DYNAMIC MODELING OF INCOMPLETE EVENT HISTORY DATA

by

Fei Heng

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

2019

Approved by:

Dr. Yanqing Sun

Dr. Jiancheng Jiang

Dr. Yang Li

Dr. Weidong Tian

©2019
Fei Heng
ALL RIGHTS RESERVED

ABSTRACT

FEI HENG. Dynamic Modeling of Incomplete Event History Data. (Under the direction of DR. YANQING SUN)

Event history analysis has important applications in many fields, such as medicine, engineering, econometrics, actuarial science and social studies. We usually encounter missing data problems in the modelling of event history data. One typical problem is that the observations of event times are censored at the end of a study or by a terminal event. Also, the covariates in the model may be subject to missingness. In the multivariate case, sometimes the record of the types of events is missing. In this dissertation, we investigate incomplete event history data including competing risks data with missing failure causes and recurrent event data under nonparametric models.

For the competing risks data, we study the Cox model with time-varying coefficients for cause-specific hazard functions when causes of failure are subject to missingness. The inverse probability weighted (IPW) and augmented inverse probability weighted (AIPW) estimators are investigated. The latter is shown as a two-stage estimator by directly utilizing the inverse probability weighted estimator and through modeling available auxiliary variables to improve efficiency. The proposed methods are illustrated using the Mashi trial data for investigating the effect of randomization to formula-feeding versus breast-feeding plus extended infant zidovudine prophylaxis on death due to mother-to-child HIV transmission in Botswana.

In the field of recurrent events, we simultaneously explore the time-varying and gap-time-varying effects of covariates on intensities under generalized nonparametric dynamic additive intensity models. The local linear kernel smooth methods are employed to estimate the mixed effects that include time-varying effects, and the effects that may depend on past histories. Furthermore, we consider a special case where the covariates with gap-time-varying effects may be missing. AIPW estimators are obtained by solving the local weighted AIPW estimating equation which utilizes the AIPW technique.

ACKNOWLEDGMENTS

First and foremost, I would like to express the most heartfelt appreciation to my dissertation advisor, Dr. Yanqing Sun for her guidance, patience, and encouragements throughout my graduate study at UNC Charlotte. Without all her support and trusts, it would not be possible to complete this dissertation. Working with her has been a fantastic experience for me. Her expertise in event history data analysis and passion for research have inspired me a lot in pursuing my academic interests and planning my career path. Her optimistic life attitude has affected me and given me the same positivity and persistence against the dilemma. Also, I am deeply grateful for her financial support.

To my committee members, Drs. Jiancheng Jiang, Yang Li, and Weidong Tian, I would like to thank you for their constructive comments and valuable suggestions. Thanks also to all the faculty and staff in the Mathematics and Statistics Department. I also need to offer special appreciation to Dr. Shaozhong Deng, who served as graduate coordinator during my time in this program.

I wish to thank Dr. Peter Gilbert at Fred Hutchinson Cancer Research Center and the University of Washington for motivating discussions on the future applications to the Malaria vaccine study.

I would like to thank the Mathematics and Statistics Department and the Graduate School for providing financial support for my study and conference trips. This research was partially supported by the National Science Foundation grant DMS1513072 and NIAID NIH award number R37AI054165.

Finally, I would like to take this chance to express my sincerest thanks to my family and friends, who have given me plenty of support and love. I would dedicate my dissertation to my wife Xinhui Cai and my son Zemo.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

Event history data is longitudinal data that records the time to the occurrence of events of interest for a sample of individuals in specific studies. Analyses of the time-to-event data often encounter a specific missing data problem, called censoring, where the time to event occurrence is not observed for every subjects. Survival data is a special case where there is a single event that may occur for each individual. We have competing risks when the individual is at risk of different mutually exclusive types of events. Sometimes the event of interest may occur repeatedly over time for an individual. Examples of recurrent event data include repeated heart attacks for coronary patients, frequent claims from auto insurance policyholders and machine breakdowns of mechanic or electronic systems.

Other types of missing data problems are also common in the event history data, such as missing failure causes and missing covariates. The complete-case (CC) analysis, using only cases with complete information, is the most straightforward approach to handle missing problems but may obtain biased and misleading results when the data are not missing completely at random (MCAR) and the complete cases are not a random sample of all cases (Little and Rubin, 2002). Weighting complete cases by the inverse of their propensity scores(Horvitz and Thompson, 1952), known as the inverse probability weighting method (IPW), is a commonly used method to correct this bias under missing at random (MAR) mechanism (Rubin, 1976). The complete cases are enlarged to represent the missing data by the inverse probabilities. However, the IPW method is statistically inefficient and sensitive to the correct modeling of the propensity score. To improve the efficiency of IPW, Robins et al. (1994) developed augmented inverse probability weighted method (AIPW) for the conditional mean model by adding an augmented term to the original IPW estimating equation when the data are missing at random. The AIPW method has been receiving much attention due to its attractive doubly robust property and statistically efficiency. Bang and Robins (2005) constructed AIPW estimators in two nonlongitudinal models, missing data model and treatment effect model, and extend their method to longitudinal marginal structural models. Gao and Tsiatis (2005), Lu and Liang (2008), Sun et al. (2012), and Hyun et al. (2012) derived the AIPW estimators for competing risks data with missing causes of failure. With application to HIV vaccine efficacy trials, Sun and Gilbert (2012) and Gilbert and Sun (2015) developed estimation approaches and hypothesis testing procedures based on the AIPW technique for stratified mark-specific proportional hazards models with missing marks.

In this dissertation, we perform a careful study of missing data problems in the survival data and recurrent event data. In Chapter 2, we develop a two-step estimation procedure for competing risks data with missing causes of failure under the cause-specific time-varying Cox model. The IPW estimator evaluated in the first stage, and available auxiliary covariates are utilized to form an efficient and robust two-stage AIPW estimator. This work is motivated by the Mashi study (Thior et al., 2006) for investigating the effect of randomization to formula-feeding (FF) versus breastfeeding plus extended infant zidovudine prophylaxis (BF+AZT) on death due to mother-to-child HIV transmission in Botswana. The causes of death are missing for 61 infants of the 111 live-born infants who died in the Mashi trial.

In Chapter 3, we propose a nonparametric dynamic intensity model for recurrent event data which allows various link functions and models varying covariate effects. The local linear kernel smooth methods are employed to estimate the time-varying effects and the effects that may depend on past histories simultaneously. Additionally, it is essential to take into consideration covariates that may be subject to missingness. For example, in vaccine clinical trials, the immune-response measurements are usually very expensive, therefore not available for all participants. In Chapter 4, we analyze recurrent event data with missing covariates under a nonparametric dynamic additive intensity model. We apply the aforementioned AIPW technique to deal with missing covariate values. Based on the local linear method for full data in the sense that covariates are not missing, we achieve consistent AIPW estimators through an iterative algorithm.

CHAPTER 2: TIME-VARYING COX MODEL FOR CAUSE-SPECIFIC HAZARD FUNCTIONS WITH MISSING CAUSES

2.1 Introduction

Survival analysis is a collection of statistical procedures for analyzing the time until the occurrence of the event of interest. The time is known as survival time because death is the event of interest in many medical studies. In a sample of individuals, survival time is usually not observed for everyone. We term incomplete observations as censored survival time which makes the survival analysis different from other standard statistical models (Aalen et al., 2008).

We usually study survival times through survival function and hazard function. Let T be the survival time. The survival function $S(t)$ is the unconditional probability that the event time is later than a specified time t. The hazard function $\lambda(t)$ is the instantanous rate that the event occurred at t given the individual has survived until t:

$$\lambda(t) = \lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} P\{t \leq T \leq t + \Delta t \,|\, T \geq t\}.$$

Survival analysis examples include assessing the treatment effects for new drugs, investigating the vaccine efficacy and effectiveness against some infectious diseases, measuring the reliability of a mechanical or electronic system, and so on. The purpose of many survival analysis is to study the effects of covariates on the hazard function. Let $Z = (Z_1, \ldots, Z_p)^\mathsf{T}$ be the p-dimensional covariate. We shall define the conditional hazard function of T at t given covariate $Z = z$ as follows:

$$\lambda(t|Z=z) = \lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} P\{t \leq T \leq t + \Delta t \,|\, T \geq t,\, Z = z\}.$$

The most widely used semiparametric survival model to model the survival time considering covariates is the Cox proportional hazards model proposed by Cox (1972). It assumes the conditional hazard function consists of an unspecified baseline hazard function $\lambda_0(t)$ and an exponential part:

$$\lambda(t|Z=z) = \lambda_0(t) \exp\{\beta^{\mathsf{T}} z\}$$

where β is a p-dimensional vector of covariate coefficients. The standard Cox model does not accommodate the situations where the covariate coefficient β may change with time. We may consider a time-varying Cox model where β possesses a functional form:

$$\lambda(t|Z=z) = \lambda_0(t) \exp\{(\beta(t))^{\mathsf{T}} z\},$$

where $\beta_0(t)$ is a p-dimensional vector of unspecified coefficient functions of t. Unless $\beta(t)$ is constant, this model represents a non-proportional hazards model. Murphy and Sen (1991) proposed a histogram sieve estimation procedure by assuming that the coefficient functions are step functions. In Martinussen et al. (2002), a one-step iterative estimation procedure for the cumulative coefficient functions is developed by using the one-step Newton-Raphson iterative algorithm based on the log-likelihood. Cai and Sun (2003) applied a local partial likelihood estimation technique to estimate the time-dependent coefficients $\beta(t)$. Tian et al. (2005) further studied the local constant partial likelihood estimator and constructed point-wise and simultaneous confidence intervals for the regression parameters.

In many contexts, competing risks are said to be present when subjects in a follow-up study potentially experience more than one type of event of interest, and the occurrence of one type of failure prevents the other types of failures from occurring. For example, if we consider death as the event of interest, subjects may die from different causes.

Let V be the cause of failure, and $Z(t)$ a possibly time-dependent p-dimensional covariate over the follow-up time period $[0, \tau]$. Let $\bar{Z}(\tau) = \{Z(t), 0 \leq t \leq \tau\}$ be the covariate history. For a typical competing risks data set, the observable random variables are $(X, \delta, \delta V, Z(\cdot))$, where $X = \min\{T, C\}$, $\delta = I(T \leq C)$, and C is a censoring random variable that is assumed to be independent of T and V conditional on $Z(\cdot)$. The cause V is only observable when $\delta = 1$, whereas if T is censored, the cause is unknown. Suppose that the conditional cause-specific hazard function for cause $V = k$ at time t, given the covariate history $\bar{Z}(\tau)$, only depends on the current value $Z(t)$, which is defined as

$$\lambda_k(t|z(t)) = \lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} P\Big(t \leq T \leq t + \Delta t, V = k \,|\, T \geq t,\ Z(t) = z(t)\Big) \tag{2.1}$$

with t ranging over a fixed interval $[0, \tau]$.

During an oral cholera vaccine efficacy trial in Bangladesh (Clemens et al., 1990), four different cholera strains circulated during the trial, and it was of interest to assess vaccine efficacy against each of these strains. Since anti-cholera antibodies induced by the vaccine tended to wane over time, there was concern that vaccine efficacy would wane, which was observed for the overall vaccine efficacy regardless of strain

(Durham et al., 1998). This suggests that it is important to allow for time-variations in the assessment of strain-specific vaccine efficacy.

Motivated by this oral cholera vaccine trial, Sun et al. (2008) proposed the following conditional strain-specific (also cause-specific) hazard function of the Cox model with time varying coefficients:

$$\lambda_k(t|z(t)) = \lambda_{k0}(t)\exp\big(\beta_k(t)^{\mathsf{T}} z(t)\big), \quad k = 1, \dots K, \tag{2.2}$$

where the strain/cause variable V is a categorical variable taking K categories, $\lambda_{k0}(t)$ is an unspecified baseline function and $\beta_k(t) = (\beta_{k1}(t), \dots, \beta_{kp}(t))$ is a p-dimensional vector of unspecified time-dependent regression coefficients.

In the analysis of competing risks data, the cause of failure is often missing due to various reasons. For example, the Mashi clinical trial was conducted among HIV-infected women and their infants to compare the effect of infant feeding strategy on two outcomes in live-born infants: HIV infection (through postnatal mother-to-child HIV transmission) and death (Thior et al., 2006). Twelve hundred HIV positive pregnant mothers were randomized to two feeding strategies for their nascent infants: 6 months of breastfeeding and zidovudine for the infant (BF+AZT, 588 live-born infants) versus 12 months of formula feeding with zidovudine for the infant for the first month of life (FF, 591 live-born infants). All mothers were instructed to wean their infants between 5 and 6 months of age, and were supplied free formula from 5 through 12 months of age to facilitate safe weaning. Infants were tested for HIV infection at birth, monthly until age 7 months, at age 9 months, and then every three month through age 18 months. Of the 111 live-born infants who died in the Mashi trial, 28 infants died of an HIV-related cause, 22 infants died of an HIV-unrelated cause, and the cause of death was missing for 61 infants. We consider a death to be HIV-related if either the study clinicians deemed the death HIV-related ($n = 4$ deaths), or the infant had at least one positive test result from the PCR assay used to test for HIV infection prior to death ($n = 24$ deaths). On the other hand, we consider a death to be HIV-unrelated if the study clinician deemed the death unrelated to HIV/AIDS ($n = 22$ deaths). The Mashi study showed that breast-feeding increased the risk of infants acquiring HIV by 7 months of age while formula-feeding increased the risk of death by 7 months of age, and these treatment effects both waned toward no effect by 18 months of age. However, the effect of randomized feeding strategy on death due to HIV infection is unknown, and it is of our interest to assess the treatment effect on HIV-related death with HIV-unrelated death as a competing risk. Therefore, in the cause-specific Cox model with time varying coefficients, we allow the cause V to be missing.

In situations where only a single cause of failure is of interest, the Cox model with time-varying regression coefficients has been studied by Zucker and Karr (1990), Murphy and Sen (1991), Martinussen et al. (2002), Cai and Sun (2003), Tian et al. (2005) and Sun et al. (2009b). Modifying the score function of the local linear partial maximum likelihood estimation of Cai and Sun (2003), we propose a two-stage efficient procedure to estimate the model (2.2) with missing causes. In the first stage, an inverse probability weighted (IPW) complete-case estimator is developed. This estimator is consistent, but not efficient. More efficient and robust estimators can be constructed by adding a second stage that models the probability of the failure cause conditional on auxiliary variables, which uses the IPW estimator as a component of the estimating equation.

The rest of this chapter is organized as follows. Notations and assumptions are first introduced in Section 2.2.1. The two-stage augmented inverse probability weighted estimator (AIPW) is developed based on the IPW estimator in Sections 2.2.2 to 2.2.4. The asymptotic results for both the IPW and AIPW estimators are presented in Section 2.3. A simulation study is conducted to examine the performances of the proposed estimators and the results are presented in Section 2.4. The proposed methods are applied to analyze the Mashi clinical trial data in Section 2.5. Some conclusion remarks are given in Chapter 5.

2.2 Two-stage estimation via local linear partial likelihood

2.2.1 Notations and assumptions

Let R be the missing cause indicator such as $R = 1$ if $\delta = 1$ and the cause V is observed or if $\delta = 0$, and $R = 0$ otherwise. The missing cause indicator $R = 0$ if the failure time is observed but the cause of failure is not available, and $R = 1$ if the failure time is censored. In addition to the covariate $Z(\cdot)$ considered in model (2.2), our procedures allow use of the auxiliary covariates $A = (\delta A^{(r)}, \delta A^{(v)})$ to improve efficiency, where $A^{(r)}$ and $A^{(v)}$ are the measurements at the failure time T for subjects with observed failures, $A^{(r)}$ can be used to predict the probability whether V is observed or $\delta = 0$, and $A^{(v)}$ can be used to inform the distribution of V.

Let $W_1 = (T, Z(T), \delta A^{(r)})$, $W_2 = (T, Z(T), \delta A^{(v)})$ and $W_3 = (T, Z(T))$. We assume the following missing at random (MAR) (Rubin, 1976) assumptions:

$$\text{MAR I: } P(R=1|V,\delta A^{(v)},\delta=1,W_1)=P(R=1|\delta=1,W_1);$$

$$\text{MAR II: } P(R=1|V,\delta=1,W_2)=P(R=1|\delta=1,W_2);$$

$$\text{MAR III: } P(R=1|V,\delta=1,W_3)=P(R=1|\delta=1,W_3).$$

Here, MAR I assumes that given $\delta = 1$ and the covariates measured for all subjects W_1, the probability of observing the cause of failure V is independent of V and $A^{(v)}$. MAR II and MAR III assume the probability of observing V is independent of V conditional on $\delta = 1$ and either W_2 or W_3. MAR II also implies that V is independent of R given W_2: $P(V = k|R = 1, \delta = 1, W_2) = P(V = k|\delta = 1, W_2)$. We denote $P(R = 1|\delta = 1, W_1)$ and $P(V = k|\delta = 1, W_2)$ by $r(W_1)$ and $\rho_k(W_2)$ respectively. Let $\pi(Q) = P(R = 1|Q)$, where $Q = (W_1, \delta)$. Then $\pi(Q) = \delta r(W_1) + (1 - \delta)$.

The observed data consist of iid replicates

$$O_i = \{X_i, \delta_i, \bar{Z}_i(\tau), R_i, R_i\delta_i V_i, \delta_i A_i^{(r)}, \delta_i A_i^{(v)}\}, \quad i = 1, \ldots, n$$

of $O = \{X, \delta, \bar{Z}(\tau), R, R\delta V, \delta A^{(r)}, \delta A^{(v)}\}$. We define the counting processes $N_{ik}(t) = I(X_i \le t, \delta_i = 1, V_i = k)$, $N_i(t) = I(X_i \le t, \delta_i = 1)$ and the at-risk process $Y_i(t) = I(X_i \ge t)$.

Throughout the paper we assume that the baseline hazard function $\lambda_{k0}(t)$ is unspecified, positive and continuous, the coefficient functions $\{\beta_{kl}(t)\}$ have continuous second derivatives in a neighborhood of t, and the covariate $Z(t)$ is a locally bounded predictable process.

2.2.2 Full data local linear partial likelihood estimator

When there are no missing causes, the local linear partial likelihood method of Cai and Sun (2003) can be used to estimate the regression coefficients model (2.2). For cause $V = k$, by the Taylor expansion for u in a neighborhood of t,

$$\beta_{kl}(u) \approx \beta_{kl}(t) + \beta'_{kl}(t)(u-t), \quad l = 1, 2, \ldots, p. \tag{2.3}$$

Let $\xi_k(t) = \big(\beta_{k1}(t), \ldots, \beta_{kp}(t), \beta'_{k1}(t), \ldots, \beta'_{kp}(t)\big)^{\mathsf{T}}$ and $\widetilde{Z}_i(u, u-t) = Z_i(u) \otimes (1, u-t)^{\mathsf{T}}$, where $\otimes$ is the Kronecker product. Let

$$S_f^{(j)}(u,\xi) = \frac{1}{n}\sum_{i=1}^{n} Y_i(u)\exp\left(\widetilde{Z}_i(u,u-t)^{\mathsf{T}}\xi(t)\right)\left(\widetilde{Z}_i(u,u-t)\right)^{\otimes j}$$

for $j = 0, 1$ and 2. Here $a^{\otimes 0} = 1$, $a^{\otimes 1} = a$, and $a^{\otimes 2} = aa^{\mathsf{T}}$ for a column vector a. Following Cai and Sun (2003), the score function for $\xi_k(t)$ is

$$U_f(t,\xi_k) = \sum_{i=1}^{n}\int_0^{\tau} K_h(u-t)\Big(\widetilde{Z}_i(u,u-t) - \overline{S_f}(u,\xi_k)\Big)\, dN_{ik}(u), \tag{2.4}$$

where $\overline{S_f}(u,\xi) = S_f^{(1)}(u,\xi)/S_f^{(0)}(u,\xi)$, $K_h(\cdot) = K(\cdot/h)/h$, $K(\cdot)$ is a symmetric kernel function with support $[-1, 1]$, and h is the bandwidth. The local linear partial maximum likelihood estimator (mle) of $\beta_k(t)$ is the vector consisting of the first p components of $\widehat{\xi}_{f,k}(t)$ that solves (2.4) with respect to ξ_k.

2.2.3 Inverse probability weighted estimator

When there are missing causes in data, a naive method for estimating $\xi_k(t)$ is to simply ignore the missing data and use the local linear partial likelihood score equation (2.4) to fit the complete data only. Such complete-case estimator is inefficient and can lead to bias. Following the idea of Horvitz and Thompson (1952), the method of inversely weighting the probability of complete-case has been commonly used in missing data problems. To do this, we need to estimate the probability of a complete case, $\pi(Q_i) = \delta_i r(W_{1,i}) + (1-\delta_i)$. Let $r(W_{1,i}, \psi)$ be a parametric model for $r(W_{1,i})$, where $W_{1,i} = (T_i, Z_i(T_i), A_i^{(r)})$ and ψ is a q-dimensional vector of parameters. The maximum likelihood estimator $\widehat{\psi}$ of ψ is obtained by maximizing the observed data likelihood,

$$\prod_{i=1}^{n} \Big(r(W_{1,i}, \psi)\Big)^{R_i\delta_i} \Big(1 - r(W_{1,i}, \psi)\Big)^{(1-R_i)\delta_i}. \tag{2.5}$$

Let

$$S_I^{(j)}(u, \xi, \psi) = \frac{1}{n}\sum_{i=1}^{n} q_i Y_i(u) \exp\left(\widetilde{Z}_i(u, u-t)^{\mathsf{T}}\xi(t)\right)\left(\widetilde{Z}_i(u, u-t)\right)^{\otimes j},$$

for $j = 0, 1, 2$, where $q_i = R_i/\pi(Q_i, \psi)$. Denote $\overline{S_I}(u, \xi, \psi) = S_I^{(1)}(u, \xi, \psi)/S_I^{(0)}(u, \xi, \psi)$. Let $\widehat{\psi}$ be the mle maximizing the observed data likelihood (2.5). The inverse probability weighted (IPW) of complete-case estimating function for $\xi_k(t)$ is given by

$$U_I(t, \xi_k, \widehat{\psi}) = \sum_{i=1}^{n}\int_0^{\tau} K_h(u-t)\Big(\widetilde{Z}_i(u, u-t) - \overline{S_I}(u, \xi_k, \widehat{\psi})\Big)\widehat{q}_i \, dN_{ik}(u), \tag{2.6}$$

where $\widehat{q}_i = R_i/\pi(Q_i, \widehat{\psi})$. The IPW estimator $\widehat{\xi}_{I,k}(t)$ of ξ_k is the solution of the estimating equation $U_I(t, \xi_k, \widehat{\psi}) = 0$. Then $\widehat{\beta}_{I,k}(t)$ is the first p components of $\widehat{\xi}_{I,k}(t)$.

The baseline function $\lambda_{k0}(t)$ can be estimated by kernel smoothing $\widehat{\lambda}_{I,k0}(t) = \int_0^{\tau} K_h(u-t)\, d\widehat{\Lambda}_{I,k0}(u)$, where

$$\widehat{\Lambda}_{I,k0}(t) = \sum_{i=1}^{n}\int_0^{\mathsf{T}} \frac{1}{nS_I^{(0)}(u, \widehat{\xi}_{I,k}, \widehat{\psi})}\widehat{q}_i \, dN_{ik}(u)$$

is the estimator of the cumulative baseline function $\Lambda_{k0}(t) = \int_0^{\mathsf{T}} \lambda_{k0}(u)\, du$.

2.2.4 Augmented inverse probability weighted estimator

Studies have shown that the IPW estimator is inefficient and relies on correct modeling of the probability $r(W_{1,i}, \psi)$, cf. Scharfstein et al. (1999), Gao and Tsiatis (2005), and Lu and Liang (2008). To increase estimation efficiency, we propose the augmented inverse probability weighted complete-case estimating function obtained by subtracting the projection term of the simple weighted estimating function onto the tangent space. This methodology was advocated by Robins et al. (1994) and has been shown to be more efficient and enjoy double robustness property in a variety of situations, cf. Gao and Tsiatis (2005), Lu and Liang (2008), and Sun et al. (2017) among others. The proposed augmented inverse probability weighted complete-case estimating equation utilizes available information for individuals with missing causes through a consistent estimator of the conditional distribution of the cause, $\rho_k(W_2)$. The IPW estimators $\widehat{\lambda}_{I,k}(t)$ and $\widehat{\xi}_{I,k}(t)$ are used in the construction of this consistent estimator.

Let $W_{3,i} = (T_i, Z_i)$, $W_{2,i} = (W_{3,i}, \delta_i A_i^{(v)})$ and $\rho_k(w) = P(V_i = k | \delta_i = 1, W_{2,i} = w)$, where $w = (t, z, a)$. Under MAR II and MAR III, by Lemma A.1 given in the Appendix, we have

$$\rho_k(w) = \frac{\lambda_k(t|z) P(A_i^{(v)} = a | R_i = 1, \delta_i = 1, V_i = k, T_i = t, Z_i = z)}{\sum_{l=1}^K \lambda_l(t|z) P(A_i^{(v)} = a | R_i = 1, \delta_i = 1, V_i = l, T_i = t, Z_i = z)}. \tag{2.7}$$

Let $h(a|k, t, z) = P(A_i^{(v)} = a | R_i = 1, \delta_i = 1, V_i = k, T_i = t, Z_i = z)$. If $A_i^{(v)}$ is independent of V_i conditional on $(\delta_i = 1, T_i, Z_i)$, then $h(a|k, t, z)$ does not depend on k, and in this case $\rho_k(w) = \lambda_k(t|z) / \sum_{l=1}^K \lambda_l(t|z)$. This relationship also holds if there is no such candidate auxiliary $A_i^{(v)}$ available. In the situation that the auxiliary variable $A_i^{(v)}$ correlates V_i conditional on $(\delta_i = 1, T_i, Z_i)$, $\rho_k(w)$ depends on $h(a|k, t, z)$ as well as $\lambda_k(t|z)$ for $k = 1, \ldots, K$. Equation (2.7) shows how auxiliary variables can be utilized to improve efficiency. Although nonparametric/semiparametric density estimation methods are available, the development in the conditional density estimation is limited, in particular if the dimension is high (Hall et al., 2004; Efromovich, 2010; Izbicki and Lee, 2016). Here, we posit a parametric model $h(a|k, t, z, \theta_k)$ for $h(a|k, t, z)$, where $h(\cdot)$ is a known function and θ_k is a vector of unknown parameters. The maximum likelihood methods can be to used to obtain the estimator $\hat{\theta}_k$ of θ_k.

Let $\widehat{\lambda}_{I,k}(t|z) = \widehat{\lambda}_{I,k0}(t) \exp\left(\{\widehat{\beta}_{I,k}(t)\}^{\mathsf{T}} z\right)$ be the IPW estimator of the conditional cause-specific hazard function. Then $\rho_k(w)$ can be estimated by

$$\widehat{\rho}_k(w) = \frac{\widehat{\lambda}_{I,k}(t|z) h(a|k, t, z, \widehat{\theta}_k)}{\sum_{l=1}^{K} \widehat{\lambda}_{I,l}(t|z) h(a|l, t, z, \widehat{\theta}_l)}. \tag{2.8}$$

The augmented inverse probability weighted estimating function for $\xi_k(t)$ is obtained by subtracting the projection term of the simple weighted estimating function onto the tangent space as follows:

$$\begin{aligned} U_A(t, \xi_k, \widehat{\psi}, \widehat{\rho}_k) = \sum_{i=1}^{n} \int_0^{\tau} K_h(u - t) \Big(\widetilde{Z}_i(u, u - t) - \overline{S_f}(u, \xi_k)\Big) \Big[\widehat{q}_i \, dN_{ik}(u) \\ + (1 - \widehat{q}_i) \widehat{\rho}_k(W_{2,i}) \, dN_i(u)\Big]. \end{aligned} \tag{2.9}$$

The augmented inverse probability weighted estimator (AIPW) of $\xi_k(t)$ is the solution to the estimating equation $U_A(t, \xi_k, \widehat{\psi}, \widehat{\rho}_k) = 0$ and is denoted by $\widehat{\xi}_{A,k}(t)$. The AIPW estimator $\widehat{\beta}_{A,k}(t)$ of $\beta_k(t)$ is the first p components of $\widehat{\xi}_{A,k}(t)$.

The baseline function $\lambda_{k0}(t)$ can be estimated by $\widehat{\lambda}_{A,k0}(t) = \int K_h(u-t)\, d\widehat{\Lambda}_{A,k0}(u)$, where

$$\widehat{\Lambda}_{A,k0}(t) = \sum_{i=1}^{n} \int_0^{\mathsf{T}} \frac{1}{nS_f^{(0)}(u, \widehat{\xi}_{A,k})} \Big[\widehat{q}_i I(V_i = k) + (1 - \widehat{q}_i) \widehat{\rho}_k(W_{2,i}) \Big] dN_i(u)$$

is the estimator of the cumulative baseline function of $\Lambda_{k0}(t)$.

The AIPW estimator $\widehat{\beta}_{A,k}(t)$ can be considered as a two-stage estimator because the augmentation term of its estimation equation directly utilizes the IPW estimator (termed the first-stage estimator) through $\widehat{\rho}_k(w)$ in (2.8). The auxiliary variables are utilized through estimation of $h(a|k, t, z, \theta_k)$. Even if auxiliary variables are not available, the first-stage IPW estimator can be utilized further to improve efficiency through $\hat{\rho}_k(w) = \widehat{\lambda}_{I,k}(t|z) / \sum_{l=1}^{K} \widehat{\lambda}_{I,l}(t|z)$.

2.3 Asymptotic properties

In this section, we investigate the asymptotic properties of the IPW estimator $\widehat{\beta}_{I,k}(t)$ and the AIPW estimator $\widehat{\beta}_{A,k}(t)$. Let $\mathcal{F}_t$ be the right continuous filtration generated by the data processes $\{N_{ik}(s), Y_i(s), Z_i(s); i = 1, \ldots, n, k = 1, \ldots, K, 0 \leq s \leq t\}$. It follows that $M_{ik}(t) = N_{ik}(t) - \int_0^{\mathsf{T}} Y_i(u)\lambda_k(u|Z_i(u))du$, $i = 1, \ldots, n, k = 1, \ldots, K$, are multivariate orthogonal martingales with respect to $\mathcal{F}_t$, cf., Aalen and Johansen (1978). To accommodate additional information introduced due to missing data, we define the augmented filtration $\mathcal{F}_t^*$ generated by the data processes $\{N_{ik}(s), Y_i(s), Z_i(s), R_i, A_i; i = 1, \ldots, n, k = 1, \ldots, K, 0 \leq s \leq t\}$. Let $\lambda_{ik}^*(t)\, dt = P(T_i \in [t, t + dt), V_i = k | X_i \geq t, Z_i(t), R_i, A_i)$. Then $Y_i(t)\lambda_{ik}^*(t)$ is the intensity of $N_{ik}(t)$ with respect to $\mathcal{F}_t^*$, and $M_{ik}^*(t) = N_{ik}(t) - \int_0^{\mathsf{T}} Y_i(u)\lambda_{ik}^*(u)du$, $i = 1, \ldots, n, k =$

$1, \ldots, K$, are multivariate orthgonal martingales with respect to $\mathcal{F}_t^*$.

