SEMIPARAMETRIC ADDITIVE HAZARDS MODELS WITH MISSING COVARIATES

by

Pramesh Subedi

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

2021

Approved by:

____

Dr. Yanqing Sun

____

Dr. Eliana Christou

____

Dr. Qingning Zhou

____

Dr. Yufeng Han

©2021
Pramesh Subedi
ALL RIGHTS RESERVED

ABSTRACT

PRAMESH SUBEDI. Semiparametric Additive Hazards Models with Missing Covariates. (Under the direction of Dr. YANQING SUN)

The case-cohort study design was originally proposed by Prentice (1986). Under this design, a random sub-cohort of individuals is selected from the cohort of study. Full covariate data are collected from all the cases in the cohort and the sub-cohort, not all the original cohort, saving time and money if measures such as biomarkers or genotypes are required. Thus, certain covariates will be missing from a large number of individuals in the cohort of study. This design has been widely used in clinical and epidemiological studies to study the effect of covariates on failure times. The Cox proportional hazards model (Cox 1972) is a popular and classical choice in such data due to its nice interpretation of regression coefficients and the availability of efficient inference procedures implemented in all statistical software packages. Few other methods allow for time varying regression coefficients. An underlying assumption of the Cox model is the so-called proportional hazards assumption, that is, the hazard ratio remains constant over time or covariates have log-linear effects on the risk of the event of interest. However, in many real datasets, covariates may exhibit much more complicated effects than log-linear effects; thus, the proportional hazards assumption may be violated, and the Cox model may not be an appropriate choice. In addition, most methods do not use the data of the non-cases that are outside of sub-cohort which results into inefficient inference. Addressing these issues, we have proposed an estimation procedure for the semiparametric additive hazards model for case-cohort data, allowing the covariates of interest to be missing for cases and for non-cases. We have considered an additive model in which effects of some covariates are time varying while the effects of some other covariates are constants. Further, we have assumed that the missing covariates have constant effect on failure time. We have proposed an Augmented Inverse Probability Weighted Estimation (AIPW) procedure. It uses auxiliary information that is correlated with missing covariates. We have established the asymptotic properties of the proposed AIPW estimation. Our simulation study shows that Augmented Inverse Probability Weighted estimation is more efficient than the widely used Inverse probability Weighed (IPW) and Complete case estimation method. This result is apparent if the sub cohort is very small. The method is applied to analyze a data from a HIV vaccine efficacy trial.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my greatest appreciation to my advisor, Dr. Yanqing Sun, who led me to this research field and patiently guided me through the dissertation process. This dissertation could not have been written without her continuous support, encouragement and unmatched mentoring. Research in biostatistics with a background in theoretical mathematics was not easy for me; the simulation part being the toughest in the journey. It was Dr. Sun's guidance and care that kept me moving forward and helped me reach this point.

I would like to express my sincere gratitude to the rest of my dissertation committee members, Dr. Eliana Christou, Dr. Qingning Zhou, and Dr. Yufeng Han, for serving as my committee members and providing me with their thoughtful comments and valuable suggestions. Thank you to all the faculty and staff members in the Mathematics and Statistics Department. I would also like to offer special appreciation to Dr. Shaozhong Deng , Dr. Mohammad A. Kazemi and Mrs. Debra Shafer for their for their proper guidance, support and mentoring . My humble gratitude also goes to everyone who has been a part of my life for the past 6 years at UNC Charlotte for the companionship, support and thoughtful wishes.

I would also like to thank the Mathematics and Statistics Department and the Graduate School for providing me with financial support for my study.This research was partially supported by the National Institutes of Health NIAID grant R37-AI054165 and by the National Science Foundation grant DMS-1915829.

Last but not the least, it would have been impossible to have completed this journey without the support and upliftment from my wife, Kamala Subedi. As well as my son and daughter, Alan and Ananya Subedi, for their round-the-clock assistance. This dissertation is dedicated to my family.

TABLE of CONTENTS

List of Tables

List of Figures

CHAPTER 1

1 Nonparametric and Semiparametric Models for Survival Data

1.1 Survival data

Survival analysis also known as failure time data analysis is the statistical analysis of data where the response variable T is the time from some well-defined origin to the occurrence of event of interest.The event of interest may occur for some individuals and may not occur for others within the study period. When the study ends and analysis begins, the data come as a mixture of complete and incomplete observations.So, the common statistical regression methods will be inappropriate to analyze survival data.

There are two basic concepts that are used in the whole theory of survival analysis; survival function and hazard rate.The survival function, S(t), gives the expected proportion of individuals for which the event has not yet happened by time t. In other words, it gives the probability that the event of interest has not happened by time t. More formally we write

$$S(t) = P(T > t)$$

Where the random variable T denotes the survival time. Often, the survival function decreases as time increases and approaches to zero because more and more individuals will experience the event of interest over time. Survival function is the unconditional probability that the event of interest has not happened by time t. On the other hand, hazard rate $\alpha(t)$ is defined by means of conditional probability. Assuming T as a continuous random variable, it is considered as the probability of experiencing the event of interest in a small time interval $[t, t + dt)$ among those individuals who have not experienced the event of interest by time t. More precisely, the hazard rate is defined as

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

and Cumulative hazard rate is defined as

$$A(t) = \int_0^t \alpha(s)ds$$

The survival function may be calculated from the hazard function by

$$S(t) = exp\Big(\int_0^t -\alpha(s)ds\Big)$$

Let $N(t)$ be the number of events that have occurred by time t,i.e. $N(t) = I(T \leq t)$ Then $N(t)$ is called the counting process. It jumps one unit at the time of each observed event and is constant between events.$N_i(t)$ is the count of the number of occurrences of the event of interest for individual i in $[0, t]$. For survival data $N_i(t) = 1$ if the event has been observed for individual i by time t, otherwise $N_i(t) = 0$. For recurrent event data $N_i(t)$ may take the value larger than 1.

The intensity process $\lambda(t)$ is defined as the conditional probability that an event occurs in $[t, t + dt)$, given all observed prior to this interval, divided by length of the interval, i.e. $\lambda(t)dt = P(dN(t) = 1|past)$

where $dN(t)$ denotes the number of jumps of the process in $[t, t + dt)$. The intensity process of the counting process $N_i(t)$ is assumed to take the form

$$\lambda_i(t) = Y_i(t)\alpha_i(t)$$

where $Y_i(t) = I(T_i \geq t)$ is at risk indicator for individual i, i.e.$Y_i(t) = 1$ if individual i is at risk just before time t and $Y_i(t) = 0$otherwise. The process $N(T) = \sum_{i=1}^n N_i(t)$ counts the total number of observed events. The aggregated counting process has the intensity process

$$\lambda(t) = \sum_{i=1}^n \lambda_i(t) = Y(t)\alpha(t)$$

Where $Y(t) = \sum_{i=1}^n Y_i(t)$ is the number at risk just before time t.

