{t_1-s}^{t_2-s} g_i(s+v)I(C_i \geq v)dN_i(v)\frac{1}{F_s(V_i \mid \mathcal{D}_i)}\Big] \\
& \quad d(\widehat{F}_s(s) - F_s(s)) + o_p(1)
\end{aligned}
$$

Let
$$H_n(s) = n^{-1}\sum_{i=1}^{n}(1-R_i)\int_{t_1-s}^{t_2-s} g_i(s+v)I(C_i \geq v)dN_i(v)\frac{1}{F_s(V_i \mid \mathcal{D}_i)}.$$
So the absolute value of the third term in (A.2) equals
$$\begin{aligned}
&\left|\int_0^L H_n(s)d(\widehat{F}_s(s)-F_s(s))\right| \\
= &\left|H_n(L)(\widehat{F}_s(L)-F_s(L)) - H_n(0)(\widehat{F}_s(0)-F_s(0)) - \int_0^L(\widehat{F}_s(s)-F_s(s))dH_n(s)\right| \\
\leq &\ |H_n(L)(\widehat{F}_s(L)-F_s(L))| + |H_n(0)(\widehat{F}_s(0)-F_s(0))| + \left|\int_0^L(\widehat{F}_s(s)-F_s(s))dH_n(s)\right| \\
\leq &\ |H_n(L)(\widehat{F}_s(L)-F_s(L))| + |H_n(0)(\widehat{F}_s(0)-F_s(0))| \\
&+ \sup_{s\in[0,L]}|\widehat{F}_s(s)-F_s(s)|\int_0^L |dH_n(s)|
\end{aligned}$$

Under Conditions (I), by the uniform consistency of $\widehat{F}_s(s)$ and the convergence of $\widehat{F}_s(s)$ at point $s = 0$, or $s = L$, the third term converges to zero in probability uniformly in s as $n \to \infty$. Therefore,
$$\begin{aligned}
(A.2) \quad &\xrightarrow{P} \quad E\left\{(1-R_i)\int_0^L \sum_{j=1}^{n_i} g_i(s+T_{ij})I(C_i \geq T_{ij})\frac{dF_s(s \mid \mathcal{D}_i)}{F_s(V_i \mid \mathcal{D}_i)}\right\} \\
&= \quad E\left\{(1-R_i)E_s\left(\int_{t_1}^{t_2} g_i(u)dN_i^c(u) \mid \mathcal{D}_i, R_i = 0\right)\right\} \\
&= \quad E\left\{I(R_i = 0)E_s\left(\int_{t_1}^{t_2} g_i(u)dN_i^c(u) \mid \mathcal{D}_i, R_i = 0\right)\right\} \\
&= \quad E\left\{E_s\left(I(R_i = 0)\int_{t_1}^{t_2} g_i(u)dN_i^c(u) \mid \mathcal{D}_i, R_i = 0\right)\right\} \\
&= \quad E\left\{(1-R_i)\int_{t_1}^{t_2} g_i(u)dN_i^c(u)\right\}
\end{aligned}$$

The proof of Lemma A.2.1 is completed. □

Based on the above lemma, we can easily prove the following lemma.

Lemma A.2.2: Let a random function $g_i(t) = g(t, X_i(t), Z_i(t), Y_i(t))$. Then under

Conditions (I), for $t \in [t_1, t_2] \subset [0, \tau]$,

$$n^{-1}\sum_{i=1}^{n} \ll \int_{t_1}^{t_2} g_i(u)dN_i^c(u) \gg_R \xrightarrow{P} E\Big\{\int_{t_1}^{t_2} g_i(u)dN_i^c(u)\Big\}$$

as $n \to \infty$.

Proof. Applying Lemma A.2.1,

$$\begin{aligned}
& n^{-1}\sum_{i=1}^{n} \ll \int_{t_1}^{t_2} g_i(u)dN_i^c(u) \gg_R \\
= \quad & n^{-1}\sum_{i=1}^{n} R_i \int_{t_1}^{t_2} g_i(u)dN_i^c(u) + n^{-1}\sum_{i=1}^{n}(1-R_i)\widehat{E}_s\Big\{\int_{t_1}^{t_2} g_i(u)dN_i^c(u) \mid \mathcal{X}\Big\} \\
= \quad & n^{-1}\sum_{i=1}^{n} R_i \int_{t_1}^{t_2} g_i(u)dN_i^c(u) \\
& + n^{-1}\sum_{i=1}^{n}(1-R_i)\widehat{E}_s\Big\{\int_{t_1}^{t_2} g_i(u)dN_i^c(u) \mid \mathcal{D}_i, R_i = 0\Big\} \\
\xrightarrow{P} \quad & E\Big\{R_i \int_{t_1}^{t_2} g_i(u)dN_i^c(u)\Big\} + E\Big\{(1-R_i)\int_{t_1}^{t_2} g_i(u)dN_i^c(u)\Big\} \\
= \quad & E\Big\{R_i \int_{t_1}^{t_2} g_i(u)dN_i^c(u) + (1-R_i)\int_{t_1}^{t_2} g_i(u)dN_i^c(u)\Big\} \\
= \quad & E\Big\{\int_{t_1}^{t_2} g_i(u)dN_i^c(u)\Big\}
\end{aligned}$$

Lemma A.2.2 is proved. □

Lemma A.2.3: Let a random function $g_i(t) = g(t, X_i(t), Z_i(t), Y_i(t))$. Then under Conditions (I), for $t \in [t_1, t_2] \subset [0, \tau]$, $\xi_i(t) = I(S_i^* + C_i \geq t)$,

$$n^{-1}\sum_{i=1}^{n} \ll \int_{t_1}^{t_2} K_h(u-t)g_i(u)dN_i^c(u) \gg_R \xrightarrow{P} E(\xi_i(t)\alpha_i(t)g_i(t))$$

as $n \to \infty$, $h \to 0$ and $nh^2 \to \infty$.

Proof. By the definition,

$$\begin{aligned}
& n^{-1}\sum_{i=1}^{n} \ll \int_{t_1}^{t_2} K_h(u-t)g_i(u)dN_i^c(u) \gg_R \\
= \quad & n^{-1}\sum_{i=1}^{n} R_i \int_{t_1}^{t_2} K_h(u-t)g_i(u)dN_i^c(u)
\end{aligned}$$

$$+n^{-1}\sum_{i=1}^{n}(1-R_i)\widehat{E}_s\Big\{\int_{t_1}^{t_2}K_h(u-t)g_i(u)dN_i^c(u)\mid\mathcal{X}\Big\}$$

By the independence of subjects, the second term can be written as

$$\begin{aligned}&n^{-1}\sum_{i=1}^{n}(1-R_i)\widehat{E}_s\Big\{\int_{t_1}^{t_2}K_h(u-t)g_i(u)dN_i^c(u)\mid\mathcal{D}_i,R_i=0\Big\}\\ =\;&n^{-1}\sum_{i=1}^{n}(1-R_i)\int_{t_1}^{t_2}K_h(u-t)d\Big(\int_0^u\widehat{E}_s\{g_i(v)dN_i^c(v)\mid\mathcal{D}_i,R_i=0\}\Big).\text{(A.3)}\end{aligned}$$

Note that the limits of integration in Lemma A.2.1 can be replaced by 0 and u, and the convergence is uniform in u. We have

$$n^{-1}\sum_{i=1}^{n}(1-R_i)\Big[\int_0^u\widehat{E}_s(g_i(v)dN_i^c(v)\mid\mathcal{D}_i,R_i=0)-\int_0^u E_s(g_i(v)dN_i^c(v)\mid\mathcal{D}_i,R_i=0)\Big]$$

converges to zero in probability uniformly in $u\in[t_1,t_2]$. So

$$\begin{aligned}(A.3)\;=\;&\int_{t_1}^{t_2}K_h(u-t)d\Big(n^{-1}\sum_{i=1}^{n}(1-R_i)\int_0^u E_s\{g_i(v)dN_i^c(v)\mid\mathcal{D}_i,R_i=0\}\Big)\\ &+o_p(1)\\ =\;&\int_{t_1}^{t_2}K_h(u-t)d\Big(E\Big[(1-R_i)\int_0^u E_s\{g_i(v)dN_i^c(v)\mid\mathcal{D}_i,R_i=0\}\Big]\Big)+o_p(1)\\ =\;&\int_{t_1}^{t_2}K_h(u-t)d\Big(\int_0^u E[E_s\{(1-R_i)g_i(v)dN_i^c(v)\mid\mathcal{D}_i,R_i=0\}]\Big)+o_p(1)\\ =\;&\int_{t_1}^{t_2}K_h(u-t)d\Big(\int_0^u E\{(1-R_i)g_i(v)dN_i^c(v)\}\Big)+o_p(1)\\ =\;&\int_{t_1}^{t_2}K_h(u-t)E\{(1-R_i)g_i(u)dN_i^c(u)\}+o_p(1).\end{aligned}$$

According to the argument on Page 37 of Sun & Wu (2005), the first term at the beginning of this proof is equal to

$$\int_{t_1}^{t_2}K_h(u-t)E\{R_ig_i(u)dN_i^c(u)\}+O_p(n^{-1/2}h^{-1}).$$

So the whole expression equals

$$\int_{t_1}^{t_2}K_h(u-t)E\{R_ig_i(u)dN_i^c(u)\}+\int_{t_1}^{t_2}K_h(u-t)E\{(1-R_i)g_i(u)dN_i^c(u)\}$$

$$
\begin{aligned}
& +O_p(n^{-1/2}h^{-1}) + o_p(1) \\
= \;& \int_{t_1}^{t_2} K_h(u-t)E\{g_i(u)dN_i^c(u)\} + O_p(n^{-1/2}h^{-1}) + o_p(1) \\
= \;& \int_{t_1}^{t_2} K_h(u-t)E\{g_i(u)\xi_i(u)dN_i^0(u)\} + O_p(n^{-1/2}h^{-1}) + o_p(1) \\
= \;& \int_{t_1}^{t_2} K_h(u-t)E\{E[\xi_i(u)g_i(u)dN_i^0(u) \mid X_i(u), Z_i(u), S_i^* + C_i \geq t]\} \\
& +O_p(n^{-1/2}h^{-1}) + o_p(1) \\
= \;& \int_{t_1}^{t_2} K_h(u-t)E\{\xi_i(u)E[g_i(u) \mid X_i(u), Z_i(u), S_i^* + C_i \geq t] \\
& \quad E[dN_i^0(u) \mid X_i(u), Z_i(u), S_i^* + C_i \geq t]\} + O_p(n^{-1/2}h^{-1}) + o_p(1) \\
= \;& \int_{t_1}^{t_2} K_h(u-t)E\{\xi_i(u)E[g_i(u) \mid X_i(u), Z_i(u)]E[dN_i^0(u) \mid X_i(u), Z_i(u)]\} \\
& +O_p(n^{-1/2}h^{-1}) + o_p(1) \\
= \;& \int_{t_1}^{t_2} K_h(u-t)E[\xi_i(u)E[g_i(u) \mid X_i(u), Z_i(u)]\alpha_i(u)du] + O_p(n^{-1/2}h^{-1}) + o_p(1) \\
= \;& \int_{t_1}^{t_2} K_h(u-t)E(E[\xi_i(u)g_i(u)\alpha_i(u)du \mid X_i(u), Z_i(u)]) + O_p(n^{-1/2}h^{-1}) + o_p(1) \\
= \;& \int_{t_1}^{t_2} K_h(u-t)E(\xi_i(u)g_i(u)\alpha_i(u)du) + O_p(n^{-1/2}h^{-1}) + o_p(1) \\
= \;& E(\xi_i(t)\alpha_i(t)g_i(t)) + O(h^2) + O_p(n^{-1/2}h^{-1}) + o_p(1) \xrightarrow{P} E(\xi_i(t)\alpha_i(t)g_i(t))
\end{aligned}
$$

as $n \to \infty$, $h \to 0$ and $nh^2 \to \infty$. Lemma A.2.3 is proved. □

Lemma A.2.4:

$$
\begin{aligned}
& n^{1/2}\int_{t_1}^{t}(\tilde{E}_{zx}(u) - e_{zx}(u))(e_{xx}(u))^{-1}du \\
= \;& n^{-1/2}\sum_{i=1}^{n}\left\{\int_{t_1-h}^{t+h}\left[d(X_{zi}^I(v) + X_{zi}^{III}(v))((e_{xx}(v))^{-1} + O(h^2))\right] + X_{zi}^{II}(t)\right\} \\
& +O_p(n^{-1/2}h^2 + n^{1/2}h^2) + o_p(1)
\end{aligned}
$$

converges weakly to a vector of mean-zero Gaussian processes with continuous paths as $n \to \infty$, $h \to 0$ and $nh^4 \to 0$. Similar results hold for

$$
n^{1/2}\int_{t_1}^{t}(\tilde{E}_{yx}(u) - e_{yx}(u))(e_{xx}(u))^{-1}du
$$

$$
\begin{aligned}
&= n^{-1/2}\sum_{i=1}^{n}\left\{\int_{t_1-h}^{t+h}\left[d(X_{yi}^{I}(v)+X_{yi}^{III}(v))((e_{xx}(v))^{-1}+O(h^2))\right]+X_{yi}^{II}(t)\right\}\\
&\quad +O_p(n^{-1/2}h^2+n^{1/2}h^2)+o_p(1),
\end{aligned}
$$

$$
\begin{aligned}
&n^{1/2}\int_{t_1}^{t}\beta^T(u)(\tilde{E}_{xx}(u)-e_{xx}(u))(e_{xx}(u))^{-1}du\\
= \; & n^{-1/2}\sum_{i=1}^{n}\left\{\int_{t_1-h}^{t+h}\left[(\beta^T(u)+O(h^2))d(X_{xi}^{I}(v)+X_{xi}^{III}(v))((e_{xx}(v))^{-1}+O(h^2))\right]\right.\\
&\left.+X_{xi}^{II}(t)\right\}+O_p(n^{-1/2}h^2+n^{1/2}h^2)+o_p(1).
\end{aligned}
$$

Proof. By the definitions,

$$
\begin{aligned}
&n^{1/2}\int_{t_1}^{t}(\tilde{E}_{zx}(u)-e_{zx}(u))(e_{xx}(u))^{-1}du\\
= \; & n^{1/2}\int_{t_1}^{t}\Bigg(n^{-1}\sum_{i=1}^{n}\ll\int_0^{\tau}K_h(v-u)Z_i(v)X_i^T(v)dN_i^c(v)\gg_R\\
&\quad -E\{\xi_i(u)\alpha_i(u)Z_i(u)X_i^T(u)\}\Bigg)(e_{xx}(u))^{-1}du\\
= \; & n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t}\Bigg[R_i\int_0^{\tau}K_h(v-u)Z_i(v)X_i^T(v)dN_i^c(v)\\
&\quad -E\{R_i\xi_i(u)\alpha_i(u)Z_i(u)X_i^T(u)\}\\
&+(1-R_i)\widehat{E}_s\left\{\int_0^{\tau}K_h(v-u)Z_i(v)X_i^T(v)dN_i^c(v)\mid\mathcal{X}\right\}\\
&\quad -E\{(1-R_i)\xi_i(u)\alpha_i(u)Z_i(u)X_i^T(u)\}\Bigg](e_{xx}(u))^{-1}du\\
= \; & n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t}\Bigg[R_i\int_0^{\tau}K_h(v-u)Z_i(v)X_i^T(v)dN_i^c(v)\\
&\quad -E\{R_i\xi_i(u)\alpha_i(u)Z_i(u)X_i^T(u)\}\\
&+(1-R_i)\widehat{E}_s\left\{\int_0^{\tau}K_h(v-u)Z_i(v)X_i^T(v)dN_i^c(v)\mid\mathcal{D}_i,R_i=0\right\}\\
&\quad -E\{(1-R_i)\xi_i(u)\alpha_i(u)Z_i(u)X_i^T(u)\}\Bigg](e_{xx}(u))^{-1}du\\
= \; & n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t}\Bigg[R_i\int_0^{\tau}K_h(v-u)Z_i(v)X_i^T(v)dN_i^c(v)\\
&\quad -E\{R_i\xi_i(u)\alpha_i(u)Z_i(u)X_i^T(u)\}\Bigg](e_{xx}(u))^{-1}du
\end{aligned}
$$

$$
\begin{aligned}
&+n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t}(1-R_i)\Big[\widehat{E}_s\Big\{\int_0^{\tau}K_h(v-u)Z_i(v)X_i^T(v)dN_i^c(v)\mid\mathcal{D}_i,R_i=0\Big\}\\
&\quad -E_s\Big\{\int_0^{\tau}K_h(v-u)Z_i(v)X_i^T(v)dN_i^c(v)\mid\mathcal{D}_i,R_i=0\Big\}\Big](e_{xx}(u))^{-1}du\\
&+n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t}\Big[(1-R_i)E_s\Big\{\int_0^{\tau}K_h(v-u)Z_i(v)X_i^T(v)dN_i^c(v)\mid\mathcal{D}_i,R_i=0\Big\}\\
&\quad -E\{(1-R_i)\xi_i(u)\alpha_i(u)Z_i(u)X_i^T(u)\}\Big](e_{xx}(u))^{-1}du\\
= \;& n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t}\Big[\int_0^{\tau}R_iK_h(v-u)Z_i(v)X_i^T(v)dN_i^c(v)\\
&\quad -E\{R_i\xi_i(u)\alpha_i(u)Z_i(u)X_i^T(u)\}\Big](e_{xx}(u))^{-1}du\\
&+n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t}(1-R_i)\int_0^{L}\sum_{j=1}^{n_i}K_h(s+T_{ij}-u)Z_{ij}X_{ij}^TI(C_i\ge T_{ij})\Big[\frac{d\widehat{F}_s(s\mid\mathcal{D}_i)}{\widehat{F}_s(V_i\mid\mathcal{D}_i)}\\
&\quad -\frac{dF_s(s\mid\mathcal{D}_i)}{F_s(V_i\mid\mathcal{D}_i)}\Big](e_{xx}(u))^{-1}du\\
&+n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t}\Big[(1-R_i)E_s\Big\{\int_0^{\tau}K_h(v-u)Z_i(v)X_i^T(v)dN_i^c(v)\mid\mathcal{D}_i,R_i=0\Big\}\\
&\quad -E\{(1-R_i)\xi_i(u)\alpha_i(u)Z_i(u)X_i^T(u)\}\Big](e_{xx}(u))^{-1}du. \qquad (A.4)
\end{aligned}
$$