Additional notations are introduced in the following. Let $\nu_0 = \int K^2(x)dx$, $\mu_2 = \int x^2 K(x)dx$ and $P(t|Z_i(t)) = \Pr(X_i \geq t|Z_i(t))$. Let $s^{(j)}(t, \beta_k) = E\big[P(t|Z_i(t)) \exp\big(Z_i(t)^\mathsf{T}\beta_k(t)\big)\ (Z_i(t))^{\otimes j}\big]$ for $j = 0, 1, 2$, and define

$$\Sigma_k(t) = \big[s^{(2)}(t, \beta_k) - \big(s^{(1)}(t, \beta_k)\big)^{\otimes 2}/s^{(0)}(t, \beta_k)\big]\lambda_{k0}(t),$$

$$\Sigma_k^*(t) = E\big[\big(Z_i(t) - s^{(1)}(t, \beta_k)/s^{(0)}(t, \beta_k)\big)^{\otimes 2} R_i \pi^{-2}(Q_i) Y_i(t) \lambda_{ik}^*(t)\big].$$

2.3.1 Asymptotic results of the IPW estimator

The consistency and asymptotic normality of $\widehat{\beta}_{I,k}(t)$, $k = 1, 2, \ldots, K$, are established in the next two theorems. To avoid the problems at the boundaries $t = 0$ and $t = \tau$, we study the asymptotic properties of $\widehat{\beta}_k(t)$ for interior values of $t \in [t_1, t_2] \subset (0, \tau)$. The proofs of Theorems 2.1 and 2.2 are placed in the Appendix.

Theorem 2.1. *Under Condition A given in the Appendix, if the model for $r(W_{1,i})$ is correctly specified, then $\widehat{\beta}_{I,k}(t) \xrightarrow{\mathcal{P}} \beta_k(t)$ uniformly in $t \in [t_1, t_2] \subset (0, \tau)$ as $n \to \infty$.*

Theorem 2.2. *Under Condition A given in the Appendix, if the model for $r(W_{1,i})$ is correctly specified, then*

$$\sqrt{nh}\Big(\widehat{\beta}_{I,k}(t) - \beta_k(t) - \frac{1}{2}\mu_2 h^2 \beta_k''(t)\Big) \xrightarrow{\mathcal{D}} N\Big(0, \nu_0 \Sigma_k^{-1}(t)\Sigma_k^*(t)\Sigma_k^{-1}(t)\Big), \tag{2.10}$$

for $t \in [t_1, t_2] \subset (0, \tau)$ as $n \to \infty$.

Let $\mathcal{I}_{I,k}(t)$ be the upper left $p \times p$ matrix of

$$\frac{1}{n}\sum_{i=1}^n \int_0^\tau K_h(u-t) J_I(u, \widehat{\xi}_{I,k}(t), \widehat{\psi})\widehat{q}_i \, dN_{ik}(u),$$

and $\widetilde{\Sigma}_{I,k}(t)$ be the upper left $p \times p$ matrix of

$$n^{-1}h\sum_{i=1}^{n}\int_0^\tau \left(K_h(u-t)\right)^2\left(\widetilde{Z}_i(u,u-t)-\overline{S_I}(u,\widehat{\xi}_{I,k}(t),\widehat{\psi})\right)^{\otimes 2}\widehat{q}_i^2\,dN_{ik}(u),$$

where $J_I(u,\xi,\psi) = S_I^{(2)}(u,\xi,\psi)/S_I^{(0)}(u,\xi,\psi) - (\overline{S_I}(u,\xi,\psi))^{\otimes 2}$. Then the asymptotic variance $\nu_0\Sigma_k^{-1}(t)\Sigma_k^*(t)\Sigma_k^{-1}(t)$ of $\widehat{\beta}_{I,k}(t)$ can be consistently estimated by $\mathcal{I}_{I,k}^{-1}(t)\widetilde{\Sigma}_{I,k}(t)$ $\mathcal{I}_{I,k}^{-1}(t)$ as $n \to \infty$.

2.3.2 Asymptotic results of the AIPW estimator

Next, we present the asymptotic properties of the AIPW estimators $\widehat{\beta}_{A,k}(t),\ k = 1,2,\ldots,K$. Theorem 2.3 shows that the AIPW estimators are consistent if either $r(W_{1,i})$ or $h(A_i^{(v)}|k,T_i,Z_i)$ is correctly specified, a double robustness property. Theorem 2.4 shows the asymptotic normality of $\widehat{\beta}_{A,k}(t),\ k = 1,2,\ldots,K$. The proofs of Theorems 2.3 and 2.4 are placed in the Appendix.

Theorem 2.3. *Under Condition A given in the Appendix, $\widehat{\beta}_{A,k}(t)\xrightarrow{\mathcal{P}}\beta_k(t)$ uniformly in $t \in [t_1,t_2] \subset (0,\tau)$ as $n \to \infty$. This consistency holds if either $r(W_{1,i})$ or $h(A_i^{(v)}|k,T_i,Z_i)$ is correctly specified.*

Theorem 2.4. *Under Condition A given in the Appendix, if both $r(W_{1,i})$ and $h(A_i^{(v)}|k,$ $T_i,Z_i)$ are correctly specified, then*

$$\sqrt{nh}\left(\widehat{\beta}_{A,k}(t)-\beta_k(t)-\frac{1}{2}\mu_2h^2\beta_k''(t)\right)\xrightarrow{\mathcal{D}}N\left(0,\nu_0\Sigma_k^{-1}(t)\Sigma_k^*(t)\Sigma_k^{-1}(t)\right), \tag{2.11}$$

for $t \in [t_1,t_2] \subset (0,\tau)$ as $n \to \infty$.

Let $\mathcal{I}_{A,k}(t)$ be the upper left $p \times p$ matrix of

$$\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} K_h(u-t)J_f(u,\widehat{\xi}_{A,k}(t))\Big[\widehat{q}_i I(V_i=k)+(1-\widehat{q}_i)\widehat{\rho}_k(w)\Big]\,dN_i(u),$$

and $\widetilde{\Sigma}_{A,k}(t)$ be the upper left $p \times p$ matrix of

$$\frac{h}{n}\sum_{i=1}^{n}\int_0^{\tau}\big(K_h(u-t)\big)^2\Big(\widetilde{Z}_i(u,u-t)-\overline{S_f}(u,\widehat{\xi}_{A,k}(t))\Big)^{\otimes 2}\Big[\widehat{q}_i I(V_i=k)+(1-\widehat{q}_i)\widehat{\rho}_k(w)\Big]^2 \times dN_i(u),$$

where $J_f(u,\xi)=S_f^{(2)}(u,\xi)/S_f^{(0)}(u,\xi)-(\overline{S_f}(u,\xi))^{\otimes 2}$. The asymptotic variance $\nu_0\Sigma_k^{-1}(t)$ $\Sigma_k^*(t)\Sigma_k^{-1}(t)$ of $\widehat{\beta}_{A,k}(t)$ can be consistently estimated by $\mathcal{I}_{A,k}^{-1}(t)\widetilde{\Sigma}_{A,k}(t)\mathcal{I}_{A,k}^{-1}(t)$ as $n \to \infty$.

2.4 Numerical results

We present a simulation study conducted to evaluate the performance of the proposed methods. The AIPW estimator with missing causes is compared to the complete data estimator (CC) where the observations with missing causes are deleted from the analysis, and to the IPW estimator. These estimators are also compared to the full data likelihood estimator (FULL), which analyzes the simulated full dataset without missing causes.

We consider two failure causes with the two cause-specific hazard functions equal to

$$\lambda_1(t|Z)=0.3(t+1)^{-1}\exp(0.2Z), \quad \lambda_2(t|Z)=0.1(t+0.1)^{-1/2}\exp((t+0.1)^{1/2}Z), \tag{2.12}$$

where Z has a uniform distribution on $[0,1]$. All failure times are censored at the administration time $\tau = 2$. In addition, the random-censoring time C is generated from a uniform distribution on $[0,\ 10]$ which yields about 50% censoring level. A logistic regression model $\text{logit}\{r(W_1, \psi)\} = W_1^{\mathsf{T}}\psi$ is considered for missing causes, where $W_1 = (1, Z)^{\mathsf{T}}$. The percentages of missing causes are approximately 30%, 40% and 50% for $\psi = (1.4, -1)^{\mathsf{T}}$, $\psi = (1, -1)^{\mathsf{T}}$ and $\psi = (0.5, -1)^{\mathsf{T}}$, which are denoted by $(r30)$, $(r40)$ and $(r50)$, respectively.

We also generate a binary auxiliary covariate $A_i^{(v)}$ for the failure cause V_i from the following models for $h(a|k, t, z)$:

$$P(A_i^{(v)} = 1|V_i = k) = \frac{e^{a_k}}{1 + e^{a_k}}, \quad k = 1, 2. \tag{2.13}$$

The models allow $A_i^{(v)}$ depend on V_i, but not T_i and Z_i conditional on V_i, thus $h(a|k, t, z) = h(a|k)$. We examine the performance of AIPW estimators under four different levels of association between $A^{(v)}$ and V, by considering the settings $(a_1, a_2) = (1, 1)$, $(a_1, a_2) = (-1, 1)$, $(a_1, a_2) = (-2, 2)$, and $(a_1, a_2) = (-3, 3)$, which result in approximate Kendall's tau values of $0, 0.45, 0.75$ and 0.90, respectively. These four settings are denoted by (A0), (A1), (A2), and (A3), respectively. We note that $A_i^{(v)}$ is independent of V_i under the setting (A0), and the association between $A_i^{(v)}$ and V_i increases from (A1) to (A3).

Let $\widehat{h}(a|k)$ be the estimator of the conditional mass function $h(a|k)$ of $A_i^{(v)}$ given $V_i = k$. If follows from (2.7) that $\rho_k(w)$ is estimated by

$$\widehat{\rho}_k(w) = \frac{\widehat{\lambda}_{I,k}(t|z)\widehat{h}(a|k)}{\widehat{\lambda}_{I,1}(t|z)\widehat{h}(a|1) + \widehat{\lambda}_{I,2}(t|z)\widehat{h}(a|2)}, \quad \text{for } k = 1, 2, \tag{2.14}$$

where $\widehat{\lambda}_{I,k}(t|z)$, $k = 1, 2$, are the first stage IPW estimators.

To study the performance of the proposed IPW and AIPW estimators under misspecifications of the models $r(W_1, \psi)$ and/or $h(a|k)$, the simulations are conducted by positing a misspecified constant model $r_0 \in (0, 1)$ for $r(W_1, \psi)$ and/or by positing (A0) while the true setting is (A2). The estimators based on the correctly specified models of $r(W_1, \psi)$ and $h(a|k)$ are compared to those obtained when at least one of the two models are misspecified. We use IPW-c to denote the IPW estimator with the correctly specified model $r(W_1, \psi)$ for missing causes, and IPW-m for the IPW estimator with misspecified model for missing causes. AIPW-A2-c stands for the AIPW estimator under the setting (A2) with the correctly specified model for $r(W_1, \psi)$, and AIPW-A2-m stands for the AIPW estimator under the setting (A2) for the misspecified model for missing causes. AIPW-mA2-c stands for the AIPW estimator with misspecified $h(a|k)$ by assuming (A0) while the true setting is (A2) but correctly specified model $r(W_1, \psi)$, while AIPW-mA2-m is the AIPW estimator where both $h(a|k)$ and $r(W_1, \psi)$ are misspecified.

We use the Epanechnikov kernel $K(x) = 3/4(1 - x^2)I\{|x| \leq 1\}$. The bandwidth is selected via a 5-fold cross validation procedure. Specifically, we randomly divide the sample into 5 equally sized groups, say $(G_1, G_2, \cdots, G_5)$. The selected bandwidth is calculated using the formula $h_{opt} = \arg\min_h \sum_{l=1}^{5} PE_l(h)$, where

$$PE_l(h) = -\sum_{k=1}^{K} \sum_{i \in G_l} \int_0^{\tau} \left[(\widehat{\beta}_{A,k}^{(-l)}(t))^{\mathsf{T}} Z_i(t) - \log\left(\sum_{s \in G_l} Y_s(t) \exp\left((\widehat{\beta}_{A,k}^{(-l)}(t))^{\mathsf{T}} Z_s(t) \right) \right) \right] \times dN_{ik}(t).$$

Here, $\widehat{\beta}_{A,k}^{(-l)}(t)$ is the AIPW estimator based on the data excluding subjects in G_l. The

$PE_l(h)$ is a cross validation measure of the prediction error based on the minus of the log-partial likelihood function, cf., Tian et al. (2005). To increase the stability, this procedure is repeated 10 times. We tried a set of bandwidths from 0.1 to 0.5 and noted that the prediction error dropped dramatically at first and then remained stable when $h \geq 0.3$ (Figure 6). Figure 7 shows the perfomance of our proposed two-stage AIPW estimators using different bandwidths for sample size $n = 1200$. The bias is not sensitive to the bandwidth selection. The standard error gets smaller when using larger bandwidth. Moreover, the coverage probabilities, especially for large bandwidths ($h = 0.4$ and $h = 0.5$), are slightly lower than the nominal level, which implies the estimated standard errors underestimate the sample standard errors.

The simulation results for $\widehat{\beta}_{I,k}(t)$ and $\widehat{\beta}_{A,k}(t)$ under various correctly specified and misspecified models, along with the CC and the FULL estimators, are reported in Figures 1 to 5. The efficiency of an estimator is examined through the relative efficiency with respect to the FULL estimator which is the empirical standard deviation of the FULL estimator divided by the empirical standard deviation of the corresponding estimator over 500 simulations.

Figure 1 to 3 show the performance of estimators for sample size $n = 800, 1200$ and 2000. The complete-case estimator (CC) for $\beta(t)$ has large biases. The biases of the IPW estimator are small when the parametric model for $r(W_1, \psi)$ is correctly specified but very large when it is misspecified. The biases of the AIPW estimators are very small for all settings, even when both the parametric models are misspecified. Furthermore, the AIPW estimators are more efficient than the IPW estimators even when the auxiliary variable $A^{(v)}$ is independent of V. The 95% empirical confidence intervals have reasonable coverage probabilities (CP), with a small amount of undercoverage for estimators, more so for the CC estimator and the IPW-m estimator. Specifically, for the simulation with $n = 800$, the CP is around 94% in the early time and lower later. To estimate time-varying coefficients, we require enough number of failures within a small window for each time point. However, in our simulation setting, the censoring rate is around 50% and the missing probability for each cause is approximately 30%, which results in only about 140 failures with complete causes. And most events occur in the early stages. For larger sample size ($n = 1200$ and $n = 2000$), the coverage probability is closer to the nominal level (95%).

Figure 4 shows the results of the AIPW estimator under the setting (A2) of model (2.13) when the probability of missing causes is 30%, 40% and 50%. It shows that higher missing rate results larger standard error and thus smaller relative efficiency of the AIPW estimator.

Figure 5 shows the performance of the AIPW estimator under four different levels of association between the auxiliary variable $A^{(v)}$ and the cause V when the probability of missing causes is 30%. It indicates that performance improves as the association strengthens.

2.5 Analysis of the Mashi data

We apply our proposed method to the Mashi clinical trial data described in the introduction. We include in the analysis the subset of live-born infants with complete covariate information at delivery, which totals 1123 live-born infants out of the 1179 total live-births (95.3%) (where for twins only the first live-born infant is included).

Of the 107 infants who died over the first 18 months of life, 28 were HIV-related deaths, 21 were HIV-unrelated deaths, and 58 had missing death cause.

Considering the 20 covariates of the babies or their mothers, we used logistic regression and all subsets model selection (with criterion Mallows C_p) to select a model for predicting $r(W_1)$ as in Sun et al. (2012). The selected model included the following covariates: the infant had birth weight < 2.5 kilograms, the second randomization assignments of mom/baby was switched from Placebo/Placebo to Placebo/Nevirapine during the trial due to the DSMB recommendation, log 10 plasma viral load level of the mom at delivery, the infant had AZT toxicity, and whether the baby was hospitalized with a serious adverse event. We took the binary covariate of whether the infant received HAART (highly active ART therapy) as an auxiliary covariate $A^{(v)}$ and used a logistic regression model for $h(A^{(v)}|V = k)$ needed for $\rho_k(W_2)$.

We considered the following cause-specific Cox model with time varying coefficient $\lambda_k(t) = \lambda_{k0}(t)\exp\{\beta_k(t)z\}$, where z is the feeding strategy (1 for BF+AZT and 0 for FF defined in the first section). Figure 8 shows both the IPW and AIPW estimators and the 95% pointwise confidence bands of $\beta_k(t)$ for each cause with $k = 1$ for HIV-related death and $k = 2$ for HIV-unrelated death. Figure 8 also includes the estimation of logarithm of the hazard ratio for all-cause death using the method of Cai and Sun (2003). The estimations are evaluated over 100 evenly distributed grid points between 0 and 365 days. The bandwidth $h = 219$ days is chosen using the 5-fold cross-validation procedure. The IPW and the AIPW estimation of $\beta_k(t)$ are close but more different in early and later time for the follow-up study, especially for $k = 1$. The confidence bands of the IPW estimator is slightly wider as expected.

Figure 8 supports that BF+AZT had an effect on reducing HIV-related deaths compared to FF until about 200 days, and BF+AZT also had an effect, albeit weak, on reducing HIV-unrelated deaths compared to FF. However, after about 300 days the data suggest that the risk of HIV-related death may have been greater for the BF+AZT arm compared to the FF arm, whereas this is not the case for HIV-unrelated death. Figure 8 (c) shows BF+AZT reduces the risk of all-cause death compared to FF before 200 days, however, it elevates the risk after 200 days. Figures 8(a) and (c) also indicate a lack of fit of the Cox model for the first year of follow-up. The results are consistent with the original study results that showed that infants assigned to formula-feeding (FF) had a higher rate of all-cause mortality by age of 7 months than those assigned to BF+AZT (Thior et al., 2006).

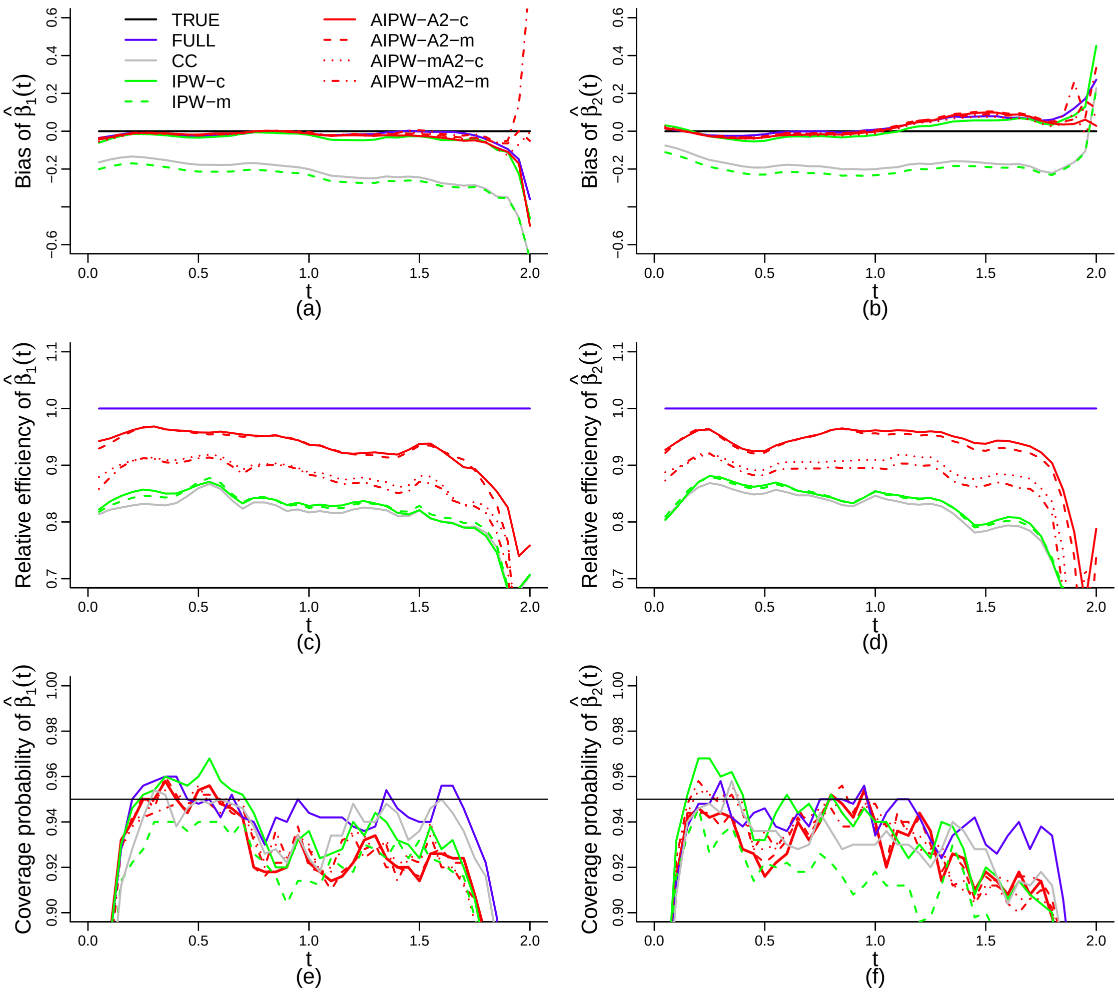


Figure 1: Bias, relative efficiency and coverage probability of the IPW and AIPW estimators for $\beta_1(t) = 0.2$ and $\beta_2(t) = \sqrt{t + 0.1}$ with 30% of missing causes under the correctly specified and misspecified models of $r(W_1, \psi)$ and $h(a|k)$, based on 500 simulations for $n = 800$ and $h = 0.3$. The legends AIPW-A2 and AIPW-mA2 refer to the AIPW estimators using the correctly specified (A2) and misspecified (A2), respectively, while -c and -m indicate the estimators using the correctly specified and misspecified model for $r(W_1)$, respectively. FULL is for the estimator based on the full data and CC is for the estimator based on the complete data only.

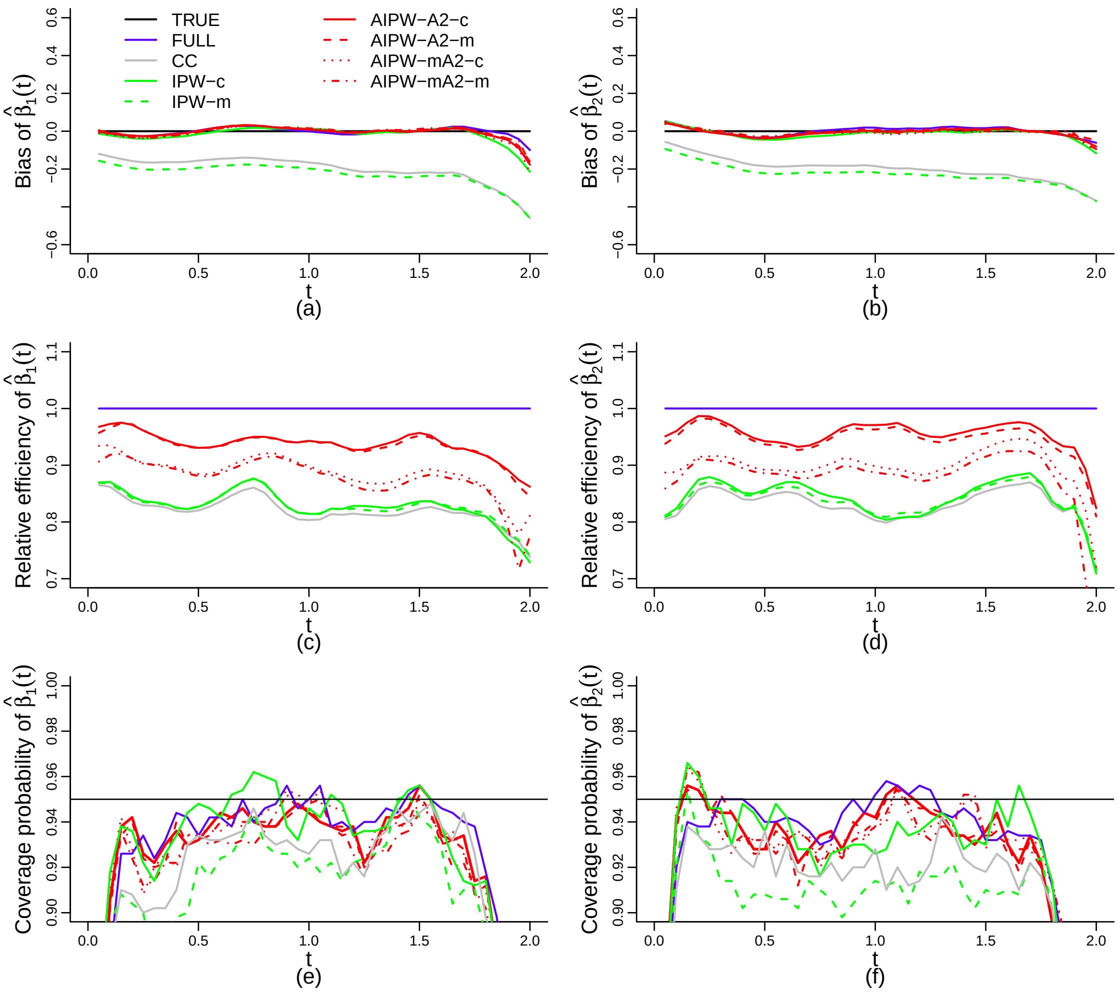


Figure 2: Bias, relative efficiency and coverage probability of the IPW and AIPW estimators for $\beta_1(t) = 0.2$ and $\beta_2(t) = \sqrt{t + 0.1}$ with 30% of missing causes under the correctly specified and misspecified models of $r(W_1, \psi)$ and $h(a|k)$, based on 500 simulations for $n = 1200$ and $h = 0.3$. The legends AIPW-A2 and AIPW-mA2 refer to the AIPW estimators using the correctly specified (A2) and misspecified (A2), respectively, while -c and -m indicate the estimators using the correctly specified and misspecified model for $r(W_1)$, respectively. FULL is for the estimator based on the full data and CC is for the estimator based on the complete data only.

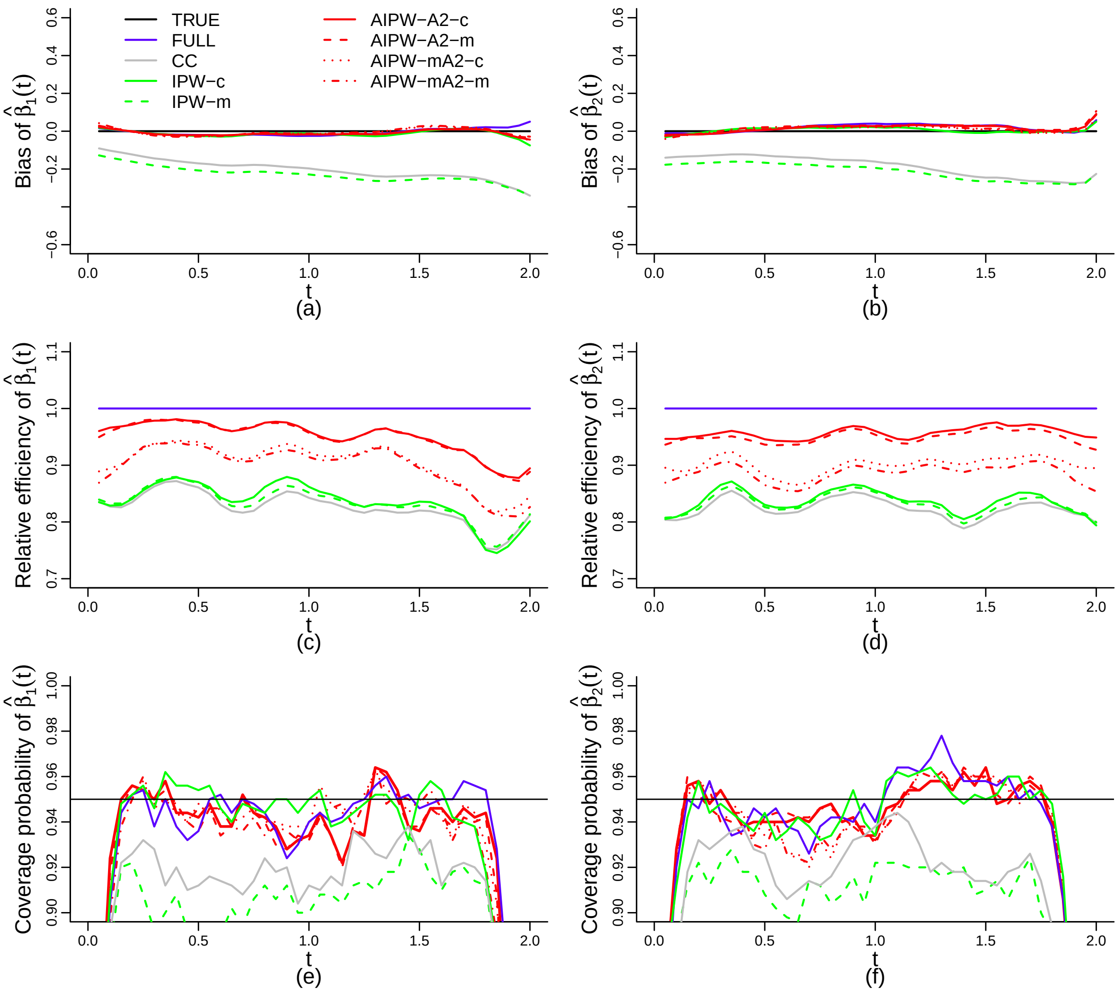


Figure 3: Bias, relative efficiency and coverage probability of the IPW and AIPW estimators for $\beta_1(t) = 0.2$ and $\beta_2(t) = \sqrt{t + 0.1}$ with 30% of missing causes under the correctly specified and misspecified models of $r(W_1, \psi)$ and $h(a|k)$, based on 500 simulations for $n = 2000$ and $h = 0.3$. The legends AIPW-A2 and AIPW-mA2 refer to the AIPW estimators using the correctly specified (A2) and misspecified (A2), respectively, while -c and -m indicate the estimators using the correctly specified and misspecified model for $r(W_1)$, respectively. FULL is for the estimator based on the full data and CC is for the estimator based on the complete data only.