1.2 Case-cohort study design

The case cohort study design for failure time analysis was first introduced by R.L. Prentice in 1986 although it was proposed earlier by Miettinen as case-base design. In epidemiologic cohort studies and disease prevention trials several thousands subjects are to be followed up for a number of years before yielding useful results, and thus can be prohibitively expensive. For example, The Multiple Risk Factor Intervention Trial (MRFIT) , one of the coronary heart disease prevention trials, was conducted at 22 US clinical centers from 1973 to 1982. 12866 men at the age of 35-57 were randomized and reported results after average follow up of 7 years. Much cost and effort in such studies is related to the analysis of raw materials to assemble covariate histories. Measurement of some of these covariates might be too expensive. If the rate of disease occurrence is low, for example only 2% of the MRFIT mean experienced the primary end point of coronary heart disease mortality, much of the covariate information on disease free subjects is redundant.

The case-cohort design is a form of two-phase sampling. At phase I, certain covariates that are available on all study subjects (e.g., treatment assignment, age, gender, and error-prone versions of the expensive true co-variates) are collected. Such data are referred to as the first-phase covariate data. Using some or all of these covariates measured in phase I, a random sample is selected, and complete covariate histories (including all of the expensive covariates not measured at the first phase) are assembled for the cases and the sub-cohort. These data are known as the second-phase covariate data. This design is especially useful in large studies with infrequent occurrence of the failure event, for which the assembly of covariate histories from all cohort members may be prohibitively expensive. The case-cohort data are a biased sample from the study population and thus applying standard methods for randomly sampled data may result in biased estimation.

1.3 Additive hazards models

The additive hazard model, or the additive Aalen model was introduced by Aalen(1980). It assumes that the intensity $\lambda(t)$ of the counting process N(t) conditional on the p-dimensional covariate,

$$X(t) = \Big(X_1(t), X_2(t), \cdots, X_p(t)\Big)^T$$

is of the form

$$\lambda(t) = Y(t)X^T(t)\beta(t)$$

where

$$\beta(t) = \Big(\beta_1(t), \beta_2(t), \cdots, \beta_p(t)\Big)^T$$

is a p-dimensional regression coefficient.This is a nonparametric model because the regression coefficients are fully time varying.It is useful for those data where the main interest is risk difference rather than relative risk. The cumulative regression coefficients defined by

$$B(t) = \int_0^t \beta(s)ds$$

are easier to calculate than the regression coefficients $\beta(t)$ themselves
Michal Kulich and D.Y. Lin (2000)demonstrated how to use case-cohort data to estimate the regression parameter of the additive hazards model. Assuming that the hazard function associated with a set of time dependent covariates $Z(\cdot)$, they proposed an additive hazards model

$$\lambda(t|Z) = \lambda_0(t) + \beta_0^T Z(t)$$

where λ_0 is an unspecified base line hazard function and $\beta_0$is a vector valued regression parameter (Cox & Oakes, 1984, breslow & Day, 197). They have discussed about constructing estimators for the model based on case-cohort data and have shown that the proposed estimators are consistent and asymptotically normal. D.Y. Lin & Zhiliang Ying (1994) constructed a semiparametric estimating function for β_0 which they is consistent and asymptotically normal. Also, they presented an estimator for the cumulative baseline hazard function.

1.4 Semiparametric additive hazards models

The additive Aalen model is very flexible with all regression coefficients being time varying. In many practical settings, however, it is of interest to investigate if the risk associated with some of the covariates is constant with time, that is, if some of the regression coefficients do not depend on time. This is of practical relevance in a number of settings when there is a desire to look more closely at the time-dynamics of covariates effects, such as for example treatment effects in medical studies. Also, when data is limited it is sometimes necessary, as well as sensible, to limit the degrees of freedom of the considered model to avoid too much variance, thus making a variance-bias trade-off to get more precise information. McKeague & Sasieni (1994) considered the semiparametric additive intensity model. It assumes that the intensity is on the form

$$\lambda(t) = Y(t)\big(X^T(t)\beta(t) + Z^T(t)\gamma\big)$$

Where $(X(t), Z(t))$ is a (p+q)- dimensional covariate, $Y(t)$ is the at risk indicator, $\beta(t)$ is a p-dimensional time-varying regression coefficient and γ is a q-dimensional time-invariant coefficient. Hence the effect of some of the covariates may change with time while the effect of others is assumed to be constant.

In some cases it may be more appropriate with models where the effect of covariates are modeled on a multiplicative scale. The multiplicative hazards models encompass the famous proportional hazards model, or the Cox model as it is also called. The Cox model was introduced by Cox (1972) in the context of survival data, and Andersen & Gill (1982) extended it to the counting process framework and gave elegant martingale proofs for the asymptotic properties of the associated estimators. Others that have contributed to establishing asymptotic results for the model are Tsiatis (1981) and Nas (1982). The Cox model assumes that the intensity is of the form

$$\lambda(t) = Y(t)\lambda_0(t)exp(X^T(t)\beta(t))$$

Where

$$X(t) = (X_1(t), X_2(t), \cdots, X_p(t))$$

is a p-dimensional bounded predictable covariate vector and $Y(t)$ is the at risk indicator. The parameters of the model are the p-dimensional regression parameter β and the nonparametric baseline intensity function $\lambda_0(t)$ that is assumed to be locally integrable.

There is an extensive literature in the analysis of case-cohort data. Most statistical methods for case-cohort studies are based on modifications of the full data partial likelihood score function for the Cox proportional hazards model, which weight the contributions of cases and subcohort members by the inverses of true or estimated sampling probabilities.We refer to Prentice (1986), Self and Prentice (1988), Kalbfleisch and Lawless (1988), Lin and Ying (1993), Barlow (1994), Chen and Lo (1999), Borgan et al. (2000), Chen (2001), Kulich and Lin (2004), and Samuelsen, Ånested and Skrondal (2007), among others.

Sun et al. (2016) proposed an estimation procedure for the semiparametric additive hazards models with case-cohort data. They assumed that phase two covariates,allowing it to be missing for cases as well as for non cases, have time varying effect on failure time . The proposed method is more efficient than widely adopted inverse probability weighted complete case estimation method.

In this paper, we considered the general semiparametric additive hazards models of Huffer and McKeague (1991) allowing some covariates have time varying effect on failure times and some covariates have time invariant effect on failure time. we have used Bernoulli two-phase sampling for the selection of subcohort and assumed that the phase two covariate, which is missing , has time invariant effect on failure times.

The approach of Kang, Cai and Chambless (2013) and most of the existing approaches for the additive hazards models under the case-cohort design are based on the inverse probability weighting of complete-case technique of Horvitz and Thompson (1952). With this approach, if a subject has a missing value for one covariate, then the observed values of other covariates together with the observed failure/censoring time of the same subject are not utilized. This leads to loss of efficiency. By adapting the idea of Robins, Rotnitzky and Zhao (1994), we propose an augmented estimating equation on the basis of the inverse probability weighting of complete cases to improve efficiency. It is well known that the augmented inverse probability weighting of complete-case method is doubly robust and is more efficient than the inverse probability weighted complete-case approach when the augmented part is correctly specified (Tsiatis, 2006). The proposed method also utilizes auxiliary variables that have the potential to influence the sampling probabilities and that may improve efficiency through their correlation with the phase-two covariates.