Now let us look at them summation by summation. The first summation of (A.4) equals

$$
\begin{aligned}
&n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t}\Big[\int_0^{\tau}K_h(v-u)R_iZ_i(v)X_i^T(v)dN_i^c(v)\\
&\quad -\int_0^{\tau}K_h(v-u)E\{R_i\xi(v)\alpha_i(v)Z_i(v)X_i^T(v)\}dv+O(h^2)\Big](e_{xx}(u))^{-1}du\\
= \;& n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t}\int_0^{\tau}K_h(v-u)[R_iZ_i(v)X_i^T(v)dN_i^c(v)\\
&\quad -E\{R_i\xi(v)\alpha_i(v)Z_i(v)X_i^T(v)\}dv](e_{xx}(u))^{-1}du+O_p(n^{1/2}h^2)\\
= \;& n^{1/2}\int_{t_1}^{t}\int_0^{\tau}K_h(v-u)n^{-1}\sum_{i=1}^{n}[R_iZ_i(v)X_i^T(v)dN_i^c(v)\\
&\quad -E\{R_i\xi(v)\alpha_i(v)Z_i(v)X_i^T(v)\}dv](e_{xx}(u))^{-1}du+O_p(n^{1/2}h^2)\\
= \;& \int_{t_1}^{t}\int_0^{\tau}K_h(v-u)d\Big(n^{-1/2}\sum_{i=1}^{n}\int_0^{v}[R_iZ_i(w)X_i^T(w)dN_i^c(w)
\end{aligned}
$$

$$-E\{R_i\xi_i(w)\alpha_i(w)Z_i(w)X_i^T(w)\}dw]\Big)(e_{xx}(u))^{-1}du + O_p(n^{1/2}h^2).$$

Let

$$X_{zn}^I(v) = n^{-1/2}\sum_{i=1}^n \int_0^v [R_iZ_i(w)X_i^T(w)dN_i^c(w) - E\{R_i\xi_i(w)\alpha_i(w)Z_i(w)X_i^T(w)\}dw].$$

Under Condition (I) $X_{zn}^I(v)$ converges to a vector of mean zero Gaussian processes, saying $X_z^I(v)$ uniformly in v. Then also by the compactness of $K(\cdot)$ and the application of the continuous mapping theorem the first summation above equals

$$\begin{aligned}
& \int_{t_1}^t \int_0^\tau K_h(v-u)dX_{zn}^I(v)(e_{xx}(u))^{-1}du + O_p(n^{1/2}h^2) \\
= & \int_{t_1-h}^{t+h} \left[dX_{zn}^I(v)\int_{t_1}^t h^{-1}K(\frac{v-u}{h})(e_{xx}(u))^{-1}du\right] + O_p(n^{1/2}h^2) \\
= & \int_{t_1-h}^{t+h} \left[dX_{zn}^I(v)((e_{xx}(v))^{-1} + O(h^2))\right] + O_p(n^{1/2}h^2) \\
\xrightarrow{D} & \int_{t_1}^t \left[dX_z^I(v)((e_{xx}(v))^{-1})\right]
\end{aligned}$$

as $n \to \infty$, $h \to 0$ and $nh^4 \to 0$.

Then the third summation in (A.4) is equal to

$$\begin{aligned}
& n^{-1/2}\sum_{i=1}^n \int_{t_1}^t \Big[\int_0^\tau K_h(v-u)E_s\{(1-R_i)Z_i(v)X_i^T(v)dN_i^c(v) \mid \mathcal{D}_i, R_i = 0\} \\
& -E\{(1-R_i)\xi_i(u)\alpha_i(u)Z_i(u)X_i^T(u)\}\Big](e_{xx}(u))^{-1}du \\
= & \; n^{-1/2}\sum_{i=1}^n \int_{t_1}^t \Big[\int_0^\tau K_h(v-u)E_s\{(1-R_i)Z_i(v)X_i^T(v)dN_i^c(v) \mid \mathcal{D}_i, R_i = 0\} \\
& -\int_0^\tau K_h(v-u)E\{(1-R_i)\xi_i(v)\alpha_i(v)Z_i(v)X_i^T(v)\}dv + O(h^2)\Big] \\
& (e_{xx}(u))^{-1}du \\
= & \int_{t_1}^t \Big[\int_0^\tau K_h(v-u)\Big\{n^{-1/2}\sum_{i=1}^n (E_s\{(1-R_i)Z_i(v)X_i^T(v)dN_i^c(v) \mid \mathcal{D}_i, R_i = 0\} \\
& -E\{(1-R_i)\xi_i(v)\alpha_i(v)Z_i(v)X_i^T(v)\}dv)\Big\}\Big](e_{xx}(u))^{-1}du + O_p(n^{1/2}h^2)
\end{aligned}$$

$$= \int_{t_1}^{t}\left[\int_0^\tau K_h(v-u)d\left\{n^{-1/2}\sum_{i=1}^{n}\int_0^v (E_s\{(1-R_i)Z_i(w)X_i^T(w)dN_i^c(w)\mid \mathcal{D}_i, R_i\right.\right.$$

$$\left.\left. = 0\} - E\{(1-R_i)\xi_i(w)\alpha_i(w)Z_i(w)X_i^T(w)\}dw)\right\}\right](e_{xx}(u))^{-1}du$$

$$+O_p(n^{1/2}h^2).$$

Let

$$\begin{aligned} X_{zn}^{III}(v) &= n^{-1/2}\sum_{i=1}^{n}\int_0^v (E_s\{(1-R_i)Z_i(w)X_i^T(w)dN_i^c(w)\mid \mathcal{D}_i, R_i = 0\} \\ &\quad -E\{(1-R_i)\xi_i(w)\alpha_i(w)Z_i(w)X_i^T(w)\}dw). \end{aligned}$$

Under Condition (I) $X_{zn}^{III}(v)$ converges to a vector of mean zero Gaussian processes, saying $X_z^{III}(v)$ uniformly in v. Now follow the argument in discussing the first summation, we know

$$\begin{aligned} &\int_{t_1}^{t}\int_0^\tau K_h(v-u)dX_{zn}^{III}(v)(e_{xx}(u))^{-1}du + O_p(n^{1/2}h^2) \\ = &\int_{t_1-h}^{t+h}\left[dX_{zn}^{III}(v)((e_{xx}(v))^{-1}+O(h^2))\right] + O_p(n^{1/2}h^2) \\ \xrightarrow{D} &\int_{t_1}^{t}\left[dX_z^{III}(v)((e_{xx}(v))^{-1})\right] \end{aligned}$$

as $n\to\infty$, $h\to 0$ and $nh^4\to 0$.

Under the assumption that $\{S_i\}$ are independent of $\mathcal{D}_i$ and defining the counting process

$$N_i^*(t) = \sum_{j=1}^{n_i} I(T_{ij}\le t)I(C_i\ge t)$$

with the mean rate

$$E\{dN_i^*(t)\mid R_i, X_i(t), Y_i(t), Z_i(t), V_i\} = \alpha_i^*(t)dt,$$

the second summation of (A.4) equals

$$n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t}(1-R_i)\int_0^L\sum_{j=1}^{n_i}K_h(s+T_{ij}-u)Z_{ij}X_{ij}^T I(C_i\ge T_{ij})\left[\frac{d\widehat{F}_s(s)}{\widehat{F}_s(V_i)}\right.$$

$$
\begin{aligned}
& \quad -\frac{dF_s(s)}{F_s(V_i)}\Bigg](e_{xx}(u))^{-1}du \\
&= n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t}(1-R_i)\int_0^L\int_0^\tau K_h(v-u)Z_i(v)X_i^T(v)dN_i^*(v-s)\Bigg[\bigg(\frac{1}{\widehat{F}_s(V_i)} \\
& \quad -\frac{1}{F_s(V_i)}\bigg)dF_s(s)+\frac{d\widehat{F}_s(s)-dF_s(s)}{\widehat{F}_s(V_i)}\Bigg](e_{xx}(u))^{-1}du \\
&= n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t}(1-R_i)\int_0^L\int_0^\tau K_h(v-u)Z_i(v)X_i^T(v)dN_i^*(v-s)\frac{F_s(V_i)-\widehat{F}_s(V_i)}{F_s^2(V_i)} \\
& \quad dF_s(s)(e_{xx}(u))^{-1}du \\
& \quad +n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t}(1-R_i)\int_0^L\int_0^\tau K_h(v-u)Z_i(v)X_i^T(v)dN_i^*(v-s) \\
& \quad \frac{d(\widehat{F}_s(s)-F_s(s))}{F_s(V_i)}(e_{xx}(u))^{-1}du+o_p(1) \\
&= n^{-1}\sum_{i=1}^{n}\int_{t_1}^{t}\int_0^L\int_0^\tau(1-R_i)K_h(v-u)Z_i(v)X_i^T(v)dN_i^*(v-s) \\
& \quad \frac{n^{1/2}(\widehat{S}_s(V_i)-S_s(V_i))}{F_s^2(V_i)}dF_s(s)(e_{xx}(u))^{-1}du \qquad (A.5) \\
& \quad -n^{-1}\sum_{i=1}^{n}\int_{t_1}^{t}(1-R_i)\int_0^L\int_0^\tau K_h(v-u)Z_i(v)X_i^T(v)dN_i^*(v-s) \\
& \quad \frac{d[n^{1/2}(\widehat{S}_s(s)-S_s(s))]}{F_s(V_i)}(e_{xx}(u))^{-1}du \qquad (A.6) \\
& \quad +o_p(1)
\end{aligned}
$$

Plugging (A.1) into both (A.5) and (A.6), we have

$$
\begin{aligned}
(A.5) &= n^{-1}\sum_{i=1}^{n}\int_{t_1}^{t}\int_0^L\int_0^\tau(1-R_i)K_h(v-u)Z_i(v)X_i^T(v)dN_i^*(v-s)\frac{n^{-1/2}F_s(V_i)}{F_s^2(V_i)} \\
& \quad \int_0^{(L-(V_i))-}\frac{dM^R(x)}{y^R(x)}dF_s(s)(e_{xx}(u))^{-1}du+o_p(1) \\
&= \int_{t_1}^{t}\int_0^L n^{-1}\sum_{i=1}^{n}\int_0^\tau(1-R_i)K_h(v-u)Z_i(v)X_i^T(v)dN_i^*(v-s)\frac{n^{-1/2}}{F_s(V_i)} \\
& \quad \int_0^\infty I(x\le(L-(V_i))-)\frac{dM^R(x)}{y^R(x)}dF_s(s)(e_{xx}(u))^{-1}du+o_p(1) \\
&= n^{-1/2}\int_0^\infty\int_0^L\int_{t_1}^{t}\int_0^\tau K_h(v-u)n^{-1}\sum_{i=1}^{n}(1-R_i)Z_i(v)X_i^T(v)dN_i^*(v-s)
\end{aligned}
$$

$$\frac{I(x \leq (L-(V_i))-)}{F_s(V_i)}(e_{xx}(u))^{-1}dudF_s(s)\frac{dM^R(x)}{y^R(x)} + o_p(1),$$

and

$$\begin{aligned}
(A.6) \quad = \quad & -n^{-1}\sum_{i=1}^{n}\int_{t_1}^{t}(1-R_i)\int_0^L\int_0^\tau K_h(v-u)\frac{Z_i(v)X_i^T(v)dN_i^*(v-s)}{F_s(V_i)} \\
& d[n^{1/2}(\widehat{S}_s(s)-S_s(s))](e_{xx}(u))^{-1}du \\
= \quad & -n^{-1}\sum_{i=1}^{n}\int_{t_1}^{t}(1-R_i)\int_0^L\int_0^\tau K_h(v-u)\frac{Z_i(v)X_i^T(v)dN_i^*(v-s)}{F_s(V_i)} \\
& d[n^{-1/2}F_s(s)\int_0^{(L-s)-}\frac{dM^R(x)}{y^R(x)}](e_{xx}(u))^{-1}du \\
= \quad & -n^{-1}\sum_{i=1}^{n}\int_{t_1}^{t}(1-R_i)\int_0^L\int_0^\tau K_h(v-u)\frac{Z_i(v)X_i^T(v)dN_i^*(v-s)}{F_s(V_i)} \\
& n^{-1/2}\int_0^{(L-s)-}\frac{dM^R(x)}{y^R(x)}dF_s(s)(e_{xx}(u))^{-1}du \\
& +n^{-1}\sum_{i=1}^{n}\int_{t_1}^{t}(1-R_i)\int_0^L\int_0^\tau K_h(v-u)\frac{Z_i(v)X_i^T(v)dN_i^*(v-s)}{F_s(V_i)}n^{-1/2}F_s(s) \\
& \frac{dM^R((L-s)-)}{y^R((L-s)-)}(e_{xx}(u))^{-1}du \\
= \quad & -n^{-1/2}\int_0^{L-}\int_0^{(L-x)-}\int_{t_1}^{t}\int_0^\tau K_h(v-u)n^{-1}\sum_{i=1}^{n}(1-R_i)\frac{Z_i(v)X_i^T(v)}{F_s(V_i)} \\
& dN_i^*(v-s)(e_{xx}(u))^{-1}du\,dF_s(s)\frac{dM^R(x)}{y^R(x)} \\
& +n^{-1/2}\int_0^L\int_{t_1}^{t}\int_0^\tau K_h(v-u)n^{-1}\sum_{i=1}^{n}(1-R_i)\frac{Z_i(v)X_i^T(v)dN_i^*(v-s)}{F_s(V_i)}F_s(s) \\
& (e_{xx}(u))^{-1}du\frac{dM^R((L-s)-)}{y^R((L-s)-)}.
\end{aligned}$$

Since

$$\begin{aligned}
& \int_0^\tau K_h(v-u)n^{-1}\sum_{i=1}^{n}(1-R_i)Z_i(v)X_i^T(v)dN_i^*(v-s)\frac{I(x \leq (L-(V_i))-)}{F_s(V_i)} \\
= \quad & \int_0^\tau K_h(v-u)d\Bigg(n^{-1}\sum_{i=1}^{n}\int_0^v(1-R_i)Z_i(w)X_i^T(w)dN_i^*(w-s) \\
& \frac{I(x \leq (L-(V_i))-)}{F_s(V_i)}\Bigg)
\end{aligned}$$

$$
\begin{aligned}
&= \int_0^\tau K_h(v-u)dE\Big\{\int_0^v (1-R_i)Z_i(w)X_i^T(w)dN_i^*(w-s)\frac{I(x\le (L-(V_i))-)}{F_s(V_i)}\Big\} \\
&\quad +o_p(1) \\
&= \int_0^\tau K_h(v-u)dE\Big\{E\Big[\int_0^v (1-R_i)Z_i(w)X_i^T(w)dN_i^*(w-s) \\
&\qquad \frac{I(x\le (L-(V_i))-)}{F_s(V_i)}\Big|R_i, X_i(t), Y_i(t), Z_i(t), V_i\Big]\Big\} + o_p(1) \\
&= \int_0^\tau K_h(v-u)dE\Big\{\int_0^v (1-R_i)Z_i(w)X_i^T(w)E[dN_i^*(w-s)\mid R_i, X_i(t), Y_i(t), \\
&\qquad Z_i(t), V_i]\frac{I(x\le (L-(V_i))-)}{F_s(V_i)}\Big\} + o_p(1) \\
&= \int_0^\tau K_h(v-u)dE\Big\{\int_0^v (1-R_i)Z_i(w)X_i^T(w)\alpha_i^*(w-s)dw\frac{I(x\le (L-(V_i))-)}{F_s(V_i)}\Big\} \\
&\quad +o_p(1) \\
&= \int_0^\tau K_h(v-u)E\Big\{(1-R_i)Z_i(v)X_i^T(v)\alpha_i^*(v-s)\frac{I(x\le (L-(V_i))-)}{F_s(V_i)}\Big\}dv + o_p(1) \\
&= E\Big\{(1-R_i)Z_i(u)X_i^T(u)\alpha_i^*(u-s)\frac{I(x\le (L-(V_i))-)}{F_s(V_i)}\Big\} + O(h^2) + o_p(1)
\end{aligned}
$$

and similarly

$$
\begin{aligned}
&\int_0^\tau K_h(v-u)n^{-1}\sum_{i=1}^n (1-R_i)\frac{Z_i(v)X_i^T(v)dN_i^*(v-s)}{F_s(V_i)} \\
&= E\Big\{(1-R_i)\frac{Z_i(u)X_i^T(u)\alpha_i^*(u-s)}{F_s(V_i)}\Big\} + O(h^2) + o_p(1),
\end{aligned}
$$

then

$$
\begin{aligned}
(A.5) &= n^{-1/2}\int_0^\infty\int_0^L\int_{t_1}^t E\Big\{(1-R_i)Z_i(u)X_i^T(u)\alpha_i^*(u-s)\frac{I(x\le (L-(V_i))-)}{F_s(V_i)}\Big\} \\
&\qquad (e_{xx}(u))^{-1}du\,dF_s(s)\frac{dM^R(x)}{y^R(x)} + O_p(n^{-1/2}h^2) + o_p(1), \\
(A.6) &= -n^{-1/2}\int_0^{L-}\int_0^{(L-x)-}\int_{t_1}^t E\Big\{(1-R_i)\frac{Z_i(u)X_i^T(u)\alpha_i^*(u-s)}{F_s(V_i)}\Big\}(e_{xx}(u))^{-1} \\
&\qquad du\,dF_s(s)\frac{dM^R(x)}{y^R(x)} \\
&\quad +n^{-1/2}\int_0^L\int_{t_1}^t E\Big\{(1-R_i)\frac{Z_i(u)X_i^T(u)\alpha_i^*(u-s)}{F_s(V_i)}\Big\}F_s(s)(e_{xx}(u))^{-1} \\
&\qquad du\frac{dM^R((L-s)-)}{y^R((L-s)-)} + O_p(n^{-1/2}h^2) + o_p(1)
\end{aligned}
$$