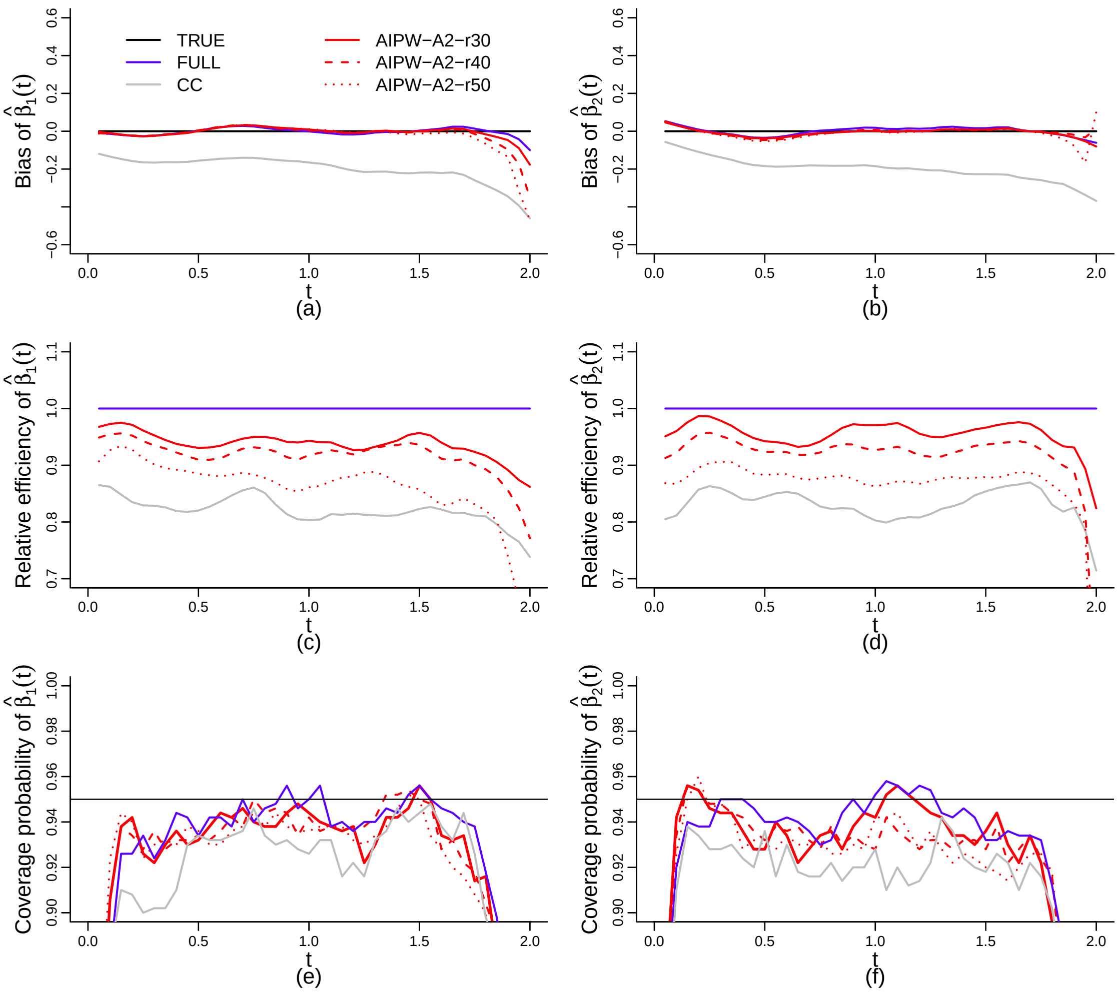


Figure 4: Bias, relative efficiency and coverage probability of the AIPW estimators for $\beta_1(t) = 0.2$ and $\beta_2(t) = \sqrt{t + 0.1}$ under the setting (A2) and when the percentages of missing causes are 30%, 40% and 50%, denoted by AIPW-A2-r30, AIPW-A2-r40, and AIPW-A2-r50, respectively, based on 500 simulations for $n = 1200$ and $h = 0.3$. FULL is for the estimator based on the full data and CC is for the estimator based on the complete data only.

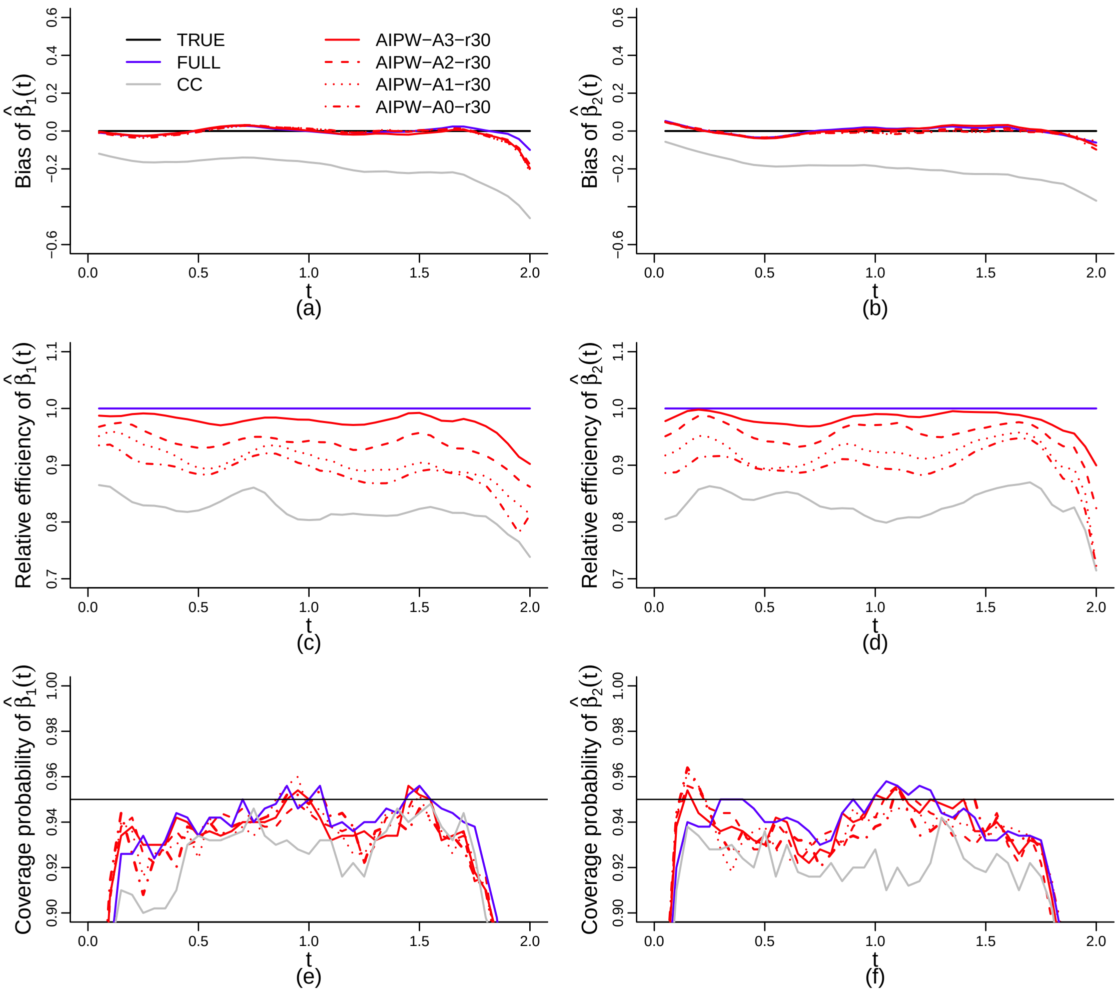


Figure 5: Bias, relative efficiency and coverage probability of the AIPW estimators for $\beta_1(t) = 0.2$ and $\beta_2(t) = \sqrt{t + 0.1}$ with 30% missing causes under the settings (A0), (A1), (A2) and (A3), based on 500 simulations for $n = 1200$ and $h = 0.3$. The legends AIPW-A0-r30, AIPW-A1-r30, AIPW-A2-r30 and AIPW-A3-r30 stand for the AIPW estimators for 30% missing causes and under the settings (A0), (A1), (A2) and (A3), respectively. FULL is for the estimator based on the full data and CC is for the estimator based on the complete data only.

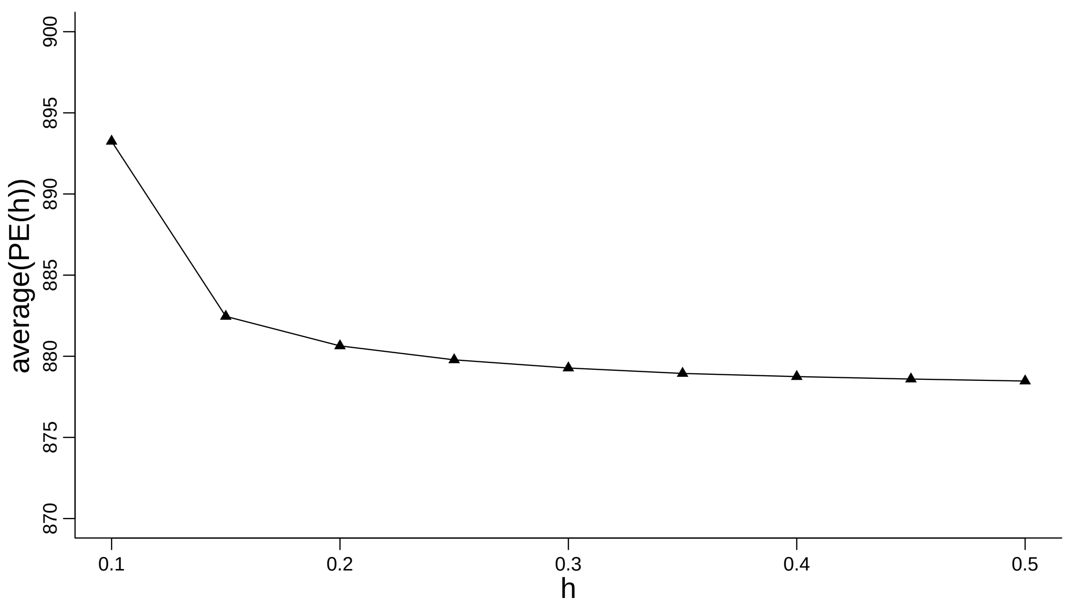


Figure 6: The average of the prediction errors from 10 simulations with sample size $n = 1200$ versus the bandwidth h. The simulation setting is (A2) and when the percentage of missing causes is 30%.

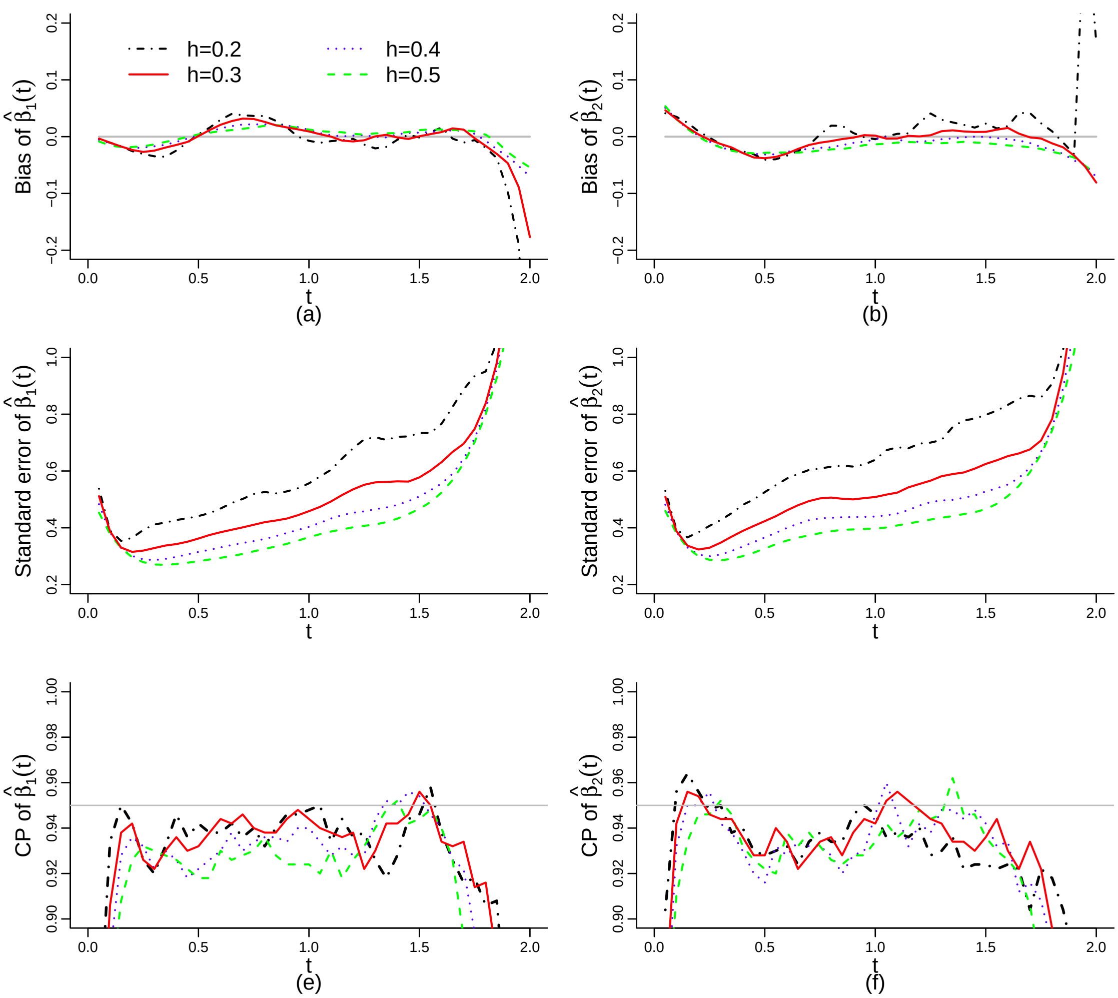
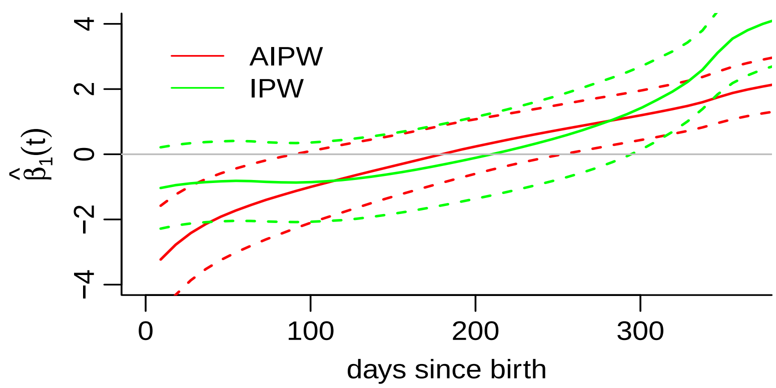
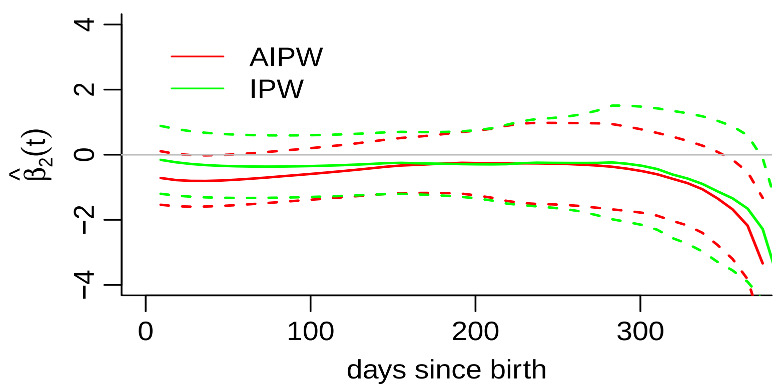


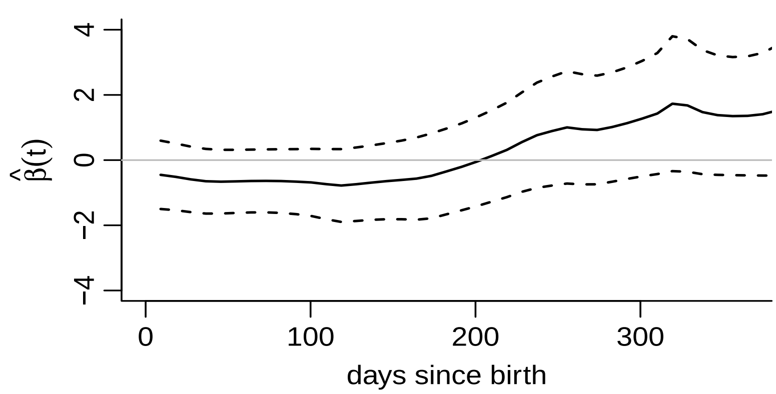
Figure 7: Bias, standard error and coverage probability of the AIPW estimators for $\beta_1(t) = 0.2$ and $\beta_2(t) = \sqrt{t+0.1}$ using different bandwidths with 30% of missing causes under the setting (A2) based on 500 simulations for $n = 1200$.



(a)



(b)



(c)

Figure 8: Estimation of $\beta_k(t)$ with 95% pointwise confidence bands for the Mashi randomized clinical trial: log hazard ratio (BF+AZT / FF) for (a) HIV-related death, (b) HIV-unrelated death and (c) all-cause death.

CHAPTER 3: GENERALIZED NONPARAMETRIC DYNAMIC INTENSITY MODELS FOR RECURRENT EVENT DATA

3.1 Introduction

The intensity-based models have been commonly used to analyze the recurrent event data. Andersen and Gill (1982) proposed a multiplicative intensity model, which is the extension of the Cox regression model to the recurrent event process. This model permits proportional effects of time-dependent covariates on the intensity function. Nielsen et al. (1992) introduced frailty variables into the multiplicative intensity model for event history data where the intensities may depend on unobservable random variables. The book by Andersen et al. (1993) is a thorough review of intensity-based models, covering both mathematical details and discussions of practical applications. Zeng and Lin (2006) developed nonparametric estimation procedures for a class of semiparametric transformation models which accommodates various cases of intensity-based models. They further studied this broad class of transformation intensity models with random effects for recurrent events (Zeng and Lin, 2007). Chen et al. (2013) proposed a semiparametric frailty model for overdispersed recurrent events data with treatment switching where EM algorithm was utilized to compute the maximum likelihood estimates.

The time since the last event, gap time, is another research interest when analyzing the recurrent event data. Many models and statistical inference methods for these models have been developed. Related literature includes, but is not limit to, Gill (1980), Gail et al. (1980), Prentice et al. (1981), Vardi (1982), Dabrowska et al. (1994), Lawless et al. (2001), Huang (2002), Chang (2004), Strawderman (2005), and Chang and Tzeng (2006). A comprehensive review of statistical models for recurrent event processes is given by Cook and Lawless (2007).

In this chapter, we simultaneously explore the time-varying and gap-time-varying effects of covariates on the intensity for recurrent event data under generalized nonparametric dynamic intensity models. This class of models is a mixture of Poisson and renewal-type models, but includes these models as specific cases. In the literature, Lawless and Thiagarajah (1996) modeled time trends and effects of past events through parametric Cox-type intensity models considering both the calendar time and gap time. Peña and Hollander (2004) represented a general and flexible class of dynamic recurrent event models that incorporate the effects of covariates, the impact on event counts, the influence of the backward recurrence time, as well as latent variables. The semiparametric estimation procedures are proposed in Peña et al. (2007) using a very general effective age process. Asymptotic properties of these semiparametric estimators are established by Peña (2016). However, there is still a need for developing nonparametric estimation procedures under a more general class of nonparametric dynamic intensity models, which includes multiplicative and additive intensity models as special cases.

Several studies, for instance Hoover et al. (1998), Cai et al. (2000), Cai and Sun (2003) and Sun and Wu (2005), have been carried out on varying-coefficient models for survival analysis and longitudinal studies. Scheike (2001) presented a nonparametric varying-coefficient survival model with two time-scales. A covariate-varying additive hazards model is proposed by Yin et al. (2008) to explore the nonlinear interactions between covariates. Qi et al. (2017) investigated a generalized semiparametric varying-coefficient model for longitudinal data that can flexibly model three types of covariate effects: time-constant effects, time-varying effects, and covariate-varying effects. They developed the estimation procedure via local linear smoothing and profile weighted least squares estimation techniques. For recurrent event process, Li et al. (2018) extended the aforementioned generalized semiparametric varying-coefficient model to the mean model. In Section 3.2.2, we employ the local linear smoothing method to estimate the mixed effects that include time-varying effects and the effects that may depend on past histories.

Our proposed model is driven by a malaria vaccine clinical trial. Malaria is a disease caused by the protozoan parasite Plasmodium, spread by infected mosquitoes. People at risk of malaria may experience recurrent malaria infections. In the trial, participants were randomized to several vaccine groups to measure the vaccine effects on preventing malaria infections. Since the protective immunity is not only induced by the malaria vaccine but also partially developed after malaria infections, we intend to investigate the time-varying change in vaccine efficacy and infection-induced immunogenicity through generalized intensity models that capture the dynamic features of recurrent event data. This clinical trial is still on the go. The method is expected to be applied to the Malaria vaccine clinical trial data once the trial is completed.

Nonparametric estimation procedure for the proposed model is developed in Section 3.2. The asymptotic properties of the proposed estimators, including the uniform consistency and weak convergence, are established in Section 3.3. Simulation studies are conducted and the results are discussed in Section 3.4.

3.2 Dynamic Statistical Models and Nonparametric Estimation

3.2.1 Notation and Model

In a random sample of n subjects, for subject i, the recurrent events occur at times $0 < T_{i,1} < T_{i,2} < \cdots$. Let $N_i(t) = \sum_{j=1}^{\infty} I(T_{i,j} \leq t)$ be the number of events recorded for the i-th subject by time t, where $I(\cdot)$ is the indicator function. Considering the censoring time C_i which is the minimum of a random censoring time and the end of follow-up τ, the observed ocurrence time is $0 < T_{i,1} < T_{i,2} < \cdots < T_{i,n_i} < C_i$, where n_i is the total number of events observed for subject i. The observed counting process can be written as $N_i^o(t) = \int_0^{\mathsf{T}} Y(s) dN_i(s)$, where $Y_i(t) = I(C_i \geq t)$ is the at-risk process. let $\mathcal{F}_t$ be the history, a σ-algebra generated by the counting process $N_i(t)$ and possible covariate processes. Conditional on $\mathcal{F}_{t^-}$, the history up to time t, the intensity function gives the instantaneous probability of an event occuring at t:

$$\lambda_i(t) = \lim_{\Delta t \downarrow 0} \frac{\Pr(\Delta N_i(t) = 1 | \mathcal{F}_{t^-})}{\Delta t},$$

where $\Delta N_i(t) = N_i(t + \Delta t^-) - N_i(t^-)$ is the number of events in the interval $[t, t + \Delta t)$. When events occurring in continuous time, we assume two events cannot occur simultaneously, therefore $E(dN_i(t)|\mathcal{F}_{t^-}) = \lambda_i(t)dt$. Let $\mathcal{G}_t = \mathcal{F}_t \vee \sigma(C_1, \ldots, C_n)$, a σ-algebra generated by $\mathcal{F}_t$ and censoring processes. We assume that the censoring is assumed to be independent in the sense that $E(dN_i(t)|\mathcal{G}_{t^-}) = E(dN_i(t)|\mathcal{F}_{t^-}) = \lambda_i(t)dt$. Let $\{\mathcal{F}_t^o\}$ is a σ-algebra generated by the observed data. Note that the at-risk processes $Y_i(t)$ are predictable with respect to $\mathcal{F}_t^o$. We assume $\lambda_i(t)$ is $\mathcal{F}_t^o$-predictable. Then, we have $E(dN_i^o(t)|\mathcal{F}_{t^-}^o) = Y_i(t)E(E(dN_i(t)|\mathcal{G}_{t^-})|\mathcal{F}_{t^-}^o) = Y_i(t)E(\lambda_i(t)dt|\mathcal{F}_{t^-}^o) = Y_i(t)\lambda_i(t)dt$.

Let $\{U_{i,k}(t), k = 1, \ldots, K\}$ be a series of $\mathcal{F}_t$-predictable covariate processes which depend on the past event history for subject i. It may relate to the number of events observed before time t, $N_i(t^-)$, or previous event times $\{T_{i,j} : T_{i,j} < t\}$. Then, we propose a general class of nonparametric dynamic intensity models for recurrent event processes:

$$\lambda_i(t) = g^{-1}\Big\{\alpha^\mathsf{T}(t)X_i(t) + \sum_{k=1}^{K} \gamma_k^\mathsf{T}(U_{i,k}(t))W_{i,k}(t)\Big\} \tag{3.1}$$

for $0 \le t \le C_i$, where $g(\cdot)$ is a known link function, $\alpha(\cdot)$ is a p_1-dimensional vector of unspecified functions and $\gamma_{i,k}(\cdot)$ are $p_{2,k}$-dimensional vectors of unspecified functions for $k = 1, \ldots, K$. $X_i(t)$ and $W_{i,k}(t)$ are corresponding time-dependent covariate processes. The class of dynamic models in Peña (2016) is included in (3.1) under these settings: (1) the link function is a logarithm function; (2) $U_{1,k}(t) = t - T_{i,N_i(t^-)}$ and $U_{2,k}(t) = N_i(t^-)$; (3) $\alpha(\cdot) = \alpha$ is a vector of time-constant coefficients; and (4) $\gamma_{i,2}(\cdot)$ takes a parametric form. Our proposed models (3.1) also contain a well-known self-exiting process, the Hawkes process:

$$\lambda_i(t) = \alpha + \sum_{k:T_{i,k}<t} \gamma_k(t - T_{i,k}).$$

More generally, similar to the idea in Zhang et al. (2013), we can extend time-dynamic coefficients $\gamma_k(U_{i,k}(t))$ to bivariate functions $\gamma_k(t, U_{i,k}(t))$ and have the following intensity models:

$$\lambda_i(t) = g^{-1}\Big\{\alpha^{\mathsf{T}}(t)X_i(t) + \sum_{k=1}^{K}\gamma_k^{\mathsf{T}}(t, U_{i,k}(t))W_{i,k}(t)\Big\}.$$

To estimate varying effects in scales of both time since a well-defined origin and gap times, we consider a special case of (3.1):

$$\lambda_i(t) = g^{-1}\Big\{\alpha^{\mathsf{T}}(t)X_i(t) + I(N_i(t^-) > 0)\gamma^{\mathsf{T}}(U_i(t))W_i(t)\Big\}, \tag{3.2}$$

where $U_i(t) = t - T_{i,N_i(t^-)}$ is the time since the last event. There is an additional indicator $I(N_i(t^-) > 0)$ here since the gap time $U_i(t)$ is only meaningful when the subject experienced at least one event in the history. Covariates $X_i(t)$ and $W_i(t)$ have time-varying and gap-time-varying effects on the occurrence of events respectively. Figure 9 gives an illustrative example of our model (3.2) in a simulation. We plot the following intensity using the solid curve:

$$\lambda_i(t) = 3 - \log(1+t) + 0.2t - \frac{0.513}{1 + t - T_{N_i(t^-)}} I(N_i(t^-) > 0).$$

It starts at $\lambda(0) = 3$ and drops 0.513 immediately after the occurrence of the first event. The decrease tends to zero as the gap time goes and is reseted to 0.513 when a new event occurred. At the end, this recurrent event process is censored by a random censoring time.

3.2.2 Local Linear Estimation

In the model (3.2), it is postulated that $\alpha(t)$ and $\gamma(u)$ have first and second derivatives $\dot{\alpha}(t)$, $\dot{\gamma}(u)$, $\ddot{\alpha}(t)$ and $\ddot{\gamma}(u)$. We locally parametrize $\alpha(t)$ in the neighborhood $\mathcal{N}_{t_0}$

of t_0, and $\gamma(u)$ in the neighborhood $\mathcal{N}_{u_0}$ of u_0 by the Taylor expansion:

$$\alpha(t) = \alpha(t_0) + \dot{\alpha}(t_0)(t - t_0) + O((t - t_0)^2)$$

and

$$\gamma(u) = \gamma(u_0) + \dot{\gamma}(u_0)(u - u_0) + O((u - u_0)^2).$$

For $t \in \mathcal{N}_{t_0}$ and $U_i(t) \in \mathcal{N}_{u_0}$, model (3.2) then can be approximated by

$$\lambda_i^*(t, \vartheta^*|t_0, u_0) = \varphi\Big\{\vartheta^{*\mathsf{T}}(t_0, u_0)\widetilde{X}_i^*(t|t_0, u_0)\Big\},$$

where $\varphi(\cdot) = g^{-1}(\cdot)$, $\vartheta^*(t_0, u_0) = (\alpha^{\mathsf{T}}(t_0), \gamma^{\mathsf{T}}(u_0), \dot{\alpha}^{\mathsf{T}}(t_0), \dot{\gamma}^{\mathsf{T}}(u_0))^{\mathsf{T}}$ and $\widetilde{X}_i^*(t|t_0, u_0) = (X_i(t)^{\mathsf{T}}, I(N_i(t^-) > 0)W_i(t)^{\mathsf{T}}, X_i(t)^{\mathsf{T}}(t - t_0), I(N_i(t^-) > 0)W_i(t)^{\mathsf{T}}(U_i(t) - u_0))^{\mathsf{T}}$.

Under the independent censoring assumption, the local log-likelihood function at t_0 and u_0 is:

$$\begin{aligned} l(\vartheta^*|t_0, u_0) =& \sum_{i=1}^n \int_0^\tau K_h(t - t_0)K_b(U_i(t) - u_0)Y_i(t)\Big[\log(\lambda_i^*(t, \vartheta^*|t_0, u_0))dN_i(t) \\ & - \lambda_i^*(t, \vartheta^*|t_0, u_0)dt\Big], \end{aligned}$$

where $K_h(\cdot) = K_1(\cdot/h)/h$ and $K_b(\cdot) = K_2(\cdot/b)/b$, $K_1(\cdot)$ and $K_2(\cdot)$ are kernel functions, and h and b are bandwidth parameters. By taking the derivative of the local log-likelihood function with respect to ϑ^*, we have the local score-type estimating equation with the bivariate kernel, a product of two univariate kernel functions:

$$\begin{aligned} U(\vartheta^*|t_0, u_0) =& \sum_{i=1}^n \int_0^\tau K_h(t - t_0)K_b(U_i(t) - u_0)Y_i(t)\frac{\dot{\lambda}_i^*(t, \vartheta^*|t_0, u_0)}{\lambda_i^*(t, \vartheta^*|t_0, u_0)}\Big\{dN_i(t) \\ & - \lambda_i^*(t, \vartheta^*|t_0, u_0)dt\Big\}\widetilde{X}_i^*(t|t_0, u_0), \end{aligned} \tag{3.3}$$

where $\dot{\lambda}_i^*(t, \vartheta^*|t_0, u_0) = \dot{\varphi}\Big\{\vartheta^{*\mathsf{T}}(t_0, u_0)\widetilde{X}_i^*(t|t_0, u_0)\Big\}$ and $\dot{\varphi}(\cdot)$ is the first derivative of $\varphi(\cdot)$ with respect to ϑ^*. The bivariate estimator $\hat{\vartheta}^*(t_0, u_0)$ can be obtained by solving $U(\vartheta^*|t_0, u_0) = 0$ through the Newton-Rapson method.