This research is motivated by RV144, a preventive vaccine efficacy trial. RV144 randomized 16,394 HIV-1 negative volunteers to receive vaccine or placebo. They were followed for 42 months for occurrence of HIV-1 infection.Vaccine recipients were distributed in the Low,Medium, and High baseline behavioral risk scores.Three HIV-1 gp120 sequences were included in the vaccine construct: 92TH023 in the ALVAC canarypox vector prime component, and A244 and MN in the AIDSVAX recombinant glycoprotein 120 (gp120) boost component. 92TH023 and A244 are subtype E HIVs, whereas MN is subtype B.

We have applied the proposed method to assess the associations of incompletely observed immune response and behavioral risk scores with the rate of subsequent HIV-1 infection. In RV144, the immune response biomarkers were measured at the Week 26 visit from 34 of 41 vaccine recipients who subsequently acquired HIV-1 infection (cases) and from 205 of 7010 vaccine recipients who completed follow-up HIV-1 uninfected (controls).

1.5 Methodology Description

1.5.1 Inverse probability weighting estimation

Suppose Y is some scalar outcome of interest and X is a set of additional variable and we want to estimate $\mu = E(Y)$. If we have a sample of full data $(Y_i, X_i), i = 1, 2, ..., N$, the unbiased estimator for μ would be the sample mean of Y, i.e.

$$\hat{\mu}^{full} = N^{-1} \sum_{i=1}^{N} Y_i$$

. Note that $\hat{\mu}^{full}$ is the solution of the estimating equation

$$\sum_{i=1}^{N} (Y_i - \mu) = 0$$

Now consider the case of missing data. Let $\delta_i = 1$ if Y_i is observed and $\delta_i = 0$ if Y_i is missing.Then the observed data of N individuals can be written as $(\delta_i, \delta_i Y_i, X_i), i = 1, 2, ..., N$. We assume that missingness of Y depends only on X and not on Y, i.e. $P(\delta = 1|Y, X) = P(\delta = 1|X) = \pi(X)$. We say Y is missing at random. The complete case estimator, the sample mean of the Y_i for the individuals on whom Y is observed,

$$\hat{\mu}^{cc} = \frac{\sum_{i=1}^{N} \delta_i Y_i}{\sum_{i=1}^{N} \delta_i}$$

, is not a consistent estimator for μ. We see that $\hat{\mu}^{cc}$ solves the estimating equation

$$\sum_{i=1}^{N} \delta_i (Y_i - \mu) = 0$$

The inverse probability weighted estimator is constructed by weighting the complete case estimating equation. The inverse probability weighted complete case estimating equation for μ is

$$\sum_{i=1}^{N} \frac{\delta_i}{\pi(X_i)} (Y_i - \mu) = 0$$

which weights the contribution of each complete case i by the reciprocal of $\pi(X_i)$. Solving this estimating equation, the IPW estimator for μ is

$$\hat{\mu}^{ipw} = \left[\sum_{i=1}^{N} \frac{\delta_i}{\pi(X_i)} \right]^{-1} \sum_{i=1}^{N} \frac{\delta_i Y_i}{\pi(X_i)}$$

1.5.2 Augmented inverse probability weighted estimator

The augmented inverse probability weighted estimating equations are formed by adding a regression model to the Inverse Probability Weighted Estimating equation.The AIPW estimating equation for μ is:

$$\sum_{i=1}^{N} \left[\frac{\delta_i}{\pi(X_i)}(Y_i - \mu) - \frac{\delta_i - \pi(X_i)}{\pi(X_i)} E(Y_i - \mu)|X_i \right] = 0$$

Using some algebra, this equation can be written as:

$$\sum_{i=1}^{N} \left[\frac{\delta_i Y_i}{\pi(X_i)} - \frac{\delta_i - \pi(X_i)}{\pi(X_i)} E(Y_i|X_i) - \mu \right] = 0$$

Solving this estimating equation, the AIPW estimator for μ is

$$\hat{\mu}^{aipw} = N^{-1} \sum_{i=1}^{N} \left[\frac{\delta_i Y_i}{\pi(X_i)} - \frac{\delta_i - \pi(X_i)}{\pi(X_i)} E(Y_i|X_i) \right]$$

The IPW estimator is an inconsistent estimator if the model $\pi(X_i)$ is misspecified. Moreover, inverse probability weight case estimators use data only from complete cases and disregard the data from individuals for whom Y_i is missing. Thus, IPW estimators are likely to result in inefficiency. The AIPW estimator for μ is a consistent estimator if the model $\pi(x;\psi)$ for $P(\delta = 1|X = x)$ is correctly specified, or the model $m(x;\xi)$ for $E(Y|X = x)$ is correctly specified. So AIPW estimators are said to be double robust.

CHAPTER 2

2 Estimation of Semiparametric Additive Hazards Models with Missing Covariates

2.1 Preliminaries

Let $U(\cdot) = \{U(t), 0 \leq t \leq \tau\}$ and $V(\cdot) = \{V(t), 0 \leq t \leq \tau\}$ be p and q-dimensional phase-one covariate processes, respectively, where $\tau < \infty$ denotes the time when follow-up ends. Suppose that $Z(\cdot) = \{Z(t), 0 \leq t \leq \tau\}$ is a r-dimensional vector of phase-two covariates. The phase-one covariates $U(\cdot)$ and $V(\cdot)$ are observed for all the cohort members but the phase-two covariates $Z(\cdot)$ are only observed for a subset (subcohort/phase-two sample) of the study subjects. We assume that the conditional hazard function of the failure time T given the covariates $\{U(t), V(t), Z(t),\ 0 \leq t \leq \tau\}$ follows the semiparametric additive hazards model

$$h(t|U(t), X(t)) = \alpha^T(t)U(t) + \beta^T V(t) + \gamma^T Z(t), \tag{1}$$

where $\alpha(t)$ is p dimensional time-varying regression coefficients, and β and γ are q and $r-$ dimensional vectors of time-invariant regression coefficients, respectively. Under model (1), the effects of the covariates $X(t) = (V^T(t), Z^T(t))^T$ are time-invariant while the effects of $U(t)$ change with time. We denote $\theta = (\beta^T, \gamma^T)^T$ as the time-invariant coefficients for $X(t)$.

Let C denote the censoring time of the subject. The observed right-censored failure time can be denoted by $(\tilde{T}, \delta)$, where $\tilde{T} = \min(T, C)$ and $\delta = I(T \leq C)$. We assume that the censoring C is independent given the covariate history in the sense that the censoring does not alter the risk of failure. This assumption is described by $E\{d\tilde{N}(t)|U[0,t], X[0,t], \tilde{T} \geq t\} = E\{dN^*(t)|U[0,t], X[0,t], T \geq t\}$, where $\tilde{N}(t) = I(\tilde{T} \leq t)$, $N^*(t) = I(T \leq t)$, and $X[0,t] = \{X(s), 0 \leq s \leq t\}$ and $Z[0,t] = \{Z(s), 0 \leq s \leq t\}$ are the covariate histories up to time t. Let ξ be the indicator of whether the subject is selected into the phase-two sample (determined via Bernoulli sampling as stated above). A subject with $\xi = 1$ has fully observed covariates $U(\cdot)$, $V(\cdot)$ and $Z(\cdot)$ while a subject with $\xi = 0$ does not have the observed values for $Z(\cdot)$.