Thus the second summation of (A.4) equals

$$
\begin{aligned}
n^{-1/2} \quad \Bigg[\quad & \int_0^\infty \int_0^L \int_{t_1}^t E\bigg\{(1-R_i)Z_i(u)X_i^T(u)\alpha_i^*(u-s)\frac{I(x \le (L-V_i)-)}{F_s(V_i)}\bigg\} \\
& \quad (e_{xx}(u))^{-1} du\, dF_s(s)\frac{dM^R(x)}{y^R(x)} \\
& - \int_0^{L-} \int_0^{(L-x)-} \int_{t_1}^t E\bigg\{(1-R_i)\frac{Z_i(u)X_i^T(u)\alpha_i^*(u-s)}{F_s(V_i)}\bigg\}(e_{xx}(u))^{-1} du \\
& \quad dF_s(s)\frac{dM^R(x)}{y^R(x)} \\
& + \int_0^L \int_{t_1}^t E\bigg\{(1-R_i)\frac{Z_i(u)X_i^T(u)\alpha_i^*(u-s)}{F_s(V_i)}\bigg\}F_s(s)(e_{xx}(u))^{-1} du \\
& \quad \frac{dM^R((L-s)-)}{y^R((L-s)-)}\Bigg] + O_p(n^{-1/2}h^2) + o_p(1).
\end{aligned}
$$

By the multivariate martingale central limit theorem, we know that the above three terms converge weakly to Wiener processes since the integrants of the martingale integral are deterministic functions.

Above all, Equation (A.4) weakly converges to a vector of mean zero Gaussian processes with continuous paths as $n \to \infty$, $h \to 0$ and $nh^4 \to 0$. $\square$

Recall the definitions in Section A.1. We can have the following lemma.

Lemma A.2.5:

$$
\begin{aligned}
& n^{1/2} \int_{t_1}^t \{\tilde{\beta}^T(u,\gamma_0) - \beta_0^T(u)\} du \\
= \quad & n^{-1/2} \sum_{i=1}^n \bigg\{ X_{yi}^{II}(t) - X_{zi}^{II}(t) - X_{xi}^{II}(t) \\
& \quad + \int_{t_1-h}^{t+h} d(X_{yi}^I(v) + X_{yi}^{III}(v) - X_{zi}^I(v) - X_{zi}^{III}(v))((e_{xx}(v))^{-1} + O(h^2)) \\
& \quad - \int_{t_1-h}^{t+h} (\beta^T(v) + O(h^2)) d(X_{xi}^I(v) + X_{xi}^{III}(v))((e_{xx}(v))^{-1} + O(h^2)) \bigg\} \\
& + O_p(n^{-1/2}h^2 + n^{1/2}h^2) + o_p(1)
\end{aligned}
$$

converges weakly to a vector of mean zero Gaussian processes with continuous paths as $n \to \infty$, $h \to 0$ and $nh^4 \to 0$.

Proof. By the definitions,

$$
\begin{aligned}
& n^{1/2}\int_{t_1}^{t}\{\tilde{\beta}^T(u,\gamma_0)-\beta_0^T(u)\}du \\
= & \int_{t_1}^{t} n^{1/2}\{\tilde{Y}_x(u)-\gamma_0^T\tilde{Z}_x(u)-(y_x(u)-\gamma_0^T z_x(u))\}du \\
= & n^{1/2}\int_{t_1}^{t}\{\tilde{Y}_x(u)-y_x(u)\}du-\gamma_0^T n^{1/2}\int_{t_1}^{t}\{\tilde{Z}_x(u)-z_x(u)\}du
\end{aligned}
$$

By the continuous mapping theorem, it is sufficient to prove that

$$
\left(n^{1/2}\int_{t_1}^{t}\{\tilde{Y}_x(u)-y_x(u)\}du,\, n^{1/2}\int_{t_1}^{t}\{\tilde{Z}_x(u)-z_x(u)\}du\right) \tag{A.7}
$$

converges weakly to a vector of mean zero Gaussian processes with continuous sample paths. And

$$
\begin{aligned}
& n^{1/2}\int_{t_1}^{t}\{\tilde{Y}_x(u)-y_x(u)\}du \\
= & n^{1/2}\int_{t_1}^{t}\{\tilde{E}_{yx}(u)(\tilde{E}_{xx}(u))^{-1}-e_{yx}(u)(e_{xx}(u))^{-1}\}du \\
= & n^{1/2}\int_{t_1}^{t}\{[\tilde{E}_{yx}(u)-e_{yx}(u)](\tilde{E}_{xx}(u))^{-1}-e_{yx}(u)(\tilde{E}_{xx}(u))^{-1}[\tilde{E}_{xx}(u) \\
& -e_{xx}(u)](e_{xx}(u))^{-1}\}du \\
= & n^{1/2}\int_{t_1}^{t}\{[\tilde{E}_{yx}(u)-e_{yx}(u)](e_{xx}(u))^{-1}-e_{yx}(u)(e_{xx}(u))^{-1}[\tilde{E}_{xx}(u) \\
& -e_{xx}(u)](e_{xx}(u))^{-1}\}du+o_p(1)
\end{aligned}
$$

$n^{1/2}\int_{t_1}^{t}\{\tilde{Y}_x(u)-y_x(u)\}du$ has a similar decomposition. Under Condition (I), applying Lemma A.1 of Lin & Ying (2001) and Lemma A.2.4 above,

$$
n^{1/2}\int_{t_1}^{t}\{\tilde{Y}_x(u)-y_x(u)\}du \quad and \quad n^{1/2}\int_{t_1}^{t}\{\tilde{Z}_x(u)-z_x(u)\}du
$$

converges weakly to a mean zero Gaussian process respectively. So using the Wald device, we could have the joint weak convergence of (A.7) which leads to the weak convergence of $n^{1/2}\int_{t_1}^{t}\{\tilde{\beta}^T(u,\gamma_0)-\beta_0^T(u)\}du$ with zero mean. This completes the proof. □

A.3 Proof of Theorems

Proof of Theorem 3.1

By the uniform convergence of $\tilde{Y}_x(t)$ and $\tilde{Z}_x(t)$, which can be proved by using Lemma A.2.3, we have

$$\tilde{\beta}(t;\gamma) = \tilde{Y}_x^T(t) - \tilde{Z}_x^T(t)\gamma \xrightarrow{P} y_x^T(t) - z_x^T(t)\gamma$$

uniformly in $t \in [t_1, t_2]$ as $n \to \infty, h \to 0$. Since $\beta_0(t) = y_x^T(t) - z_x^T(t)\gamma_0$, by using (2.6), replace $\beta(s)$ in (2.3) and Applying Lemma A.2.2 We have $n^{-1}\tilde{l}(\gamma)$ equals

$$\begin{aligned}
& n^{-1}\sum_{i=1}^{n} R_i \int_0^{\tau} W_i(s)\{Y_i(s) - (\tilde{Y}_x(s) - \gamma^T\tilde{Z}_x(s))X_i(s) - \gamma^T Z_i(s)\}^2\, dN_i^c(s) \\
& +n^{-1}\sum_{i=1}^{n}(1-R_i)\widehat{E}_S\Big[\int_0^{\tau} W_i(s)\{Y_i(s) - (\tilde{Y}_x(s) - \gamma^T\tilde{Z}_x(s))X_i(s) \\
& \qquad -\gamma^T Z_i(s)\}^2\, dN_i^c(s) \mid \mathcal{X}\Big] \\
= \;& n^{-1}\sum_{i=1}^{n} \ll \int_0^{\tau} W_i(s)\{Y_i(s) - (\tilde{Y}_x(s) - \gamma^T\tilde{Z}_x(s))X_i(s) - \gamma^T Z_i(s)\}^2\, dN_i^c(s) \gg_R \\
= \;& n^{-1}\sum_{i=1}^{n} \ll \int_0^{\tau} W_i(s)\{Y_i(s) - \tilde{Y}_x(s)X_i(s) + \gamma^T(\tilde{Z}_x(s)X_i(s) \\
& \qquad -Z_i(s))\}^2\, dN_i^c(s) \gg_R
\end{aligned}$$

where

$$\begin{aligned}
& \int_0^{\tau} W_i(s)\{Y_i(s) - \tilde{Y}_x(s)X_i(s) + \gamma^T(\tilde{Z}_x(s)X_i(s) - Z_i(s))\}^2\, dN_i^c(s) \\
= \;& \int_0^{\tau} W_i(s)[\{Y_i(s) - \tilde{Y}_x(s)X_i(s) + \gamma^T(\tilde{Z}_x(s)X_i(s) - Z_i(s))\}^2 - \{Y_i(s) \\
& \qquad -y_x(s)X_i(s) + \gamma^T(z_x(s)X_i(s) - Z_i(s))\}^2]\, dN_i^c(s) \\
& + \int_0^{\tau} W_i(s)\{Y_i(s) - y_x(s)X_i(s) + \gamma^T(z_x(s)X_i(s) - Z_i(s))\}^2\, dN_i^c(s) \\
= \;& \int_0^{\tau} \{-(\tilde{Y}_x(s) - y_x(s)) + \gamma^T(\tilde{Z}_x(s) - z_x(s))\}X_i(s)W_i(s)[2Y_i(s) \\
& \qquad -(\tilde{Y}_x(s) + y_x(s))X_i(s) + \gamma^T\{(\tilde{Z}_x(s) + z_x(s))X_i(s) - 2Z_i(s)\}]dN_i^c(s) \\
& + \int_0^{\tau} W_i(s)\{Y_i(s) - y_x(s)X_i(s) + \gamma^T(z_x(s)X_i(s) - Z_i(s))\}^2\, dN_i^c(s)
\end{aligned}$$

$$
\begin{aligned}
&= \int_0^\tau \{-(\tilde{Y}_x(s)-y_x(s))+\gamma^T(\tilde{Z}_x(s)-z_x(s))\}X_i(s)W_i(s)\{-(\tilde{Y}_x(s)-y_x(s))X_i(s)\\
&\qquad +\gamma^T(\tilde{Z}_x(s)-z_x(s))X_i(s)+2y_x(s)X_i(s)+2Y_i(s)\\
&\qquad +\gamma^T(2z_x(s)X_i(s)-2Z_i(s))\}dN_i^c(s)\\
&\quad +\int_0^\tau W_i(s)\{Y_i(s)-y_x(s)X_i(s)+\gamma^T(z_x(s)X_i(s)-Z_i(s))\}^2\,dN_i^c(s)\\
&= \int_0^\tau [\{-(\tilde{Y}_x(s)-y_x(s))+\gamma^T(\tilde{Z}_x(s)-z_x(s))\}X_i(s)]^2W_i(s)dN_i^c(s)\\
&\quad +\int_0^\tau 2\{-(\tilde{Y}_x(s)-y_x(s))+\gamma^T(\tilde{Z}_x(s)-z_x(s))\}X_i(s)W_i(s)\{y_x(s)X_i(s)+Y_i(s)\\
&\qquad +\gamma^T(z_x(s)X_i(s)-Z_i(s))\}dN_i^c(s)\\
&\quad +\int_0^\tau W_i(s)\{Y_i(s)-y_x(s)X_i(s)+\gamma^T(z_x(s)X_i(s)-Z_i(s))\}^2\,dN_i^c(s)
\end{aligned}
$$

So by the linearity of the operation $\ll\gg_R$,

$$
\begin{aligned}
n^{-1}\tilde{l}(\gamma) &= n^{-1}\sum_{i=1}^n \ll \int_0^\tau [\{-(\tilde{Y}_x(s)-y_x(s))+\gamma^T(\tilde{Z}_x(s)-z_x(s))\}X_i(s)]^2W_i(s)\\
&\qquad dN_i^c(s) \gg_R\\
&\quad +n^{-1}\sum_{i=1}^n \ll \int_0^\tau 2\{-(\tilde{Y}_x(s)-y_x(s))+\gamma^T(\tilde{Z}_x(s)-z_x(s))\}X_i(s)W_i(s)\\
&\qquad \{y_x(s)X_i(s)+Y_i(s)+\gamma^T(z_x(s)X_i(s)-Z_i(s))\}dN_i^c(s) \gg_R\\
&\quad +n^{-1}\sum_{i=1}^n \ll \int_0^\tau W_i(s)\{Y_i(s)-y_x(s)X_i(s)+\gamma^T(z_x(s)X_i(s)-Z_i(s))\}^2\\
&\qquad dN_i^c(s) \gg_R + o_p(1)
\end{aligned}
$$

The first term equals

$$
\begin{aligned}
&n^{-1}\sum_{i=1}^n R_i\int_0^\tau [\{-(\tilde{Y}_x(s)-y_x(s))+\gamma^T(\tilde{Z}_x(s)-z_x(s))\}X_i(s)]^2W_i(s)dN_i^c(s)\\
&+n^{-1}\sum_{i=1}^n(1-R_i)\widehat{E}_s\Big\{\int_0^\tau [\{-(\tilde{Y}_x(s)-y_x(s))+\gamma^T(\tilde{Z}_x(s)-z_x(s))\}X_i(s)]^2W_i(s)\\
&\qquad dN_i^c(s)\mid\mathcal{X}\Big\}\\
&= n^{-1}\sum_{i=1}^n R_i\int_0^\tau [\{-(\tilde{Y}_x(s)-y_x(s))+\gamma^T(\tilde{Z}_x(s)-z_x(s))\}X_i(s)]^2W_i(s)dN_i^c(s)
\end{aligned}
$$

$$
\begin{aligned}
&+n^{-1}\sum_{i=1}^{n}(1-R_i)E_s\bigg\{\int_0^\tau[\{-(\tilde{Y}_x(s)-y_x(s))+\gamma^T(\tilde{Z}_x(s)-z_x(s))\}X_i(s)]^2W_i(s)\\
&\quad dN_i^c(s)\mid\mathcal{X}\bigg\}+o_p(1)\\
= &\int_0^\tau\{-(\tilde{Y}_x(s)-y_x(s))+\gamma^T(\tilde{Z}_x(s)-z_x(s))\}\bigg(n^{-1}\sum_{i=1}^{n}R_iX_i(s)X_i(s)^TW(s)\\
&\quad dN_i^c(s)\bigg)\{-(\tilde{Y}_x(s)-y_x(s))+\gamma^T(\tilde{Z}_x(s)-z_x(s))\}^T\\
&+E_s\bigg\{\int_0^\tau\{-(\tilde{Y}_x(s)-y_x(s))+\gamma^T(\tilde{Z}_x(s)-z_x(s))\}\bigg(n^{-1}\sum_{i=1}^{n}(1-R_i)X_i(s)X_i(s)^T\\
&\quad W_i(s)dN_i^c(s)\bigg)\{-(\tilde{Y}_x(s)-y_x(s))+\gamma^T(\tilde{Z}_x(s)-z_x(s))\}^T\mid\mathcal{X}\bigg\}+o_p(1)\\
= &\int_0^\tau\{-(\tilde{Y}_x(s)-y_x(s))+\gamma^T(\tilde{Z}_x(s)-z_x(s))\}d\bigg(n^{-1}\sum_{i=1}^{n}\int_0^sR_iX_i(u)X_i(u)^TW_i(u)\\
&\quad dN_i^c(u)\bigg)\{-(\tilde{Y}_x(s)-y_x(s))+\gamma^T(\tilde{Z}_x(s)-z_x(s))\}^T\\
&+E_s\bigg\{\int_0^\tau\{-(\tilde{Y}_x(s)-y_x(s))+\gamma^T(\tilde{Z}_x(s)-z_x(s))\}d\bigg(n^{-1}\sum_{i=1}^{n}\int_0^s(1-R_i)X_i(u)\\
&\quad X_i(u)^TW_i(u)dN_i^c(u)\bigg)\{-(\tilde{Y}_x(s)-y_x(s))+\gamma^T(\tilde{Z}_x(s)-z_x(s))\}^T\mid\mathcal{X}\bigg\}\\
&+o_p(1).
\end{aligned}
$$

Since

$$
\begin{aligned}
&n^{-1}\sum_{i=1}^{n}\int_0^sR_iX_i(u)X_i(u)^TW_i(u)dN_i^c(u)\\
\xrightarrow{P}\ &E\bigg\{\int_0^sR_iX_i(u)X_i(u)^TW(u)dN_i^c(u)\bigg\},\\
&n^{-1}\sum_{i=1}^{n}\int_0^s(1-R_i)X_i(u)X_i(u)^TW_i(u)dN_i^c(u)\\
\xrightarrow{P}\ &E\bigg\{\int_0^s(1-R_i)X_i(u)X_i(u)^TW_i(u)dN_i^c(u)\bigg\}
\end{aligned}
$$

and by the uniform convergence of $\tilde{Y}_x(s)$ and $\tilde{Z}_x(s)$ which lead to $-(\tilde{Y}_x(s)-y_x(s))+\gamma^T(\tilde{Z}_x(s)-z_x(s))\xrightarrow{P}0$, the first term converges to zero in probability.