Let $\hat{\vartheta}(t_0, u_0)$ be first $p_1 + p_2$ components of $\hat{\vartheta}^*(t_0, u_0)$. Let $\hat{\alpha}(t_0, u_0)$ include the first p_1 elements of $\hat{\vartheta}(t_0, u_0)$ and let $\hat{\gamma}(t_0, u_0)$ be the vector including the elements of $\hat{\vartheta}(t_0, u_0)$ from $p_1 + 1$ to $p_1 + p_2$. Thus, $\hat{\vartheta}(t_0, u_0) = (\hat{\alpha}^{\mathsf{T}}(t_0, u_0), \hat{\gamma}^{\mathsf{T}}(t_0, u_0))^{\mathsf{T}}$. More efficient estimators for $\alpha^{\mathsf{T}}(t_0)$ and $\gamma^{\mathsf{T}}(u_0)$ can be achieved by aggregating the estimated bivariate functions $\hat{\alpha}(t_0, u_0)$ and $\hat{\gamma}(t_0, u_0)$ along each direction:

$$\hat{\alpha}(t_0) = n^{-1}\sum_{i=1}^{n}\hat{\alpha}(t_0, U_i(t_0)), \qquad \hat{\gamma}(u_0) = n_{u_0}^{-1}\sum_{t_{u_0}\in\mathcal{V}_{u_0}}\hat{\gamma}(t_{u_0}, u_0), \tag{3.4}$$

where $\mathcal{V}_{u_0} = \cup_{i=1}^{n} U_i^{-1}(u_0)$, $U_i^{-1}(u_0) = \{t : U_i(t) = u_0\}$, and $n_{u_0} = |\mathcal{V}_{u_0}|$, the cardinality of $\mathcal{V}_{u_0}$.

3.2.3 Bandwidth Selection

In nonparametric estimation procedures, the bandwidth selection requires caution. We select the optimal bandwidths via the K-fold cross validation method:

$$(h_{opt}, b_{opt}) = \arg\min_{h,b}\sum_{l=1}^{K} PE_l(h, b).$$

The cross validation measure of the prediction error, $PE_l(h, b)$, is constructed based on the minus of the log-partial likelihood function (Tian et al., 2005):

$$PE_l(h, b) = -\sum_{i\in G_l}\int_0^{\tau} Y_i(t)\Big\{\log(\widehat{\lambda}_i^{(-l)}(t))dN_i(t) - \widehat{\lambda}_i^{(-l)}(t)dt\Big\},$$

where $(G_1, G_2, \cdots, G_K)$ are K equally-divided subsamples, $\widehat{\lambda}_i^{(-l)}(t) = \widehat{\alpha}^{(-l)\mathsf{T}}(t)X_i(t) + I(N_i(t^-) > 0)\widehat{\gamma}^{(-l)\mathsf{T}}(U_i(t))W_i(t)$, and $\widehat{\alpha}^{(-l)}(\cdot)$ and $\widehat{\gamma}^{(-l)}(\cdot)$ are local linear estimators based on the data excluding subjects in G_l. We use the Epanechnikov kernel $K(x) = 3/4(1-x^2)I\{|x| \leq 1\}$ in numerical studies.

3.3 Uniform Consistency and Weak Convergence

In this section, we focus on the large-sample properties of local linear estimators for the parameters of the generaliazed dynamic intensity model described in Section 3.2. Let $\alpha_0(\cdot)$ and $\gamma_0(\cdot)$ be the true parameter vector. $[t_1, t_2]$ is a subinterval of $(0, \tau)$. Let $\mathcal{I}_1 = \{\mathcal{I}_{jk}\}$ be a $p_1 \times (p_1 + p_2)$ matrix with $\mathcal{I}_{jk} = 1$ for $j = 1, \ldots, p_1$ and $k = j$, and $\mathcal{I}_{jk} = 0$ otherwise. Let $\mathcal{I}_2 = \{\mathcal{I}_{jk}\}$ be a $p_2 \times (p_1 + p_2)$ matrix with $\mathcal{I}_{jk} = 1$ for $j = 1, \ldots, p_2$ and $k = j + p_1$, and $\mathcal{I}_{jk} = 0$ otherwise. Define $\widetilde{X}_i(t) = (X_i(t)^\mathsf{T}, I(N_i(t^-) > 0)W_i(t)^\mathsf{T})^\mathsf{T}$, $\hat{\lambda}_i(t) = \varphi\Big\{\hat{\alpha}(t)^\mathsf{T} X_i(t) + I(N_i(t^-) > 0)\hat{\gamma}(U_i(t))^\mathsf{T} W_i(t)\Big\}$, and $\hat{\dot{\lambda}}_i(t) = \dot{\varphi}\Big\{\hat{\alpha}(t)^\mathsf{T} X_i(t) + I(N_i(t^-) > 0)\hat{\gamma}(U_i(t))^\mathsf{T} W_i(t)\Big\}$. Let $\lambda_{i,0}(t,u) = \varphi\Big\{\alpha_0(t)^\mathsf{T} X_i(t) + I(N_i(t^-) > 0)\gamma_0(u)^\mathsf{T} W_i(t)\Big\}$, $\dot{\lambda}_{i,0}(t,u) = \dot{\varphi}\Big\{\alpha_0(t)^\mathsf{T} X_i(t) + I(N_i(t^-) > 0)\gamma_0(u)^\mathsf{T} W_i(t)\Big\}$, and $\ddot{\lambda}_{i,0}(t,u) = \ddot{\varphi}\Big\{\alpha_0(t)^\mathsf{T} X_i(t) + I(N_i(t^-) > 0)\gamma_0(u)^\mathsf{T} W_i(t)\Big\}$. Then $dM_i(t) = dN_i(t) - \lambda_{i,0}(t, U_i(t))dt$ is an $\mathcal{F}_t$-martingale under independent censoring assumption. Finally, we define

$$D(t,u) = E\left[Y_i(t) \frac{\dot{\lambda}_{i,0}(t,u)^2}{\lambda_{i,0}(t,u)} \widetilde{X}_i(t)^{\otimes 2} \middle| U_i(t) = u \right] f_U(t,u),$$

where $f_U(t,u)$ is the density function of the gap time process $U_i(t)$ at u.

The following theorems establish the uniform consistency and weak convergence of our proposed estimators. Conditions and proofs are outlined in the Appendix.

Theorem 3.1. *Under Condition B in the Appendix B, we have that:*

(a) $\sup_{t \in [t_1, t_2]} |\hat{\alpha}(t) - \alpha_0(t)| = o_p(1)$*;*

(b) $\sqrt{nh}(\hat{\alpha}(t)-\alpha_0(t)-\frac{1}{2}h^2\mu_2\ddot{\alpha}(t))\xrightarrow{\mathcal{D}}N\,(0,\Sigma_\alpha(t))$, *for* $t\in[t_1,t_2]$,

where $\Sigma_\alpha(t)=\nu_0\mathcal{I}_1E\{D^{-1}(t,U_i(t))\}\mathcal{I}_1^T$, $\mu_2=\int u^2K(u)\,du$ *and* $\nu_0=\int K^2(u)\,du$.

The covariance matrix $\Sigma_\alpha(t)$ can be consistently estimated by

$$\hat{\Sigma}_\alpha(t)=\frac{h}{n}\sum_{i=1}^{n}\int_0^\tau K_h(s-t)^2Y_i(s)\frac{\hat{\dot{\lambda}}_i(s)^2}{\hat{\lambda}_i(s)^2}\left\{\mathcal{I}_1\hat{D}^{-1}(t,U_i(s))\widetilde{X}_i(s)\right\}^{\otimes 2}dN_i(s),$$

where

$$\hat{D}(t_0,u_0)=\frac{1}{n}\sum_{i=1}^{n}\int_0^\tau K_h(t-t_0)K_b(U_i(t)-u_0)Y_i(t)\frac{\hat{\dot{\lambda}}_i(t)^2}{\hat{\lambda}_i(t)}\widetilde{X}_i(t)^{\otimes 2}dt.$$

Theorem 3.2. *Under Condition B in the Appendix B, we have that:*

(a) $\sup_{u\in[u_1,u_2]}\|\hat{\gamma}(u)-\gamma_0(u)\|=o_p(1)$;

(b) $\sqrt{nb}(\hat{\gamma}(u)-\gamma_0(u)-\frac{1}{2}b^2\mu_2\ddot{\gamma}(u))\xrightarrow{\mathcal{D}}N\,(0,\Sigma_\gamma(u))$, *for* $u\in[u_1,u_2]$,

where

$$\Sigma_\gamma(u)=\nu_0\mathcal{I}_2\Big\{\int_0^\tau D^{-1}(t,u)dt\Big\}\mathcal{I}_2^T.$$

The asymptotic covariance matrix $\Sigma_\gamma(u)$ can be consistently estimated by

$$\hat{\Sigma}_\gamma(u)=\frac{b}{n}\sum_{i=1}^{n}\int_0^\tau K_b(U_i(s)-u)^2Y_i(s)\frac{\hat{\dot{\lambda}}_i(s)^2}{\hat{\lambda}_i(s)^2}\left\{\mathcal{I}_2\hat{D}^{-1}(s,u)\widetilde{X}_i(s)\right\}^{\otimes 2}dN_i(s).$$

3.4 Simulation

We conduct simulation studies to examine the finite sample properties of our proposed estimators. For subject i, the censoring time $C_i=\min(\tau,C_i^*)$, where the follow-up time $\tau=4$ and C_i^* is from the Uniform$(3,8)$ distribution. We consider time-constant covariates here for simplicity. Recurrent events are generated from the following intensity function:

$$\lambda_i(t) = g^{-1}\Big\{\alpha_0(t) + \alpha_1(t)X_i + \gamma(t - T_{N_i(t^-)})I(N_i(t^-) > 0)W_i\Big\} \tag{3.5}$$

for $0 < t < C_i$, where X_i follows a Bernoulli distribution with probability of success $p = 0.5$ and W_i is a uniform random variable on $[0, 1]$. Our generation algorithm (1) for recurrent event process is based on the thinning method (Lewis and Shedler, 1979):

Algorithm 1: Simulation of generating recurrent events under intensity function 3.5 on $(0, C_i]$

Generate covariates X_i and W_i;

Set the highest intensity $\bar{\lambda} = \sup_{0 \le t \le T} g^{-1}\{\alpha_0(t) + \alpha_1(t)X_i\}$;

Initialize $n = 0, T_0 = 0, u = 0, s \sim exponential(\bar{\lambda})$;

/* n: the number of events, T_k: the occurrence time of the $k-$th event, u: time since last event, s: the event time generated from $\bar{\lambda}$. */

while $s \le C_i$ **do**

  $u = s - T_n$;

  Calculate $\lambda_i(s) = g^{-1}\{\alpha_0(s) + \alpha_1(t)X_i + \gamma(u)I(u > 0)W_i\}$;

  Generate $D \sim uniform(0, 1)$;

  if $D \le \lambda(s)/\bar{\lambda}$ **then**

    $n = n + 1$;

    $T_n = s$;

    $\bar{\lambda} = \sup_{T_n \le t \le T; 0 \le u \le T - T_n} g^{-1}\{\alpha_0(t) + \alpha_1(t)X_i + \gamma(u)I(u > 0)W_i\}$;

  end

  Generate $w \sim exponential(\bar{\lambda})$;

  $s = s + w$;

end

return *arrival times* $\{T_k\}_{k=1,2,\ldots,n}$

Under model (3.5), we consider two different link functions, the identity link and the logarithm link which yield additive intensity model and multiplicative intensity model respectively:

$$\lambda_i(t) = \alpha_0(t) + \alpha_1(t)X_i + \gamma(t - T_{N_i(t^-)})I(N_i(t^-) > 0)W_i. \tag{3.6}$$

$$\lambda_i(t) = \exp\{\alpha_0(t) + \alpha_1(t)X_i + \gamma(t - T_{N_i(t^-)})I(N_i(t^-) > 0)W_i\}, \tag{3.7}$$

In the additive intensity model (3.6), we consider $\alpha_0(t) = 4-\log(1+t)$, $\alpha_1(t) = -1+0.2t$, and $\gamma(u) = -1/(1+u)$. The average number of recurrent events for each subject is around 6.2 for $X_i = 1$ and around 9.7 for $X_i = 0$. Figure 10 summarizes biases (Bias), empirical standard errors (SEE), average estimated standard errors (ESE) and the 95% empirical coverage probabilities (CP) of our proposed estimators $\hat{\alpha}_0(t)$, $\hat{\alpha}_1(t)$, and $\hat{\gamma}(u)$ under model (3.6) for $n = 400, 600$ and 800 based on 500 simulations. The estimators have small biases, the average estimated standard errors are close to the empirical standard errors, and the 95% empirical coverage probabilities are around the nominal level.

Because the logarithm link yields a larger intensity than the identity link, we chose $\alpha_0(t) = 1 - \log(1 + 0.2\log(1 + t))$, $\alpha_1(t) = -0.5 + 0.1t$, and $\gamma(u) = -0.3/(1 + u)$ in the multiplicative intensity model (3.7). Under this setting, the average number of recurrent events for each subject is around 4.1 for $X_i = 1$ and around 9.5 for $X_i = 0$. Figure 11 summarizes Bias, SEE, ESE and 95% CP of $\hat{\alpha}_0(t)$, $\hat{\alpha}_1(t)$, and $\hat{\gamma}(u)$ under model (3.7) for $n = 400, 600$ and 800 based on 500 simulations. These estimators perform reasonably well, especially for large sample size.

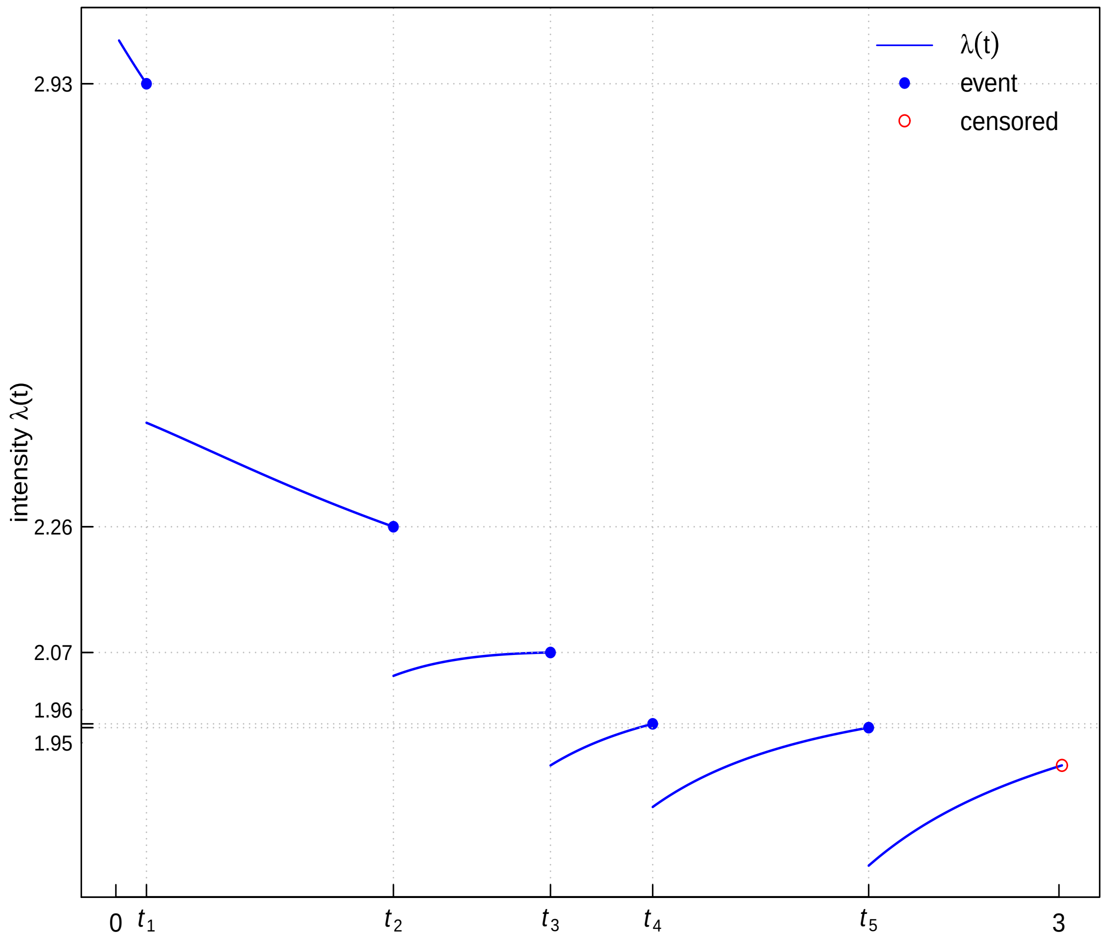


Figure 9: An illustrative example of simulating recurrent event process with the intensity function $\lambda_i(t) = 3 - \log(1+t) + 0.2t - 0.513/(1+t-T_{N_i(t^-)})I(N_i(t^-) > 0)$.

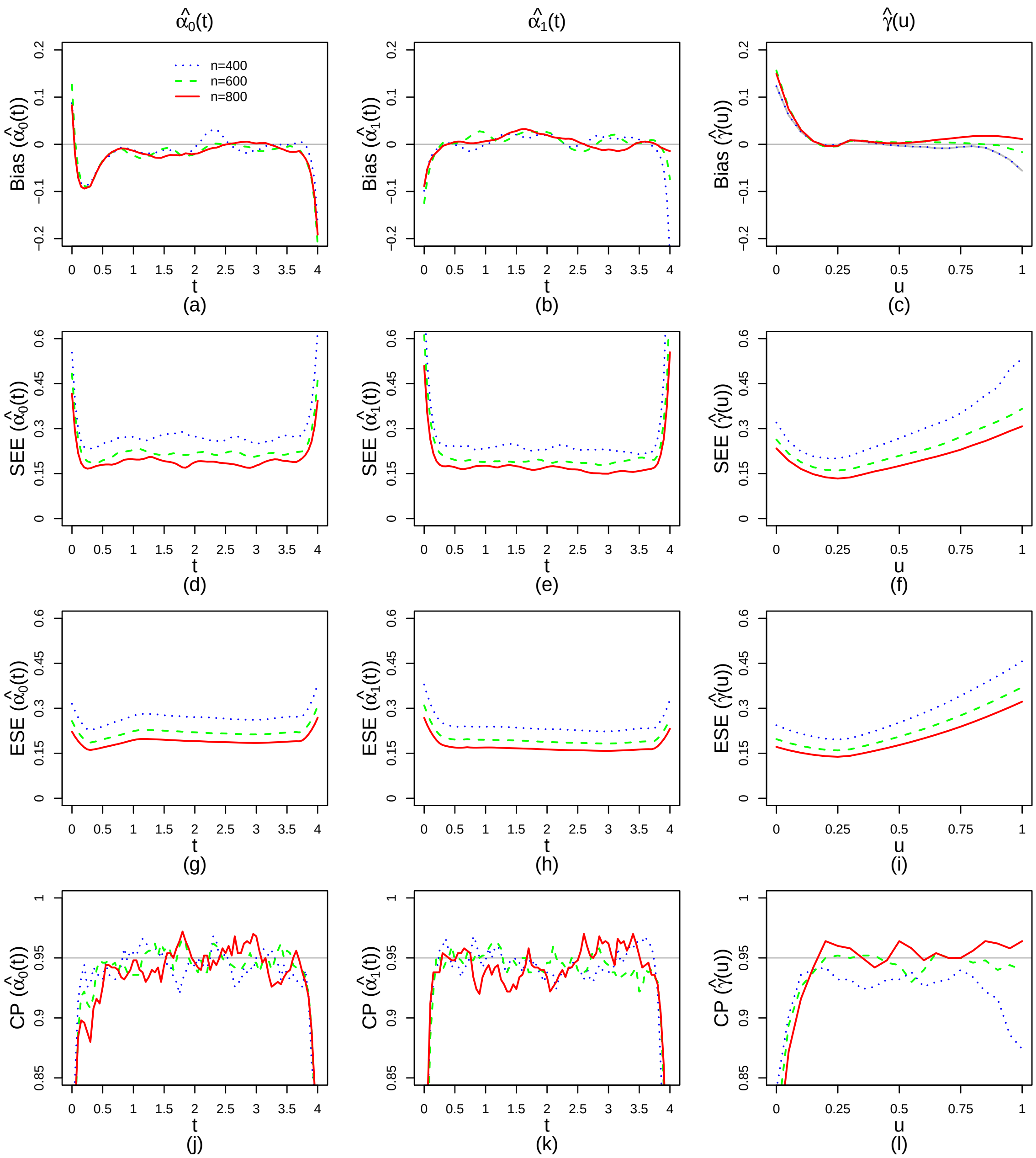


Figure 10: Plots of Bias, SEE, ESE and CP of $\hat{\alpha}_0(t)$, $\hat{\alpha}_1(t)$ and $\hat{\gamma}(u)$ under model (3.6) for $n = 400, 600$ and 800 using bandwidths $(h, b) = (0.3, 0.3)$ based on 500 simulations. The left panel is for $\hat{\alpha}_0(t)$, the middle panel is for $\hat{\alpha}_1(t)$, and the right panel is for $\hat{\gamma}(u)$. The blue dotted line is for $n = 400$, the green dashed line is for $n = 600$ and the red solid line is for $n = 800$.

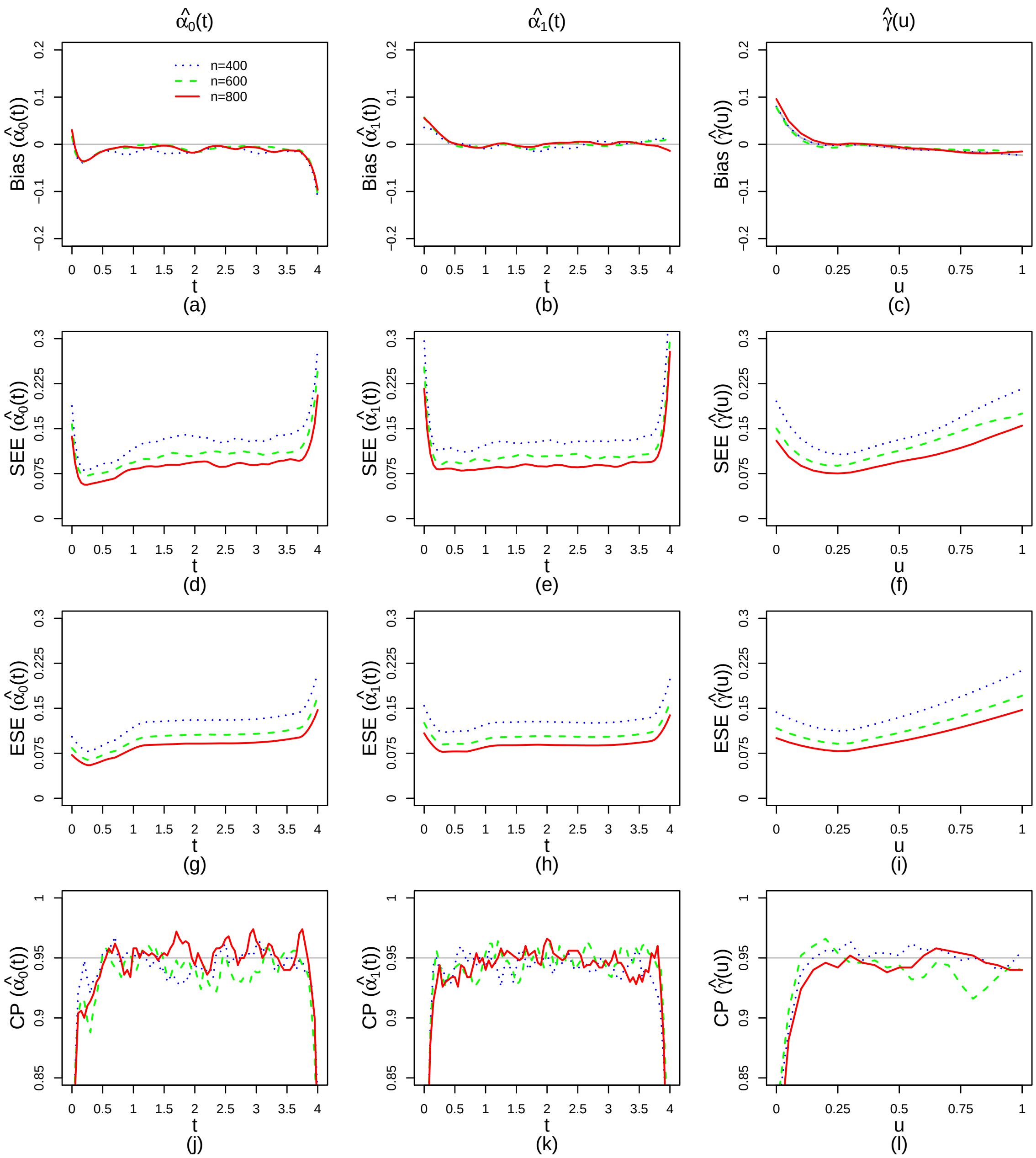


Figure 11: Plots of Bias, SEE, ESE and CP of $\hat{\alpha}_0(t)$, $\hat{\alpha}_1(t)$ and $\hat{\gamma}(u)$ under model (3.7) for $n = 400, 600$ and 800 using bandwidths $(h, b) = (0.3, 0.3)$ based on 500 simulations. The left panel is for $\hat{\alpha}_0(t)$, the middle panel is for $\hat{\alpha}_1(t)$, and the right panel is for $\hat{\gamma}(u)$. The blue dotted line is for $n = 400$, the green dashed line is for $n = 600$ and the red solid line is for $n = 800$.

CHAPTER 4: NONPARAMETRIC DYNAMIC ADDITIVE INTENSITY MODELS FOR RECURRENT EVENT DATA WITH MISSING COVARIATES

In Chapter 3, we propose nonparametric estimation procedures for a broad class of dynamic intensity models (3.2) which capture the dynamic features of recurrent event data. Missing covariates are frequently encountered in epidemiologic and clinical research. In this chapter, we investigate a specific missing covariate problem in recurrent event data under a nonparametric dynamic additive intensity model where the covariate with gap-time varying effects is subject to missingness. We organize Chapter 4 as follows: In Section 4.1, we propose a weighted estimating equation instead of the local score-type function and then develop the AIPW estimators. We establish the asymptotic properties of our proposed estimators in Section 3. The simulation studies are reported in Section 4.

4.1 Methodology

4.1.1 Notation and Model

Suppose that a subject i in a random sample may experience a sequence of n_i recurrent events at times $\{T_{i,j} : j = 1, 2, \ldots, n_i\}$ in an observation window $[0, C_i]$, where $T_{i,j}$ is the j-th recurrent event time and C_i is the censoring time. Let $N_i(t) = \sum_{j=1}^{\infty} I(T_{i,j} \leq t)$ be the counting process and $Y_i(t) = I(C_i \geq t)$ the at-risk process. Define $U_i(t) = t - T_{N_i(t^-)}$. We consider the following dynamic additive intensity model:

$$\lambda_i(t) = \alpha(t)X_i + \gamma(U_i(t))I(N_i(t^-) > 0)W_i, \tag{4.1}$$

where $\alpha(\cdot)$ and $\gamma(\cdot)$ are p_1-dimensional and p_2-dimensional vectors of unspecified functions respectively. $\alpha(t)$ represents the time varying effects of covariate X_i which is fully observed. $\gamma(U_i(t))$, the effect of potentially missing covariate W_i, depends on the time since the last event. Let R_i be the missing indicator that equals to zero if W_i is missing, and one otherwise. $\{N_i(t), Y_i(t), X_i, W_i, A_i, R_i\}$ are independent and identically distributed (iid) for subject $i = 1, 2, \ldots, n$, where A_i denotes the auxiliary covariate that may help to predict the missing probability and the missing covariate W_i. We assume the censoring is independent of the counting process given the covariate history. Let (Ω_i, R_iW_i, R_i) be the observed data, where $\Omega_i = (n_i, T_{i,1}, T_{i,2}, \ldots, T_{i,n_i}, C_i, X_i, A_i)$. That is, W_i is only observed for a subject with $R_i = 1$. We assume W_i is missing at random (MAR): $P(R_i = 1|W_i, \Omega_i) = P(R_i = 1|\Omega_i)$. Denote $P(R_i = 1|\Omega_i)$ by $\pi(\Omega_i)$.

4.1.2 Estimation Procedure

In Chapter 3, we propose a local linear partial likelihood estimation method for the full data. Under the identity link, the local score-type estimating equation for model (3.2) is:

$$\begin{aligned} U(\vartheta^*|t_0, u_0) = &\sum_{i=1}^{n} \int_0^\tau K_h(t - t_0)K_b(U_i(t) - u_0)Y_i(t)\frac{1}{\lambda_i^*(t, \vartheta^*|t_0, u_0)}\Big\{dN_i(t) \\ &- \lambda_i^*(t, \vartheta^*|t_0, u_0)dt\Big\}\widetilde{X}_i^*(t|t_0, u_0), \end{aligned} \tag{4.2}$$

where $\lambda_i^*(t, \vartheta^*|t_0, u_0) = \vartheta^{*\mathsf{T}}(t_0, u_0)\widetilde{X}_i^*(t|t_0, u_0)$, $\vartheta^*(t_0, u_0) = (\alpha^\mathsf{T}(t_0), \gamma^\mathsf{T}(u_0), \dot{\alpha}^\mathsf{T}(t_0),$ $\dot{\gamma}^\mathsf{T}(u_0))^\mathsf{T}$ and $\widetilde{X}_i^*(t|t_0, u_0) = (X_i(t)^\mathsf{T}, I(N_i(t^-) > 0)W_i(t)^\mathsf{T},\ X_i(t)^\mathsf{T}(t - t_0), I(N_i(t^-) >$

$0)W_i(t)^{\mathsf{T}}(U_i(t)-u_0))^{\mathsf{T}}$.