Let $\Omega = (\tilde{T}, \delta, U(\cdot), V(\cdot), S)$ be the fully observed part of the data, where S denotes possible auxiliary variables that have the potential to influence the sampling probabilities and may predict the phase-two covariates. We assume that the missingness pattern of $Z(\cdot)$ is non-informative, i.e., not dependent on the unobserved values. This assumption can be expressed as $P(\xi = 1|Z(\cdot), \Omega) = P(\xi = 1|\Omega)$, termed as the missing at random (MAR) assumption in Rubin (1976). However, the sampling probability may depend on any of the phase-one information, Ω.

Let $(\Omega_i, Z_i(\cdot), \xi_i)$, $i = 1, \ldots, n$, be independent identically distributed (iid) copies of $(\Omega, Z(\cdot), \xi)$, where $\Omega_i = (\tilde{T}_i, \delta_i, U_i(\cdot), V_i(\cdot), S_i)$. The observed data are $\{\Omega_i, \xi_i Z_i(\cdot), \xi_i, i = 1, \ldots, n\}$. That is, $\{\tilde{T}_i, \delta_i, X_i(\cdot), U_i(\cdot), S_i\}$ are observed for a subject with $\xi_i = 1$, and $\{\tilde{T}_i, \delta_i, U_i(\cdot), V_i(\cdot), S_i\}$ are observed if $\xi_i = 0$, where $X_i(\cdot) = (V_i^T(\cdot), Z_i^T(\cdot))^T$. The sampling probability, $\theta_i = P(\xi_i = 1|\Omega_i)$, is the conditional probability that $Z_i(\cdot)$ is observed. In particular, this sampling probability depends on the censoring indicator. Under the classical case-cohort Bernoulli sampling design, $\theta_i = 1$ if $\delta_i = 1$ (known as a case) and $\theta_i = P(\xi_i = 1|\Omega_i, \delta_i = 0) < 1$ if $\delta_i = 0$ (known as a non-case). In this paper, we study the semiparametric additive hazards regression model (1) where the covariates can be missing for the cases as well as for the non-cases. Moreover, we assume Bernoulli two-phase sampling such that within each level of a specified phase-one discrete stratification variable defined by δ, $U(\cdot)$, and/or $V(\cdot)$, subjects are selected for measurement of $Z(\cdot)$ based on a random draw from a Bernoulli distribution.

2.2 Inverse probability weighted complete-case estimation

Let $N_i(t) = I(\tilde{T}_i \leq t, \delta_i = 1)$, $Y_i(t) = I(\tilde{T}_i \geq t)$ and $\lambda_i(t) = Y_i(t)h(t|U_i(t), X_i(t))$. Following Horvitz and Thompson (1952), the inverse probability weighting of the complete cases has been commonly used in missing data problems. Suppose that the probability of complete-case $\theta_i = P(\xi_i = 1|\Omega_i)$ is known. Let $A(t) = \int_0^t \alpha(s)\, ds$. Modifying the estimation equations of McKeague and Sasieni (1994) for the fully observed covariates, model (1) can be estimated based on the following inverse probability weighted estimating equations for $A(t)$ and θ:

$$\sum_{i=1}^{n}[q_i Y_i(t)U_i(t)W_i(t)\,(dN_i(t) - \lambda_i(t)dt)] = 0 \tag{2}$$

$$\sum_{i=1}^{n}\int_0^{\tau}[q_i Y_i(t)X_i(t)W_i(t)\,(dN_i(t) - \lambda_i(t)dt)] = 0 \tag{3}$$

where $q_i = \xi_i/\theta_i$ and $W_i(t)$ is a weight process depending only on phase-one variables. The integrals concerned here and later are the Lebesgue integrals defined at each sample point. The integrals are random variables whose values at each sample point are the values of the Lebesgue integrals. In practice, the sampling probability θ_i is unknown. Let $\hat{\theta}_i$ be an estimate of θ_i, say, based on a parametric model such as logistic regression. The inverse probability weighting of the complete case (IPW) estimators of θ and $A(t)$ can be obtained by solving (2) and (3) with q_i replaced by $\hat{q}_i = \xi_i/\hat{\theta}_i$. The estimator for $\alpha(t)$ can be obtained by kernel smoothing the estimator of $A(t)$.

Under the classical case-cohort design, the sampling probability θ_i is 1 for all the cases and equals $\theta_i = P(\xi_i = 1|\Omega_i, \delta_i = 0)$ for subcohort members that are not cases. Hence $q_i = \delta_i + (1-\delta_i)\xi_i/\theta_i$.

Note from (2) and (3) that if a subject has a missing value for $Z_i(\cdot)$, then the observed failure times and the values of $U_i(\cdot)$ and $V_i(\cdot)$ from the same subject are not fully utilized except through the sampling probability θ_i. Hence the inverse probability weighting of complete cases approach is inefficient. In the following we describe an improved estimation procedure to remedy this potential inefficiency.

2.3 Augmented inverse probability weighted estimation

We adapt the idea of Robins, Rotnizky and Zhao (1994) and propose an augmented estimation procedure for model (1) with the case-cohort/two-phase sampling data. The procedure augments the inverse probability weighting of complete cases with auxiliary predictors of the first and second moments of the missing values of phase-two covariates. The new procedure utilizes the information on the conditional distribution of the missing covariates and is thus more efficient.

2.3.1 Estimation with known θ_i and known $E(Z_i(t)|\Omega_i)$ and $E\{Z_i(t)Z_i^T(t)|\Omega_i\}$

First we assume that the sampling probability θ_i and the conditional expectations $E(Z_i(t)|\Omega_i)$ and $E(Z_i(t)Z_i^T(t)|\Omega_i)$ are known for those with missing values of $Z_i(t)$. Let $de_{i,x}(t)$ and $de_{i,u}(t)$ be the conditional expectations of $Y_i(t)X_i(t)W_i(t)\{dN_i(t) - \lambda_i(t)dt\}$ and $Y_i(t)U_i(t)$ $W_i(t)\{dN_i(t) - \lambda_i(t)dt\}$ given Ω_i, respectively. Since $\Omega_i = (\tilde{T}_i, \delta_i, U_i(\cdot), V_i(\cdot), S_i)$ are observed phase-one data, the quantities $de_{i,x}(t)$ and $de_{i,u}(t)$ depend only on Ω_i and $(\theta, \alpha(\cdot))$. Following the augmentation theory of Robins, Rotnizky and Zhao (1994), we propose the following estimating equations for $(\alpha(\cdot), \theta)$:

$$\tilde{U}_1(\alpha(t), \theta) = \sum_{i=1}^{n} \Big[q_i Y_i(t) U_i(t) W_i(t) \{dN_i(t) - \lambda_i(t)dt\} + (1 - q_i)\, de_{i,u}(t) \Big] = 0 \quad (4)$$

$$\tilde{U}_2(\alpha(t), \theta) = \sum_{i=1}^{n} \int_0^{\tau} \Big[q_i Y_i(t) X_i(t) W_i(t) \{dN_i(t) - \lambda_i(t)dt\} + (1 - q_i)\, de_{i,x}(t) \Big] = 0 \quad (5)$$

The contribution to equation (5) from subject i with $\xi_i = 1$ is the weighted average of the observed residual $Y_i(t)X_i(t)W_i(t)\{dN_i(t) - \lambda_i(t)dt\}$ and its conditional expectation $de_{i,x}(t) = E[Y_i(t)X_i(t)W_i(t)\{dN_i(t) - \lambda_i(t)dt\}|\Omega_i]$ with weights q_i and $1-q_i$, respectively. The first part of the contribution, $q_i Y_i(t)X_i(t)W_i(t)\{dN_i(t) - \lambda_i(t)dt\}$, represents the inverse probability weighting of complete-case. The second part, $(1 - q_i)\, de_{i,x}(t)$, is the augmentation to the first part with the knowledge of the conditional expectations $E(Z_i(t)|\Omega_i)$ and $E(Z_i(t)Z_i^T(t)|\Omega_i)$ for the missing covariates. The contribution from subject i with $\xi_i = 0$ only involves the conditional expectation $de_{i,x}(t)$. A similar interpretation applies to equation (4).