The second term equals

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^{n} R_i \int_0^{\tau} 2\{-(\tilde{Y}_x(s)-y_x(s))+\gamma^T(\tilde{Z}_x(s)-z_x(s))\}X_i(s)W_i(s)\{y_x(s)X_i(s) \\
& \qquad +Y_i(s)+\gamma^T[z_x(s)X_i(s)-Z_i(s)]\}dN_i^c(s) \\
& +n^{-1}\sum_{i=1}^{n}(1-R_i)\widehat{E}_s\bigg\{\int_0^{\tau} 2\{-(\tilde{Y}_x(s)-y_x(s))+\gamma^T(\tilde{Z}_x(s)-z_x(s))\}X_i(s)W_i(s) \\
& \qquad \{y_x(s)X_i(s)+Y_i(s)+\gamma^T[z_x(s)X_i(s)-Z_i(s)]\}dN_i^c(s) \mid \mathcal{X}\bigg\} \\
= \;& \int_0^{\tau} 2\{-(\tilde{Y}_x(s)-y_x(s))+\gamma^T(\tilde{Z}_x(s)-z_x(s))\}d\bigg(n^{-1}\sum_{i=1}^{n}\int_0^{s} R_iX_i(u)W_i(u) \\
& \qquad \{y_x(u)X_i(u)+Y_i(u)+\gamma^T[z_x(u)X_i(u)-Z_i(u)]\}dN_i^c(u)\bigg) \\
& +E_s\bigg\{\int_0^{\tau} 2\{-(\tilde{Y}_x(s)-y_x(s))+\gamma^T(\tilde{Z}_x(s)-z_x(s))\}d\bigg(n^{-1}\sum_{i=1}^{n}\int_0^{s}(1-R_i)X_i(u) \\
& \qquad W_i(u)\{y_x(u)X_i(u)+Y_i(u)+\gamma^T[z_x(u)X_i(u)-Z_i(u)]\}dN_i^c(u)\bigg) \mid \mathcal{X}\bigg\} \\
& +o_p(1).
\end{aligned}
$$

Also

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^{n}\int_0^{s} R_iX_i(u)W_i(u)\{y_x(u)X_i(u)+Y_i(u) \\
& \qquad\qquad +\gamma^T[z_x(u)X_i(u)-Z_i(u)]\}dN_i^c(u) \\
\xrightarrow{P} \;& E\bigg\{\int_0^{s} R_iX_i(u)W_i(u)\{y_x(u)X_i(u)+Y_i(u) \\
& \qquad\qquad +\gamma^T[z_x(u)X_i(u)-Z_i(u)]\}dN_i^c(u)\bigg\}, \\
& n^{-1}\sum_{i=1}^{n}\int_0^{s}(1-R_i)X_i(u)W(u)\{y_x(u)X_i(u)+Y_i(u) \\
& \qquad\qquad +\gamma^T[z_x(u)X_i(u)-Z_i(u)]\}dN_i^c(u) \\
\xrightarrow{P} \;& E\bigg\{\int_0^{s}(1-R_i)X_i(u)W(u)\{y_x(u)X_i(u)+Y_i(u) \\
& \qquad\qquad +\gamma^T[z_x(u)X_i(u)-Z_i(u)]\}dN_i^c(u)\bigg\}.
\end{aligned}
$$

Similarly to the first term, the second term converges to zero in probability.

Therefore according to our lemma A.2.2,

$$
\begin{aligned}
n^{-1}\tilde{l}(\gamma) &= n^{-1}\sum_{i=1}^{n} \ll \int_0^\tau W_i(s)\{Y_i(s) - y_x(s)X_i(s) + \gamma^T(z_x(s)X_i(s) \\
&\qquad -Z_i(s))\}^2 \, dN_i^c(s) \gg_R + o_p(1) \\
&\xrightarrow{P} E\Big\{ \int_0^\tau w(s)\{Y_i(s) - y_x(s)X_i(s) + \gamma^T(z_x(s)X_i(s) - Z_i(s))\}^2 \, dN_i^c(s) \Big\} \\
&= E\Big\{ \int_0^\tau w(s)\{Y_i(s) - (y_x(s) - \gamma_0^T z_x(s))X_i(s) - \gamma_0^T Z_i(s) \\
&\qquad +(\gamma-\gamma_0)^T(z_x(s)X_i(s) - Z_i(s))\}^2 \, dN_i^c(s) \Big\} \\
&= E\Big\{ \int_0^\tau w(s)\{\epsilon_i(s) + (\gamma-\gamma_0)^T(z_x(s)X_i(s) - Z_i(s))\}^2 \, dN_i^c(s) \Big\} \\
&= E\Big\{ \int_0^\tau w(s)\{\epsilon_i^2(s) + 2\epsilon_i(s)[(\gamma-\gamma_0)^T(z_x(s)X_i(s) - Z_i(s))] \\
&\qquad +[(\gamma-\gamma_0)^T(z_x(s)X_i(s) - Z_i(s))]^2\} \, dN_i^c(s) \Big\} \\
&= E\Big\{ \int_0^\tau w(s)\{\epsilon_i^2(s) + [(\gamma-\gamma_0)^T(z_x(s)X_i(s) - Z_i(s))]^2\} \, dN_i^c(s) \Big\} \\
&\quad + E\Big\{ \int_0^\tau 2w(s)\epsilon_i(s)(\gamma-\gamma_0)^T(z_x(s)X_i(s) - Z_i(s)) dN_i^c(s) \Big\} \\
&= E\Big\{ \int_0^\tau w(s)\{\epsilon_i^2(s) + [(\gamma-\gamma_0)^T(z_x(s)X_i(s) - Z_i(s))]^2\} \, dN_i^c(s) \Big\} \\
&\quad + \int_0^\tau E\{E[2w(s)\epsilon_i(s)(\gamma-\gamma_0)^T(z_x(s)X_i(s) - Z_i(s)) dN_i^c(s) \mid X_i(s), \\
&\qquad Z_i(s)]\} \\
&= E\Big\{ \int_0^\tau w(s)\{\epsilon_i^2(s) + [(\gamma-\gamma_0)^T(z_x(s)X_i(s) - Z_i(s))]^2\} \, dN_i^c(s) \Big\} \\
&\quad + \int_0^\tau E\{2w(s)(\gamma-\gamma_0)^T(z_x(s)X_i(s) - Z_i(s)) E[\epsilon_i(s) dN_i^c(s) \mid X_i(s), \\
&\qquad Z_i(s)]\} \\
&= E\Big\{ \int_0^\tau w(s)\{\epsilon_i^2(s) + [(\gamma-\gamma_0)^T(z_x(s)X_i(s) - Z_i(s))]^2\} \, dN_i^c(s) \Big\} \\
&\quad + \int_0^\tau E\{2w(s)(\gamma-\gamma_0)^T(z_x(s)X_i(s) - Z_i(s)) E[\epsilon_i(s) \mid X_i(s), Z_i(s)] \\
&\qquad E[dN_i^c(s) \mid X_i(s), Z_i(s)]\} \\
&= E\Big\{ \int_0^\tau w(s)\{\epsilon_i^2(s) + [(\gamma-\gamma_0)^T(z_x(s)X_i(s) - Z_i(s))]^2\} \, dN_i^c(s) \Big\}
\end{aligned}
$$

$$\equiv \quad l_0(\gamma) \geq l_0(\gamma_0) \equiv E\bigg\{ \int_0^\tau w(s)\epsilon_i^2(s)dN_i^c(s) \bigg\},$$

uniformly in γ in Γ. Let $d(\gamma, \gamma_0)$ be the Euclidean distance between γ and γ_0. Therefore, for every $\epsilon > 0$,

$$\begin{aligned}
& \sup_{\gamma:d(\gamma,\gamma_0)\geq\epsilon} (-l_0(\gamma)) = -\inf_{\gamma:d(\gamma,\gamma_0)\geq\epsilon} l_0(\gamma) \\
= & -\inf_{\gamma:d(\gamma,\gamma_0)\geq\epsilon} E\bigg\{ \int_0^\tau w(s)\{\epsilon_i^2(s) + [(\gamma-\gamma_0)^T(z_x(s)X_i(s) - Z_i(s))]^2\}\, dN_i^c(s) \bigg\} \\
< & -\inf_{\gamma:d(\gamma,\gamma_0)\geq\epsilon} E\bigg\{ \int_0^\tau w(s)\{\epsilon_i^2(s)dN_i^c(s) \bigg\} = -\inf_{\gamma:d(\gamma,\gamma_0)\geq\epsilon} l_0(\gamma_0) \\
= & \sup_{\gamma:d(\gamma,\gamma_0)\geq\epsilon} (-l_0(\gamma_0)).
\end{aligned}$$

Then according to Theorem 5.7 of van der Vaart (1998), we have $\widehat{\gamma} \xrightarrow{P} \gamma_0$. □

Proof of Theorem 3.2

By continuous mapping theorem, the asymptotic uniform consistency of $\widehat{\beta}(t)$ on $[t_1, t_2]$ can be easily obtained by the consistency of $\widehat{\gamma}$, the uniform consistency of $\tilde{Y}_x(t)$ and $\tilde{Z}_x(t)$ since $\widehat{\beta}(t) = \tilde{Y}_x^T(t) - \tilde{Z}_x^T(t)\widehat{\gamma}$. □

Proof of Theorem 3.3

Recall the score function $U(\gamma)$ and the Taylor expansion of $U(\widehat{\gamma})$ at γ_0

$$n^{1/2}(\widehat{\gamma} - \gamma_0) = -\bigg(n^{-1}\frac{\partial U(\gamma^*)}{\partial \gamma^T}\bigg)^{-1}[n^{-1/2}U(\gamma_0)], \tag{A.8}$$

where γ^* is on the line segment between $\widehat{\gamma}$ and γ_0.

By plugging (2.6) into the score function (2.7) we will have

$$\begin{aligned}
U(\gamma) &= \sum_{i=1}^n \int_{t_1}^{t_2} \ll W_i(t)\{Z_i(t) - \tilde{Z}_x(t)X_i(t)\}\{Y_i(t) - X_i^T(t)(\tilde{Y}_x^T(t) \\
&\qquad -\tilde{Z}_x^T(t)\gamma) - Z_i^T(t)\gamma\}\, dN_i^c(t) \gg_R \\
&= \sum_{i=1}^n \int_{t_1}^{t_2} \ll W_i(t)\{Z_i(t) - \tilde{Z}_x(t)X_i(t)\}\{Y_i(t) - X_i^T(t)\tilde{Y}_x^T(t) + (X_i^T(t)\tilde{Z}_x^T(t) \\
&\qquad -Z_i^T(t))\gamma\}\, dN_i^c(t) \gg_R .
\end{aligned}$$

Then take the partial derivative with respect to γ, we get

$$n^{-1}\frac{\partial U(\gamma^*)}{\partial \gamma^T} = -n^{-1}\sum_{i=1}^{n}\int_{t_1}^{t_2} \ll W_i(t)\{Z_i(t) - \tilde{Z}_x(t)X_i(t)\}^{\otimes 2}\, dN_i^c(t) \gg_R . \qquad \text{(A.9)}$$

According to the similar argument we discussed in the proof of consistency of $\widehat{\gamma}$, $\tilde{Z}_x(t)$ and $W_i(t)$ can be replaced by their limits $z_x(t)$ and $w(t)$ respectively, and this change only contributes a $o_p(1)$ difference to the above equation. Thus by Lemma A.2.2

$$\begin{aligned} n^{-1}\frac{\partial U(\gamma^*)}{\partial \gamma^T} &= -n^{-1}\sum_{i=1}^{n} \ll \int_{t_1}^{t_2} w(t)\{Z_i(t) - z_x(t)X_i(t)\}^{\otimes 2}\, dN_i^c(t) \gg_R + o_p(1) \\ &\xrightarrow{P} -E\Bigg(\int_{t_1}^{t_2} w(t)\{Z_i(t) - z_x(t)X_i(t)\}^{\otimes 2}\, dN_i^c(t)\Bigg) = -D. \end{aligned}$$

Now we define $\mathcal{B}(t) = \int_{t_1}^{t} \beta_0(s)ds$ and a mean zero process

$$M_i(t;\mathcal{B},\gamma,\alpha) = \int_{t_1}^{t}\{[Y_i(s) - \gamma^T Z_i(s)]dN_i^c(s) - \xi_i(s)\alpha_i(s)X_i^T(s)d\mathcal{B}(s)\}. \qquad \text{(A.10)}$$

For simplicity, we use $M_i(t) = M_i(t;\mathcal{B},\gamma_0,\alpha)$. Also let $O_i(t) = N_i^c(t) - \int_0^t \xi_i(s)\alpha_i(s)ds$. Hence

$$\begin{aligned} n^{-1/2}U(\gamma_0) &= n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2} \ll W_i(t)\{Z_i(t) - \tilde{Z}_x(t)X_i(t)\}\{Y_i(t) - X_i^T(t)\tilde{\beta}(t;\gamma_0) \\ &\qquad -Z_i^T(t)\gamma_0\}\, dN_i^c(t) \gg_R \\ &= n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2} \ll W_i(t)\{Z_i(t) - \tilde{Z}_x(t)X_i(t)\}\{dM_i(t) \\ &\qquad +\xi_i(t)\alpha_i(t)X_i^T(t)d\mathcal{B}(t)\} \gg_R \\ &\quad -n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2} \ll W_i(t)\{Z_i(t) - \tilde{Z}_x(t)X_i(t)\}X_i^T(t)\tilde{\beta}(t;\gamma_0)\, dN_i^c(t) \gg_R \\ &= n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2} \ll W_i(t)\{Z_i(t) - \tilde{Z}_x(t)X_i(t)\}\{dM_i(t) \\ &\qquad -\beta_0^T(t)X_i(t)dO_i(t)\} \gg_R \\ &\quad -n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2} \ll W_i(t)\{Z_i(t) - \tilde{Z}_x(t)X_i(t)\}\{\tilde{\beta}^T(t;\gamma_0) \\ &\qquad -\beta_0^T(t)\}X_i(t)\, dN_i^c(t) \gg_R \\ &\quad +n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2} \ll W_i(t)\{Z_i(t) - \tilde{Z}_x(t)X_i(t)\}\{\beta_0^T(t)X_i^T(t)dO_i(t) \end{aligned}$$

$$+\xi_i(t)\alpha_i(t)X_i^T(t)\beta(t)dt - \beta_0^T(t)X_i^T(t)\,dN_i^c(t)\} \gg_R$$

By the definition of $O_i(t)$, the third term above is equal to zero. Let η be the second term. Hence

$$\begin{aligned}
\eta &= n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2} \ll W_i(t)\{Z_i(t)-\tilde{Z}_x(t)X_i(t)\}\{\tilde{\beta}^T(t;\gamma_0) \\
&\quad -\beta_0^T(t)\}X_i(t)\,dN_i^c(t) \gg_R \\
&= n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2} \ll W_i(t)\{Z_i(t)-\tilde{Z}_x(t)X_i(t)\}\{\tilde{\beta}^T(t;\gamma_0)-\beta_0^T(t)\}X_i(t)[dO_i(t) \\
&\quad +\xi_i(t)\alpha_i(t)dt] \gg_R \\
&= n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2} \ll W_i(t)\{\xi_i(t)\alpha_i(t)Z_i(t)X_i^T(t)-\tilde{Z}_x(t)\xi_i(t)\alpha_i(t)X_i(t)X_i^T(t)\} \\
&\quad \{\tilde{\beta}(t;\gamma_0)-\beta_0(t)\}dt \gg_R \\
&\quad +n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2} \ll W_i(t)\{Z_i(t)-\tilde{Z}_x(t)X_i(t)\}X_i^T(t)\{\tilde{\beta}(t;\gamma_0) \\
&\quad -\beta_0(t)\}dO_i(t) \gg_R .
\end{aligned}$$