In this section, we propose estimation methods for the dynamic additive intensity model (4.1) that utilize the augmented inverse probability weighting technique to handle missing covariates (Robins et al., 1994). Since the desired intensity is in the denominator of (4.2), the calculation for the estimates of ϑ^* is often not stable, especially for a small sample. Also, it is unmanageable to apply the AIPW method. Considering the numerical stability, we replace the inverse of the intensity in (4.2), $1/\lambda_i^*(t,\vartheta^*|t_0,u_0)$, by a general weight function $w_i(t)$ to give a local linear weighted estimating equation for the full data:

$$\begin{aligned} U(\vartheta^*|t_0,u_0) = &\sum_{i=1}^{n}\int_0^{\tau} K_h(t-t_0)K_b(U_i(t)-u_0)Y_i(t)w_i(t)\Big\{dN_i(t)-\lambda_i^*(t,\vartheta^*|t_0,u_0)dt\Big\} \\ &\times \widetilde{X}_i^*(t|t_0,u_0). \end{aligned} \tag{4.3}$$

Based on (4.3), we propose the local AIPW estimating function for ϑ^*:

$$\begin{aligned} U_{A0}(\vartheta^*|t_0,u_0) = &\sum_{i=1}^{n}\int_0^{\tau} K_h(t-t_0)K_b(U_i(t)-u_0)Y_i(t)w_i(t)\Big[q_i\Big\{dN_i(t) \\ &-\lambda_i^*(t,\vartheta^*|t_0,u_0)dt\Big\}\widetilde{X}_i^*(t|t_0,u_0)+(1-q_i)d\epsilon_i^*(t|t_0,u_0)\Big], \end{aligned} \tag{4.4}$$

where $q_i = R_i/\pi(\Omega_i)$ is the inverse probability weight and $d\epsilon_i^*(t,\vartheta^*|t_0,u_0) = E\Big[\Big\{dN_i(t)-\lambda_i^*(t,\vartheta^*|t_0,u_0)dt\Big\}\widetilde{X}_i^*(t|t_0,u_0)\Big|\Omega_i\Big]$.

In practice, $\pi(\Omega_i)$, $E(W_i|\Omega_i)$ and $E(W_i^{\otimes 2}|\Omega_i)$ may be unknown. As discussed in Sun et al. (2017), we estimate them with working models $\pi(\Omega_i,\psi)$, $\mu_1(\Omega_i,\phi_1)$ and $\mu_2(\Omega_i,\phi_2)$ and denote the M-estimators (van der Vaart, 1998) of ψ,ϕ_1 and ϕ_2 by $\hat{\psi},\hat{\phi}_1$ and $\hat{\phi}_2$, respectively. To define $d\hat{\epsilon}_i^*(t,\vartheta^*|t_0,u_0)$, the estimator of the expectation term $d\epsilon_i^*(t|t_0,u_0)$, we introduce additional notations:

$$\hat{M}_{11,i}(t|t_0,u_0) = \begin{pmatrix} X_i^{\otimes 2} & X_i\mu_1(\Omega_i,\hat{\phi}_1)^\intercal I(N_i(t^-)>0) \\ \mu_1(\Omega_i,\hat{\phi}_1)X_i^\intercal I(N_i(t^-)>0) & \mu_2(\Omega_i,\hat{\phi}_2)I(N_i(t^-)>0) \end{pmatrix},$$

$$\hat{M}_{12,i}(t|t_0,u_0) = \hat{M}_{11,i}(t|t_0,u_0) \circ \begin{pmatrix} (t-t_0)\mathbb{1}_{p_1\times p_1} & (U_i(t)-u_0)\mathbb{1}_{p_1\times p_2} \\ (t-t_0)\mathbb{1}_{p_2\times p_1} & (U_i(t)-u_0)\mathbb{1}_{p_2\times p_2} \end{pmatrix},$$

$$\begin{aligned} &\hat{M}_{22,i}(t|t_0,u_0) \\ =&\hat{M}_{11,i}(t|t_0,u_0) \circ \begin{pmatrix} (t-t_0)^2\mathbb{1}_{p_1\times p_1} & (t-t_0)(U_i(t)-u_0)\mathbb{1}_{p_1\times p_2} \\ (t-t_0)(U_i(t)-u_0)\mathbb{1}_{p_2\times p_1} & (U_i(t)-u_0)^2\mathbb{1}_{p_2\times p_2} \end{pmatrix}, \end{aligned}$$

where $\circ$ is the Hadamard product and $\mathbb{1}_{m\times n}$ is an $m\times n$ all ones matrix:

$$\begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & 1 & 1 & \dots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \dots & 1 \end{bmatrix}_{m\times n}.$$

Then,

$$\begin{aligned} &d\hat{\epsilon}_i^*(t,\vartheta^*|t_0,u_0) \\ =\ & dN_i(t)\begin{pmatrix} X_i \\ \mu_1(\Omega_i,\hat{\phi}_1)I(N_i(t^-)>0) \\ X_i(t-t_0) \\ \mu_1(\Omega_i,\hat{\phi}_1)I(N_i(t^-)>0)(U_i(t)-u_0) \end{pmatrix} \\ & - \begin{pmatrix} \hat{M}_{11,i}(t|t_0,u_0) & \hat{M}_{12,i}(t|t_0,u_0) \\ \hat{M}_{12,i}(t|t_0,u_0)^\intercal & \hat{M}_{22,i}(t|t_0,u_0) \end{pmatrix} \vartheta^*(t_0,u_0)dt. \end{aligned}$$

Replacing q_i and $d\epsilon_i^*(t, \vartheta^*|t_0, u_0)$ in (4.4) by $\hat{q}_i = R_i/\pi(\Omega_i, \hat{\psi})$ and $d\hat{\epsilon}_i^*(t, \vartheta^*|t_0, u_0)$ respectively, we have the following estimating function for ϑ^*:

$$
\begin{aligned}
U_A(\vartheta^*|t_0, u_0) = \sum_{i=1}^{n} \int_0^{\tau} & K_h(t - t_0) K_b(U_i(t) - u_0) Y_i(t) w_i(t) \Big[\hat{q}_i \Big\{ dN_i(t) \\
& - \lambda_i^*(t, \vartheta^*|t_0, u_0) dt \Big\} \widetilde{X}_i^*(t|t_0, u_0) + (1 - \hat{q}_i) d\hat{\epsilon}_i^*(t|t_0, u_0) \Big]. \qquad (4.5)
\end{aligned}
$$

For simplicity, we take the identity weight $\omega_i(t) \equiv 1$ and denote the solution to $U_A(\vartheta^*, t_0, u_0)|_{\omega_i(t)\equiv 1} = 0$ by $\hat{\vartheta}_A^*(t_0, u_0)$, the AIPW estimator for ϑ^*. Let $\hat{\vartheta}_A(t_0, u_0) = (\hat{\alpha}_A(t_0, u_0)^{\mathsf{T}}, \hat{\gamma}_A(t_0, u_0)^{\mathsf{T}})^{\mathsf{T}}$, where $\hat{\alpha}_A(t_0, u_0)$ and $\hat{\gamma}_A(t_0, u_0)$ are vectors including the first p_1 elements and elements from $p_1 + 1$ to $p_1 + p_2$ of $\hat{\vartheta}_A(t_0, u_0)$ respectively. Note that more efficient AIPW estimators for $\alpha^{\mathsf{T}}(t_0)$ and $\gamma^{\mathsf{T}}(u_0)$ can be calculated in a similar way as (3.4):

$$
\hat{\alpha}_A(t_0) = n^{-1} \sum_{i=1}^{n} \hat{\alpha}_A(t_0, U_i(t_0)), \qquad \hat{\gamma}_A(u_0) = n_{u_0}^{-1} \sum_{t_{u_0} \in \mathcal{V}_{u_0}} \hat{\gamma}_A(t_{u_0}, u_0).
$$

4.2 Uniform Consistency and Weak Convergence

Let $\mu_2 = \int u^2 K(u)\, du$ and $\nu_0 = \int K^2(u)\, du$. We define

$$
\begin{aligned}
\Sigma_a(t, u) &= E\Big[Y_i(t) w_i(t)^2 \lambda_{i,0}(t, u) \widetilde{X}_i(t)^{\otimes 2} \Big| U_i(t) = u \Big] f_U(t, u), \\
D_a(t, u) &= E\Big[Y_i(t) w_i(t) \widetilde{X}_i(t)^{\otimes 2} \Big| U_i(t) = u \Big] f_U(t, u),
\end{aligned}
$$

where $f_U(t, u)$ is the density function of $U_i(t)$ at u.

The following theorems characterize the uniform consistency and weak convergence of the proposed local AIPW estimators. The proofs are outlined in the Appendix.

Theorem 4.1. *Under Condition C in the Appendix C, we have that:*

(a) $\sup_{t\in[t_1,t_2]} |\hat{\alpha}_A(t)-\alpha_0(t)| = o_p(1)$ *when* $\pi(\Omega_i)$ *and/or both* $E\{W_i|\Omega_i\}$ *and* $E\{W_i^{\otimes 2}|\Omega_i\}$ *are correctly specified;*

(b) $\sqrt{nh}(\hat{\alpha}_A(t) - \alpha_0(t) - \frac{1}{2}h^2\mu_2\ddot{\alpha}(t)) \xrightarrow{\mathcal{D}} N\left(0, \Sigma_{a,\alpha}(t)\right)$, *for* $t \in [t_1, t_2]$, *when* $\pi(\Omega_i)$, $E\{W_i|\Omega_i\}$ *and* $E\{W_i^{\otimes 2}|\Omega_i\}$ *are all correctly specified,*

where

$$\Sigma_{a,\alpha}(t) = \nu_0 \mathcal{I}_1 E\big\{D_a^{-1}(t, U_i(t))\Sigma_a(t, U_i(t))D_a^{-1}(t, U_i(t))\big\}\mathcal{I}_1^{\mathsf{T}}.$$

The covariance matrix $\Sigma_{a,\alpha}(t)$ can be consistently estimated by

$$\hat{\Sigma}_{a,\alpha}(t) = \frac{h}{n}\sum_{i=1}^{n}\int_0^{\tau} K_h(s-t)^2 Y_i(s) w_i(t)^2 \left\{\mathcal{I}_1 \hat{D}_a^{-1}(t, U_i(s))\widetilde{X}_i(s)\right\}^{\otimes 2} dN_i(s),$$

where

$$\hat{D}_a(t_0, u_0) = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} K_h(t-t_0)K_b(U_i(t)-u_0)Y_i(t)w_i(t)\widetilde{X}_i(t)^{\otimes 2} dt.$$

Theorem 4.2. *Under Condition C in the Appendix C, we have that:*

(a) $\sup_{u\in[u_1,u_2]} \|\hat{\gamma}_A(u)-\gamma_0(u)\| = o_p(1)$ *when* $\pi(\Omega_i)$ *and/or both* $E\{W_i|\Omega_i\}$ *and* $E\{W_i^{\otimes 2}|\Omega_i\}$ *are correctly specified;*

(b) $\sqrt{nb}(\hat{\gamma}_A(u) - \gamma_0(u) - \frac{1}{2}b^2\mu_2\ddot{\gamma}(u)) \xrightarrow{\mathcal{D}} N\left(0, \Sigma_{a,\gamma}(u)\right)$, *for* $u \in [u_1, u_2]$, *when* $\pi(\Omega_i)$, $E\{W_i|\Omega_i\}$ *and* $E\{W_i^{\otimes 2}|\Omega_i\}$ *are all correctly specified,*

where

$$\Sigma_{a,\gamma}(u) = \nu_0 \mathcal{I}_2 \Big\{\int_0^{\tau} D_a^{-1}(t,u)\Sigma_a(t,u)D_a^{-1}(t,u)dt\Big\}\mathcal{I}_2^{\mathsf{T}}$$

The asymptotic covariance matrix $\Sigma_{a,\gamma}(u)$ can be consistently estimated by

$$\hat{\Sigma}_{a,\gamma}(u) = \frac{b}{n}\sum_{i=1}^{n}\int_0^{\tau} K_b(U_i(s)-u)^2 Y_i(s) w_i(t)^2 \left\{\mathcal{I}_2 \hat{D}_a^{-1}(s, u)\widetilde{X}_i(s)\right\}^{\otimes 2} dN_i(s).$$

4.3 Simulation

We conduct simulation studies to examine the finite sample properties of our proposed estimators. For subject i, we generate covariate X_i from the Bernoulli distribution with probability of success $p = 0.5$. The potentially missing covariate W_i is generated from Uniform(0,1). In our simulation study, we consider $\pi(\Omega_i) = \frac{1}{1+\exp(-a_1 - a_2 X_i)}$, where (a_1, a_2) are parameters leading different missing probabilites. We simulate the censoring time $C_i = \min(\tau, C_i^*)$ which yields 25% censoring, where $\tau = 4$ and C_i^* is from the Uniform$(3, 8)$ distribution. By the thinning method (Lewis and Shedler, 1979), recurrent event times are generated from the non-homogeneous poisson process with the following additive intensity:

$$\lambda_i(t) = \alpha_0(t) + \alpha_1(t)X_i + \gamma(U_i(t))I(N_i(t^-) > 0)W_i \tag{4.6}$$

for $0 < t < C_i$, where $U_i(t) = t - T_{N_i(t^-)}$, $\alpha_0(t) = 4 - \log(1+t)$, $\alpha_1(t) = -1 + 0.2t$, and $\gamma(u) = -1/(1+u)$. The average number of recurrent events for each subject is around 6.2 for $X_i = 1$ and around 9.7 for $X_i = 0$.

Conditional on W_i, we generate the auxiliary variable $A_i = (W_i + \theta\zeta_i)/(1+\theta)$, where ζ_i is a random variable following the Uniform$(0, 1)$ distribution and θ is a parameter indicating the correlation between A_i and the missing covariate. Three different settings for θ are considered in this study: the correlation coefficient ρ between A_i and W_i is 0.5 for $\theta = 1.7321$, 0.7 for $\theta = 1.0202$, and 0.9 for $\theta = 0.4843$, denoted by $(A1)$, $(A2)$, and $(A3)$, respectively.

In Figure 12 and Figure 13, we report the finite sample performance of our proposed AIPW estimators under different levels of association between A_i and W_i. The FULL estimator and complete case (CC) estimator are solutions of the local weighted estimating equation $U(\vartheta^*|t_0, u_0)$ (4.3) and $U_{CC}(\vartheta^*|t_0, u_0)$ repectively, where

$$U_{CC}(\vartheta^*|t_0, u_0) = \sum_{i=1}^{n} \int_0^\tau K_h(t - t_0) K_b(U_i(t) - u_0) Y_i(t) w_i(t) R_i \times \Big\{ dN_i(t) - \lambda_i^*(t, \vartheta^*|t_0, u_0) dt \Big\} \widetilde{X}_i^*(t|t_0, u_0). \quad (4.7)$$

Figure 12 summarizes biases (Bias), empirical standard errors (SEE), average estimated standard errors (ESE) and the 95% empirical coverage probabilities (CP) of estimators for $\alpha_0(t)$, $\alpha_1(t)$, and $\gamma(u)$ under model (4.6) with $(a_1, a_2) = (0, 2.1972)$ for $n = 800$ based on 500 simulations. This setting leads to 30% missing covariates. The missing probabilities for $X_i = 1$ and $X_i = 0$ are 10% and 50% respectively. The results show that the bias of FULL, CC, IPW and AIPW estimators are all small except for boundary points corresponding to the boundary effect of kernel method. The AIPW estimators with $\rho = 0.7$ and 0.9 are more efficient than other estimators. The AIPW estimator with small correlation coefficient, $\rho = 0.5$, and the IPW estimator are less efficient than CC estimator. That is because we use the identity weight not the optimal weight in our simulations. The 95% empirical coverage probabilities are close to the nominal level.

Figure 13 shows the performance of our estimators in a heavier missing setting: $(a_1, a_2) = (-0.8473, 0.8473)$, which leads to 60% missingness. The missing probabilities for $X_i = 1$ and $X_i = 0$ are 50% and 70% respectively. Estimators listed in Figure 13 exhibit small bias. We observe that the SEE of our proposed AIPW estimators with $\rho = 0.5, 0.7$ and 0.9 are smaller than CC and IPW estimators. The 95% empirical coverage probabilities are almost all in the range of 93% and 97% showing average estimated standard errors are close to SEE.

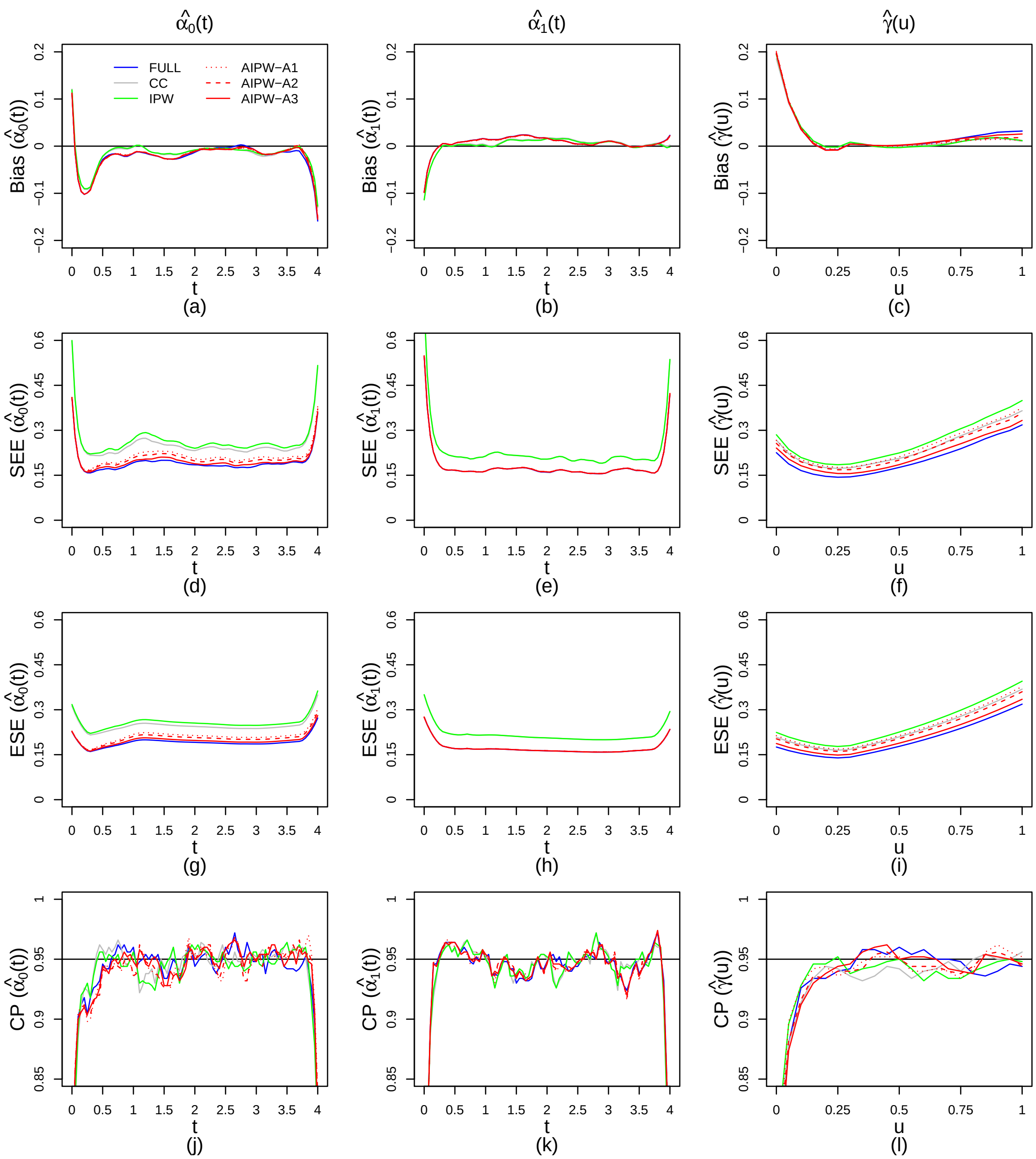


Figure 12: Bias, SEE, ESE and CP of estimators for $\alpha_0(t)$, $\alpha_1(t)$, and $\gamma(u)$ under model (4.6) for sample size $n = 800$ when using the Epanechnikov kernel with $h = b = 0.3$ based on 500 replications. $(a_1, a_2) = (0, 2.1972)$ leads to 30% missingness of covariate W.

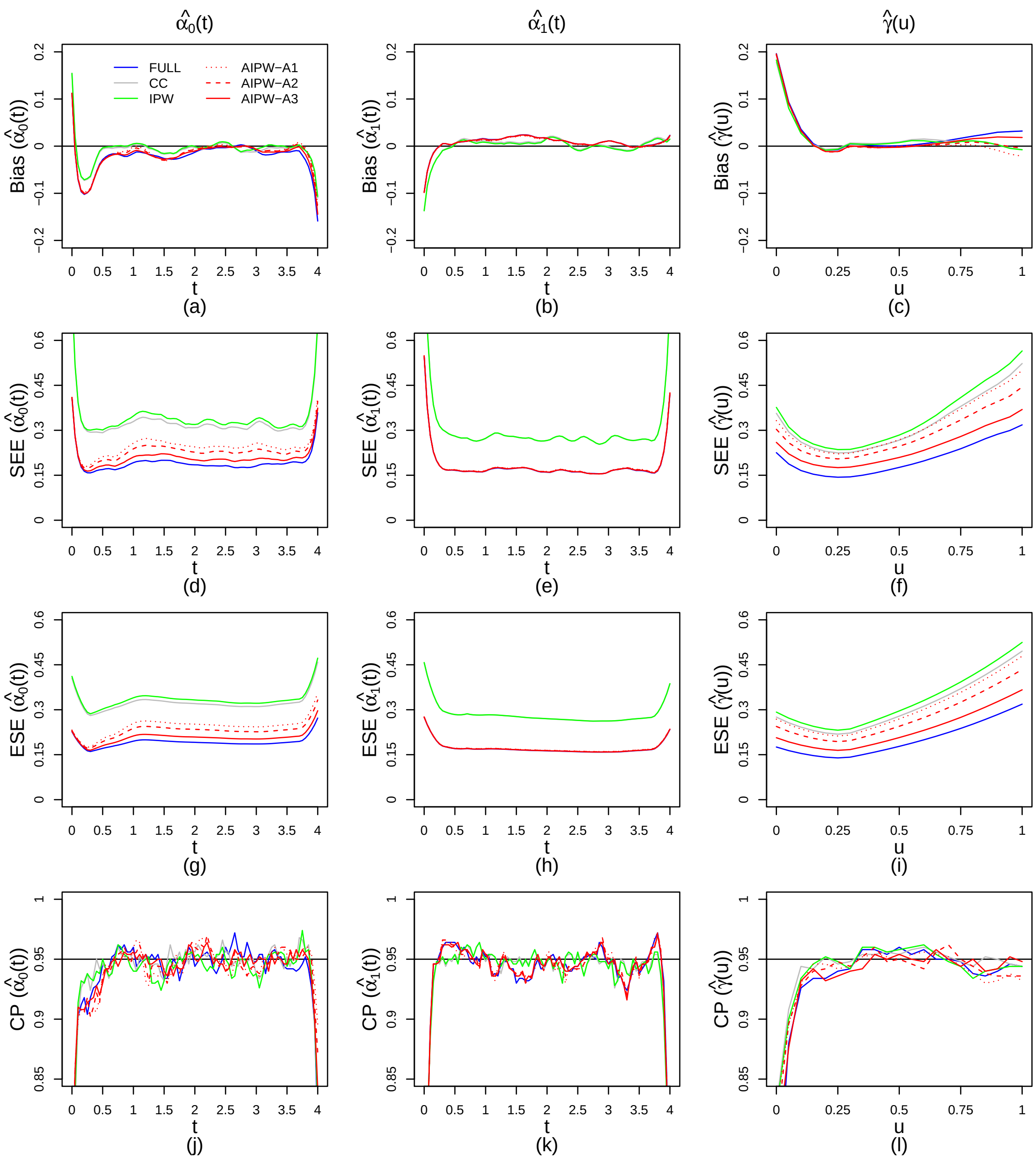


Figure 13: Bias, SEE, ESE and CP of estimators for $\alpha_0(t)$, $\alpha_1(t)$, and $\gamma(u)$ under model (4.6) for sample size $n = 800$ when using the Epanechnikov kernel with $h = b = 0.3$ based on 500 replications. $(a_1, a_2) = (-0.8473, 0.8473)$ leads to 60% missingness of covariate W.

CHAPTER 5: CONCLUSION AND FUTURE WORK

In Chapter 2, we developed the IPW and AIPW estimation methods for the cause-specific hazard regression models with missing causes, where the Cox models with time-varying coefficients are utilized to examine the cause-specific covariate effects. The AIPW estimating equation is obtained by subtracting the projection term of the IPW estimating equation onto the nuisance tangent space and is shown as a two-stage estimator by directly utilizing the inverse probability weighted estimator and through modeling available auxiliary variables to improve efficiency. Based on our asymptotic theoretical results and numerical simulation studies, the AIPW estimators are more efficient and robust than the IPW estimators. The proposed AIPW estimators can utilize auxiliary information that are not included in the models to improve estimation efficiency. We demonstrate that the performance of the AIPW estimators improve as the association between the auxiliary variable and the cause of failure strengthens. In addition, the AIPW estimators are more efficient than the IPW estimators even when the auxiliary variables are not available due to the more efficient construction of the AIPW estimating equation. The proposed estimators are very useful in the analysis of the Mashi clinical trial to examine the treatment effects for HIV-related and HIV-unrelated infant deaths where causes of deaths are missing for significant number of infant deaths and the treatment effects are demonstrated to vary over time.

In Chapter 3, we construct local score-type estimating equation (3.3) for the general class of nonparametric dynamic intensity models given in (3.2) for recurrent events. Our estimators are calculated by solving (3.3) by Newton-Raphson method. We derive asymptotic results and explore the finite sample behaviors of proposed estimators. In a special case where covariates X_i and W_i share common vectors, the identifiability problem may exist (Scheike, 2001). In this case, our method only works when the first occurence of recurrent event in data is relatively evenly distributed in the domain of time due to the existence of an additional indicator function $I(N_i(t^-) > 0)$. Another challenge of our estimation procedure is computational instability. As we discussed in Chapter 4, a local weighted estimating equation is an alternative of the local score-type estimating equation (3.3) with a cost of efficiency.

In Chapter 4, we investigate a specific missing covariate problem under a nonparametric dynamic additive intensity model (4.1) where the covariate with gap-time varying effect may be missing. We propose a local weight estimating equation for the fully observed data (4.3) and then develop the local AIPW estimating equation (4.5) by applying the AIPW technique for missing data. In our simulation study, we show that the AIPW estimators are consistent and more efficient than IPW and CC estimators especially when there is a strong correlation between the missing covariate and auxiliary covariates. To improve the efficiency of our proposed estimators, a weight selection method is desired. One simple approach is that we utilize an iterative algorithm that updates the weight function at each step. In details, we use the unit weight function as our initial weight function to get an estimator call it $\hat{\theta}_A^{(1)}$ by the estimation procedure in Section 4.1.2. After updating the weight by the inverse of the AIPW estimator for intensity $\hat{\lambda}_A^{(1)}(t) = \hat{\theta}_A^{(1)}(t, U_i(t))^\top \{\hat{q}_i \widetilde{X}_i(t) + (1 - \hat{q}_i)\hat{E}(\widetilde{X}_i(t))\}$, we get the AIPW estimator $\hat{\theta}_A^{(2)}$ in the second step. We repeat the above steps iteratively until satisfying certain criteria and denote the AIPW estimator in the final step by $\hat{\theta}_A$.

In next projects, we will develop hypothesis testing procedures for time-varying and gap-time-varying parameters including $\beta_j(t)$ in Chapter 2, $\alpha(t)$ and $\gamma(u)$ in Chapter 3 and 4. For example, in Chapter 3, we may have interest in testing the hypotheses: H_0 : $\gamma(u) = 0$, for $u \in [u_1, u_2]$, versus $H_a : \gamma(u) \neq 0$, for some $u \in [u_1, u_2]$. For covariates with time-constant effects, we can consider the following generalized semiparametric dynamic intensity model:

$$\lambda_i(t) = g^{-1}\Big\{\alpha^{\mathsf{T}}(t)X_i(t) + \beta Z_i(t) + I(N_i(t^-) > 0)\gamma^{\mathsf{T}}(t - T_{N_i(t^-)})W_i(t)\Big\}.$$

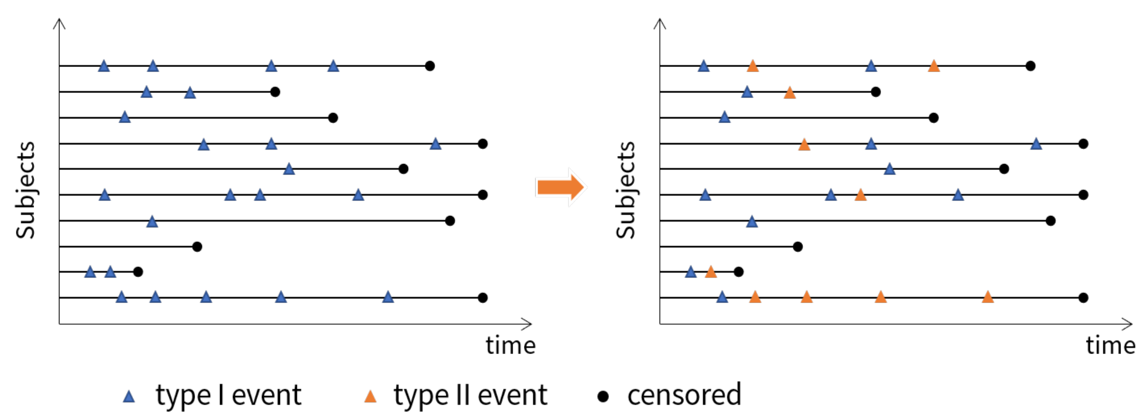


Figure 14: Extension to the multivariate recurrent event process

Considering different types of events, we are also interested in multivariate recurrent event processes (Figure 14). In the Malaria vaccine clinical trial, there are five types of Malaria infections. We intend to explore the genotype-specific vaccine efficacy and interactions among different types of Malaria infections. Other examples include different failure modes in automobile warranty claims and cause-specific hospitalizations. We may deal with more complicated missing data problems, such as the data has missing event types as well as missing covariates.

REFERENCES

Aalen, O. O., Borgan, O. r., and Gjessing, H. k. K. (2008). *Survival and event history analysis a process point of view*. Statistics for biology and health. Springer, New York, NY.

Aalen, O. O. and Johansen, S. (1978). An empirical transition matrix for non-homogeneous markov chains based on censored observations. *Scandinavian Journal of Statistics*, 5(3):141–150.

Andersen, P., Borgan, O., Gill, R., and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer Series in Statistics. Springer New York.

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

Bang, H. and Robins, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics*, 61(4):962–973.

Cai, Z., Fan, J., and Li, R. (2000). Efficient estimation and inferences for varying-coefficient models. *Journal of the American Statistical Association*, 95(451):888–902.

Cai, Z. and Sun, Y. (2003). Local linear estimation for time-dependent coefficients in cox's regression models. *Scandinavian Journal of Statistics*, 30(1):93–111.

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

Chang, S.-H. and Tzeng, S.-J. (2006). Nonparametric estimation of sojourn time distributions for truncated serial event data—a weight-adjusted approach. *Lifetime Data Analysis*, 12(1):53–67.