Let

$$\begin{aligned}
E_{ux}(t) &= n^{-1}\sum_{i=1}^n W_i(t)Y_i(t)U_i(t)\{q_iX_i^T(t) + (1-q_i)E(X_i^T(t)|\Omega_i)\},\\
E_{uu}(t) &= n^{-1}\sum_{i=1}^n W_i(t)Y_i(t)U_i(t)U_i^T(t), \qquad (6)\\
E_{xx}(t) &= n^{-1}\sum_{i=1}^n W_i(t)Y_i(t)\{q_iX_i(t)X_i^T(t) + (1-q_i)E(X_i(t)X_i^T(t)|\Omega_i)\},\\
E_{un}(t) &= n^{-1}\sum_{i=1}^n \int_0^t W_i(s)Y_i(s)U_i(s)\,dN_i(s),\\
E_{xn}(t) &= n^{-1}\sum_{i=1}^n \int_0^t W_i(s)Y_i(s)\{q_iX_i(s) + (1-q_i)E(X_i(s)|\Omega_i)\}\,dN_i(s).
\end{aligned}$$

Note that

$$E\{X_i(t)|\Omega_i\} = \begin{pmatrix} V_i(t) \\ E(Z_i(t)|\Omega_i) \end{pmatrix}, \qquad (7)$$

$$E\{X_i(t)X_i^T(t)|\Omega_i\} = \begin{pmatrix} V_i(t)V_i^T(t) & V_i(t)E\{Z_i^T(t)|\Omega_i\} \\ E(Z_i(t)|\Omega_i)V_i^T(t) & E(Z_i(t)Z_i^T(t)|\Omega_i) \end{pmatrix}. \qquad (8)$$

If the sampling probability θ_i and the conditional expectations $E(Z_i(t)|\Omega_i)$ and $E(Z_i(t)Z_i^T(t)|\Omega_i)$ for the phase-two covariates are known, then $E_{ux}(t)$, $E_{uu}(t)$, $E_{xx}(t)$, $E_{un}(t)$ and $E_{xn}(t)$ depend only on the observed two-phase data. The estimators for θ and $A(t)$, denoted by $\tilde{\theta}$ and $\tilde{A}(t)$, are based on the estimating equations (4) and (5) and are given explicitly in the following theorem. The proof of Theorem 1 is given at the end of this paper.

Theorem 1. *Assume that the sampling probability θ_i and the conditional expectations $E(Z_i(t)|\Omega_i)$ and $E(Z_i(t)Z_i^T(t)|\Omega_i)$ for the phase-two covariates are known. The estimators of θ and $A(t)$*

obtained by solving (4) and (5) are respectively given by

$$\tilde{\theta} = \left\{\int_0^\tau \left[E_{xx}(t) - E_{xu}(t)E_{uu}^{-1}(t)E_{xu}^T(t)\right] dt\right\}^{-1}\left\{E_{xn}(\tau) - \int_0^\tau E_{xu}(t)E_{uu}^{-1}(t)\, dE_{un}(t)\right\} \quad (9)$$

$$\tilde{A}(t) = \int_0^t E_{uu}^{-1}(s)\, dE_{un}(s) - \int_0^t E_{uu}^{-1}(s)E_{ux}(s)\, ds\, \tilde{\theta} \quad (10)$$

If the sampling probability $\theta_i = 1$ for all subjects, then $q_i = 1$. The estimators $\tilde{\theta}$ and $\tilde{A}(t)$ become the estimators of McKeague and Sasieni (1994) for the full cohort. The estimators $\tilde{\theta}$ and $\tilde{A}(t)$ can be viewed as the expectation maximization (EM) estimators based on the estimating functions of McKeague and Sasieni (1994) for the full cohort. The estimators $\tilde{\theta}$ and $\tilde{A}(t)$ are obtained by replacing the unobserved components $Z_i(t)$ and $Z_i(t)Z_i^T(t)$ in the estimators of McKeague and Sasieni (1994) by their conditional expectations $E(Z_i(t)|\Omega_i)$ and $E(Z_i(t)Z_i^T(t)|\Omega_i)$, respectively. We notice from (6), (7) and (8) that all observations on the covariates $V_i(\cdot)$ and $U_i(\cdot)$ are utilized even for those individuals with missing values of $Z_i(\cdot)$. This estimation procedure utilizes all observed information including $(\widetilde{T}_i, \delta_i, U_i(\cdot), V_i(\cdot))$ for those subjects with unobserved covariates $Z_i(\cdot)$. More efficiency can be achieved with better predictions of $Z_i(\cdot)$ and $Z_i(\cdot)Z_i^T(\cdot)$.

2.3.2 Estimation of θ_i, $E(Z_i(t)|\Omega_i)$, and $E(Z_i(t)Z_i^T(t)|\Omega_i)$ and the AIPW estimator

Application of the estimators $\tilde{\theta}$ and $\tilde{A}(t)$ require knowledge of the sampling probabilities θ_i and/or of the conditional expectations $E(Z_i(t)|\Omega_i)$ and $E(Z_i(t)Z_i^T(t)|\Omega_i)$ of the phase-two covariates, which may be unknown in practice. However, these quantities can be readily estimated under the MAR assumption. Appropriate modelling of the conditional expectations $E(Z_i(t)|\Omega_i)$ and $E(Z_i(t)Z_i^T(t)|\Omega_i)$ for the phase-two covariates can lead to improved efficiency as we will see later with further discussions and in simulations. It is convenient to estimate the terms in question with some well-established parametric methods. However, they can also be estimated with semi- or nonparametric methods.

Assume that $\pi(\Omega_i, \psi)$ is the working parametric model for the probability of complete-case, $\theta_i = P(\xi_i = 1|\Omega_i)$, where ψ is a m-dimensional vector of parameters belonging to a compact set Θ_ψ. For example, with only case status the phase-one stratification variable, one can assume the logistic model with $\text{logit}(\pi(\Omega_i, \psi)) = \psi_1^T \Omega_i$ for those with $\delta_i = 1$ and a different logistic model with $\text{logit}(\pi(\Omega_i, \psi)) = \psi_2^T \Omega_i$ for those with $\delta_i = 0$. In this case, $\psi = (\psi_1, \psi_2)$. The parameter ψ can be estimated by the M-estimator (Huber, 1981), $\hat{\psi}$, that maximizes $\log\left[\prod_{i=1}^n \{\pi(\Omega_i, \psi)\}^{\xi_i}\{1 - \pi(\Omega_i, \psi)\}^{1-\xi_i}\right]$. Therefore, we can estimate $\theta_i = \pi(\Omega_i, \psi)$ by $\hat{\theta}_i = \pi(\Omega_i, \hat{\psi})$. The M-estimators converge even if the true model is not a member of the assumed parametric family (van der Vaart, 1998).