Denote

$$\begin{aligned}
\eta_1 &= n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2} \ll W_i(t)\{\xi_i(t)\alpha_i(t)Z_i(t)X_i^T(t)-\tilde{Z}_x(t)\xi_i(t)\alpha_i(t)X_i(t)X_i^T(t)\} \\
&\quad \{\tilde{\beta}(t;\gamma_0)-\beta_0(t)\}dt \gg_R, \\
\eta_2 &= n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2} \ll W_i(t)\{Z_i(t)-\tilde{Z}_x(t)X_i(t)\}X_i^T(t)\{\tilde{\beta}(t;\gamma_0)-\beta_0(t)\}dO_i(t) \gg_R .
\end{aligned}$$

In the following statement we will prove that both terms converge to zero in probability.

$$\begin{aligned}
\eta_1 &= n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2} \ll W_i(t)\{\xi_i(t)\alpha_i(t)Z_i(t)X_i^T(t)-z_x(t)\xi_i(t)\alpha_i(t)X_i(t)X_i^T(t)\} \\
&\quad \{\tilde{\beta}(t;\gamma_0)-\beta_0(t)\}dt \gg_R \\
&\quad -n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2} \ll W_i(t)(\tilde{Z}_x(t)-z_x(t))\xi_i(t)\alpha_i(t)X_i(t)X_i^T(t)(\tilde{\beta}(t;\gamma_0) \\
&\quad -\beta_0(t))dt \gg_R
\end{aligned}$$

$$
\begin{aligned}
&= n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2} R_i W(t)\{\xi_i(t)\alpha_i(t)Z_i(t)X_i^T(t) - z_x(t)\xi_i(t)\alpha_i(t)X_i(t)X_i^T(t)\} \\
&\quad \{\tilde{\beta}(t;\gamma_0) - \beta_0(t)\}dt \\
&\quad + n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2} (1-R_i)\widehat{E}_s[W_i(t)\{\xi_i(t)\alpha_i(t)Z_i(t)X_i^T(t) - z_x(t)\xi_i(t)\alpha_i(t)X_i(t) \\
&\quad X_i^T(t)\}\{\tilde{\beta}(t;\gamma_0) - \beta_0(t)\}dt \mid \mathcal{X}] \\
&\quad - n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2} R_i W(t)(\tilde{Z}_x(t) - z_x(t))\xi_i(t)\alpha_i(t)X_i(t)X_i^T(t)(\tilde{\beta}(t;\gamma_0) - \beta_0(t))dt \\
&\quad - n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2} (1-R_i)\widehat{E}_s[W_i(t)(\tilde{Z}_x(t) - z_x(t))\xi_i(t)\alpha_i(t)X_i(t)X_i^T(t)(\tilde{\beta}(t;\gamma_0) \\
&\quad - \beta_0(t))dt \mid \mathcal{X}].
\end{aligned}
$$

By the $\mathcal{X}$-measurability of the random functions $\tilde{\beta}(\cdot;\gamma_0)$, $\tilde{Z}_x(\cdot)$, $X_i(\cdot)$, $Z_i(\cdot)$ R_i and $\xi_i(\cdot)$, then

$$
\begin{aligned}
\eta_1 &= n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2} R_i W(t)\{\xi_i(t)\alpha_i(t)Z_i(t)X_i^T(t) - z_x(t)\xi_i(t)\alpha_i(t)X_i(t)X_i^T(t)\} \\
&\quad \{\tilde{\beta}(t;\gamma_0) - \beta_0(t)\}dt \\
&\quad + n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2} (1-R_i)W_i(t)\{\xi_i(t)\alpha_i(t)Z_i(t)X_i^T(t) - z_x(t)\xi_i(t)\alpha_i(t)X_i(t) \\
&\quad X_i^T(t)\}\{\tilde{\beta}(t;\gamma_0) - \beta_0(t)\}dt \\
&\quad - n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2} R_i W(t)(\tilde{Z}_x(t) - z_x(t))\xi_i(t)\alpha_i(t)X_i(t)X_i^T(t)(\tilde{\beta}(t;\gamma_0) - \beta_0(t))dt \\
&\quad - n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2} (1-R_i)W_i(t)(\tilde{Z}_x(t) - z_x(t))\xi_i(t)\alpha_i(t)X_i(t)X_i^T(t)(\tilde{\beta}(t;\gamma_0) \\
&\quad - \beta_0(t))dt \\
&= n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2} W_i(t)\{\xi_i(t)\alpha_i(t)Z_i(t)X_i^T(t) - z_x(t)\xi_i(t)\alpha_i(t)X_i(t)X_i^T(t)\} \\
&\quad \{\tilde{\beta}(t;\gamma_0) - \beta_0(t)\}dt \\
&\quad - n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2} W_i(t)(\tilde{Z}_x(t) - z_x(t))\xi_i(t)\alpha_i(t)X_i(t)X_i^T(t)(\tilde{\beta}(t;\gamma_0) - \beta_0(t))dt \\
&= \int_{t_1}^{t_2} W_i(t)n^{-1}\sum_{i=1}^{n}\{\xi_i(t)\alpha_i(t)Z_i(t)X_i^T(t) - z_x(t)\xi_i(t)\alpha_i(t)X_i(t)X_i^T(t)\} \\
&\quad d\left(n^{1/2}\int_{t_1}^{t}(\tilde{\beta}(s;\gamma_0) - \beta_0(s))ds\right)
\end{aligned}
$$

$$-\int_{t_1}^{t_2} W_i(t)(\tilde{Z}_x(t)-z_x(t))n^{-1}\sum_{i=1}^{n}\xi_i(t)\alpha_i(t)X_i(t)X_i^T(t)d\Bigg(n^{1/2}\int_{t_1}^{t}(\tilde{\beta}(s;\gamma_0)$$
$$-\beta_0(s))ds\Bigg).$$

By the consistency of the $\tilde{Z}_x(t)$, the convergence of $W_i(t)$, the application of Lemma A.2.5 and Lemma A.1 of Lin & Ying (2001), and the facts that

$$\begin{aligned}
& n^{-1}\sum_{i=1}^{n}\{\xi_i(t)\alpha_i(t)Z_i(t)X_i^T(t)-z_x(t)\xi_i(t)\alpha_i(t)X_i(t)X_i^T(t)\} \\
\xrightarrow{P} \quad & E\{\xi_i(t)\alpha_i(t)Z_i(t)X_i^T(t)-z_x(t)\xi_i(t)\alpha_i(t)X_i(t)X_i^T(t)\} \\
= \quad & E\{\xi_i(t)\alpha_i(t)Z_i(t)X_i^T(t)\}-z_x(t)E\{\xi_i(t)\alpha_i(t)X_i(t)X_i^T(t)\} \\
= \quad & e_{zx}(t)-z_x(t)e_{xx}(t)=e_{zx}(t)-e_{zx}(t)(e_{xx}(t))^{-1}e_{xx}(t)=0
\end{aligned}$$

and

$$n^{-1}\sum_{i=1}^{n}\xi_i(t)\alpha_i(t)X_i(t)X_i^T(t)\xrightarrow{P}E\{\xi_i(t)\alpha_i(t)X_i(t)X_i^T(t)\}=e_{xx}(t),$$

we have $\eta_1 \xrightarrow{P} 0$.

$$\begin{aligned}
\eta_2 \quad = \quad & n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2}R_iW_i(t)\{Z_i(t)-\tilde{Z}_x(t)X_i(t)\}X_i^T(t)\{\tilde{\beta}(t;\gamma_0)-\beta_0(t)\}dO_i(t) \\
& +n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2}(1-R_i)\widehat{E}_s\{W_i(t)\{Z_i(t)-\tilde{Z}_x(t)X_i(t)\}X_i^T(t)\{\tilde{\beta}(t;\gamma_0) \\
& \quad -\beta_0(t)\}dO_i(t)\mid\mathcal{X}\} \\
= \quad & n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2}[R_iW_i(t)\{Z_i(t)-\tilde{Z}_x(t)X_i(t)\}X_i^T(t)dO_i(t)\{\tilde{\beta}(t;\gamma_0)-\beta_0(t)\}] \\
& +n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2}[(1-R_i)\widehat{E}_s\{W_i(t)\{Z_i(t)-\tilde{Z}_x(t)X_i(t)\}X_i^T(t)dO_i(t)\mid\mathcal{X}\} \\
& \quad \{\tilde{\beta}(t;\gamma_0)-\beta_0(t)\}].
\end{aligned}$$

The first term of η_2

$$n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2}[R_iW_i(t)\{Z_i(t)-\tilde{Z}_x(t)X_i(t)\}X_i^T(t)dO_i(t)\{\tilde{\beta}(t;\gamma_0)-\beta_0(t)\}]$$

$$
\begin{aligned}
&= \int_{t_1}^{t_2} W_i(t) n^{-1/2} \sum_{i=1}^{n} R_i\{Z_i(t) - \tilde{Z}_x(t) X_i(t)\} X_i^T(t) dO_i(t) \{\tilde{\beta}(t;\gamma_0) - \beta_0(t)\} \\
&= \int_{t_1}^{t_2} W_i(t) n^{-1/2} \sum_{i=1}^{n} R_i\{Z_i(t) - z_x(t) X_i(t)\} X_i^T(t) dO_i(t) \{\tilde{\beta}(t;\gamma_0) - \beta_0(t)\} \\
&\quad - \int_{t_1}^{t_2} W_i(t) n^{-1/2} \sum_{i=1}^{n} R_i\{\tilde{Z}_x(t) - z_x(t)\} X_i(t) X_i^T(t) dO_i(t) \{\tilde{\beta}(t;\gamma_0) - \beta_0(t)\} \\
&= \int_{t_1}^{t_2} d\Big(n^{-1/2} \sum_{i=1}^{n} \int_{t_1}^{t} R_i\{Z_i(s) - z_x(s) X_i(s)\} X_i^T(s) dO_i(s) \Big) W_i(t) \{\tilde{\beta}(t;\gamma_0) \\
&\qquad -\beta_0(t)\} \\
&\quad - \int_{t_1}^{t_2} \{\tilde{Z}_x(t) - z_x(t)\} d\Big(n^{-1/2} \sum_{i=1}^{n} \int_{t_1}^{t} R_i X_i(s) X_i^T(s) dO_i(s) \Big) W_i(t) \{\tilde{\beta}(t;\gamma_0) \\
&\qquad -\beta_0(t)\}.
\end{aligned}
$$

Under the condition (I) and by Lemma 1 of Sun & Wu (2005), both

$$
n^{-1/2} \sum_{i=1}^{n} \int_{t_1}^{t} R_i\{Z_i(s) - z_x(s) X_i(s)\} X_i^T(s) dO_i(s)
$$

and

$$
n^{-1/2} \sum_{i=1}^{n} \int_{t_1}^{t} R_i X_i(s) X_i^T(s) dO_i(s)
$$

converge weakly to vectors of mean zero Gaussian processes with continuous sample paths respectively. And from the early derivation, $W_i(t)\{\tilde{\beta}(t;\gamma_0) - \beta_0(t)\}$ and $\tilde{Z}_x(t) - z_x(t)$ are of bounded variations and both converge to zero in probability uniformly in t. Hence by Lemma A.1 of Lin & Ying (2001), the first term converges to zero in probability.

As the second term of η_2

$$
\begin{aligned}
& n^{-1/2} \sum_{i=1}^{n} \int_{t_1}^{t_2} [(1 - R_i) \widehat{E}_s\{W_i(t)\{Z_i(t) - \tilde{Z}_x(t) X_i(t)\} X_i^T(t) dO_i(t) \mid \mathcal{X}\} \{\tilde{\beta}(t;\gamma_0) \\
&\qquad -\beta_0(t)\}] \\
= \; & n^{-1/2} \sum_{i=1}^{n} \int_{t_1}^{t_2} [(1 - R_i) \widehat{E}_s\{W_i(t)\{Z_i(t) - z_x(t) X_i(t)\} X_i^T(t) dO_i(t) \mid \mathcal{X}\} \{\tilde{\beta}(t;\gamma_0) \\
&\qquad -\beta_0(t)\}]
\end{aligned}
$$

$$
\begin{aligned}
& -n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2}[(1-R_i)\widehat{E}_s\{W_i(t)\{\tilde{Z}_x(t)-z_x(t)\}X_i(t)X_i^T(t)dO_i(t)\mid \mathcal{X}\}\{\tilde{\beta}(t;\gamma_0) \\
& \quad -\beta_0(t)\}] \\
= \ & n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2}[(1-R_i)\widehat{E}_s\{W_i(t)\{Z_i(t)-z_x(t)X_i(t)\}X_i^T(t)dO_i(t)\mid \mathcal{X}_i\}\{\tilde{\beta}(t;\gamma_0) \\
& \quad -\beta_0(t)\}] \\
& -\widehat{E}_s\{n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2}W_i(t)(1-R_i)\{\tilde{Z}_x(t)-z_x(t)\}X_i(t)X_i^T(t)dO_i(t)\{\tilde{\beta}(t;\gamma_0) \\
& \quad -\beta_0(t)\}\mid \mathcal{X}\} \\
= \ & n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2}[(1-R_i)\widehat{E}_s\{W_i(t)\{Z_i(t)-z_x(t)X_i(t)\}X_i^T(t)dO_i(t)\mid \mathcal{D}_i, R_i=0\} \\
& \quad \{\tilde{\beta}(t;\gamma_0)-\beta_0(t)\}] \\
& -\widehat{E}_s\{\int_{t_1}^{t_2}W_i(t)\{\tilde{Z}_x(t)-z_x(t)\}d\Bigg(n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t}(1-R_i)X_i(u)X_i^T(u)dO_i(u)\Bigg) \\
& \quad \{\tilde{\beta}(t;\gamma_0)-\beta_0(t)\}\mid \mathcal{X}\},
\end{aligned}
$$

also by Lemma 1 of Sun & Wu (2005) $n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t}(1-R_i)X_i(u)X_i^T(u)dO_i(u)$ converges weakly to a vector of mean zero Gaussian processes with continuous sample paths. Then from the early derivation, $W_i(t)\{\tilde{\beta}(t;\gamma_0)-\beta_0(t)\}$ is of bounded variations and converges to zero in probability uniformly in t. Hence by Lemma A.1 of Lin & Ying (2001),

$$
\int_{t_1}^{t_2}W_i(t)\{\tilde{Z}_x(t)-z_x(t)\}d\Bigg(n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t}(1-R_i)X_i(u)X_i^T(u)dO_i(u)\Bigg)\xrightarrow{P}0.
$$

Also using the similar argument in Lemma A.2.1, the second term of η_2 equals to

$$
\begin{aligned}
& n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2}[(1-R_i)E_s\{W_i(t)\{Z_i(t)-z_x(t)X_i(t)\}X_i^T(t)dO_i(t)\mid \mathcal{D}_i, R_i=0\} \\
& \quad \{\tilde{\beta}(t;\gamma_0)-\beta_0(t)\}]+o_p(1) \\
= \ & \int_{t_1}^{t_2}\Bigg[n^{-1/2}\sum_{i=1}^{n}(1-R_i)\{Z_i(t)-z_x(t)X_i(t)\}X_i^T(t)E_s\{dO_i(t)\mid \mathcal{D}_i, R_i=0\}W_i(t) \\
& \quad \{\tilde{\beta}(t;\gamma_0)-\beta_0(t)\}\Bigg]+o_p(1) \\
= \ & \int_{t_1}^{t_2}\Bigg[d\Bigg(n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t}(1-R_i)\{Z_i(u)-z_x(u)X_i(u)\}X_i^T(u)E_s\{dO_i(u)\mid \mathcal{D}_i,
\end{aligned}
$$

$$R_i = 0\}\Big) W_i(t)\{\tilde{\beta}(t;\gamma_0) - \beta_0(t)\}\Big] + o_p(1)$$

Now apply Lemma 1 of Sun & Wu (2005) again.

$$n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t}(1-R_i)\{Z_i(u) - z_x(u)X_i(u)\}X_i^T(u)E_s\{dO_i(u) \mid \mathcal{D}_i, R_i = 0\}$$

converges weakly to a vector of mean zero Gaussian processes with continuous sample paths. Also from the early derivation, $W_i(t)\{\tilde{\beta}(t;\gamma_0) - \beta_0(t)\}$ is of bounded variations and converges to zero in probability uniformly in t. Hence by Lemma A.1 of Lin & Ying (2001) the second term of $\eta_2 \xrightarrow{P} 0$. Then $\eta = \eta_1 + \eta_2 \xrightarrow{P} 0$. Thus $n^{-1/2}U(\gamma_0)$ equals

$$n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2} \ll W_i(t)\{Z_i(t) - \tilde{Z}_x(t)X_i(t)\}\{dM_i(t) - \beta_0^T(t)X_i(t)dO_i(t)\} \gg_R .$$