Chen, Q., Zeng, D., Ibrahim, J. G., Akacha, M., and Schmidli, H. (2013). Estimating time-varying effects for overdispersed recurrent events data with treatment switching. *Biometrika*, 100(2):339–354.

Clemens, J., Sack, D., Harris, J., van Loon, F., Chakraborty, J., Ahmed, F., Rao, M., Khan, M., Yunus, M., Huda, N., Stanton, B., Kay, B., Eeckels, R., Clemens, J., Rao, M., Kay, B., Sack, D., Harris, J., Stanton, B., Walter, S., Eeckels, R., Svennerholm, A.-M., and Holmgren, J. (1990). Field trial of oral cholera vaccines in bangladesh: results from three-year follow-up. *The Lancet*, 335(8684):270–273. Originally published as Volume 1, Issue 8684.

Cook, R. and Lawless, J. (2007). *The Statistical Analysis of Recurrent Events*. Statistics for Biology and Health. Springer New York.

Cox, D. (1972). Regression models and life table. *Journal of the Royal Statistical Society. Series B*, 34(2):187–220.

Dabrowska, D. M., Sun, G.-W., and Horowitz, M. M. (1994). Cox regression in a markov renewal model: An application to the analysis of bone marrow transplant data. *Journal of the American Statistical Association*, 89(427):867–877.

Durham, L. K., Longini, Jr., I. M., Halloran, M. E., Clemens, J. D., Azhar, N., and Rao, M. (1998). Estimation of vaccine efficacy in the presence of waning: Application to cholera vaccines. *American Journal of Epidemiology*, 147(10):948–959.

Efromovich, S. (2010). Dimension reduction and adaptation in conditional density estimation. *Journal of the American Statistical Association*, 105(490):761–774.

Gail, M. H., Santner, T. J., and Brown, C. C. (1980). An analysis of comparative carcinogenesis experiments based on multiple times to tumor. *Biometrics*, 36(2):255–266.

Gao, G. and Tsiatis, A. A. (2005). Semiparametric estimators for the regression coefficients in the linear transformation competing risks model with missing cause of failure. *Biometrika*, 92(4):875–891.

Gilbert, P. and Sun, Y. (2015). Inferences on relative failure rates in stratified mark-specific proportional hazards models with missing marks, with application to hiv vaccine efficacy trials. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 64(1):49–73.

Gill, R. (1980). Nonparametric estimation based on censored observations of a markov renewal process. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 53(1):97–116.

Hall, P., Racine, J., and Li, Q. (2004). Cross-validation and the estimation of conditional probability densities. *Journal of the American Statistical Association*, 99(468):1015–1026.

Hoover, D. R., Rice, J. A., Wu, C. O., and Yang, L.-P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika*, 85(4):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(260):663–685.

Huang, Y. (2002). Censored regression with the multistate accelerated sojourn times model. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(1):17–29.

Hyun, S., Lee, J., and Sun, Y. (2012). Proportional hazards model for competing risks data with missing cause of failure. *Journal of Statistical Planning and Inference*, 142(7):1767–1779.

Izbicki, R. and Lee, A. B. (2016). Nonparametric conditional density estimation in a high-dimensional regression setting. *Journal of Computational and Graphical Statistics*, 25(4):1297–1316.

Lawless, J. F. and Thiagarajah, K. (1996). A point-process model incorporating renewals and time trends, with application to repairable systems. *Technometrics*, 38(2):131–138.

Lawless, J. F., Wigg, M. B., Tuli, S., Drake, J., and Lamberti-Pasculli, M. (2001). Analysis of repeated failures or durations, with application to shunt failures for patients with paediatric hydrocephalus. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 50(4):449–465.

Lewis, P. A. W. and Shedler, G. S. (1979). Simulation of nonhomogeneous poisson processes by thinning. *Naval Research Logistics Quarterly*, 26(3):403–413.

Li, Y., Qi, L., and Sun, Y. (2018). Semiparametric varying-coefficient regression analysis of recurrent events with applications to treatment switching. *Statistics in Medicine*, 37(27):3959–3974.

Little, R. J. A. and Rubin, D. B. (2002). *Statistical analysis with missing data*. Wiley Series in Probability and Statistics. John Wiley and Sons, Inc., Hoboken, New Jersey, 2nd ed. edition.

Lu, W. and Liang, Y. (2008). Analysis of competing risks data with missing cause of failure under additive hazards model. *Statistica Sinica*, 18(1):219–234.

Martinussen, T., Scheike, T. H., Skovgaard, I. M., and Matinerssen, T. (2002). Efficient estimation of fixed and time-varying covariate effects in multiplicative intensity models. *Scandinavian Journal of Statistics*, 29(1):57–74.

Murphy, S. and Sen, P. (1991). Time-dependent coefficients in a cox-type regression model. *Stochastic Processes and their Applications*, 39(1):153–180.

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

Peña, E. A. (2016). Asymptotics for a class of dynamic recurrent event models. *Journal of Nonparametric Statistics*, 28(4):716–735.

Peña, E. A. and Hollander, M. (2004). Models for recurrent events in reliability and survival analysis. In Soyer, R., Mazzuchi, T. A., and Singpurwalla, N. D., editors, *Mathematical Reliability: An Expository Perspective*, pages 105–123. Springer US, Boston, MA.

Peña, E. A., Slate, E. H., and González, J. R. (2007). Semiparametric inference for a general class of models for recurrent events. *Journal of Statistical Planning and Inference*, 137(6):1727–1747.

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

Qi, L., Sun, Y., and Gilbert, P. B. (2017). Generalized semiparametric varying-coefficient model for longitudinal data with applications to adaptive treatment randomizations. *Biometrics*, 73(2):441–451.

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(427):846–866.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3):581–592.

Scharfstein, D. O., Rotnitzky, A., and Robins, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models: Rejoinder. *Journal of the American Statistical Association*, 94(448):1135–1146.

Scheike, T. H. (2001). A generalized additive regression model for survival times. *The Annals of Statistics*, 29(5):1344–1360.

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

Sun, Y. and Gilbert, P. B. (2012). Estimation of stratified mark-specific proportional hazards models with missing marks. *Scandinavian Journal of Statistics*, 39(1):34–52.

Sun, Y., Gilbert, P. B., and McKeague, I. W. (2009a). Proportional hazards models with continuous marks. *The Annals of Statistics*, 37(1):394–426.

Sun, Y., Hyun, S., and Gilbert, P. (2008). Testing and estimation of time-varying cause-specific hazard ratios with covariate adjustment. *Biometrics*, 64(4):1070–1079.

Sun, Y., Qian, X., Shou, Q., and Gilbert, P. B. (2017). Analysis of two-phase sampling data with semiparametric additive hazards models. *Lifetime data analysis*, 23(3):377–399.

Sun, Y., Sundaram, R., and Zhao, Y. (2009b). Empirical likelihood inference for the cox model with time-dependent coefficients via local partial likelihood. *Scandinavian Journal of Statistics*, 36(3):444–462.

Sun, Y., Wang, H. J., and Gilbert, P. B. (2012). Quantile regression for competing risks data with missing cause of failure. *Statistica Sinica*, 22(2):703–728.

Sun, Y. and Wu, H. (2005). Semiparametric time-varying coefficients regression model for longitudinal data. *Scandinavian Journal of Statistics*, 32(1):21–47.

Thior, I., Lockman, S., M Smeaton, L., L Shapiro, R., Wester, C., Jody Heymann, S., Gilbert, P., Stevens, L., Peter, T., Kim, S., Van Widenfelt, E., Moffat, C., Ndase, P., Arimi, P., Kebaabetswe, P., Mazonde, P., Makhema, J., McIntosh, K., Novitsky, V., and Essex, M. (2006). Breastfeeding plus infant zidovudine prophylaxis for 6 months vs formula feeding plus infant zidovudine for 1 month to reduce mother-to-child hiv transmission in botswana: A randomized trial: the mashi study. *JAMA*, 296(7):794–805.

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(469):172–183.

van der Vaart, A. W. (1998). Asymptotic statistics.

Vardi, Y. (1982). Nonparametric estimation in renewal processes. *Ann. Statist.*, 10(3):772–785.

Yin, G., Li, H., and Zeng, D. (2008). Partially linear additive hazards regression with varying coefficients. *Journal of the American Statistical Association*, 103(483):1200–1213.

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

Zeng, D. and Lin, D. Y. (2007). Semiparametric transformation models with random effects for recurrent events. *Journal of the American Statistical Association*, 102(477):167–180.

Zhang, X., Park, B. U., and Wang, J.-L. (2013). Time-varying additive models for longitudinal data. *Journal of the American Statistical Association*, 108(503):983–998.

Zucker, D. M. and Karr, A. F. (1990). Nonparametric survival analysis with time-dependent covariate effects: A penalized partial likelihood approach. *The Annals of Statistics*, 18(1):329–353.

APPENDIX A: PROOFS OF THE THEOREMS IN CHAPTER 2

Let $H = diag[I_p, hI_p]$. For $k = 1, \ldots, K$ and $j = 0, 1, 2$, we define the following notations:

$$S^{(j)}(t, \beta_k) = \frac{1}{n} \sum_{i=1}^{n} Y_i(t) \exp\left(Z_i(t)^{\mathsf{T}} \beta_k(t)\right) Z_i(t)^{\otimes j},$$

$$S_I^{*(j)}(t, \beta_k, \psi) = \frac{1}{n} \sum_{i=1}^{n} q_i Y_i(t) \exp\left(Z_i(t)^{\mathsf{T}} \beta_k(t)\right) Z_i(t)^{\otimes j}.$$

Let $s^{(j)}(t, \beta_k) = ES^{(j)}(t, \beta_k)$ and $s_I^{*(j)}(t, \beta_k, \psi) = ES_I^{*(j)}(t, \beta_k, \psi)$. $s^{(j)}(t, \beta_k) = s_I^{*(j)}(t, \beta_k, \psi)$ if the model $r(W_{1,i}, \psi)$ is correctly specified.

Condition A.

(A.1) For $k = 1, \ldots, K$, $\beta_k(t)$ has componentwise second derivatives on $[0, \tau]$. The sample path of the covariate process $Z_i(t)$ is left continuous and of bounded variation, and satisfies the moment condition $E[||Z_i(t)||^4 \exp(2M||Z_i(t)||)] < \infty$, where M is a constant such that $(t, \beta_k(t)) \in [0, \tau] \times [-M, M]^p$ for all t and $||A|| = \max_{k,l} |a_{kl}|$ for a matrix $A = (a_{kl})$.

(A.2) The kernel function $K(\cdot)$ is bounded and symmetric with bounded support $[-1, 1]$. The bandwidth h satisfies $nh^2 \to \infty$ and nh^5 is bounded as $n \to \infty$.

(A.3) The function $\pi(Q_i, \psi)$ is twice differentiable with respect to ψ on the compact set Θ_ψ, $\pi^{'}(Q_i, \psi) = \partial \pi(Q_i, \psi)/\partial \psi$ is uniformly bounded, and there is an $\varepsilon > 0$ such that $\pi(Q_i, \psi) \geq \varepsilon$ for all i.

(A.4) $s^{(j)}(t, \beta_k)$ and $s_I^{*(j)}(t, \beta_k, \psi)$ are componentwise continuous on $t \in [0, \tau], \beta_k \in [-M, M]^p, \psi \in \Theta_\psi$ for $j = 0, 1, 2$. $\sup_{t \in [0,\tau], \beta_k \in [-M,M]^p} ||S^{(j)}(t, \beta_k) - s^{(j)}(t, \beta_k)|| =$

$O_p(n^{-1/2})$, and $\sup_{t\in[0,\tau],\beta_k\in[-M,M]^p,\psi\in\Theta_\psi} ||S_I^{*(j)}(t,\beta_k,\psi) - s_I^{*(j)}(t,\beta_k,\psi)|| =$ $O_p(n^{-1/2})$.

(A.5) The matrix $\Sigma_k(t)$ is positive definite for all $t \in [0,\tau]$.

Lemma A.1. *Let* $W_{2,i} = (T_i, Z_i, A_i^{(v)})$*,* $W_{3,i} = (T_i, Z_i)$*,* $w_3 = (t,z)$ *and* $\rho_k(w) = P(V_i = k|\delta_i = 1, W_{2,i} = w)$ *where* $w = (t,z,a)$*. Under MAR II and MAR III, we have*

$$\rho_k(w) = \frac{\lambda_k(t|z)P(A_i^{(v)} = a|R_i = 1, \delta_i = 1, V_i = k, W_{3,i} = w_3)}{\sum_{l=1}^K \lambda_l(t|z)P(A_i^{(v)} = a|R_i = 1, \delta_i = 1, V_i = l, W_{3,i} = w_3)}.$$

Proof of Lemma A.1.

By the definition of $\rho_k(w)$,

$$\begin{aligned}\frac{\rho_k(w)}{1-\rho_k(w)} &= \frac{P(V_i = k|\delta_i = 1, T_i = t, Z_i = z, A_i^{(v)} = a)}{1 - P(V_i = k|\delta_i = 1, T_i = t, Z_i = z, A_i^{(v)} = a)} \\ &= \frac{P(V_i = k, \delta_i = 1, T_i = t, Z_i = z, A_i^{(v)} = a)}{\sum_{l\neq k} P(V_i = l, \delta_i = 1, T_i = t, Z_i = z, A_i^{(v)} = a)} \\ &= \frac{\lambda_k(t|z)P(A_i^{(v)} = a|\delta_i = 1, V_i = k, W_{3,i} = w_3)}{\sum_{l\neq k} \lambda_l(t|z)P(A_i^{(v)} = a|\delta_i = 1, V_i = l, W_{3,i} = w_3)},\end{aligned}$$

where the last equation is obtained since the censoring time C_i is independent of T_i and V_i conditional on Z_i.

Hence,

$$\rho_k(w) = \frac{\lambda_k(t|z)P(A_i^{(v)} = a|\delta_i = 1, V_i = k, W_{3,i} = w_3)}{\sum_{l=1}^K \lambda_l(t|z)P(A_i^{(v)} = a|\delta_i = 1, V_i = l, W_{3,i} = w_3)}. \tag{A.1}$$

Note that

$$
\begin{aligned}
&P(A_i^{(v)} = a | R_i = 1, \delta_i = 1, V_i = k, W_{3,i} = w_3) \\
&\quad = \frac{P(A_i^{(v)} = a, R_i = 1 | \delta_i = 1, V_i = k, W_{3,i} = w_3)}{P(R_i = 1 | \delta_i = 1, V_i = k, W_{3,i} = w_3)} \\
&\quad = \frac{P(R_i = 1 | \delta_i = 1, V_i = k, W_{2,i} = w) P(A_i^{(v)} = a | \delta_i = 1, V_i = k, W_{3,i} = w_3)}{P(R_i = 1 | \delta_i = 1, V_i = k, W_{3,i} = w_3)},
\end{aligned}
$$

$P(R_i = 1|\delta_i = 1, V_i = k, W_{2,i} = w) = P(R_i = 1|\delta_i = 1, W_{2,i} = w)$ under MAR II, and $P(R_i = 1|\delta_i = 1, V_i = k, W_{3,i} = w_3) = P(R_i = 1|\delta_i = 1, W_{3,i} = w_3)$ under MAR III.

We have

$$
\begin{aligned}
&P(A_i^{(v)} = a | \delta_i = 1, V_i = k, W_{3,i} = w_3) \\
&= \frac{P(R_i = 1 | \delta_i = 1, W_{3,i} = w_3) P(A_i^{(v)} = a | R_i = 1, \delta_i = 1, , V_i = k, W_{3,i} = w_3)}{P(R_i = 1 | \delta_i = 1, W_{2,i} = w)}.
\end{aligned}
\tag{A.2}
$$

By (A.1) and (A.2),

$$
\begin{aligned}
\rho_k(w) &= \frac{\lambda_k(t|z) P(A_i^{(v)} = a | \delta_i = 1, V_i = k, W_{3,i} = w_3)}{\sum_{l=1}^{K} \lambda_l(t|z) P(A_i^{(v)} = a | \delta_i = 1, V_i = l, W_{3,i} = w_3)} \\
&= \frac{\lambda_k(t|z) P(A_i^{(v)} = a | R_i = 1, \delta_i = 1, V_i = k, W_{3,i} = w_3)}{\sum_{l=1}^{K} \lambda_l(t|z) P(A_i^{(v)} = a | R_i = 1, \delta_i = 1, V_i = l, W_{3,i} = w_3)}.
\end{aligned}
$$

□

Proof of Theorem 2.1.

Let ψ_0 be the true value of ψ such that $r(W_{1,i}) = r(W_{1,i}, \psi_0)$ under the correctly specified model for $r(W_{1,i})$. Because the estimator $\widehat{\psi}$ of ψ_0 is an M-estimator, by Theorems 5.2 and 5.7 in van der Vaart (1998), we have $\widehat{\psi} - \psi_0 = O_p(n^{-1/2})$.

Let $q_{i0} = R_i/\pi(Q_i, \psi_0)$ and $\mathcal{D}_i = \dfrac{-R_i}{(\pi(Q_i, \psi_0))^2} \left(\dfrac{\partial \pi(Q_i, \psi_0)}{\partial \psi} \right)^{\mathsf{T}}$. By the Taylor expansion,

$$\widehat{q}_i - q_{i0} = \mathcal{D}_i(\widehat{\psi} - \psi_0) + o_p(n^{-1/2}). \tag{A.3}$$

Note that

$$\begin{aligned}
&n^{-1}H^{-1}\Big(U_I(t,\xi_k,\widehat{\psi}) - U_I(t,\xi_k,\psi_0)\Big) \\
&= n^{-1}\sum_{i=1}^{n}\int_0^{\tau} K_h(u-t)H^{-1}\left(\widetilde{Z}_i(u,u-t) - \overline{S_I}(u,\xi_k,\psi_0)\right)(\widehat{q}_i - q_{i0})\; dN_{ik}(u) \\
&\quad - n^{-1}\sum_{i=1}^{n}\int_0^{\tau} K_h(u-t)H^{-1}\left(\overline{S_I}(u,\xi_k,\widehat{\psi}) - \overline{S_I}(u,\xi_k,\psi_0)\right) q_{i0}\, dN_{ik}(u) \\
&\quad - n^{-1}\sum_{i=1}^{n}\int_0^{\tau} K_h(u-t)H^{-1}\left(\overline{S_I}(u,\xi_k,\widehat{\psi}) - \overline{S_I}(u,\xi_k,\psi_0)\right)(\widehat{q}_i - q_{i0})\; dN_{ik}(u).
\end{aligned} \tag{A.4}$$

Under Condition A, by the definition of the martingale and (A.3), we obtain

$$\begin{aligned}
&n^{-1}\sum_{i=1}^{n}\int_0^{\tau} K_h(u-t)H^{-1}\left(\widetilde{Z}_i(u,u-t) - \overline{S_I}(u,\xi_k,\psi_0)\right)(\widehat{q}_i - q_{i0})\; dN_{ik}(u) \\
&= n^{-1}\sum_{i=1}^{n}\left[\int_0^{\tau} K_h(u-t)H^{-1}\left(\widetilde{Z}_i(u,u-t) - \overline{S_I}(u,\xi_k,\psi_0)\right)\mathcal{D}_i\, dN_{ik}(u)\right](\widehat{\psi} - \psi_0) \\
&\quad + o_p(n^{-1/2}) \\
&= n^{-1}\sum_{i=1}^{n}\left[\int_0^{\tau} K_h(u-t)H^{-1}\left(\widetilde{Z}_i(u,u-t) - \overline{S_I}(u,\xi_k,\psi_0)\right)\mathcal{D}_i\, dM_{ik}(u)\right](\widehat{\psi} - \psi_0) \\
&\quad + n^{-1}\sum_{i=1}^{n}\left[\int_0^{\tau} K_h(u-t)H^{-1}\left(\widetilde{Z}_i(u,u-t) - \overline{S_I}(u,\xi_k,\psi_0)\right)\mathcal{D}_i\, Y_i(u)\lambda_k(u|Z_i(u))du\right] \\
&\quad \times (\widehat{\psi} - \psi_0) + o_p(n^{-1/2}) \\
&= O_p(n^{-1/2}).
\end{aligned}$$

Similarly, let

$$S_D^{(j)}(u,\xi_k,\psi_0) = n^{-1}\sum_{i=1}^{n} Y_i(u)\exp\left(\widetilde{Z}_i(u,u-t)^{\mathsf{T}}\xi(t)\right)\left(\widetilde{Z}_i(u,u-t)\right)^{\otimes j}\mathcal{D}_i$$

for $j = 0, 1$, we obtain

$$\begin{aligned}
& n^{-1}\sum_{i=1}^{n}\int_0^{\tau} K_h(u-t)H^{-1}\left(\overline{S_I}(u,\xi_k,\widehat{\psi}) - \overline{S_I}(u,\xi_k,\psi_0)\right) q_{i0}\, dN_{ik}(u) \\
&= n^{-1}\sum_{i=1}^{n}\Bigg[\int_0^{\tau} K_h(u-t)H^{-1}\Bigg(\frac{1}{S_I^{(0)}(u,\xi_k,\psi_0)} S_D^{(1)}(u,\xi_k,\psi_0) \\
&\quad - \frac{S_I^{(1)}(u,\xi_k,\psi_0)}{(S_I^{(0)}(u,\xi_k,\psi_0))^2} S_D^{(0)}(u,\xi_k,\psi_0)\Bigg) q_{i0}\, dN_{ik}(u)\Bigg](\widehat{\psi}-\psi_0) + o_p(n^{-1/2}) \\
&= O_p(n^{-1/2}).
\end{aligned}$$

Since the order of the third term in (A.4) is less than other terms, it follows that

$$n^{-1}H^{-1}\Big(U_I(t,\xi_k,\widehat{\psi}) - U_I(t,\xi_k,\psi_0)\Big) = O_p(n^{-1/2}).$$

Let $\widetilde{\xi_k}(t)$ be the running parameter in $U_I(t,\xi_k,\widehat{\psi})$ and $\xi_k(t) = (\beta_k^{\mathsf{T}}(t), (\beta_k^{'}(t))^{\mathsf{T}})^{\mathsf{T}}$ be the true parameter vector. Let $\theta = H(\widetilde{\xi_k}(t) - \xi_k(t))$ and $\widehat{\theta} = H(\widehat{\xi}_{I,k}(t) - \xi_k(t))$.

Under Condition A, by a Taylor expansion and the Glivenko-Cantelli theorem,

$$
\begin{aligned}
& n^{-1}H^{-1}\Big(U_I(t,\xi_k(t)+H^{-1}\theta,\widehat{\psi})-U_I(t,\xi_k(t),\psi_0)\Big) \\
= \; & n^{-1}H^{-1}\Big(U_I(t,\xi_k(t)+H^{-1}\theta,\psi_0)-U_I(t,\xi_k(t),\psi_0)\Big)+o_p(1) \\
= \; & -n^{-1}\sum_{i=1}^{n}\int_0^{\tau}K_h(u-t)H^{-1}\Big(\overline{S_I}(u,\xi_k(t)+H^{-1}\theta,\psi_0)-\overline{S_I}(u,\xi_k(t),\psi_0)\Big)q_{i0}\,dN_{ik}(u) \\
& +o_p(1) \\
= \; & -\int_0^{\tau}K_h(u-t)\left(s^{(2)}(u,\beta_k(u))-\frac{s^{(1)}(u,\beta_k(u))^{\otimes 2}}{s^{(0)}(u,\beta_k(u))}\right)\otimes\begin{pmatrix}1 & \frac{u-t}{h}\\ \frac{u-t}{h} & \left(\frac{u-t}{h}\right)^2\end{pmatrix}\theta\lambda_{k0}(u)du \\
& +o_p(1) \\
= \; & -\Sigma_k(t)\otimes\begin{pmatrix}1 & 0\\ 0 & \mu_2\end{pmatrix}\theta+o_p(1)
\end{aligned}
$$

uniformly in $t\in[t_1,t_2]$ and $\theta\in\mathcal{N}_0$, a neighborhood of $\mathbf{0}_{2p}$, as $n\to\infty$ and $nh^2\to\infty$, where $\mathbf{0}_{2p}$ is a $2p\times 1$ vector of zeros. The right side of the equation has a unique root at $\theta=\mathbf{0}_{2p}$. By the Glivenko-Cantelli theorem again, we have $n^{-1}H^{-1}U_I(t,\xi_k(t),\psi_0)\xrightarrow{\mathcal{P}}\mathbf{0}_{2p}$.

It follows from Lemma 2 of Sun et al. (2012) that $\widehat{\beta}_{I,k}(t)\xrightarrow{\mathcal{P}}\beta_k(t)$ uniformly in $t\in[t_1,t_2]$. $\square$

Proof of Theorem 2.2.

First, since $\widehat{\xi}_{I,k}(t)$ is the root of $U_I(t,\xi_k,\widehat{\psi})$, by a Taylor expansion, we note that

$$
\begin{aligned}
& n^{1/2}h^{1/2}H\left(\widehat{\xi}_{I,k}(t)-\xi_k(t)\right) \\
= & -\Big(n^{-1}H^{-1}U_I'(t,\xi_k^*(t),\widehat{\psi})H^{-1}\Big)^{-1}n^{-1/2}h^{1/2}H^{-1}U_I(t,\xi_k(t),\widehat{\psi}),
\end{aligned}
\tag{A.5}
$$

where $\xi_k^*(t)$ is on the line segment between $\widehat{\xi}_{I,k}(t)$ and $\xi_k(t)$. By the uniform consistency of $\widehat{\xi}_{I,k}(t)$ on $t \in [t_1, t_2]$ and the Glivenko-Cantelli theorem, we have

$$-n^{-1}H^{-1}U_I'(t, \xi_k^*(t), \widehat{\psi})H^{-1} \xrightarrow{\mathcal{P}} \Sigma_k(t) \otimes \begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix} \tag{A.6}$$

uniformly in $t \in [t_1, t_2]$ as $n \to \infty$ and $nh^2 \to \infty$.

From the proof of Therorem 1, we know

$$n^{-1/2}h^{1/2}H^{-1}\Big(U_I(t, \xi_k, \widehat{\psi}) - U_I(t, \xi_k, \psi_0)\Big) = O_p(h^{1/2}).$$

Note that

$$n^{-1/2}h^{1/2}H^{-1}U_I(t, \xi_k, \psi_0) = n^{1/2}h^{1/2}A_n(t, \xi_k, \psi_0) + n^{1/2}h^{1/2}B_n(t, \xi_k, \psi_0),$$

where

$$A_n(t, \xi_k, \psi_0) = n^{-1}\sum_{i=1}^{n}\int_0^\tau K_h(u-t)H^{-1}\Big(\widetilde{Z}_i(u, u-t) - \overline{S_I}(u, \xi_k, \psi_0)\Big)q_{i0}\ dM_{ik}(u)$$

and

$$B_n(t, \xi_k, \psi_0) = n^{-1}\sum_{i=1}^{n}\int_0^\tau K_h(u-t)H^{-1}\Big(\widetilde{Z}_i(u, u-t) - \overline{S_I}(u, \xi_k, \psi_0)\Big)q_{i0}\ \lambda_k(u|Z_i(u))du$$

Under Condition A, by the definition of martingale, we have

$$n^{1/2}h^{1/2}A_n(t, \xi_k, \psi_0) = n^{1/2}h^{1/2}W_n(t, \beta_k, \psi_0) + h^{1/2}\int_0^\tau K_h(u-t)\delta_n(du) + o_p(h^{1/2}), \tag{A.7}$$

where

$$W_n(t, \beta_k, \psi_0) = n^{-1}\sum_{i=1}^{n}\int_0^\tau K_h(u-t)\Big(Z_i(u) - \frac{s_I^{*(1)}(u, \beta_k, \psi_0)}{s_I^{*(0)}(u, \beta_k, \psi_0)}\Big) \otimes \begin{pmatrix} 1 \\ \frac{u-t}{h} \end{pmatrix} q_{i0}\ dM_{ik}^*(u)$$

and

$$\delta_n(t) = n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\Big(Z_i(u) - \frac{s_I^{*(1)}(u,\beta_k,\psi_0)}{s_I^{*(0)}(u,\beta_k,\psi_0)}\Big)\otimes\begin{pmatrix}1\\ \\ \frac{u-t}{h}\end{pmatrix} q_{i0}\Big(\lambda_{ik}^*(u) - \lambda_k(u|Z_i(u))\Big)du.$$

Note that the expectation of each term in the summand of $\delta_n(t)$ is zero. By Lemma 1 of Sun and Wu (2005), $\delta_n(t), 0 \le t \le \tau$, converges weakly to a mean-zero Gaussian process. Therefore, the second term in (A.7) converges to zero in probability.

Let $\widetilde{W}_n(t,\beta_k,\psi_0)$ and $\widetilde{B}_n(t,\xi_k,\psi_0)$ be the first p components of $W_n(t,\beta_k,\psi_0)$ and $B_n(t,\xi_k,\psi_0)$, respectively. Following the same arguments in the proof of theorem 2 of Cai and Sun (2003), we have

$$n^{1/2}h^{1/2}\widetilde{W}_n(t,\beta_k,\psi_0)\xrightarrow{\mathcal{D}}N(0,\nu_0\Sigma_k^*(t)) \tag{A.8}$$

and

$$\widetilde{B}_n(t,\xi_k,\psi_0) = \frac{1}{2}\mu_2 h^2\Sigma_k(t)\beta_k''(t) + o_p(h^2). \tag{A.9}$$

Combining (A.5)-(A.9), we conclude that

$$\sqrt{nh}\Big(\widehat{\beta}_{I,k}(t) - \beta_k(t) - \frac{1}{2}\mu_2 h^2\beta_k''(t)\Big)\xrightarrow{\mathcal{D}}N\Big(0,\nu_0\Sigma_k^{-1}(t)\Sigma_k^*(t)\Sigma_k^{-1}(t)\Big).$$

□

Proof of Theorem 2.3.