We estimate $E(Z_i(t)|\Omega_i)$ and $E\{Z_i(t)Z_i^T(t)\ |\Omega_i\}$ with the working models $\mu_1(\Omega_i, \varphi_1)$ and $\mu_2(\Omega_i, \varphi_2)$, respectively, where φ_1 and φ_2 are k_1 and k_2 dimensional vectors of parameters belonging to the compact sets Θ_{φ_1} and Θ_{φ_2}, respectively. For example, one can choose $\mu_1(\cdot, \varphi_1)$ and $\mu_2(\cdot, \varphi_2)$ as the first order or second order linear functions of the variables in Ω_i or their transformations. In this case, the parameters φ_1 and φ_2 can be estimated by the M-estimators based on the least squares regressions of $Z_i(t)$ on Ω_i and $Z_i(t)Z_i^T(t)$ on Ω_i, respectively, based on the observations with $\xi_i = 1$ (i.e., those with observed $Z_i(t)$). We denote the estimators of φ_1 and φ_2 by $\hat{\varphi}_1$ and $\hat{\varphi}_2$, respectively.

Let $\hat{E}_{ux}(t)$, $\hat{E}_{xx}(t)$ and $\hat{E}_{xn}(t)$ be the counterparts of $E_{ux}(t)$, $E_{xx}(t)$ and $E_{xn}(t)$ defined in (6), obtained by replacing q_i with $\hat{q}_i = \xi_i/\pi(\Omega_i, \hat{\psi})$, and by replacing $E(Z_i(t)|\Omega_i)$ and $E(Z_i(t)Z_i^T(t)|\Omega_i)$ with $\mu_1(\Omega_i, \hat{\varphi}_1)$ and $\mu_2(\Omega_i, \hat{\varphi}_2)$, respectively. Replacing $E_{ux}(t)$, $E_{xx}(t)$ and $E_{xn}(t)$ by $\hat{E}_{ux}(t)$, $\hat{E}_{xx}(t)$ and $\hat{E}_{xn}(t)$, respectively, in $\tilde{\theta}$ and $\tilde{A}(t)$ defined in (9) and (10), we obtain the following augmented inverse probability weighted complete-case (AIPW) estimators for θ and $A(t)$:

$$\hat{\theta} = \left\{\int_0^\tau \big[\hat{E}_{xx}(t) - \hat{E}_{xu}(t)E_{uu}^{-1}(t)\hat{E}_{xu}^T(t)\big]\,dt\right\}^{-1}\left\{\hat{E}_{xn}(\tau) - \int_0^\tau \hat{E}_{xu}(t)E_{uu}^{-1}(t)\,dE_{un}(t)\right\} \quad (11)$$

$$\hat{A}(t) = \int_0^t E_{uu}^{-1}(s)\,dE_{un}(s) - \int_0^t E_{uu}^{-1}(s)\hat{E}_{ux}(s)\,ds\,\hat{\theta}. \quad (12)$$

The estimators $\hat{\alpha}(t)$ for $\alpha(t)$ can be obtained by using the kernel smoothing for the estimator $\hat{A}(t)$.

CHAPTER 3

3 Asymptotic Properties

This section investigates the asymptotic properties of the proposed estimators. Since the weights q_i and $\hat{q}_i$ are not generally predictable, the asymptotic properties are investigated using the empirical process theory which does not require predictability. Suppose that $\alpha_0(t)$ and θ_0 are the true values of $\alpha(t)$ and θ under model (1). Let

$A_0(t) = \int_0^t \alpha_0(s)\, ds.$

$e_{ux}(t) = E\{W_i(t)Y_i(t)U_i(t)X_i^T(t)\}$, $e_{xu}(t) = e_{ux}^T(t)$

$e_{uu}(t) = E\{W_i(t)Y_i(t)U_i(t)U_i^T(t)\}$

$e_{xx}(t) = E\{W_i(t)Y_i(t)X_i(t)X_i^T(t)\}$

$e_{un}(t) = E\{\int_0^t W_i(s)Y_i(s)U_i(s)\, dN_i(s)\}$

$e_{xn}(t) = E\{\int_0^t W_i(s)Y_i(s)X_i(s)\, dN_i(s)\}$.

and $A = \int_0^\tau \{e_{xx}(t) - e_{xu}(t)e_{uu}^{-1}(t)e_{xu}^T(t)\}\, dt$.

The regularity conditions for the asymptotic results are stated in Condition A given in the Appendix (which includes the assumption of Bernoulli two-phase sampling). Suppose that $\pi(\Omega_i, \psi)$ is the working model for $P(\xi_i = 1|\Omega_i)$, and $\mu_1(\Omega_i, \varphi_1)$ and $\mu_2(\Omega_i, \varphi_2)$ are working models for $E(Z_i(t)|\Omega_i)$ and $E\{Z_i(t)Z_i^T(t)|\Omega_i\}$, respectively. The asymptotic results of the M-estimators are established in Theorem 5.7 and Theorem 5.2 in van der Vaart (1998). The conditions of Theorem 5.7 and Theorem 5.21 in van der Vaart (1998) can be easily checked when $\text{logit}(\pi(\Omega_i, \psi)) = \psi^T \Omega_i$ and $\hat{\psi}$ is the maximizer of the log likelihood function under this working model. The conditions can also be easily checked when $\mu_1(\Omega_i, \varphi_1)$ and $\mu_2(\Omega_i, \varphi_2)$ are linear regression models and $\hat{\varphi}_1$ and $\hat{\varphi}_2$ are the ordinary least squares estimators.

Let ψ^*, φ_1^* and φ_2^* be the limits of the M-estimators $\hat{\psi}$, $\hat{\varphi}_1$ and $\hat{\varphi}_2$, respectively. Let $q_i^* = \xi_i/\pi(\Omega_i, \psi^*)$ and let $E^*\{Z_i(t)|\Omega_i\} = \mu_1(\Omega_i, \varphi_1^*)$ and $E^*\{Z_i(t)Z_i^T(t)|\Omega_i\} = \mu_2(\Omega_i, \varphi_2^*)$. Let $E^*(X_i(t)|\Omega_i)$ and $E^*(X_i(t)X_i^T(t)|\Omega_i)$ correspond to $E(X_i(t)|\Omega_i)$ and $E(X_i(t)\ X_i^T(t)|\Omega_i)$ defined in (7) and (8) with $E(Z_i(t)|\Omega_i)$ and $E(Z_i(t)Z_i^T(t)|\Omega_i)$ replaced by $E^*\{Z_i(t)|\Omega_i\}$ and

$E^*\{Z_i(t)Z_i^T(t)|\Omega_i\}$, respectively. Replacing q_i, $E(X_i(t)|\Omega_i)$ and $E(X_i(t)X_i^T(t)|\Omega_i)$ by q_i^*, $E^*(X_i(t)|\Omega_i)$ and $E^*(X_i(t)X_i^T(t)|\Omega_i)$ in (6) to get $E_{ux}^*(t)$, $E_{xx}^*(t)$ and $E_{xn}^*(t)$ in place of $E_{ux}(t)$, $E_{xx}(t)$ and $E_{xn}(t)$, respectively.