Since

$$\begin{aligned}
& dM_i(t) - \beta_0^T(t)X_i(t)dO_i(t) \\
= \; & [Y_i(t) - \gamma_0^T Z_i(t)]dN_i^c(t) - \xi_i(t)\alpha_i(t)X_i^T(t)d\mathcal{B}(t) - \beta_0^T(t)X_i(t)dN_i^c(t) \\
& \quad +\beta_0^T(t)X_i(t)\xi_i(t)\alpha_i(t)dt \\
= \; & [Y_i(t) - \gamma_0^T Z_i(t) - \beta_0^T(t)X_i(t)]dN_i^c(t) - \xi_i(t)\alpha_i(t)X_i^T(t)\beta_0(t)d(t) \\
& \quad +\beta_0^T(t)X_i(t)\xi_i(t)\alpha_i(t)dt \\
= \; & \epsilon_i(t)dN_i^c(t),
\end{aligned}$$

$$\begin{aligned}
n^{-1/2}U(\gamma_0) \;=\; & n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2} \ll W_i(t)\{Z_i(t) - \tilde{Z}_x(t)X_i(t)\}\epsilon_i(t)dN_i^c(t) \gg_R \\
=\; & n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2} R_i W(t)\{Z_i(t) - \tilde{Z}_x(t)X_i(t)\}\epsilon_i(t)dN_i^c(t) \qquad \text{(A.11)} \\
& +n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2}(1-R_i)\widehat{E}_s\{W_i(t)\{Z_i(t) \\
& \quad -\tilde{Z}_x(t)X_i(t)\}\epsilon_i(t)dN_i^c(t) \mid \mathcal{X}\}. \qquad \text{(A.12)}
\end{aligned}$$

$$\begin{aligned}
(A.11) \quad = \quad & n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2} R_i W(t)\{Z_i(t) - z_x(t)X_i(t)\}\epsilon_i(t)dN_i^c(t) \\
& -n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2} R_i W(t)\{\tilde{Z}_x(t) - z_x(t)\}X_i(t)\epsilon_i(t)dN_i^c(t) \\
= \quad & n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2} R_i W(t)\{Z_i(t) - z_x(t)X_i(t)\}\epsilon_i(t)dN_i^c(t) \\
& -\int_{t_1}^{t_2} W_i(t)\{\tilde{Z}_x(t) - z_x(t)\}n^{-1/2}\sum_{i=1}^{n} R_i X_i(t)[dM_i(t) - \beta_0^T(t)X_i(t)dO_i(t)] \\
= \quad & n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2} R_i W(t)\{Z_i(t) - z_x(t)X_i(t)\}\epsilon_i(t)dN_i^c(t) \\
& -\int_{t_1}^{t_2} W_i(t)\{\tilde{Z}_x(t) - z_x(t)\}d\Big(n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t} R_i X_i(u)dM_i(u)\Big) \\
& +\int_{t_1}^{t_2} W_i(t)\{\tilde{Z}_x(t) - z_x(t)\}n^{-1/2}\sum_{i=1}^{n} R_i X_i(t)X_i^T(t)dO_i(t)\beta_0(t) \\
= \quad & n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2} R_i W(t)\{Z_i(t) - z_x(t)X_i(t)\}\epsilon_i(t)dN_i^c(t) \\
& -\int_{t_1}^{t_2} W_i(t)\{\tilde{Z}_x(t) - z_x(t)\}d\Big(n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t} R_i X_i(u)dM_i(u)\Big) \\
& +\int_{t_1}^{t_2} W_i(t)\{\tilde{Z}_x(t) - z_x(t)\}d\Big(n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t} R_i X_i(u)X_i^T(u)dO_i(u)\Big)\beta_0(t) \\
= \quad & n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2} R_i W(t)\{Z_i(t) - z_x(t)X_i(t)\}\epsilon_i(t)dN_i^c(t) + o_p(1).
\end{aligned}$$

The last equality holds because of the joint weak convergence of

$$\Big(n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t} R_i X_i(u)dM_i(u),\ n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t} R_i X_i(u)X_i^T(u)dO_i(u)\Big)$$

by Lemma 1 of Sun & Wu (2005), the consistency of $W_i(t)\{\tilde{Z}_x(t) - z_x(t)\}$ and Lemma A.1 of Lin & Ying (2001).

$$\begin{aligned}
(A.12) \quad = \quad & n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2}(1-R_i)\widehat{E}_s\{W_i(t)\{Z_i(t) - \tilde{Z}_x(t)X_i(t)\}\epsilon_i(t)dN_i^c(t) \mid \mathcal{X}\} \\
= \quad & n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2}(1-R_i)\widehat{E}_s\{W_i(t)\{Z_i(t) - z_x(t)X_i(t)\}\epsilon_i(t)dN_i^c(t) \mid \mathcal{X}\}
\end{aligned}$$

$$
\begin{aligned}
& -n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2}(1-R_i)\widehat{E}_s\{W_i(t)\{\tilde{Z}_x(t)-z_x(t)\}X_i(t)[dM_i(t) \\
& \quad -\beta_0^T(t)X_i(t)dO_i(t)] \mid \mathcal{X}\} \\
= \; & n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2}(1-R_i)\widehat{E}_s\{W_i(t)\{Z_i(t)-z_x(t)X_i(t)\}\epsilon_i(t)dN_i^c(t) \mid \mathcal{X}_i\} \\
& -\widehat{E}_s\{\int_{t_1}^{t_2} W_i(t)\{\tilde{Z}_x(t)-z_x(t)\}d\Bigg(n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t}(1-R_i)X_i(u)dM_i(t)\Bigg)|\mathcal{X}\} \\
& +\widehat{E}_s\{\int_{t_1}^{t_2} W_i(t)\{\tilde{Z}_x(t)-z_x(t)\}d\Bigg(n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t}(1-R_i)X_i(u)X_i^T(t) \\
& \quad dO_i(t)\Bigg)\beta_0^T(t) \mid \mathcal{X}\} \\
= \; & n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2}(1-R_i)\widehat{E}_s\{W_i(t)\{Z_i(t)-z_x(t)X_i(t)\}\epsilon_i(t)dN_i^c(t) \mid \mathcal{D}_i, \\
& \quad R_i=0\} + o_p(1).
\end{aligned}
$$

The last equality holds also because of the weak convergence of

$$
n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t}(1-R_i)X_i(u)dM_i(u) \;\; and \;\; n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t}(1-R_i)X_i(u)X_i^T(u)dO_i(u)
$$

by Lemma 1 of Sun & Wu (2005), the consistency of $W_i(t)\{\tilde{Z}_x(t)-z_x(t)\}$ and Lemma A.1 of Lin & Ying (2001). Similarly the $W_i(t)$ can be replaced by its limit $w(t)$. Then

$$
\begin{aligned}
(A.12) \; = \; & n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2}(1-R_i)E_s\{w(t)\{Z_i(t)-z_x(t)X_i(t)\}\epsilon_i(t)dN_i^c(t) \mid \mathcal{D}_i, \\
& \quad R_i=0\} \\
& +n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2}(1-R_i)\widehat{E}_s\{w(t)\{Z_i(t)-z_x(t)X_i(t)\}\epsilon_i(t)dN_i^c(t) \mid \mathcal{D}_i, \\
& \quad R_i=0\} \\
& -n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2}(1-R_i)E_s\{w(t)\{Z_i(t)-z_x(t)X_i(t)\}\epsilon_i(t)dN_i^c(t) \mid \mathcal{D}_i, \\
& \quad R_i=0\} + o_p(1) \\
= \; & n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2}(1-R_i)E_s\{w(t)\{Z_i(t)-z_x(t)X_i(t)\}\epsilon_i(t)dN_i^c(t) \mid \mathcal{D}_i, \\
& \quad R_i=0\}
\end{aligned}
$$

$$+n^{-1/2}\sum_{i=1}^{n}(1-R_i)\int_0^L\sum_{j=1}^{n_i}I(t_1\le s+T_{ij}\le t_2)w(s+T_{ij})\{Z_{ij}$$

$$-z_x(s+T_{ij})X_{ij}\}\epsilon_i(s+T_{ij})I(C_i\ge T_{ij})\left[\frac{d\widehat{F}_s(s)}{\widehat{F}_s(V_i)}-\frac{dF_s(s)}{F_s(V_i)}\right] \quad (A.13)$$

$$+o_p(1)$$

Referring to the argument in Lemma A.2.4, (A.13) has the following decomposition.

$$\begin{aligned}
(A.13) \quad = \quad & n^{-1/2}\int_0^\infty\int_0^L E\Big\{\int_{t_1}^{t_2}(1-R_i)w(v)(Z_i(v)-z_x(v)X_i(v))\epsilon_i(v)dN_i^*(v-s) \\
& \frac{I(x<(L-(V_i)))}{F_s(V_i)}\Big\}dF_s(s)\frac{dM^R(x)}{y^R(x)} \\
& +n^{-1/2}\int_0^L\int_0^{(L-x)-} E\Big\{\int_{t_1}^{t_2}(1-R_i)w(v)(Z_i(v)-z_x(v)X_i(v))\epsilon_i(v) \\
& \frac{dN_i^*(v-s)}{F_s(V_i)}\Big\}dF_s(s)\frac{dM^R(x)}{y^R(x)} \\
& +n^{-1/2}\int_0^L E\Big\{\int_{t_1}^{t_2}(1-R_i)w(v)(Z_i(v)-z_x(v)X_i(v))\epsilon_i(v)\frac{dN_i^*(v-s)}{F_s(V_i)}\Big\} \\
& F_s(s)\frac{dM^R(L-s)-}{y^R(L-s)-}+o_p(1) \\
= \quad & n^{-1/2}\int_0^\infty\int_0^L E\Big\{E\Big[\int_{t_1}^{t_2}(1-R_i)w(v)(Z_i(v)-z_x(v)X_i(v))\epsilon_i(v) \\
& dN_i^*(v-s)\frac{I(x<(L-(V_i)))}{F_s(V_i)}\Big|X_i(\cdot),Z_i(\cdot),N_i(\cdot),S_i,V_i,C_i\Big]\Big\} \\
& dF_s(s)\frac{dM^R(x)}{y^R(x)} \\
& +n^{-1/2}\int_0^L\int_0^{(L-x)-} E\Big\{E\Big[\int_{t_1}^{t_2}(1-R_i)w(v)(Z_i(v)-z_x(v)X_i(v))\epsilon_i(v) \\
& \frac{dN_i^*(v-s)}{F_s(V_i)}\Big|X_i(\cdot),Z_i(\cdot),N_i(\cdot),S_i,V_i,C_i\Big]\Big\}dF_s(s)\frac{dM^R(x)}{y^R(x)} \\
& +n^{-1/2}\int_0^L E\Big\{E\Big[\int_{t_1}^{t_2}(1-R_i)w(v)(Z_i(v)-z_x(v)X_i(v))\epsilon_i(v)\frac{dN_i^*(v-s)}{F_s(V_i)} \\
& \Big|X_i(\cdot),Z_i(\cdot),N_i(\cdot),S_i,V_i,C_i\Big]\Big\}F_s(s)\frac{dM^R(L-s)-}{y^R(L-s)-}+o_p(1) \\
= \quad & n^{-1/2}\int_0^\infty\int_0^L E\Big\{\int_{t_1}^{t_2}(1-R_i)w(v)(Z_i(v)-z_x(v)X_i(v))E[\epsilon_i(v)\mid X_i(\cdot), \\
& Z_i(\cdot),N_i(\cdot),S_i,V_i,C_i]dN_i^*(v-s)\frac{I(x<(L-(V_i)))}{F_s(V_i)}\Big\}dF_s(s)
\end{aligned}$$

$$
\begin{aligned}
&\frac{dM^R(x)}{y^R(x)} \\
&+n^{-1/2}\int_0^L\int_0^{(L-x)-} E\Big\{\int_{t_1}^{t_2}(1-R_i)w(v)(Z_i(v)-z_x(v)X_i(v))E[\epsilon_i(v) \\
&\quad \mid X_i(\cdot),Z_i(\cdot),N_i(\cdot),S_i,V_i,C_i]\frac{dN_i^*(v-s)}{F_s(V_i)}\Big\}dF_s(s)\frac{dM^R(x)}{y^R(x)} \\
&+n^{-1/2}\int_0^L E\Big\{\int_{t_1}^{t_2}(1-R_i)w(v)(Z_i(v)-z_x(v)X_i(v))E[\epsilon_i(v)\mid X_i(\cdot), \\
&\quad Z_i(\cdot),N_i(\cdot),S_i,V_i,C_i]\frac{dN_i^*(v-s)}{F_s(V_i)}\Big\}F_s(s)\frac{dM^R(L-s)-}{y^R(L-s)-}+o_p(1).
\end{aligned}
$$

Under the assumption that $E\{Y_i(t)|X_i(\cdot),Z_i(\cdot),N_i(\cdot),S_i,V_i,C_i\} = E\{Y_i(t)|X_i(\cdot),Z_i(\cdot)\}$,

$$
E[\epsilon_i(v)\mid X_i(\cdot),Z_i(\cdot),N_i(\cdot),S_i,V_i,C_i] = E[\epsilon_i(v)\mid X_i(\cdot),Z_i(\cdot)] = 0.
$$

Then $(A.13) = 0 + o_p(1) \xrightarrow{P} 0$. Hence

$$
\begin{aligned}
n^{-1/2}U(\gamma_0) &= n^{-1/2}\sum_{i=1}^n\int_{t_1}^{t_2} R_i w(t)\{Z_i(t)-z_x(t)X_i(t)\}\epsilon_i(t)dN_i^c(t) \\
&\quad +n^{-1/2}\sum_{i=1}^n\int_{t_1}^{t_2}(1-R_i)E_s\{w(t)\{Z_i(t)-z_x(t)X_i(t)\}\epsilon_i(t)dN_i^c(t)\mid \mathcal{D}_i, \\
&\qquad R_i=0\}+o_p(1) \\
&= n^{-1/2}\sum_{i=1}^n\int_{t_1}^{t_2} w(t)\{Z_i(t)-z_x(t)X_i(t)\}\epsilon_i(t)R_i dN_i^c(t) \\
&\quad +n^{-1/2}\sum_{i=1}^n\int_{t_1}^{t_2} w(t)\{Z_i(t)-z_x(t)X_i(t)\}\epsilon_i(t)E_s\{(1-R_i)dN_i^c(t)\mid \mathcal{D}_i, \\
&\qquad R_i=0\}+o_p(1) \\
&= n^{-1/2}\sum_{i=1}^n\int_{t_1}^{t_2} w(t)\{Z_i(t)-z_x(t)X_i(t)\}\epsilon_i(t)[R_i dN_i^c(t) \\
&\qquad +E_s\{(1-R_i)dN_i^c(t)\mid \mathcal{D}_i,R_i=0\}]+o_p(1).
\end{aligned}
$$

Applying theorem 5.21 (van der Vaart, 1998) to the score function, (A.8) becomes

$$
n^{1/2}(\widehat{\gamma}-\gamma_0) = D^{-1}[n^{-1/2}U(\gamma_0)]+o_p(1).
$$

Hence $n^{1/2}(\widehat{\gamma}-\gamma_0) \xrightarrow{D} \mathcal{N}(0, D^{-1}VD^{-1})$. □

Proof of Theorem 3.4

By the definitions, we have

$$\begin{aligned}
\tilde{\beta}(t;\gamma_0) - \beta^*(t) &= \tilde{Y}_x^T(t) - \tilde{Z}_x^T(t)\gamma_0 - [\tilde{y}_x^T(t) - \tilde{z}_x^T(t)\gamma_0] \\
&= (\tilde{E}_{xx}(t))^{-1}\tilde{E}_{xy}(t) - (\tilde{E}_{xx}(t))^{-1}\tilde{E}_{xz}(t)\gamma_0 - (\tilde{e}_{xx}(t))^{-1}\tilde{e}_{xy}(t) \\
&\quad +(\tilde{e}_{xx}(t))^{-1}\tilde{e}_{xz}(t)\gamma_0 \\
&= (\tilde{E}_{xx}(t))^{-1}[(\tilde{E}_{xy}(t) - \tilde{e}_{xy}(t)) - (\tilde{E}_{xz}(t) - \tilde{e}_{xz}(t))\gamma_0] \\
&\quad -(\tilde{e}_{xx}(t))^{-1}[\tilde{E}_{xx}(t) - \tilde{e}_{xx}(t)](\tilde{E}_{xx}(t))^{-1}[\tilde{e}_{xy}(t) - \tilde{e}_{xz}(t)\gamma_0] \\
&= (e_{xx}(t))^{-1}[(\tilde{E}_{xy}(t) - \tilde{e}_{xy}(t)) - (\tilde{E}_{xz}(t) - \tilde{e}_{xz}(t))\gamma_0] \\
&\quad -(e_{xx}(t))^{-1}[\tilde{E}_{xx}(t) - \tilde{e}_{xx}(t)](e_{xx}(t))^{-1}[e_{xy}(t) - e_{xz}(t)\gamma_0] + o_p(1).
\end{aligned}$$