Let θ_{k0} be the true value of θ_k. If $h(A_i^{(v)}|k,T_i,Z_i)$ is correctly specified, then $h(A_i^{(v)}|k,T_i,Z_i) = h(A_i^{(v)}|k,T_i,Z_i,\theta_{k0})$. Let

$$\rho_k(W_{2,i}) = \frac{\lambda_k(T_i|Z_i)h(A_i^{(v)}|k,T_i,Z_i,\theta_{k0})}{\sum_{l=1}^{K}\lambda_l(T_i|Z_i)h(A_i^{(v)}|l,T_i,Z_i,\theta_{l0})}.$$

Since $\widehat{\rho}_k(W_{2,i}) - \rho_k(W_{2,i}) = O_p(n^{-1/2}h^{-1/2})$ and (A.3), we have

$$
\begin{aligned}
& n^{-1}H^{-1}U_A(t, \xi_k, \widehat{\psi}, \widehat{\rho}_k) \\
= \; & n^{-1}H^{-1}U_A(t, \xi_k, \psi_0, \rho_k) \\
& + n^{-1}\sum_{i=1}^{n}\int_0^{\tau} K_h(u-t)H^{-1}\Big(\widetilde{Z}_i(u, u-t) - \overline{S_f}(u, \xi_k)\Big)(\widehat{q}_i - q_{i0}) \\
& \times \Big(dN_{ik}(u) - \rho_k(W_{2,i})dN_i(u)\Big) \\
& + n^{-1}\sum_{i=1}^{n}\int_0^{\tau} K_h(u-t)H^{-1}\Big(\widetilde{Z}_i(u, u-t) - \overline{S_f}(u, \xi_k)\Big)(1 - q_{i0}) \\
& \times \Big(\widehat{\rho}_k(W_{2,i}) - \rho_k(W_{2,i})\Big)dN_i(u) \\
& - n^{-1}\sum_{i=1}^{n}\int_0^{\tau} K_h(u-t)H^{-1}\Big(\widetilde{Z}_i(u, u-t) - \overline{S_f}(u, \xi_k)\Big)(\widehat{q}_i - q_{i0}) \\
& \times \Big(\widehat{\rho}_k(W_{2,i}) - \rho_k(W_{2,i})\Big)dN_i(u) \\
\xrightarrow{\mathcal{P}} \; & n^{-1}H^{-1}U_A(t, \xi_k, \psi_0, \rho_k)
\end{aligned}
$$

uniformly in $t \in [t_1, t_2]$ if either $r(W_{1,i})$ or $h(A_i^{(v)}|k, T_i, Z_i)$ is correctly specified.

Let $\widetilde{\xi_k}(t)$ be the running parameter in $U_A(t, \xi_k, \widehat{\psi})$ and $\xi_k(t) = (\beta_k^{\mathsf{T}}(t), (\beta_k^{'}(t))^{\mathsf{T}})^{\mathsf{T}}$ be the true parameter vector. Let $\theta = H(\widetilde{\xi_k}(t) - \xi_k(t))$ and $\widehat{\theta} = H(\widehat{\xi}_{A,k}(t) - \xi_k(t))$.

Under Condition A, by a Taylor expansion and Glivenko-Cantelli theorem,

$$
\begin{aligned}
& n^{-1}H^{-1}\Big(U_A(t,\xi_k(t)+H^{-1}\theta,\psi_0,\rho_k)-U_A(t,\xi_k(t),\psi_0,\rho_k)\Big)\\
= & -n^{-1}\sum_{i=1}^{n}\int_0^{\tau}K_h(u-t)H^{-1}\Big(\overline{S_f}(u,\xi_k(t)+H^{-1}\theta)-\overline{S_f}(u,\xi_k(t))\Big)dN_{ik}(u)\\
& +n^{-1}\sum_{i=1}^{n}\int_0^{\tau}K_h(u-t)H^{-1}\Big(\overline{S_f}(u,\xi_k(t)+H^{-1}\theta)-\overline{S_f}(u,\xi_k(t))\Big)(1-q_{i0})\\
& \qquad\qquad \times\Big(dN_{ik}(u)-\rho_k(W_{2,i})dN_i(u)\Big)\\
= & -\int_0^{\tau}K_h(u-t)\left(s^{(2)}(u,\beta_k(u))-\frac{s^{(1)}(u,\beta_k(u))^{\otimes 2}}{s^{(0)}(u,\beta_k(u))}\right)\otimes\begin{pmatrix}1 & \frac{u-t}{h}\\ \frac{u-t}{h} & \left(\frac{u-t}{h}\right)^2\end{pmatrix}\theta\lambda_{k0}(u)du\\
& +o_p(1)\\
= & -\Sigma_k(t)\otimes\begin{pmatrix}1 & 0\\ 0 & \mu_2\end{pmatrix}\theta+o_p(1)
\end{aligned}
$$

uniformly in $t\in[t_1,t_2]$ and $\theta\in\mathcal{N}_0$, a neighborhood of $\mathbf{0}_{2p}$, if either $r(W_{1,i})$ or $h(A_i^{(v)}|k,T_i,Z_i)$ is correctly specified. The right side of the equation has a unique root at $\theta=\mathbf{0}_{2p}$. By the Glivenko-Cantelli theorem again, we have

$$
n^{-1}H^{-1}U_A(t,\xi_k(t),\psi_0,\rho_k)\xrightarrow{\mathcal{P}}\mathbf{0}_{2p}.
$$

It follows from Lemma 2 of Sun et al. (2012) that $\widehat{\beta}_{A,k}(t)\xrightarrow{\mathcal{P}}\beta_k(t)$ uniformly in $t\in[t_1,t_2]$. $\square$

Proof of Theorem 2.4.

First, since $\widehat{\xi}_{A,k}(t)$ is the root of $U_A(t,\xi_k,\widehat{\psi},\widehat{\rho}_k)$, by a Taylor expansion, we note that

$$
\begin{aligned}
&n^{1/2}h^{1/2}H\left(\widehat{\xi}_{A,k}(t)-\xi_k(t)\right)\\
=&-\left(n^{-1}H^{-1}U_A'(t,\xi_k^*(t),\widehat{\psi},\widehat{\rho}_k)H^{-1}\right)^{-1}n^{-1/2}h^{1/2}H^{-1}U_A(t,\xi_k(t),\widehat{\psi},\widehat{\rho}_k),
\end{aligned}
$$

where $\xi_k^*(t)$ is on the line segment between $\widehat{\xi}_{A,k}(t)$ and $\xi_k(t)$. By the uniform consistency of $\widehat{\xi}_{A,k}(t)$ on $t\in[t_1,t_2]$ and the Glivenko-Cantelli theorem, we have

$$
-n^{-1}H^{-1}U_A'(t,\xi_k^*(t),\widehat{\psi},\widehat{\rho}_k)H^{-1}\xrightarrow{\mathcal{P}}\Sigma_k(t)\otimes\begin{pmatrix}1 & 0\\ 0 & \mu_2\end{pmatrix}
$$

uniformly in $t\in[t_1,t_2]$ as $n\to\infty$ and $nh^2\to\infty$.

Following the same arguments as in the proof of Theorem 1,

$$
n^{-1/2}h^{1/2}H^{-1}U_A(t,\xi_k,\widehat{\psi},\widehat{\rho}_k)=n^{-1/2}h^{1/2}H^{-1}U_A(t,\xi_k,\psi_0,\rho_k)+O_p(h^{1/2}).
$$

By the definition of a martingale, we have

$$
\begin{aligned}
&n^{-1/2}h^{1/2}H^{-1}U_A(t,\xi_k,\psi_0,\rho_k)\\
=&n^{1/2}h^{1/2}A_n(t,\xi_k,\psi_0)+n^{1/2}h^{1/2}C_n(t,\xi_k,\psi_0)+n^{1/2}h^{1/2}B_n(t,\xi_k),
\end{aligned}
$$

where

$$
\begin{aligned}
&A_n(t, \xi_k, \psi_0) \\
=&n^{-1} \sum_{i=1}^{n} \int_0^\tau K_h(u-t) H^{-1}\Big(\widetilde{Z}_i(u, u-t) - \overline{S_f}(u, \xi_k)\Big) q_{i0}\ dM_{ik}(u), \\
&C_n(t, \xi_k, \psi_0) \\
=&n^{-1} \sum_{i=1}^{n} \int_0^\tau K_h(u-t) H^{-1}\Big(\widetilde{Z}_i(u, u-t) - \overline{S_f}(u, \xi_k)\Big)(1 - q_{i0}) \\
&\times E(dM_{ik}(u) | \delta_i = 1, W_{2,i}), \\
&B_n(t, \xi_k) \\
=&n^{-1} \sum_{i=1}^{n} \int_0^\tau K_h(u-t) H^{-1}\Big(\widetilde{Z}_i(u, u-t) - \overline{S_f}(u, \xi_k)\Big) \lambda_k(u | Z_i(u)) du.
\end{aligned}
$$

Similar to the proof of Theorem 2,

$$
n^{1/2} h^{1/2} A_n(t, \xi_k, \psi_0) = n^{1/2} h^{1/2} W_n(t, \beta_k, \psi_0) + o_p(1),
$$

where

$$
W_n(t, \beta_k, \psi_0) = n^{-1} \sum_{i=1}^{n} \int_0^\tau K_h(u-t) \Big(Z_i(u) - \frac{s^{(1)}(u, \beta_k)}{s^{(0)}(u, \beta_k)} \Big) \otimes \begin{pmatrix} 1 \\ \\ \frac{u-t}{h} \end{pmatrix} q_{i0}\ dM^*_{ik}(u)
$$

Let $\widetilde{W}_n(t, \beta_k, \psi_0)$, $\widetilde{C}_n(t, \beta_k, \psi_0, \rho_k)$ and $\widetilde{B}_n(t, \beta_k)$ be the first p components of $W_n(t, \beta_k, \psi_0)$, $C_n(t, \xi_k, \psi_0)$ and $B_n(t, \xi_k)$, respectively. Following the same arguments as in the proof of theorem 2 of Cai and Sun (2003), we have

$$
n^{1/2} h^{1/2} \widetilde{W}_n(t, \beta_k, \psi_0) \xrightarrow{\mathcal{D}} N(0, \nu_0 \Sigma^*_k(t))
$$

and

$$
\widetilde{B}_n(t, \xi_k) = \frac{1}{2} \mu_2 h^2 \Sigma_k(t) \beta''_k(t) + o_p(h^2).
$$

Also note that $n^{1/2}h^{1/2}\widetilde{C}_n(t, \beta_k, \psi_0) = o_p(1)$. Therefore,

$$\sqrt{nh}\Big(\widehat{\beta}_{A,k}(t) - \beta_k(t) - \frac{1}{2}\mu_2 h^2 \beta_k''(t)\Big) \xrightarrow{\mathcal{D}} N\Big(0, \nu_0 \Sigma_k^{-1}(t)\Sigma_k^{*}(t)\Sigma_k^{-1}(t)\Big).$$

$\square$

APPENDIX B: PROOFS OF THE THEOREMS IN CHAPTER 3

Let $p = p_1 + p_2$, $H = diag[I_p, hI_{p_1}, bI_{p_2}]$ and $\mathbf{0}_d$ be a $d \times 1$ vector of zeros. Let $\vartheta_0^*(t,u) = (\alpha_0^{\mathsf{T}}(t), \gamma_0^{\mathsf{T}}(u), \dot{\alpha}_0^{\mathsf{T}}(t), \dot{\gamma}_0^{\mathsf{T}}(u))^{\mathsf{T}}$ and $\vartheta_0(t,u) = (\alpha_0^{\mathsf{T}}(t), \gamma_0^{\mathsf{T}}(u))^{\mathsf{T}}$ where $\alpha_0(\cdot), \gamma_0(\cdot), \dot{\alpha}_0(\cdot), \dot{\gamma}_0(\cdot)$ are true values of $\alpha(\cdot), \gamma(\cdot)$ and their first derivatives.

Condition B.

(B.1) The inverse function of the link function $\varphi(\cdot) = g^{-1}(\cdot)$ is twice differentiable;

(B.2) The processes $X_i(t)$, $W_i(t)$, $U_i(t)$ and $\lambda_i(t)$, $0 \le t \le \tau$, are bounded and their total variations are bounded by a constant;

(B.3) The kernel function $K(\cdot)$ is symmetric with compact support on $[-1,1]$ and Lipschitz continuous; Bandwidths $h \asymp b$; $h \to 0$; $nh^2 \to \infty$ and nh^5 is bounded;

(B.4) $\alpha_0(t)$ and $\gamma_0(u)$ are twice differentiable on $t \in [0,\tau]$ and a compact support $\mathcal{U}$ respectively; $D^{-1}(t,u)$ is positive definite for all $(t,u) \in [0,\tau] \times \mathcal{U}$;

(B.5) (Condition (C.6) in Yin et al. (2008)) The conditional density of $(X_i(t), W_i(t))$ given $U_i(t) = u_0$ is twice continuously differentiable with respect to u_0. $f_U(t, u0)$ is twice continuously differentiable with respect to u_0 and satisfies

$$\inf_{t \in [0,\tau], u_0 \in \mathcal{U}} f_U(t, u_0) > 0.$$

Lemma B.1. *Under Condition B, we have that*

$$H\hat{\vartheta}^*(t,u) \xrightarrow{\mathcal{P}} (\vartheta_0(t,u)^{\mathsf{T}}, \mathbf{0}_p)^{\mathsf{T}}$$

uniformly in $t \in [t_1, t_2], u \in [u_1, u_2]$ *as* $n \to \infty$.

Proof of Lemma B.1.

Let $\eta(t,u) = H(\vartheta^*(t,u) - \vartheta_0^*(t,u))$. To prove $H\vartheta_0^*(t,u) \xrightarrow{\mathcal{P}} (\vartheta_0(t,u)^\mathsf{T}, \mathbf{0}_p)^\mathsf{T}$ is equivalent to showing $\hat{\eta}(t,u) = H(\hat{\vartheta}^*(t,u) - \vartheta_0^*(t,u)) \xrightarrow{\mathcal{P}} \mathbf{0}_{2p}$. In the following, for simplicity, we denote $\eta(t_0,u_0)$, $\hat{\eta}(t_0,u_0)$ and $\vartheta_0^*(t_0,u_0)$ as η, $\hat{\eta}$ and ϑ_0^* respectively. Let $\lambda_i^*(t,\theta) = \varphi\Big\{\theta^\mathsf{T}\widetilde{X}_i^*(t|t_0,u_0)\Big\}$, $\dot{\lambda}_i^*(t,\theta) = \dot{\varphi}\Big\{\theta^\mathsf{T}\widetilde{X}_i^*(t|t_0,u_0)\Big\}$, and $\ddot{\lambda}_i^*(t,\theta) = \ddot{\varphi}\Big\{\theta^\mathsf{T}\widetilde{X}_i^*(t|t_0,u_0)\Big\}$ where θ is an arbitrary $2p$-dimentional column vector. Then, by equation (3.3), $\hat{\eta}$ is the solution of $U_\eta(\eta|t_0,u_0) = 0$, where

$$
\begin{aligned}
U_\eta(\eta|t_0,u_0) = & \sum_{i=1}^n \int_0^\tau K_h(t-t_0)K_b(U_i(t)-u_0)Y_i(t)\frac{\dot{\lambda}_i^*(t,H^{-1}\eta+\vartheta_0^*)}{\lambda_i^*(t,H^{-1}\eta+\vartheta_0^*)} \\
& \times \Big\{dN_i(t) - \lambda_i^*(t,H^{-1}\eta+\vartheta_0^*)dt\Big\}\widetilde{X}_i^*(t|t_0,u_0).
\end{aligned}
$$

First, we consider

$$
\begin{aligned}
& n^{-1}H^{-1}(U_\eta(\eta|t_0,u_0) - U_\eta(\mathbf{0}_{2p}|t_0,u_0)) \\
= & n^{-1}\sum_{i=1}^n \int_0^\tau K_h(t-t_0)K_b(U_i(t)-u_0)Y_i(t)\frac{\dot{\lambda}_i^*(t,H^{-1}\eta+\vartheta_0^*)}{\lambda_i^*(t,H^{-1}\eta+\vartheta_0^*)} \\
& \times \Big\{dN_i(t) - \lambda_i^*(t,H^{-1}\eta+\vartheta_0^*)dt\Big\}H^{-1}\widetilde{X}_i^*(t|t_0,u_0) \\
& - n^{-1}\sum_{i=1}^n \int_0^\tau K_h(t-t_0)K_b(U_i(t)-u_0)Y_i(t)\frac{\dot{\lambda}_i^*(t,\vartheta_0^*)}{\lambda_i^*(t,\vartheta_0^*)} \\
& \times \Big\{dN_i(t) - \lambda_i^*(t,\vartheta_0^*)dt\Big\}H^{-1}\widetilde{X}_i^*(t|t_0,u_0).
\end{aligned}
$$

Let $D_{22}(t_0,u_0) = D(t_0,u_0) \circ \mathrm{diag}(\mu_2\mathbb{1}_{p_1\times p_1}, \mu_2\mathbb{1}_{p_2\times p_2})$, where $\circ$ is the Hadamard product, $\mathbb{1}_{m\times n}$ is a $m \times n$ all ones matrix:

$$\begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & 1 & 1 & \dots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \dots & 1 \end{bmatrix}_{m\times n}.$$

Under Condition B, by the Taylor expansion and the Lemma A.1 in Yin et al. (2008), we have

$$\begin{aligned} & n^{-1}H^{-1}(U_\eta(\eta|t_0,u_0) - U_\eta(\mathbf{0}_{2p}|t_0,u_0)) \\ =& n^{-1}\sum_{i=1}^{n}\int_0^\tau K_h(t-t_0)K_b(U_i(t)-u_0)Y_i(t)\left[\left\{\frac{\ddot{\lambda}_i^*(t,\vartheta_0^*)}{\lambda_i^*(t,\vartheta_0^*)} - \frac{\dot{\lambda}_i^*(t,\vartheta_0^*)^2}{\lambda_i^*(t,\vartheta_0^*)^2}\right\}dN_i(t) \right. \\ & \left. - \ddot{\lambda}_i^*(t,\vartheta_0^*)dt\right]\{H^{-1}\widetilde{X}_i^*(t|t_0,u_0)\}^{\otimes 2}\eta + o_p(\eta) \\ =& -\operatorname{diag}(D(t_0,u_0), D_{22}(t_0,u_0))\eta + o_p(\eta) \end{aligned}$$

uniformly in $t_0 \in [t_1,t_2]$, $u_0 \in [u_1,u_2]$ and $\eta \in \mathcal{N}_0$, a neighborhood of $\mathbf{0}_{2p}$. Furthermore,

$$\begin{aligned} & n^{-1}H^{-1}U_\eta(\mathbf{0}_{2p}|t_0,u_0) \\ =& n^{-1}\sum_{i=1}^{n}\int_0^\tau K_h(t-t_0)K_b(U_i(t)-u_0)Y_i(t)\frac{\dot{\lambda}_{i,0}(t,U_i(t))}{\lambda_{i,0}(t,U_i(t))}H^{-1}\widetilde{X}_i^*(t|t_0,u_0)dM_i(t) + o_p(1). \end{aligned}$$

By applying Lemma 1 of Zhang et al. (2013), $n^{-1}H^{-1}U_\eta(\mathbf{0}_{2p}|t_0,u_0)\xrightarrow{\mathcal{P}}\mathbf{0}_{2p}$. By Lemma 2 of Sun et al. (2012), we conclude that $\hat{\eta}(t,u)\xrightarrow{\mathcal{P}}\mathbf{0}_{2p}$, thus $H\hat{\vartheta}^*(t,u)\xrightarrow{\mathcal{P}}(\vartheta_0(t,u)^\mathsf{T},\mathbf{0}_p)^\mathsf{T}$ uniformly in $t \in [t_1,t_2], u \in [u_1,u_2]$ as $n \to \infty$.

□

Lemma B.2. *Under Condition B, we have*

$$\sqrt{nhb}\{\hat{\vartheta}(t_0,u_0)-\vartheta_0(t_0,u_0)-\frac{1}{2}h^2\mu_2 D^{-1}(t_0,u_0)b_\alpha(t_0,u_0)$$
$$-\frac{1}{2}b^2\mu_2 D^{-1}(t_0,u_0)b_\gamma(t_0,u_0)\}=D^{-1}(t_0,u_0)\sqrt{nhb}\mathbf{A}_n(t_0,u_0)+o_p(1)$$

uniformly in $t_0\in[t_1,t_2]$ *and* $u_0\in[u_1,u_2]$ *as* $nh^6=O_p(1)$, *where*

$$\begin{aligned}
b_\alpha(t_0,u_0)=&E\left\{Y_i(t_0)\frac{\dot{\lambda}_{i,0}(t_0,u_0)^2}{\lambda_{i,0}(t_0,u_0)}\widetilde{X}_i(t_0)X_i^{\mathsf{T}}(t_0)\middle|U_i(t_0)=u_0\right\}f_U(t_0,u_0)\ddot{\alpha}(t_0),\\
b_\gamma(t_0,u_0)=&E\left\{Y_i(t_0)\frac{\dot{\lambda}_{i,0}(t_0,u_0)^2}{\lambda_{i,0}(t_0,u_0)}\widetilde{X}_i(t_0)W_i^{\mathsf{T}}(t_0)I(N_i(t^-)>0)\middle|U_i(t_0)=u_0\right\}\\
&\times f_U(t_0,u_0)\ddot{\gamma}(u_0),\\
\mathbf{A}_n(t_0,u_0)=&n^{-1}\sum_{i=1}^n\int_0^\tau K_h(t-t_0)K_b(U_i(t)-u_0)Y_i(t)\frac{\dot{\lambda}_{i,0}(t,U_i(t))}{\lambda_{i,0}(t,U_i(t))}\widetilde{X}_i(t)dM_i(t).
\end{aligned}$$

Proof of Lemma B.2.

Because $U(\hat{\vartheta}^*|t_0,u_0)=0$, we have

$$\hat{\vartheta}^*(t_0,u_0)-\vartheta_0(t_0,u_0)=-\left\{\frac{\partial U(\vartheta_0^*|t_0,u_0)}{\partial\vartheta^*}\right\}^{-1}U(\vartheta_0^*|t_0,u_0)+o_p(1).$$

We consider the first p components of ϑ^*. Since

$$\frac{\partial U(\vartheta_0^*|t_0,u_0)}{\partial\vartheta^*}\xrightarrow{\mathcal{P}}-\operatorname{diag}(D(t_0,u_0),D_{22}(t_0,u_0)),$$

it yields

$$\hat{\vartheta}(t_0,u_0)-\vartheta_0(t_0,u_0)=D^{-1}(t_0,u_0)U_1(\vartheta_0^*|t_0,u_0)+o_p(1),$$

where

$$
\begin{aligned}
&U_1(\vartheta_0^*|t_0, u_0)\\
=&n^{-1}\sum_{i=1}^{n}\int_0^{\tau} K_h(t-t_0)K_b(U_i(t)-u_0)Y_i(t)\frac{\dot{\lambda}_i^*(t,\vartheta_0^*)}{\lambda_i^*(t,\vartheta_0^*)}\Big\{dN_i(t)-\lambda_i^*(t,\vartheta_0^*)dt\Big\}\widetilde{X}_i(t).
\end{aligned}
$$

By the Taylor expansion, we have

$$
U_1(\vartheta_0^*|t_0, u_0) = \mathbf{A}_n(t_0, u_0) + \mathbf{B}_n(t_0, u_0) + \mathbf{C}_n(t_0, u_0) + o_p(h^2 + b^2),
$$

where

$$
\begin{aligned}
&\mathbf{B}_n(t_0, u_0)\\
=&-\frac{1}{2n}\sum_{i=1}^{n}\int_0^{\tau} K_h(t-t_0)K_b(U_i(t)-u_0)Y_i(t)\left[\left\{\frac{\ddot{\lambda}_{i,0}(t,U_i(t))}{\lambda_{i,0}(t,U_i(t))}-\frac{\dot{\lambda}_{i,0}(t,U_i(t))^2}{\lambda_{i,0}(t,U_i(t))^2}\right\}\right.\\
&\left.\times\, dN_i(t)-\ddot{\lambda}_i^*(t,\vartheta_0^*)dt\right]\ddot{\alpha}(t_0)^{\mathsf{T}}X_i(t)(t-t_0)^2\widetilde{X}_i(t),\\
&\mathbf{C}_n(t_0, u_0)\\
=&-\frac{1}{2n}\sum_{i=1}^{n}\int_0^{\tau} K_h(t-t_0)K_b(U_i(t)-u_0)Y_i(t)\left[\left\{\frac{\ddot{\lambda}_{i,0}(t,U_i(t))}{\lambda_{i,0}(t,U_i(t))}-\frac{\dot{\lambda}_{i,0}(t,U_i(t))^2}{\lambda_{i,0}(t,U_i(t))^2}\right\}\right.\\
&\left.\times\, dN_i(t)-\ddot{\lambda}_i^*(t,\vartheta_0^*)dt\right]\ddot{\gamma}(u_0)^{\mathsf{T}}W_i(t)I(N_i(t^-)>0)(U_i(t)-u_0)^2\widetilde{X}_i(t).
\end{aligned}
$$

Following the arguments in Lemma A.1 in Yin et al. (2008), we conclude that

$$
\frac{1}{h^2}\mathbf{B}_n(t_0, u_0)\xrightarrow{\mathcal{P}}\frac{1}{2}\mu_2 b_{\alpha}(t_0, u_0)
$$

and

$$
\frac{1}{b^2}\mathbf{C}_n(t_0, u_0)\xrightarrow{\mathcal{P}}\frac{1}{2}\mu_2 b_{\gamma}(t_0, u_0).
$$

Therefore, Lemma B.2 holds. □

Proof of Theorem 3.1.

(a) By Lemma B.1, $\hat{\alpha}(t,u)\xrightarrow{\mathcal{P}}\alpha_0(t)$ uniformly in $t \in [t_1,t_2], u \in [u_1,u_2]$ as $n \to \infty$. Then,

$$\begin{aligned}
&\sup_{t\in[t_1,t_2]} |\hat{\alpha}(t)\xrightarrow{\mathcal{P}}\alpha_0(t)| \\
=&\sup_{t\in[t_1,t_2]} |n^{-1}\sum_{i=1}^{n}\{\hat{\alpha}(t,U_i(t)) - \alpha_0(t)\} \\
\le&\sup_{t\in[t_1,t_2],u\in[u_1,u_2]} |\hat{\alpha}(t,u) - \alpha_0(t)| \\
=&o_p(1).
\end{aligned}$$

(b) Under Condition B, by Lemma B.2, we have

$$\begin{aligned}
&\sqrt{nh}\{\hat{\alpha}(t_0,u_0) - \alpha_0(t_0) - \frac{1}{2}h^2\mu_2\ddot{\alpha}_0(t_0)\} \\
=&\mathcal{I}_1 D^{-1}(t_0,u_0)\sqrt{\frac{h}{n}}\sum_{i=1}^{n}\int_0^{\tau} K_h(t-t_0)K_b(U_i(t)-u_0)Y_i(t)\frac{\dot{\lambda}_{i,0}(t,U_i(t))}{\lambda_{i,0}(t,U_i(t))}\widetilde{X}_i(t)dM_i(t) \\
&+ o_p(1).
\end{aligned}$$

Then, by Lemma A.1 in Yin et al. (2008), we obtain

$$\begin{aligned}
&\sqrt{nh}\{\hat{\alpha}(t_0) - \alpha_0(t_0) - \frac{1}{2}h^2\mu_2\ddot{\alpha}(t_0)\} \\
=&\sqrt{\frac{h}{n}}\sum_{i=1}^{n}\int_0^{\tau} K_h(t-t_0)Y_i(t)\frac{\dot{\lambda}_{i,0}(t,U_i(t))}{\lambda_{i,0}(t,U_i(t))} \\
&\times \mathcal{I}_1\left\{\frac{1}{n}\sum_{j=1}^{n} K_b(U_i(t)-U_j(t_0))D^{-1}(t_0,U_j(t_0))\right\}\widetilde{X}_i(t)dM_i(t) + o_p(1) \\
=&\sqrt{nh}\mathbf{A}_n^{(\alpha)}(t_0) + o_p(1),
\end{aligned}$$

where

$$\mathbf{A}_n^{(\alpha)}(t_0) = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} K_h(t-t_0)Y_i(t)\frac{\dot{\lambda}_{i,0}(t,U_i(t))}{\lambda_{i,0}(t,U_i(t))}\mathcal{I}_1 D^{-1}(t_0,U_i(t))\widetilde{X}_i(t)dM_i(t).$$

$\sqrt{nh}\mathbf{A}_n^{(\alpha)}(t_0)$ is a sum of local square integrable martingales, and

$$\begin{aligned} & nh\langle \mathbf{A}_n^{(\alpha)}, \mathbf{A}_n^{(\alpha)}\rangle(t_0) \\ =& \frac{h}{n}\sum_{i=1}^{n}\int_0^{\tau} K_h(t-t_0)^2 Y_i(t)\frac{\dot{\lambda}_{i,0}(t,U_i(t))^2}{\lambda_{i,0}(t,U_i(t))}\left\{\mathcal{I}_1 D^{-1}(t_0,U_i(t))\widetilde{X}_i(t)\right\}^{\otimes 2} dt \\ \xrightarrow{\mathcal{P}}& \Sigma_\alpha(t_0), \end{aligned}$$

where $\Sigma_\alpha(t_0) = \nu_0 \mathcal{I}_1 E\{D^{-1}(t_0,U_i(t_0))\}\mathcal{I}_1^{\mathsf{T}}$.

Moreover, for any $\epsilon > 0$,

$$\begin{aligned} & nh\langle \mathbf{A}_{n,\epsilon}^{(\alpha)}, \mathbf{A}_{n,\epsilon}^{(\alpha)}\rangle(t_0) \\ =& \frac{h}{n}\sum_{i=1}^{n}\int_0^{\tau} K_h(t-t_0)^2 Y_i(t)\frac{\dot{\lambda}_{i,0}(t,U_i(t))^2}{\lambda_{i,0}(t,U_i(t))}X_{i,j}(t,t_0)^{\otimes 2} \\ & \times I\left(\sqrt{\frac{h}{n}}\left|K_h(t-t_0)Y_i(t)\frac{\dot{\lambda}_{i,0}(t,U_i(t))}{\lambda_{i,0}(t,U_i(t))}X_{i,j}(t,t_0)\right| > \epsilon\right)dt \\ \xrightarrow{\mathcal{P}}& 0, \end{aligned}$$

where $X_{i,j}(t,t_0)$ is the jth component of $\mathcal{I}_1 D^{-1}(t_0,U_i(t))\widetilde{X}_i(t)$. By Theorem 5.1.1 in Flemming (1991), we conclude that as $n \to \infty$,

$$\sqrt{nh}(\hat{\alpha}(t) - \alpha_0(t) - \frac{1}{2}h^2\mu_2\ddot{\alpha}(t)) \xrightarrow{\mathcal{D}} N\left(0, \Sigma_\alpha(t)\right), \text{for} t \in [t_1, t_2].$$

□

Proof of Theorem 3.2.

(a) Following the proof of Theorem 3.1 (a), we can apply Lemma B.1 to prove the consistency of $\hat{\gamma}(u_0)$ similarly.