The following theorems show that the AIPW estimators possess the double robustness property, wherein the AIPW estimators $\hat{\theta}$ and $\hat{A}(t)$ are asymptotically unbiased if the sampling probability $P(\xi_i = 1|\Omega_i)$ and/or both the conditional expectations $E(Z_i(t)|\Omega_i)$ and $E(Z_i(t)Z_i^T(t)|\Omega_i)$ are modeled correctly. The asymptotic weak convergence results presented in Theorem 2 and 3 for $n^{1/2}(\hat{\theta}-\theta_0)$ and $G_n(t) = n^{1/2}(\hat{A}(t) - A_0(t))$ over $t \in [0,\tau]$ are useful for construction of a confidence interval for θ, and confidence bands for $A(t)$. The asymptotic weak convergence result in Theorem 3 is also useful for developing hypothesis testing procedures for $\alpha(t)$. The proofs of Theorem 2 and 3 are in appindix A.

Let

$$\begin{aligned}\eta_{u,i} \;=\; & \int_0^\tau e_{xu}(t)e_{uu}^{-1}(t)W_i(t)Y_i(t)\big[U_i(t)\,dN_i(t) \\ & -U_i(t)\{q_i^*X_i^T(t)+(1-q_i^*)E^*(X_i^T(t)|\Omega_i)\}\,\theta_0\,dt - U_i(t)U_i^T(t)\alpha_0(t)\,dt\big].\end{aligned} \tag{13}$$

$$\begin{aligned}\eta_{x,i} \;=\; & \int_0^\tau W_i(t)Y_i(t)\big[\{q_i^*X_i(t)+(1-q_i^*)E^*(X_i(t)|\Omega_i)\}\,dN_i(t) \\ & -[q_i^*X_i(t)X_i^T(t)+(1-q_i^*)E^*\{X_i(t)X_i^T(t)|\Omega_i\}]\,\theta_0\,dt \\ & -\{q_i^*X_i(t)+(1-q_i^*)E^*(X_i(t)|\Omega_i)\}U_i^T(t)\alpha_0(t)\,dt\big].\end{aligned} \tag{14}$$

Theorem 2. *Assuming Condition A, if the sampling probability* $P(\xi_i = 1|\Omega_i) = \pi(\Omega_i, \psi)$, *and/or both the conditional expectations* $E(Z_i(t)|\Omega_i) = \mu_1(\Omega_i, \varphi_1)$ *and* $E(Z_i(t)Z_i^T(t)|\Omega_i) = \mu_2(\Omega_i, \varphi_2)$ *are correctly specified, then the following assertions hold:*

(a) $\hat{\theta} \xrightarrow{P} \theta_0$;

(b) $n^{1/2}(\hat{\theta} - \theta_0) \xrightarrow{\mathcal{D}} N(0, W)$ *as* $n \to \infty$, *where* $W = A^{-1}\Sigma A^{-1}$, $\Sigma = E\{(\eta_{x,i} - \eta_{u,i} + \varepsilon_{\Phi,i} + \varepsilon_{\Psi,i})(\eta_{x,i} - \eta_{u,i} + \varepsilon_{\Phi,i} + \varepsilon_{\Psi,i})^T\}$, *and* $\varepsilon_{\Phi,i}$ *and* $\varepsilon_{\Psi,i}$ *are given in (63) and (64);*

(c) In addition, if $P(\xi_i = 1|\Omega_i) = \pi(\Omega_i, \psi)$ *is correctly specified then* $\varepsilon_{\Phi,i} = 0$; *and if* $E(Z_i(t)|\Omega_i) = \mu_1(\Omega_i, \varphi_1)$ *and* $E(Z_i(t)Z_i^T(t)|\Omega_i) = \mu_2(\Omega_i, \varphi_2)$ *are modelled correctly, then* $\varepsilon_{\Psi,i} = 0$.

The matrix A can be consistently estimated by

$$\hat{A} = \int_0^\tau \{\hat{E}_{xx}(t) - \hat{E}_{xu}(t)E_{uu}^{-1}(t)\hat{E}_{xu}^T(t)\}\, dt,$$

and Σ can be consistently estimated by

$$\hat{\Sigma} = n^{-1}\sum_{i=1}^n (\hat{\eta}_{x,i} - \hat{\eta}_{u,i} + \hat{\varepsilon}_{\Phi,i} + \hat{\varepsilon}_{\Psi,i})(\hat{\eta}_{x,i} - \hat{\eta}_{u,i} + \hat{\varepsilon}_{\Phi,i} + \hat{\varepsilon}_{\Psi,i})^T,$$

where $\hat{\eta}_{u,i}$ and $\hat{\eta}_{x,i}$ are the empirical counterparts of $\eta_{u,i}$ and $\eta_{x,i}$ obtained by replacing θ_0, $A_0(t)$, q_i^*, $e_{xu}(t)$ and $e_{uu}(t)$ with $\hat{\theta}$, $\hat{A}(t)$, $\hat{q}_i = \xi_i/\pi(\Omega_i, \hat{\psi})$, $\hat{E}_{xu}(t)$ and $\hat{E}_{uu}(t)$, respectively, and by replacing the unknown quantities $E^*(Z_i(t)|\Omega_i)$ and $E^*(Z_i(t)Z_i^T(t)|\Omega_i)$ in $E^*(X_i(t)|\Omega_i)$ and $E^*(X_i(t)X_i^T(t)|\Omega_i)$ with $\mu_1(\Omega_i, \hat{\varphi}_1)$ and $\mu_2(\Omega_i, \hat{\varphi}_2)$, respectively. Similarly, $\hat{\varepsilon}_{\Phi,i}$ and $\hat{\varepsilon}_{\Psi,i}$ are the empirical counterparts of $\varepsilon_{\Phi,i}$ and $\varepsilon_{\Psi,i}$ given in (63) and (64).Let

$$\begin{aligned}\zeta_i(t) \;=\; & \int_0^t e_{uu}^{-1}(s)W_i(s)Y_i(s)\big[U_i(s)\, dN_i(s) - U_i(s)U_i^T(s)\, dA_0(s) \\ & -U_i(s)\{q_i^* X_i^T(s) + (1-q_i^*)E^*(X_i^T(s)|\Omega_i)\}\theta_0\, ds\big]. \qquad (15)\end{aligned}$$

Theorem 3. *Assuming Condition A, if the sampling probability* $P(\xi_i = 1|\Omega_i) = \pi(\Omega_i, \psi)$, *and/or both the conditional expectations* $E(Z_i(t)|\Omega_i) = \mu_1(\Omega_i, \varphi_1)$ *and* $E(Z_i(t)Z_i^T(t)|\Omega_i) = \mu_2(\Omega_i, \varphi_2)$ *are correctly specified, then the following assertions hold:*