The last equality holds by Slutsky's theorem. Then

$$\begin{aligned}
&\tilde{\beta}(t;\gamma_0) - \beta^*(t) \\
= \ & (e_{xx}(t))^{-1}[(\tilde{E}_{xy}(t) - \tilde{e}_{xy}(t)) - (\tilde{E}_{xz}(t) - \tilde{e}_{xz}(t))\gamma_0] \\
& -(e_{xx}(t))^{-1}[\tilde{E}_{xx}(t) - \tilde{e}_{xx}(t)][y_x^T(t) - z_x^T(t)\gamma_0] + o_p(1) \\
= \ & (e_{xx}(t))^{-1}[(\tilde{E}_{xy}(t) - \tilde{e}_{xy}(t)) - (\tilde{E}_{xz}(t) - \tilde{e}_{xz}(t))\gamma_0] \\
& -(e_{xx}(t))^{-1}[\tilde{E}_{xx}(t) - \tilde{e}_{xx}(t)]\beta_0(t) + o_p(1) \\
= \ & (e_{xx}(t))^{-1}[(\tilde{E}_{xy}(t) - \tilde{e}_{xy}(t)) - (\tilde{E}_{xz}(t) - \tilde{e}_{xz}(t))\gamma_0 - (\tilde{E}_{xx}(t) - \tilde{e}_{xx}(t))\beta_0(t)] \\
& +o_p(1) \\
= \ & (e_{xx}(t))^{-1}\Bigg(n^{-1}\sum_{i=1}^{n} R_i \int_0^\tau K_h(u-t)X_i(u)[Y_i(u) - Z_i^T(u)\gamma_0 \\
& \qquad -X_i^T(u)\beta_0(u)]dN_i^c(u) \\
& \quad +n^{-1}\sum_{i=1}^{n}(1-R_i)\widehat{E}_s\Bigg\{\int_0^\tau K_h(u-t)X_i(u)[Y_i(u) - Z_i^T(u)\gamma_0 \\
& \qquad -X_i^T(u)\beta_0(u)]dN_i^c(u) \mid \mathcal{X}\Bigg\} \\
& \quad -\int_0^\tau K_h(u-t)E\{\xi_i(u)\alpha_i(u)X_i(u)[Y_i(u) - Z_i^T(u)\gamma_0
\end{aligned}$$

$$
\begin{aligned}
& \qquad -X_i^T(u)\beta_0(u)]\}du\Big) \\
& -(e_{xx}(t))^{-1}\Big(n^{-1}\sum_{i=1}^n R_i\int_0^\tau K_h(u-t)X_i(u)X_i^T(u)[\beta_0(t)-\beta_0(u)]dN_i^c(u) \\
& \quad +n^{-1}\sum_{i=1}^n(1-R_i)\widehat{E}_s\Big\{\int_0^\tau K_h(u-t)X_i(u)X_i^T(u)[\beta_0(t)-\beta_0(u)] \\
& \qquad dN_i^c(u)\mid\mathcal{X}\Big\} \\
& \quad -\int_0^\tau K_h(u-t)E\{\xi_(u)\alpha_i(u)X_i(u)X_i^T(u)\}[\beta_0(t)-\beta_0(u)]du\Big)+o_p(1) \\
= \ & (e_{xx}(t))^{-1}\Big(n^{-1}\sum_{i=1}^n R_i\int_0^\tau K_h(u-t)X_i(u)\epsilon_i(u)dN_i^c(u) \\
& \quad +n^{-1}\sum_{i=1}^n(1-R_i)\widehat{E}_s\Big\{\int_0^\tau K_h(u-t)X_i(u)\epsilon_i(u)dN_i^c(u)\mid\mathcal{X}\Big\} \\
& \quad -\int_0^\tau K_h(u-t)E\{\xi_i(u)\alpha_i(u)X_i(u)\epsilon_i(u)\}du\Big) \\
& -(e_{xx}(t))^{-1}\Big(\int_0^\tau K_h(u-t)d\Big[n^{-1}\sum_{i=1}^n\int_0^u R_iX_i(w)X_i^T(w)dN_i^c(w)\Big][\beta_0(t)-\beta_0(u)] \\
& \quad +\widehat{E}_s\Big\{\int_0^\tau K_h(u-t)d\Big[n^{-1}\sum_{i=1}^n\int_0^u(1-R_i)X_i(w)X_i^T(w)dN_i^c(w)\Big] \\
& \qquad [\beta_0(t)-\beta_0(u)]\mid\mathcal{X}\Big\} \\
& \quad -\int_0^\tau K_h(u-t)E\{\xi_(u)\alpha_i(u)X_i(u)X_i^T(u)\}[\beta_0(t)-\beta_0(u)]du\Big)+o_p(1)
\end{aligned}
$$

We know that

$$
\begin{aligned}
& \int_0^\tau K_h(u-t)E\{\xi_i(u)\alpha_i(u)X_i(u)\epsilon_i(u)\}du \\
= \ & \int_0^\tau K_h(u-t)E\{E[\xi_i(u)\alpha_i(u)X_i(u)\epsilon_i(u)\mid X_i(\cdot),Z_i(\cdot)]\}du \\
= \ & \int_0^\tau K_h(u-t)E\{\xi_i(u)\alpha_i(u)X_i(u)E[\epsilon_i(u)\mid X_i(\cdot),Z_i(\cdot)]\}du=0.
\end{aligned}
$$

Therefore,

$$
(nh)^{1/2}(\tilde{\beta}(t;\gamma_0)-\beta^*(t))
$$

$$
\begin{aligned}
= \quad & (nh)^{1/2}(e_{xx}(t))^{-1}\Bigg(n^{-1}\sum_{i=1}^{n} \ll \int_0^\tau K_h(u-t)X_i(u)\epsilon_i(u)dN_i^c(u) \gg_R \Bigg) \\
& -(e_{xx}(t))^{-1}\Bigg(\int_0^\tau h^{1/2}K_h(u-t)d\Bigg[n^{-1/2}\sum_{i=1}^{n}\int_0^u R_iX_i(w)X_i^T(w)dN_i^c(w)\Bigg][\beta_0(t) \\
& \qquad -\beta_0(u)] \\
& \qquad +\widehat{E}_s\Bigg\{\int_0^\tau h^{1/2}K_h(u-t)d\Bigg[n^{-1/2}\sum_{i=1}^{n}\int_0^u(1-R_i)X_i(w)X_i^T(w)dN_i^c(w)\Bigg] \\
& \qquad [\beta_0(t)-\beta_0(u)] \mid \mathcal{X}\Bigg\} \\
& \qquad -\int_0^\tau (nh)^{1/2}K_h(u-t)E\{\xi_(u)\alpha_i(u)X_i(u)X_i^T(u)\}[\beta_0(t)-\beta_0(u)]du\Bigg) \\
& +o_p(1),
\end{aligned}
$$

Applying the substitution $x = \frac{u-t}{h}$,

$$
\begin{aligned}
& \int_{-1}^{1} h^{1/2}K(u-t)d\Bigg[n^{-1/2}\sum_{i=1}^{n}\int_0^u R_iX_i(w)X_i^T(w)dN_i^c(w)\Bigg][\beta_0(t)-\beta_0(u)] \\
= \quad & \int_{-1}^{1} h^{1/2}K(x)d\Bigg[n^{-1/2}\sum_{i=1}^{n}\int_0^{x+th} R_iX_i(w)X_i^T(w)dN_i^c(w)\Bigg]\frac{\beta_0(t)-\beta_0(t+xh)}{h} \\
= \quad & -\int_{-1}^{1} h^{1/2}K(x)d\Bigg[n^{-1/2}\sum_{i=1}^{n}\int_0^{x+th} R_iX_i(w)X_i^T(w)dN_i^c(w)\Bigg][x\beta_0'(t)+O(h)] \\
\xrightarrow{P} \quad & 0
\end{aligned}
$$

since $n^{-1/2}\sum_{i=1}^{n}\int_0^{x+th} R_iX_i(w)X_i^T(w)dN_i^c(w)$ converges weakly as $h \to 0$ and $n \to \infty$. Similarly,

$$
\int_0^\tau h^{1/2}K_h(u-t)d\Bigg[n^{-1/2}\sum_{i=1}^{n}\int_0^u(1-R_i)X_i(w)X_i^T(w)dN_i^c(w)\Bigg][\beta_0(t)-\beta_0(u)] \xrightarrow{P} 0
$$

as $h \to 0$ and $n \to \infty$. And

$$
\begin{aligned}
& \int_0^\tau (nh)^{1/2}K_h(u-t)E\{\xi_(u)\alpha_i(u)X_i(u)X_i^T(u)\}[\beta_0(t)-\beta_0(u)]du \\
= \quad & \int_{-1}^{1}(nh)^{1/2}K(x)e_{xx}(t+xh)[\beta_0(t)-\beta_0(t+xh)]dx \\
= \quad & -\int_{-1}^{1}(nh)^{1/2}K(x)e_{xx}(t+xh)[xh\beta_0'(t)+(1/2)x^2h^2\beta_0''(t)+o(h^2)]dx
\end{aligned}
$$

$$
\begin{aligned}
&= -\int_{-1}^{1}(nh)^{1/2}K(x)[e_{xx}(t)+xhe'_{xx}(t)+(1/2)x^2h^2e''_{xx}(t)+o(h^2)][xh\beta_0'(t)\\
&\qquad +(1/2)x^2h^2\beta_0''(t)+o(h^2)]dx\\
&= -(nh)^{1/2}\int_{-1}^{1}K(x)[e_{xx}(t)xh\beta_0'(t)+x^2h^2e'_{xx}(t)\beta_0'(t)+(1/2)x^2h^2e_{xx}(t)\beta_0''(t)\\
&\qquad +o(h^2)]dx\\
&= -(nh^3)^{1/2}\int_{-1}^{1}xK(x)dxe_{xx}(t)\beta_0'(t)-(nh^5)^{1/2}[e'_{xx}(t)\beta_0'(t)\\
&\qquad +(1/2)e_{xx}(t)\beta_0''(t)]\int_{-1}^{1}x^2K(x)dx+o_p((nh^5)^{1/2})\\
&= -0-(nh^5)^{1/2}[e'_{xx}(t)\beta_0'(t)+(1/2)e_{xx}(t)\beta_0''(t)]\int_{-1}^{1}x^2K(x)dx+o_p((nh^5)^{1/2})\\
&= -(nh^5)^{1/2}[e'_{xx}(t)\beta_0'(t)+(1/2)e_{xx}(t)\beta_0''(t)]\int_{-1}^{1}x^2K(x)dx+o_p((nh^5)^{1/2})
\end{aligned}
$$

as $nh^5 = O(1)$. Thus

$$
\begin{aligned}
&(nh)^{1/2}\bigg(\tilde{\beta}(t;\gamma_0)-\beta^*(t)+h^2(e_{xx}(t))^{-1}[e'_{xx}(t)\beta_0'(t)\\
&\qquad +(1/2)e_{xx}(t)\beta_0''(t)]\int_{-1}^{1}x^2K(x)dx\bigg)\\
&= (nh)^{1/2}(e_{xx}(t))^{-1}\bigg(n^{-1}\sum_{i=1}^{n}\ll\int_0^\tau K_h(u-t)X_i(u)\epsilon_i(u)dN_i^c(u)\gg_R\bigg) \qquad \text{(A.14)}\\
&= (nh)^{1/2}(e_{xx}(t))^{-1}\bigg(n^{-1}\sum_{i=1}^{n}R_i\int_0^\tau K_h(u-t)X_i(u)\epsilon_i(u)dN_i^c(u)\\
&\qquad +n^{-1}\sum_{i=1}^{n}(1-R_i)\widehat{E}_s\Big\{\int_0^\tau K_h(u-t)X_i(u)\epsilon_i(u)dN_i^c(u)\mid\mathcal{X}\Big\}\bigg)\\
&= (nh)^{1/2}(e_{xx}(t))^{-1}\bigg(n^{-1}\sum_{i=1}^{n}R_i\int_0^\tau K_h(u-t)X_i(u)\epsilon_i(u)dN_i^c(u)\\
&\qquad +n^{-1}\sum_{i=1}^{n}(1-R_i)\widehat{E}_s\Big\{\int_0^\tau K_h(u-t)X_i(u)\epsilon_i(u)dN_i^c(u)\mid\mathcal{X}_i\Big\}\bigg)\\
&= (nh)^{1/2}(e_{xx}(t))^{-1}\bigg(n^{-1}\sum_{i=1}^{n}R_i\int_0^\tau K_h(u-t)X_i(u)\epsilon_i(u)dN_i^c(u)\\
&\qquad +n^{-1}\sum_{i=1}^{n}(1-R_i)\widehat{E}_s\Big\{\int_0^\tau K_h(u-t)X_i(u)\epsilon_i(u)dN_i^c(u)\mid\mathcal{D}_i,R_i=0\Big\}\bigg)\\
&= (nh)^{1/2}(e_{xx}(t))^{-1}\bigg(n^{-1}\sum_{i=1}^{n}R_i\int_0^\tau K_h(u-t)X_i(u)\epsilon_i(u)dN_i^c(u)
\end{aligned}
$$

$$
\begin{aligned}
&\quad +n^{-1}\sum_{i=1}^{n}(1-R_i)E_s\Big\{\int_0^{\tau} K_h(u-t)X_i(u)\epsilon_i(u)dN_i^c(u)\mid \mathcal{D}_i, R_i=0\Big\} \\
&\quad +n^{-1}\sum_{i=1}^{n}(1-R_i)\widehat{E}_s\Big\{\int_0^{\tau} K_h(u-t)X_i(u)\epsilon_i(u)dN_i^c(u)\mid \mathcal{D}_i, R_i=0\Big\} \\
&\quad -n^{-1}\sum_{i=1}^{n}(1-R_i)E_s\Big\{\int_0^{\tau} K_h(u-t)X_i(u)\epsilon_i(u)dN_i^c(u)\mid \mathcal{D}_i, R_i=0\Big\}\Big) \\
= &\ (nh)^{1/2}(e_{xx}(t))^{-1}\Big(n^{-1}\sum_{i=1}^{n}R_i\int_0^{\tau} K_h(u-t)X_i(u)\epsilon_i(u)dN_i^c(u) \\
&\quad +n^{-1}\sum_{i=1}^{n}(1-R_i)E_s\Big\{\int_0^{\tau} K_h(u-t)X_i(u)\epsilon_i(u)dN_i^c(u)\mid \mathcal{D}_i, R_i=0\Big\}\Big) \\
&+h^{1/2}(e_{xx}(t))^{-1}\Big\{n^{-1/2}\Big[\int_0^{\infty}\int_0^{L} E\Big((1-R_i)X_i(u)\epsilon_i(u)\alpha_i^*(s-u) \\
&\qquad \frac{I(x<L-(V_i))}{F_s(V_i)}\Big)dF_s(s)\frac{dM^R(x)}{y^R(x)} \\
&\quad +\int_0^{L-}\int_0^{(L-x)-} E\Big(\frac{(1-R_i)X_i(u)\epsilon_i(u)\alpha_i^*(s-u)}{F_s(V_i)}\Big)dF_s(s)\frac{dM^R(x)}{y^R(x)} \\
&\quad +\int_0^{L} E\Big(\frac{(1-R_i)X_i(u)\epsilon_i(u)\alpha_i^*(s-u)}{F_s(V_i)}\Big)F_s(s)\frac{dM^R((L-s)-)}{y^R((L-s)-)}\Big] \\
&+o_p(1)+O_p(n^{-1/2}h^2)\Big\} \\
= &\ (nh)^{1/2}(e_{xx}(t))^{-1}\Big(n^{-1}\sum_{i=1}^{n}R_i\int_0^{\tau} K_h(u-t)X_i(u)\epsilon_i(u)dN_i^c(u) \\
&\quad +n^{-1}\sum_{i=1}^{n}(1-R_i)E_s\Big\{\int_0^{\tau} K_h(u-t)X_i(u)\epsilon_i(u)dN_i^c(u)\mid \mathcal{D}_i, R_i=0\Big\}\Big) \\
&+O(h^{1/2})+o_p(h^{1/2})+O_p(n^{-1/2}h^{5/2})
\end{aligned}
$$

which for each fixed time point t, converges in distribution to a multivariate distribution with mean 0 and covariance matrix $\mu_0\Sigma(t)$ by Lindeberg-Feller theorem.