(b) Under Condition B, by Lemma B.2, we have

$$
\begin{aligned}
&\sqrt{nb}\{\hat{\gamma}(t_0,u_0)-\gamma_0(u_0)-\frac{1}{2}b^2\mu_2\ddot{\gamma}_0(u_0)\}\\
=&\mathcal{I}_2D^{-1}(t_0,u_0)\sqrt{\frac{b}{n}}\sum_{i=1}^{n}\int_0^\tau K_h(t-t_0)K_b(U_i(t)-u_0)Y_i(t)\frac{\dot{\lambda}_{i,0}(t,U_i(t))}{\lambda_{i,0}(t,U_i(t))}\widetilde{X}_i(t)dM_i(t)\\
&+o_p(1).
\end{aligned}
$$

Then, by Lemma A.1 in Yin et al. (2008), we obtain

$$
\begin{aligned}
&\sqrt{nb}\{\hat{\gamma}(u_0)-\gamma_0(u_0)-\frac{1}{2}b^2\mu_2\ddot{\gamma}(u_0)\}\\
=&\sqrt{\frac{b}{n}}\sum_{i=1}^{n}\int_0^\tau K_b(U_i(t)-u_0)Y_i(t)\frac{\dot{\lambda}_{i,0}(t,U_i(t))}{\lambda_{i,0}(t,U_i(t))}\\
&\times\mathcal{I}_2\left\{n_{u_0}^{-1}\sum_{t_{u_0}\in\mathcal{V}_{u_0}}K_h(t-t_{u_0})D^{-1}(t_{u_0},u_0)\right\}\widetilde{X}_i(t)dM_i(t)+o_p(1)\\
=&\sqrt{nb}\mathbf{A}_n^{(\gamma)}(u_0)+o_p(1),
\end{aligned}
$$

if $n_{u_0}\asymp n$, where

$$
\mathbf{A}_n^{(\gamma)}(u_0)=\frac{1}{n}\sum_{i=1}^{n}\int_0^\tau K_b(U_i(t)-u_0)Y_i(t)\frac{\dot{\lambda}_{i,0}(t,U_i(t))}{\lambda_{i,0}(t,U_i(t))}\mathcal{I}_2D^{-1}(t,u_0)\widetilde{X}_i(t)dM_i(t).
$$

$\sqrt{nb}\mathbf{A}_n^{(\gamma)}(u_0)$ is a sum of local square integrable martingales, and

$$
\begin{aligned}
&nb\langle\mathbf{A}_n^{(\gamma)},\mathbf{A}_n^{(\gamma)}\rangle(u_0)\\
=&\frac{b}{n}\sum_{i=1}^{n}\int_0^\tau K_b(U_i(t)-u_0)^2Y_i(t)\frac{\dot{\lambda}_{i,0}(t,U_i(t))^2}{\lambda_{i,0}(t,U_i(t))}\left\{\mathcal{I}_2D^{-1}(t,u_0)\widetilde{X}_i(t)\right\}^{\otimes 2}dt\\
\xrightarrow{\mathcal{P}}&\Sigma_\gamma(u_0),
\end{aligned}
$$

where

$$\Sigma_\gamma(u_0) = \nu_0 \mathcal{I}_2 \Big\{ \int_0^\tau D^{-1}(t, u_0) dt \Big\} \mathcal{I}_2^\mathsf{T}.$$

Moreover, for any $\epsilon > 0$,

$$\begin{aligned}
& nb\langle \mathbf{A}_{n,\epsilon}^{(\gamma)}, \mathbf{A}_{n,\epsilon}^{(\gamma)} \rangle (u_0) \\
=& \frac{b}{n} \sum_{i=1}^{n} \int_0^\tau K_b(U_i(t) - u_0)^2 Y_i(t) \frac{\dot{\lambda}_{i,0}(t, U_i(t))^2}{\lambda_{i,0}(t, U_i(t))} X_{i,j}(t, u_0)^{\otimes 2} \\
& \times I\left(\sqrt{\frac{b}{n}} \left| K_b(U_i(t) - u_0) Y_i(t) \frac{\dot{\lambda}_{i,0}(t, U_i(t))}{\lambda_{i,0}(t, U_i(t))} X_{i,j}(t, u_0) \right| > \epsilon \right) dt \\
\xrightarrow{\mathcal{P}} & 0,
\end{aligned}$$

where $X_{i,j}(t, u_0)$ is the jth component of $\mathcal{I}_2 D^{-1}(t, u_0) \widetilde{X}_i(t)$. By Theorem 5.1.1 in Flemming 1991, we conclude that as $n \to \infty$,

$$\sqrt{nb}(\hat{\gamma}(u) - \gamma_0(u) - \frac{1}{2} b^2 \mu_2 \ddot{\gamma}_0(u)) \xrightarrow{\mathcal{D}} N\left(0, \Sigma_\gamma(u)\right), \text{for} u \in [u_1, u_2].$$

□

APPENDIX C: PROOFS OF THE THEOREMS IN CHAPTER 4

Condition C.

(C.1) Conditions (B.2)-(B.5) in Appendix B;

(C.2) (Condition (A.7) in Sun et al. (2017)) The function $\pi(\Omega_i, \psi)$ is twice differentiable with respect to ψ on the compact set Θ_ψ, $\pi^{'}(\Omega_i, \psi) = \partial \pi(\Omega_i, \psi)/\partial \psi$ is uniformly bounded, and there is an $\varepsilon > 0$ such that $\pi(\Omega_i, \psi) \geq \varepsilon$ for all i;

(C.3) (Condition (A.8) in Sun et al. (2017)) The functions $\mu_1(\Omega_i, \phi_1)$ and $\mu_2(\Omega_i, \phi_2)$ are twice differentiable with respect to ϕ_1 and ϕ_2 on the compact sets Θ_{ϕ_1} and Θ_{ϕ_2}, repectively.

Lemma C.1. *Under Condition C, if $\pi(\Omega_i)$ and/or both $E\{W_i|\Omega_i\}$ and $E\{W_i^{\otimes 2}|\Omega_i\}$ are correctly specified, we have that*

$$H\hat{\vartheta}_A^*(t,u) \xrightarrow{\mathcal{P}} (\vartheta_0(t,u)^{\mathsf{T}}, \mathbf{0}_p)^{\mathsf{T}}$$

uniformly in $t \in [t_1, t_2], u \in [u_1, u_2]$ as $n \to \infty$.

Proof of Lemma C.1.

Let ψ_0, ϕ_{10} and ϕ_{20} be the true values of ψ, ϕ_1 and ϕ_2 such that $\pi(\Omega_i) = \pi(\Omega_i, \psi_0)$, $E(W_i|\Omega_i) = \mu_1(\Omega_i, \phi_{10})$ and $E(W_i^{\otimes 2}|\Omega_i) = \mu_2(\Omega_i, \phi_{20})$ under the correctly specified models for $\pi(\Omega_i)$, $E(W_i|\Omega_i)$ and $E(W_i^{\otimes 2}|\Omega_i)$, respectively.

We define

$$M_{11,i0}(t|t_0,u_0) = \begin{pmatrix} X_i^{\otimes 2} & X_i\mu_1(\Omega_i,\phi_{10})^\mathsf{T} I(N_i(t^-)>0) \\ \mu_1(\Omega_i,\phi_{10})X_i^\mathsf{T} I(N_i(t^-)>0) & \mu_2(\Omega_i,\phi_{20})I(N_i(t^-)>0) \end{pmatrix},$$

$$M_{12,i0}(t|t_0,u_0) = M_{11,i0}(t|t_0,u_0) \circ \begin{pmatrix} (t-t_0)\mathbb{1}_{p_1\times p_1} & (U_i(t)-u_0)\mathbb{1}_{p_1\times p_2} \\ (t-t_0)\mathbb{1}_{p_2\times p_1} & (U_i(t)-u_0)\mathbb{1}_{p_2\times p_2} \end{pmatrix},$$

$$\begin{aligned} & M_{22,i0}(t|t_0,u_0) \\ =& M_{11,i0}(t|t_0,u_0) \circ \begin{pmatrix} (t-t_0)^2\mathbb{1}_{p_1\times p_1} & (t-t_0)(U_i(t)-u_0)\mathbb{1}_{p_1\times p_2} \\ (t-t_0)(U_i(t)-u_0)\mathbb{1}_{p_2\times p_1} & (U_i(t)-u_0)^2\mathbb{1}_{p_2\times p_2} \end{pmatrix}. \end{aligned}$$

Let $q_{i0} = R_i/\pi(\Omega_i,\psi_0)$ and

$$\begin{aligned} & d\epsilon^*_{i0}(t,\vartheta^*|t_0,u_0) \\ =& dN_i(t) \begin{pmatrix} X_i \\ \mu_1(\Omega_i,\phi_{10})I(N_i(t^-)>0) \\ X_i(t-t_0) \\ \mu_1(\Omega_i,\phi_{10})I(N_i(t^-)>0)(U_i(t)-u_0) \end{pmatrix} \\ & - \begin{pmatrix} M_{11,i0}(t|t_0,u_0) & M_{12,i0}(t|t_0,u_0) \\ M_{12,i0}(t|t_0,u_0)^\mathsf{T} & M_{22,i0}(t|t_0,u_0) \end{pmatrix} \vartheta^*(t_0,u_0)dt. \end{aligned}$$

Since $\hat{\psi}, \hat{\phi}_1$, and $\hat{\phi}_2$ are M-estimators, by Theorems 5.2 and 5.7 in van der Vaart

(1998), we have

$$
\begin{aligned}
& n^{-1}H^{-1}U_A(\vartheta^*, \hat{\psi}, \hat{\phi}_1, \hat{\phi}_2|t_0, u_0) \\
= \; & n^{-1}H^{-1}U_A(\vartheta^*, \psi_0, \phi_{10}, \phi_{20}|t_0, u_0) \\
& + n^{-1}\sum_{i=1}^{n}\int_0^{\tau} K_h(t-t_0)K_b(U_i(t)-u_0)Y_i(t)w_i(t)(\hat{q}_i - q_{i0}) \\
& \qquad\qquad \times \Big[\Big\{dN_i(t) - \lambda_i^*(t, \vartheta^*|t_0, u_0)dt\Big\}\widetilde{X}_i^*(t|t_0, u_0) - d\epsilon_{i0}^*(t, \vartheta^*|t_0, u_0)\Big] \\
& + n^{-1}\sum_{i=1}^{n}\int_0^{\tau} K_h(t-t_0)K_b(U_i(t)-u_0)Y_i(t)w_i(t)(1 - q_{i0}) \\
& \qquad\qquad \times \Big\{d\hat{\epsilon}_i^*(t, \vartheta^*|t_0, u_0) - d\epsilon_{i0}^*(t, \vartheta^*|t_0, u_0)\Big\} \\
& - n^{-1}\sum_{i=1}^{n}\int_0^{\tau} K_h(t-t_0)K_b(U_i(t)-u_0)Y_i(t)w_i(t)(\hat{q}_i - q_{i0}) \\
& \qquad\qquad \times \Big\{d\hat{\epsilon}_i^*(t, \vartheta^*|t_0, u_0) - d\epsilon_{i0}^*(t, \vartheta^*|t_0, u_0)\Big\} \\
= \; & n^{-1}H^{-1}U_A(\vartheta^*, \psi_0, \phi_{10}, \phi_{20}|t_0, u_0) + o_p(1)
\end{aligned} \tag{C.1}
$$

uniformly in $t_0 \in [t_1, t_2], u_0 \in [u_1, u_2]$ if $\pi(\Omega_i)$ and/or both $E\{W_i|\Omega_i\}$ and $E\{W_i^{\otimes 2}|\Omega_i\}$ are correctly specified.

Let $\eta(t, u) = H(\vartheta^*(t, u) - \vartheta_0^*(t, u))$ and $\hat{\eta}_A(t, u) = H(\hat{\vartheta}_A^*(t, u) - \vartheta_0^*(t, u))$. For simplicity, we denote $\eta(t_0, u_0)$, $\hat{\eta}_A(t_0, u_0)$ and $\vartheta_0^*(t_0, u_0)$ by η, $\hat{\eta}_A$ and ϑ_0^* respectively. Let $\lambda_i^*(t, \theta) = \varphi\Big\{\theta^{\mathsf{T}}\widetilde{X}_i^*(t|t_0, u_0)\Big\}$, $\dot{\lambda}_i^*(t, \theta) = \dot{\varphi}\Big\{\theta^{\mathsf{T}}\widetilde{X}_i^*(t|t_0, u_0)\Big\}$, and $\ddot{\lambda}_i^*(t, \theta) = \ddot{\varphi}\Big\{\theta^{\mathsf{T}}\widetilde{X}_i^*(t|t_0, u_0)\Big\}$ where θ is an arbitrary $2p$-dimentional column vector. Then, $\hat{\eta}_A$ is the solution of $U_A(H^{-1}\eta + \vartheta_0^*, \hat{\psi}, \hat{\phi}_1, \hat{\phi}_2|t_0, u_0) = 0$.

Let $D_{a,22}(t_0, u_0) = D_a(t_0, u_0) \circ \mathrm{diag}(\mu_2 \mathbb{1}_{p_1 \times p_1}, \mu_2 \mathbb{1}_{p_2 \times p_2})$. Under Condition C, by the

Taylor expansion and the Lemma A.1 in Yin et al. (2008), we have

$$
\begin{aligned}
& n^{-1}H^{-1}\Big\{U_A(H^{-1}\eta+\vartheta_0^*,\psi_0,\phi_{10},\phi_{20}|t_0,u_0)-U_A(\vartheta_0^*,\psi_0,\phi_{10},\phi_{20}|t_0,u_0)\Big\} \\
=& -n^{-1}\sum_{i=1}^{n}\int_0^{\tau}K_h(t-t_0)K_b(U_i(t)-u_0)Y_i(t)w_i(t)\Big[q_{i0}(H^{-1}\widetilde{X}_i^*(t|t_0,u_0))^{\otimes 2} \\
& \qquad +(1-q_{i0})E\Big\{(H^{-1}\widetilde{X}_i^*(t|t_0,u_0))^{\otimes 2}\Big|\Omega_i\Big\}\Big]\eta dt \\
=& -\operatorname{diag}(D_a(t_0,u_0),D_{a,22}(t_0,u_0))\eta+o_p(\eta) \qquad\qquad \text{(C.2)}
\end{aligned}
$$

uniformly in $t_0\in[t_1,t_2]$, $u_0\in[u_1,u_2]$ and $\eta\in\mathcal{N}_0$, a neighborhood of $\mathbf{0}_{2p}$. Furthermore, by applying Lemma 1 of Zhang et al. (2013), we have

$$
\begin{aligned}
& n^{-1}H^{-1}U_A(\vartheta_0^*,\psi_0,\phi_{10},\phi_{20}|t_0,u_0) \\
=& n^{-1}\sum_{i=1}^{n}\int_0^{\tau}K_h(t-t_0)K_b(U_i(t)-u_0)Y_i(t)w_i(t)H^{-1}\widetilde{X}_i^*(t|t_0,u_0)dM_i(t)+o_p(1) \\
=& o_p(1) \qquad\qquad \text{(C.3)}
\end{aligned}
$$

From (C.1)-(C.3), by Lemma 2 of Sun et al. (2012), we conclude that

$$
H\hat{\vartheta}_A^*(t,u)\xrightarrow{\mathcal{P}}(\vartheta_0(t,u)^{\mathsf{T}},\mathbf{0}_p)^{\mathsf{T}}
$$

uniformly in $t\in[t_1,t_2],u\in[u_1,u_2]$ as $n\to\infty$ if $\pi(\Omega_i)$ and/or both $E\{W_i|\Omega_i\}$ and $E\{W_i^{\otimes 2}|\Omega_i\}$ are correctly specified.

$\square$

Lemma C.2. *Under Condition C, if* $\pi(\Omega_i)$, $E\{W_i|\Omega_i\}$ *and* $E\{W_i^{\otimes 2}|\Omega_i\}$ *are all cor-*

rectly specified, we have

$$\sqrt{nhb}\{\hat{\vartheta}_A(t_0,u_0) - \vartheta_0(t_0,u_0) - \frac{1}{2}h^2\mu_2 D_a^{-1}(t_0,u_0)b_{a,\alpha}(t_0,u_0)$$
$$- \frac{1}{2}b^2\mu_2 D_a^{-1}(t_0,u_0)b_{a,\gamma}(t_0,u_0)\} = D_a^{-1}(t_0,u_0)\sqrt{nhb}\mathbf{A}_{a,n}(t_0,u_0) + o_p(1)$$

uniformly in $t_0 \in [t_1,t_2]$ *and* $u_0 \in [u_1,u_2]$ *as* $nh^6 = O_p(1)$*, where*

$$b_{a,\alpha}(t_0,u_0) = E\left\{ Y_i(t_0)w_i(t)\widetilde{X}_i(t_0)X_i^{\mathsf{T}}(t_0) \middle| U_i(t_0) = u_0 \right\} f_U(t_0,u_0)\ddot{\alpha}(t_0),$$
$$b_{a,\gamma}(t_0,u_0) = E\left\{ Y_i(t_0)w_i(t)\widetilde{X}_i(t_0)W_i^{\mathsf{T}}(t_0)I(N_i(t^-) > 0) \middle| U_i(t_0) = u_0 \right\}$$
$$\times f_U(t_0,u_0)\ddot{\gamma}(u_0),$$
$$\mathbf{A}_{a,n}(t_0,u_0) = n^{-1}\sum_{i=1}^{n}\int_0^{\tau} K_h(t-t_0)K_b(U_i(t)-u_0)Y_i(t)w_i(t)\widetilde{X}_i(t)dM_i(t).$$

Proof of Lemma C.2.

If $\pi(\Omega_i)$, $E\{W_i|\Omega_i\}$ and $E\{W_i^{\otimes 2}|\Omega_i\}$ are all correctly specified, we have

$$\hat{\vartheta}_A(t_0,u_0) - \vartheta_0(t_0,u_0) = D_a^{-1}(t_0,u_0)U_{A,1}(\vartheta_0^*,\psi_0,\phi_{10},\phi_{20}|t_0,u_0) + o_p(1),$$

where

$$U_{A,1}(\vartheta_0^*,\psi_0,\phi_{10},\phi_{20}|t_0,u_0)$$
$$= n^{-1}\sum_{i=1}^{n}\int_0^{\tau} K_h(t-t_0)K_b(U_i(t)-u_0)Y_i(t)w_i(t)\Big[q_{i0}\big(dN_i(t) - \lambda_i^*(t,\vartheta_0^*)dt\big)\widetilde{X}_i(t)$$
$$+ (1-q_{i0})E\Big\{\big(dN_i(t) - \lambda_i^*(t,\vartheta_0^*)dt\big)\widetilde{X}_i(t)\big|\Omega_i\Big\}\Big].$$

By the Taylor expansion, we have

$$
\begin{aligned}
& U_{A,1}(\vartheta_0^*|t_0,u_0) \\
=& \mathbf{A}_{a,n}(t_0,u_0)+\mathbf{B}_{a,n}(t_0,u_0)+\mathbf{C}_{a,n}(t_0,u_0)+\delta_{a,n}(t_0,u_0)+o_p(h^2+b^2),
\end{aligned}
$$

where

$$
\begin{aligned}
& \mathbf{B}_{a,n}(t_0,u_0) \\
=& -\frac{1}{2n}\sum_{i=1}^{n}\int_0^{\tau} K_h(t-t_0)K_b(U_i(t)-u_0)Y_i(t)w_i(t)\ddot{\alpha}(t_0)^{\mathsf{T}}X_i(t)(t-t_0)^2\widetilde{X}_i(t), \\
& \mathbf{C}_{a,n}(t_0,u_0) \\
=& -\frac{1}{2n}\sum_{i=1}^{n}\int_0^{\tau} K_h(t-t_0)K_b(U_i(t)-u_0)Y_i(t)w_i(t)\ddot{\gamma}(u_0)^{\mathsf{T}}W_i(t)I(N_i(t^-)>0) \\
& \qquad \times (U_i(t)-u_0)^2\widetilde{X}_i(t), \\
& \delta_{a,n}(t_0,u_0) \\
=& -\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} K_h(t-t_0)K_b(U_i(t)-u_0)Y_i(t)w_i(t)(1-q_{i0})\Big[\Big(dN_i(t)-\lambda_i^*(t,\vartheta_0^*)dt\Big)\widetilde{X}_i(t) \\
& \qquad -E\Big\{\Big(dN_i(t)-\lambda_i^*(t,\vartheta_0^*)dt\Big)\widetilde{X}_i(t)\big|\Omega_i\Big\}\Big].
\end{aligned}
$$

$\sqrt{nhb}\ \delta_{a,n}(t_0,u_0)=o_p(1)$. Again, following the arguments in Lemma A.1 in Yin et al. (2008), we have that

$$
\frac{1}{h^2}\mathbf{B}_{a,n}(t_0,u_0)\xrightarrow{\mathcal{P}}\frac{1}{2}\mu_2 b_{a,\alpha}(t_0,u_0)
$$

and

$$
\frac{1}{b^2}\mathbf{C}_{a,n}(t_0,u_0)\xrightarrow{\mathcal{P}}\frac{1}{2}\mu_2 b_{a,\gamma}(t_0,u_0).
$$

Therefore, Lemma C.2 holds. $\square$

Proof of Theorem 4.1.

(a) By Lemma C.1, $\hat{\alpha}_A(t,u) \xrightarrow{\mathcal{P}} \alpha_0(t)$ uniformly in $t \in [t_1, t_2], u \in [u_1, u_2]$ as $n \to \infty$ if $\pi(\Omega_i)$ and/or both $E\{W_i|\Omega_i\}$ and $E\{W_i^{\otimes 2}|\Omega_i\}$ are correctly specified. Then,

$$
\begin{aligned}
&\sup_{t\in[t_1,t_2]} |\hat{\alpha}_A(t) \xrightarrow{\mathcal{P}} \alpha_0(t)| \\
=&\sup_{t\in[t_1,t_2]} |n^{-1}\sum_{i=1}^{n}\{\hat{\alpha}_A(t,U_i(t)) - \alpha_0(t)\} \\
\le&\sup_{t\in[t_1,t_2],u\in[u_1,u_2]} |\hat{\alpha}_A(t,u) - \alpha_0(t)| \\
=&o_p(1).
\end{aligned}
$$

(b) Under Condition C, by Lemma C.2, we have

$$
\begin{aligned}
&\sqrt{nh}\{\hat{\alpha}_A(t_0,u_0) - \alpha_0(t_0) - \frac{1}{2}h^2\mu_2\ddot{\alpha}_0(t_0)\} \\
=&\mathcal{I}_1 D_a^{-1}(t_0,u_0)\sqrt{\frac{h}{n}}\sum_{i=1}^{n}\int_0^{\tau} K_h(t-t_0)K_b(U_i(t)-u_0)Y_i(t)w_i(t)\widetilde{X}_i(t)dM_i(t) + o_p(1).
\end{aligned}
$$

Then, by Lemma A.1 in Yin et al. (2008), we obtain

$$
\begin{aligned}
&\sqrt{nh}\{\hat{\alpha}_A(t_0) - \alpha_0(t_0) - \frac{1}{2}h^2\mu_2\ddot{\alpha}(t_0)\} \\
=&\sqrt{\frac{h}{n}}\sum_{i=1}^{n}\int_0^{\tau} K_h(t-t_0)Y_i(t)w_i(t)\mathcal{I}_1\left\{\frac{1}{n}\sum_{j=1}^{n}K_b(U_i(t)-U_j(t_0))D_a^{-1}(t_0,U_j(t_0))\right\} \\
&\times \widetilde{X}_i(t)dM_i(t) + o_p(1) \\
=&\sqrt{nh}\mathbf{A}_{a,n}^{(\alpha)}(t_0) + o_p(1),
\end{aligned}
$$

where

$$
\mathbf{A}_{a,n}^{(\alpha)}(t_0) = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} K_h(t-t_0)Y_i(t)w_i(t)\mathcal{I}_1 D_a^{-1}(t_0,U_i(t))\widetilde{X}_i(t)dM_i(t).
$$

$\sqrt{nh}\mathbf{A}_{a,n}^{(\alpha)}(t_0)$ is a sum of local square integrable martingales, and

$$\begin{aligned}
&nh\langle \mathbf{A}_{a,n}^{(\alpha)}, \mathbf{A}_{a,n}^{(\alpha)}\rangle(t_0)\\
=&\frac{h}{n}\sum_{i=1}^{n}\int_0^\tau K_h(t-t_0)^2 Y_i(t)w_i(t)^2\lambda_{i,0}(t,U_i(t))\left\{\mathcal{I}_1 D_a^{-1}(t_0,U_i(t))\widetilde{X}_i(t)\right\}^{\otimes 2}dt\\
\xrightarrow{\mathcal{P}}&\nu_0\mathcal{I}_1 E\big\{D_a^{-1}(t_0,U_i(t_0))\Sigma_a(t_0,U_i(t_0))D_a^{-1}(t_0,U_i(t_0))\big\}\mathcal{I}_1^{\mathsf{T}},
\end{aligned}$$

where

$$\Sigma_a(t,u) = E\left[Y_i(t)w_i(t)^2\lambda_{i,0}(t,u)\widetilde{X}_i(t)^{\otimes 2}\middle| U_i(t)=u\right]f_U(t,u).$$

Moreover, for any $\epsilon > 0$,

$$\begin{aligned}
&nh\langle \mathbf{A}_{a,n,\epsilon}^{(\alpha)}, \mathbf{A}_{a,n,\epsilon}^{(\alpha)}\rangle(t_0)\\
=&\frac{h}{n}\sum_{i=1}^{n}\int_0^\tau K_h(t-t_0)^2 Y_i(t)w_i(t)^2\lambda_{i,0}(t,U_i(t))X_{i,j}(t,t_0)^{\otimes 2}\\
&\times I\left(\sqrt{\frac{h}{n}}\left|K_h(t-t_0)Y_i(t)w_i(t)^2\lambda_{i,0}(t,U_i(t))X_{i,j}(t,t_0)\right|>\epsilon\right)dt\\
\xrightarrow{\mathcal{P}}&0,
\end{aligned}$$

where $X_{i,j}(t,t_0)$ is the jth component of $\mathcal{I}_1 D_a^{-1}(t_0,U_i(t))\widetilde{X}_i(t)$. By Theorem 5.1.1 in Flemming (1991), we conclude that as $n\to\infty$,

$$\sqrt{nh}(\hat{\alpha}_A(t)-\alpha_0(t)-\frac{1}{2}h^2\mu_2\ddot{\alpha}(t))\xrightarrow{\mathcal{D}}N\left(0,\Sigma_{a,\alpha}(t)\right), \text{for } t\in[t_1,t_2],$$

where

$$\Sigma_{a,\alpha}(t)=\nu_0\mathcal{I}_1 E\big\{D_a^{-1}(t,U_i(t))\Sigma_a(t,U_i(t))D_a^{-1}(t,U_i(t))\big\}\mathcal{I}_1^{\mathsf{T}}.$$

□

Proof of Theorem 4.2.

(a) Following the proof of Theorem 4.1 (a), we can apply Lemma C.1 to prove the consistency of $\hat{\gamma}_A(u_0)$ similarly.

(b) Under Condition C, by Lemma C.2, we have

$$\begin{aligned}
&\sqrt{nb}\{\hat{\gamma}_A(t_0,u_0)-\gamma_0(u_0)-\frac{1}{2}b^2\mu_2\ddot{\gamma}_0(u_0)\}\\
=&\mathcal{I}_2 D_a^{-1}(t_0,u_0)\sqrt{\frac{b}{n}}\sum_{i=1}^{n}\int_0^{\tau}K_h(t-t_0)K_b(U_i(t)-u_0)Y_i(t)w_i(t)\widetilde{X}_i(t)dM_i(t)+o_p(1).
\end{aligned}$$

Then, by Lemma A.1 in Yin et al. (2008), we obtain

$$\begin{aligned}
&\sqrt{nb}\{\hat{\gamma}_A(u_0)-\gamma_0(u_0)-\frac{1}{2}b^2\mu_2\ddot{\gamma}(u_0)\}\\
=&\sqrt{\frac{b}{n}}\sum_{i=1}^{n}\int_0^{\tau}K_b(U_i(t)-u_0)Y_i(t)w_i(t)\mathcal{I}_2\left\{n_{u_0}^{-1}\sum_{t_{u_0}\in\mathcal{V}_{u_0}}K_h(t-t_{u_0})D_a^{-1}(t_{u_0},u_0)\right\}\\
&\times\widetilde{X}_i(t)dM_i(t)+o_p(1)\\
=&\sqrt{nb}\mathbf{A}_{a,n}^{(\gamma)}(u_0)+o_p(1),
\end{aligned}$$

if $n_{u_0}\asymp n$, where

$$\mathbf{A}_{a,n}^{(\gamma)}(u_0)=\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}K_b(U_i(t)-u_0)Y_i(t)w_i(t)\mathcal{I}_2 D_a^{-1}(t,u_0)\widetilde{X}_i(t)dM_i(t).$$

$\sqrt{nb}\mathbf{A}_{a,n}^{(\gamma)}(u_0)$ is a sum of local square integrable martingales, and

$$\begin{aligned}
&nb\langle\mathbf{A}_{a,n}^{(\gamma)},\mathbf{A}_{a,n}^{(\gamma)}\rangle(u_0)\\
=&\frac{b}{n}\sum_{i=1}^{n}\int_0^{\tau}K_b(U_i(t)-u_0)^2Y_i(t)w_i(t)^2\lambda_{i,0}(t,U_i(t))\left\{\mathcal{I}_2 D_a^{-1}(t,u_0)\widetilde{X}_i(t)\right\}^{\otimes 2}dt\\
\xrightarrow{\mathcal{P}}&\nu_0\mathcal{I}_2\Big\{\int_0^{\tau}D_a^{-1}(t,u_0)\Sigma_a(t,u_0)D_a^{-1}(t,u_0)dt\Big\}\mathcal{I}_2^{\intercal}.
\end{aligned}$$

Moreover, for any $\epsilon > 0$,

$$
\begin{aligned}
& nb\langle \mathbf{A}^{(\gamma)}_{a,n,\epsilon}, \mathbf{A}^{(\gamma)}_{a,n,\epsilon}\rangle(u_0) \\
=& \frac{b}{n}\sum_{i=1}^{n}\int_0^{\tau} K_b(U_i(t)-u_0)^2 Y_i(t) w_i(t)^2 \lambda_{i,0}(t, U_i(t)) X_{i,j}(t,u_0)^{\otimes 2} \\
& \times I\left(\sqrt{\frac{b}{n}}\left| K_b(U_i(t)-u_0) Y_i(t) w_i(t)^2 \lambda_{i,0}(t,U_i(t)) X_{i,j}(t,u_0)\right| > \epsilon\right) dt \\
\xrightarrow{\mathcal{P}} & 0,
\end{aligned}
$$

where $X_{i,j}(t,u_0)$ is the jth component of $\mathcal{I}_2 D_a^{-1}(t,u_0)\widetilde{X}_i(t)$. By Theorem 5.1.1 in Flemming 1991, we conclude that as $n \to \infty$,

$$
\sqrt{nb}(\hat{\gamma}(u) - \gamma_0(u) - \frac{1}{2}b^2\mu_2\ddot{\gamma}_0(u)) \xrightarrow{\mathcal{D}} N\left(0, \Sigma_{a,\gamma}(u)\right), \text{for} u \in [u_1, u_2].
$$

where

$$
\Sigma_{a,\gamma}(u) = \nu_0 \mathcal{I}_2 \Big\{ \int_0^{\tau} D_a^{-1}(t,u)\Sigma_a(t,u) D_a^{-1}(t,u) dt \Big\} \mathcal{I}_2^{\mathsf{T}}
$$

□