We derive the asymptotic covariance matrix in the following way.

$$
\begin{aligned}
&cov\Big[(nh)^{1/2}\Big(\tilde{\beta}(t;\gamma_0)-\beta^*(t)+h^2(e_{xx}(t))^{-1}[e'_{xx}(t)\beta'_0(t) \\
&\qquad +(1/2)e_{xx}(t)\beta''_0(t)]\int_{-1}^{1}x^2K(x)dx\Big)\Big] \\
= &\ cov\Big[(nh)^{1/2}(e_{xx}(t))^{-1}\Big(n^{-1}\sum_{i=1}^{n}R_i\int_0^{\tau} K_h(u-t)X_i(u)\epsilon_i(u)dN_i^c(u)
\end{aligned}
$$

$$+n^{-1}\sum_{i=1}^{n}(1-R_i)E_s\Big\{\int_0^{\tau}K_h(u-t)X_i(u)\epsilon_i(u)dN_i^c(u)\mid\mathcal{D}_i,R_i=0\Big\}\Big)\Big]$$

$$= n^{-1}h(e_{xx}(t))^{-1}cov\Big[\Big(\sum_{i=1}^{n}R_i\int_0^{\tau}K_h(u-t)X_i(u)\epsilon_i(u)dN_i^c(u)$$

$$+\sum_{i=1}^{n}(1-R_i)E_s\Big\{\int_0^{\tau}K_h(u-t)X_i(u)\epsilon_i(u)dN_i^c(u)\mid\mathcal{D}_i,R_i=0\Big\}\Big)\Big](e_{xx}(t))^{-1}$$

Note that all the subjects are i.i.d. and that R_i is an indicator,

$$cov\Big[(nh)^{1/2}\Big(\tilde{\beta}(t;\gamma_0)-\beta^*(t)+h^2(e_{xx}(t))^{-1}[e'_{xx}(t)\beta'_0(t)$$

$$+(1/2)e_{xx}(t)\beta''_0(t)]\int_{-1}^{1}x^2K(x)dx\Big)\Big]$$

$$= n^{-1}h(e_{xx}(t))^{-1}\sum_{i=1}^{n}\Big[cov\Big(R_i\int_0^{\tau}K_h(u-t)X_i(u)\epsilon_i(u)dN_i^c(u)\Big)$$

$$+cov\Big((1-R_i)E_s\Big\{\int_0^{\tau}K_h(u-t)X_i(u)\epsilon_i(u)dN_i^c(u)\mid\mathcal{D}_i,R_i=0\Big\}\Big)\Big](e_{xx}(t))^{-1}$$

$$= h(e_{xx}(t))^{-1}\Big[cov\Big(\int_0^{\tau}K_h(u-t)R_iX_i(u)\epsilon_i(u)dN_i^c(u)\Big)$$

$$+cov\Big(E_s\Big\{\int_0^{\tau}K_h(u-t)(1-R_i)X_i(u)\epsilon_i(u)dN_i^c(u)\mid\mathcal{D}_i,R_i=0\Big\}\Big)\Big](e_{xx}(t))^{-1}.$$

By the Doob-Meyer decomposition of $N_i^c(t)$, $N_i^c(t)=\int_0^t Y_i^c(s)\alpha_i^c(s)ds+M_i^c(t)$. Let $Y_i^c(t)=\sum_{j=1}^{n_i}I(T_{ij}^0\geq t)$. So

$$h(e_{xx}(t))^{-1}cov\Big(\int_0^{\tau}K_h(u-t)R_iX_i(u)\epsilon_i(u)dN_i^c(u)\Big)(e_{xx}(t))^{-1}$$

$$= h(e_{xx}(t))^{-1}cov\Big(\int_0^{\tau}K_h(u-t)R_iX_i(u)\epsilon_i(u)dM_i^c(u)\Big)(e_{xx}(t))^{-1}$$

$$+2h(e_{xx}(t))^{-1}cov\Big(\int_0^{\tau}K_h(u-t)R_iX_i(u)\epsilon_i(u)dM_i^c(u),$$

$$\int_0^{\tau}K_h(u-t)R_iX_i(u)\epsilon_i(u)Y_i^c(u)\alpha_i^c(u)du\Big)(e_{xx}(t))^{-1}$$

$$+h(e_{xx}(t))^{-1}cov\Big(\int_0^{\tau}K_h(u-t)R_iX_i(u)\epsilon_i(u)Y_i^c(u)\alpha_i^c(u)du\Big)(e_{xx}(t))^{-1}.$$

R_i, $X_i(t)$ and $\epsilon_i(t)$ are $\mathcal{F}_t^c$-predictable. This leads the first term above to

$$h(e_{xx}(t))^{-1}cov\Big(\int_0^{\tau}K_h(u-t)R_iX_i(u)\epsilon_i(u)dM_i^c(u)\Big)(e_{xx}(t))^{-1}$$

$$
\begin{aligned}
&= h(e_{xx}(t))^{-1}E\bigg(\int_0^\tau K_h^2(u-t)R_i^2X_i(u)X_i^T(u)\epsilon_i^2(u)d<M>_i^c(u)\bigg)(e_{xx}(t))^{-1}\\
&= h(e_{xx}(t))^{-1}E\bigg(\int_0^\tau K_h^2(u-t)R_i^2X_i(u)X_i^T(u)\epsilon_i^2(u)Y_i^c(u)\alpha_i^c(u)du\bigg)(e_{xx}(t))^{-1}\\
&= (e_{xx}(t))^{-1}E\bigg(\int_{-1}^1 K^2(x)R_i^2X_i(t+xh)X_i^T(t+xh)\epsilon_i^2(t+xh)Y_i^c(t+xh)\\
&\qquad \alpha_i^c(t+xh)dx\bigg)(e_{xx}(t))^{-1}\\
&= (e_{xx}(t))^{-1}E\bigg(R_i^2X_i(t)X_i^T(t)\epsilon_i^2(t)Y_i^c(t)\alpha_i^c(t)\int_{-1}^1 K^2(x)dx+O(h^2)\bigg)(e_{xx}(t))^{-1}\\
&= \mu_0(e_{xx}(t))^{-1}E[R_i^2X_i(t)X_i^T(t)\epsilon_i^2(t)Y_i^c(t)\alpha_i^c(t)](e_{xx}(t))^{-1}+O(h^2).
\end{aligned}
$$

And

$$
\begin{aligned}
& h(e_{xx}(t))^{-1}cov\bigg(\int_0^\tau K_h(u-t)R_iX_i(u)\epsilon_i(u)Y_i^c(u)\alpha_i^c(u)du\bigg)(e_{xx}(t))^{-1}\\
&= h(e_{xx}(t))^{-1}cov\bigg(\int_{-1}^1 K(x)R_iX_i(t+xh)\epsilon_i(t+xh)Y_i^c(t+xh)\alpha_i^c(t+xh)dx\bigg)\\
&\qquad \cdot(e_{xx}(t))^{-1}\\
&= h(e_{xx}(t))^{-1}cov[R_iX_i(t)\epsilon_i(t)Y_i^c(t)\alpha_i^c(t)+O(h^2)](e_{xx}(t))^{-1}=O(h).
\end{aligned}
$$

Since

$$
\begin{aligned}
& E\bigg(\int_0^\tau K_h(u-t)R_iX_i(u)\epsilon_i(u)dM_i^c(u)\bigg)\\
&= E\bigg[E\bigg(\int_0^\tau K_h(u-t)R_iX_i(u)\epsilon_i(u)dM_i^c(u)\bigg|X_i(\cdot),Z_i(\cdot),N_i(\cdot),S_i,V_i,C_i\bigg)\bigg]\\
&= E\bigg[\int_0^\tau K_h(u-t)R_iX_i(u)E(\epsilon_i(u)|X_i(\cdot),Z_i(\cdot),N_i(\cdot),S_i,V_i,C_i)dM_i^c(u)\bigg]\\
&= E\bigg[\int_0^\tau K_h(u-t)R_iX_i(u)E(\epsilon_i(u)|X_i(\cdot),Z_i(\cdot))dM_i^c(u)\bigg]=0
\end{aligned}
$$

and

$$
\begin{aligned}
& E\bigg(\int_0^\tau K_h(u-t)R_iX_i(u)\epsilon_i(u)Y_i^c(u)\alpha_i^c(u)du\bigg)\\
&= E\bigg[E\bigg(\int_0^\tau K_h(u-t)R_iX_i(u)\epsilon_i(u)Y_i^c(u)\alpha_i^c(u)du\bigg|X_i(\cdot),Z_i(\cdot),N_i(\cdot),S_i,V_i,C_i\bigg)\bigg]
\end{aligned}
$$

$$= E\bigg[\int_0^\tau K_h(u-t)R_iX_i(u)E(\epsilon_i(u)|X_i(\cdot),Z_i(\cdot),N_i(\cdot),S_i,V_i,C_i)Y_i^c(u)\alpha_i^c(u)du\bigg]$$
$$= E\bigg[\int_0^\tau K_h(u-t)R_iX_i(u)E(\epsilon_i(u)|X_i(\cdot),Z_i(\cdot))Y_i^c(u)\alpha_i^c(u)du\bigg] = 0,$$

we have

$$\begin{aligned} & cov\bigg(\int_0^\tau K_h(u-t)R_iX_i(u)\epsilon_i(u)dM_i^c(u)\bigg) \\ = \;& E\bigg(\int_0^\tau K_h(u-t)R_iX_i(u)\epsilon_i(u)dM_i^c(u)\bigg)^{\otimes 2}, \end{aligned}$$

$$\begin{aligned} & cov\bigg(\int_0^\tau K_h(u-t)R_iX_i(u)\epsilon_i(u)Y_i^c(u)\alpha_i^c(u)du\bigg) \\ = \;& E\bigg(\int_0^\tau K_h(u-t)R_iX_i(u)\epsilon_i(u)Y_i^c(u)\alpha_i^c(u)du\bigg)^{\otimes 2}, \end{aligned}$$

$$\begin{aligned} & cov\bigg(\int_0^\tau K_h(u-t)R_iX_i(u)\epsilon_i(u)dM_i^c(u), \int_0^\tau K_h(u-t)R_iX_i(u)\epsilon_i(u)Y_i^c(u) \\ & \quad \alpha_i^c(u)du\bigg) \\ = \;& E\bigg[\bigg(\int_0^\tau K_h(u-t)R_iX_i(u)\epsilon_i(u)dM_i^c(u)\bigg)\bigg(\int_0^\tau K_h(u-t)R_iX_i(u)\epsilon_i(u)Y_i^c(u) \\ & \quad \alpha_i^c(u)du\bigg)^T\bigg]. \end{aligned}$$

Then by the Cauchy-Schwarz inequality,

$$\begin{aligned} & h\,cov\bigg(\int_0^\tau K_h(u-t)R_iX_i(u)\epsilon_i(u)dM_i^c(u), \int_0^\tau K_h(u-t)R_iX_i(u)\epsilon_i(u)Y_i^c(u) \\ & \quad \alpha_i^c(u)du\bigg) \\ = \;& hE\bigg[\bigg(\int_0^\tau K_h(u-t)R_iX_i(u)\epsilon_i(u)dM_i^c(u)\bigg)\bigg(\int_0^\tau K_h(u-t)R_iX_i(u)\epsilon_i(u)Y_i^c(u) \\ & \quad \alpha_i^c(u)du\bigg)^T\bigg] \\ \leq \;& h\bigg\{\bigg[E\bigg(\int_0^\tau K_h(u-t)R_iX_i(u)\epsilon_i(u)dM_i^c(u)\bigg)^{\otimes 2}\bigg]\bigg[E\bigg(\int_0^\tau K_h(u-t)R_iX_i(u)\epsilon_i(u) \\ & \quad Y_i^c(u)\alpha_i^c(u)du\bigg)^{\otimes 2}\bigg]\bigg\}^{1/2} \end{aligned}$$

$$
\begin{aligned}
&= h\Bigg\{\Bigg[cov\Bigg(\int_0^\tau K_h(u-t)R_iX_i(u)\epsilon_i(u)dM_i^c(u)\Bigg)\Bigg]\Bigg[cov\Bigg(\int_0^\tau K_h(u-t)R_iX_i(u) \\
&\qquad \epsilon_i(u)Y_i^c(u)\alpha_i^c(u)du\Bigg)\Bigg]\Bigg\}^{1/2} \\
&= \Bigg\{\Bigg[hcov\Bigg(\int_0^\tau K_h(u-t)R_iX_i(u)\epsilon_i(u)dM_i^c(u)\Bigg)\Bigg]\Bigg[hcov\Bigg(\int_0^\tau K_h(u-t)R_iX_i(u) \\
&\qquad \epsilon_i(u)Y_i^c(u)\alpha_i^c(u)du\Bigg)\Bigg]\Bigg\}^{1/2} \\
&= \{[\mu_0(e_{xx}(t))^{-1}E[R_i^2X_i(t)X_i^T(t)\epsilon_i^2(t)Y_i^c(t)\alpha_i^c(t)](e_{xx}(t))^{-1}+O(h^2)][O(h)]\}^{1/2} \\
&= O(h)
\end{aligned}
$$

Hence

$$
\begin{aligned}
& h(e_{xx}(t))^{-1}cov\Bigg(\int_0^\tau K_h(u-t)R_iX_i(u)\epsilon_i(u)dN_i^c(u)\Bigg)(e_{xx}(t))^{-1} \\
= \ & \mu_0(e_{xx}(t))^{-1}E[R_i^2X_i(t)X_i^T(t)\epsilon_i^2(t)Y_i^c(t)\alpha_i^c(t)](e_{xx}(t))^{-1}+O(h^2)+O(h).
\end{aligned}
$$

Note that

$$
\begin{aligned}
\tilde{e}_{xy}(t) &= \int_0^\tau K_h(s-t)e_{xy}(s)ds = \int_{t-h}^{t+h} h^{-1}K(\frac{s-t}{h})e_{xy}(s)ds \\
&= \int_{-1}^1 K(x)e_{xy}(t+xh)dx \\
&= \int_{-1}^1 K(x)(e_{xy}(t)+hxe'_{xy}(t)+(1/2)h^2x^2e''_{xy}(t)+o(h^2))dx \\
&= e_{xy}(t)\int_{-1}^1 K(x)dx + he'_{xy}(t)\int_{-1}^1 xK(x)dx + (1/2)h^2e''_{xy}(t)\int_{-1}^1 x^2K(x)dx \\
&\quad +o(h^2) \\
&= e_{xy}(t)+(1/2)h^2e''_{xy}(t)\int_{-1}^1 x^2K(x)dx+o(h^2).
\end{aligned}
$$

Similar results hold for $\tilde{e}_{xx}(t)$ and $\tilde{e}_{xz}(t)$. Let $\mu_2 = \int_{-1}^1 x^2K(x)dx$. So by the long division of functions

$$
\begin{aligned}
\tilde{y}_x^T(t) &= (\tilde{e}_{xx}(t))^{-1}\tilde{e}_{xy}(t) \\
&= (e_{xx}(t)+(1/2)\mu_2h^2e''_{xx}(t)+o(h^2))^{-1}(e_{xy}(t)+(1/2)\mu_2h^2e''_{xy}(t)+o(h^2))
\end{aligned}
$$

$$= \quad y_x^T(t) + (1/2)\mu_2 h^2 (e_{xx}(t))^{-1}[e''_{xy}(t) - e''_{xx}(t)(e_{xx}(t))^{-1} e_{xy}(t)] + o(h^2).$$

Also

$$\tilde{z}_x^T(t) = z_x^T(t) + (1/2)\mu_2 h^2 (e_{xx}(t))^{-1}[e''_{xz}(t) - e''_{xx}(t)(e_{xx}(t))^{-1} e_{xz}(t)] + o(h^2).$$

Then

$$\begin{aligned}
\beta^*(t) &= \tilde{y}_x^T(t) - \tilde{z}_x^T(t)\gamma_0 \\
&= y_x^T(t) - z_x^T(t)\gamma_0 + (1/2)\mu_2 h^2 (e_{xx}(t))^{-1}[e''_{xy}(t) - e''_{xx}(t)(e_{xx}(t))^{-1} e_{xy}(t)] \\
&\quad -(1/2)\mu_2 h^2 (e_{xx}(t))^{-1}[e''_{xz}(t) - e''_{xx}(t)(e_{xx}(t))^{-1} e_{xz}(t)]\gamma_0 + o(h^2) \\
&= y_x^T(t) - z_x^T(t)\gamma_0 + (1/2)\mu_2 h^2 (e_{xx}(t))^{-1}[e''_{xy}(t) - e''_{xx}(t) y_x^T(t)] \\
&\quad -(1/2)\mu_2 h^2 (e_{xx}(t))^{-1}[e''_{xz}(t) - e''_{xx}(t) z_x^T(t)]\gamma_0 + o(h^2) \\
&= \beta_0(t) + (1/2)\mu_2 h^2 (e_{xx}(t))^{-1}[e''_{xy}(t) - e''_{xz}(t)\gamma_0 - e''_{xx}(t)\beta(t)] + o(h^2).
\end{aligned}$$

So

$$\begin{aligned}
& (nh)^{1/2}(\widehat{\beta}(t) - \beta^*(t)) \\
&= (nh)^{1/2}(\tilde{\beta}(t;\widehat{\gamma}) - \beta^*(t)) \\
&= (nh)^{1/2}(\tilde{\beta}(t;\gamma_0) - \beta^*(t)) + (nh)^{1/2}(\widehat{\gamma} - \gamma_0)\frac{\partial \tilde{\beta}(t;\gamma_0)}{\partial \gamma} + O_p(n^{-1/2}h^{1/2}) \\
&= (nh)^{1/2}(\tilde{\beta}(t;\gamma_0) - \beta^*(t)) + O(h^{1/2}) + O_p(n^{-1/2}h^{1/2})
\end{aligned}$$

since $n^{1/2}(\widehat{\gamma} - \gamma_0) \xrightarrow{D} \mathcal{N}(0, D^{-1}VD^{-1})$ and

$$\frac{\partial \tilde{\beta}(t;\gamma_0)}{\partial \gamma} = -\tilde{Z}_x(t) \xrightarrow{P} -z_x(t).$$

Therefore,

$$(nh)^{1/2}(\widehat{\beta}(t) - \beta_0(t) - \beta_{Bias}(t)) \xrightarrow{D} \mathcal{N}(0, \mu_0\Sigma(t)),$$

as $n \to \infty$, $h \to 0$, $nh \to \infty$, $nh^5 = O(1)$. $\square$