(a) $\sup_{t\in[0,\tau]} |\hat{A}(t) - A_0(t)| \xrightarrow{P} 0$ *as* $n \to \infty$*;*

(b) The process $G_n(t) = n^{1/2}(\hat{A}(t) - A_0(t))$ *converges weakly to a zero-mean Gaussian process* $G(t)$ *on* $[0, \tau]$ *with the covariance matrix*

$$\Sigma_G(t) = \left\{\zeta_i(t) - \int_0^t e_{uu}^{-1}(s)e_{ux}(s)\,ds\,A^{-1}(\eta_{x,i} - \eta_{u,i}) + \upsilon_{\Phi,i}(t) + \upsilon_{\Psi,i}(t)\right\}^{\otimes 2},$$

where $\eta_{u,i}$ *and* $\eta_{x,i}$ *are defined in (13) and (14), respectively,* $\zeta_i(t)$ *is defined in (15), and the expressions for* $\upsilon_{\Phi,i}(t)$ *and* $\upsilon_{\Psi,i}(t)$ *are given in (76) and (77);*

(c) In addition, if $P(\xi_i = 1|\Omega_i) = \pi(\Omega_i, \psi)$ *is correctly specified then* $\upsilon_{\Phi,i}(t) = 0$ *and* $\varepsilon_{\Phi,i} = 0$*; and if* $E(Z_i(t)|\Omega_i) = \mu_1(\Omega_i, \varphi_1)$ *and* $E(Z_i(t)Z_i^T(t)|\Omega_i) = \mu_2(\Omega_i, \varphi_2)$ *are modelled correctly, then* $\upsilon_{\Psi,i}(t) = 0$ *and* $\varepsilon_{\Psi,i} = 0$*.*

Under Theorem 3, if the sampling probability $P(\xi_i = 1|\Omega_i) = \pi(\Omega_i, \psi)$ and both the conditional expectations $E(Z_i(t)|\Omega_i) = \mu_1(\Omega_i, \varphi_1)$ and $E(Z_i(t)Z_i^T(t)|\Omega_i) = \mu_2(\Omega_i, \varphi_2)$ are correctly specified, then $\upsilon_{\Psi,i}(t) = 0$, $\upsilon_{\Phi,i}(t) = 0$, $\varepsilon_{\Psi,i} = 0$ and $\varepsilon_{\Phi,i} = 0$. The asymptotic covariance matrix of $G_n(t)$ can be estimated consistently by

$$\hat{\Sigma}_G(t) = n^{-1}\sum_{i=1}^{n}\left\{\hat{\zeta}_i(t) - \int_0^t \hat{E}_{uu}^{-1}(s)\hat{E}_{ux}(s)\,ds\,\hat{A}^{-1}(\hat{\eta}_{x,i} - \hat{\eta}_{u,i}) + \hat{\upsilon}_{\Phi,i}(t) + \hat{\upsilon}_{\Psi,i}(t)\right\}^{\otimes 2} \quad (16)$$

where $\hat{\zeta}_i(t)$, $\hat{\eta}_{u,i}$ and $\hat{\eta}_{x,i}$ are the empirical counterparts of $\zeta_i(t)$, $\eta_{u,i}$ and $\eta_{x,i}$, obtained by replacing q_i^* with $\hat{q}_i = \xi_i/\pi(\Omega_i, \hat{\psi})$, and by replacing $E^*(Z_i(t)|\Omega_i)$ and $E^*(Z_i(t)Z_i^T(t)|\Omega_i)$ with $\mu_1(\Omega_i, \hat{\varphi}_1)$ and $\mu_2(\Omega_i, \hat{\varphi}_2)$, respectively. Here $\hat{\upsilon}_{\Phi,i}(t)$ and $\hat{\upsilon}_{\Psi,i}(t)$ are the empirical counterparts of $\upsilon_{\Phi,i}(t)$ and $\upsilon_{\Psi,i}(t)$, respectively.

CHAPTER 4

4 Simulation Study

To study the finite sample properties of the proposed methods, we conducted simulation studies. Let U_1 and $V = (V_1, V_2)^T$ be the phase-one co-variates, where U_1 is a normal random variable with μ =1 and σ = 0.2, V_1 is a uniform random variable on (0,1) and V_2 is a Bernoulli random variable with $P(V_2 = 1) = 0.5$. Let Z be a phase-two co-variate following the uniform distribution on (0,1). The random variables U_1, V_1, V_2 and Z are independent. We consider the following hazard regression model for the failure time T:

$$h(t|X, Z) = \alpha_0(t) + \alpha_1(t)U_1 + \beta_1 V_1 + \beta_2 V_2 + \gamma Z, \quad 0 \leq t \leq \tau, \tag{17}$$

where
$\alpha_0(t) = (0.05 + 0.04t)$, $\alpha_1(t) = (0.03 + 0.05t)$, $\beta_1 = 0.03$, $\beta_2 = 0.02$, $\gamma = 0.02$ and $\tau = 1.5$.
Let S be an auxiliary variable for the phase-two covariate Z with the relationship $S = (Z + \theta\zeta)$, where ζ is a uniform random variable on $(0, 1)$ and θ is a parameter dictating the association between Z and S. The values $\theta = 1.73, 0.8875, 0.3278$ yield correlation coefficients $\rho = 0.50, 0.750.95$ between Z and S, respectively.The AIPW estimators with $\rho = 0.50, 0.75, 0.95$ are denoted by AIPW-R50, AIPW-R75, AIPW-R95, respectively.

Let C^* follow an exponential distribution with mean equal to 10. The censoring time is taken as $C = C^* \wedge \tau$, yielding about 80% censoring of the failure time. Let T be the failure time (time to event) given by model 17 and $\tilde{T} = T \wedge C$ be the observed time, and $\delta = I(T \leq C)$.Let ξ be the indicator of whether is selected in phase-two,i.e. $\xi = 1$ indicates that Z is observed and $\xi = 0$ indicates that Z is missing. The phase-one data for subject i are $\Omega_i = (\delta_i, U_{1i}, V_{1i}, V_{2i}, S_i)$. We consider two scenarios of the two-phase sampling. The first is the classical case-cohort design where the phase-two covariate Z is sampled for all cases and for a selected subset of non-cases. For the non-cases, we assume that the sampling probability $\alpha_i = P(\xi_i = 1|\Omega_i, \delta_i = 0)$ follows a logistic regression model $\text{logit}(\alpha_i) = \pi_0 + \pi_1 S_i + \pi_2 U_{1i} + \pi_3 V_{1i} + \pi_4 V_{2i}$ based on phase-one data. Three different sampling probabilities are considered. The choices of $(\pi_0, \pi_1, \pi_2, \pi_3, \pi_4) = (-0.57, -0.5, -0.5, -0.5, -0.5)$, $(0.5, -0.5, -0.5, -0.5, -0.5)$ and $(0.5, 0.36, -0.5, -0.5, -0.5)$ correspond to the average sampling probabilities $p_0 = 0.10$,0.25 and 0.50 for the non-cases, respectively. We use the linear model with response variable Z_i and predictors $S_i, U_{1i}, V_{1i}, V_{2i}$ and $\log(\tilde{T}_i)$ to estimate $E(Z_i|\Omega_i)$, based on the observations for which Z is not missing and with observed values of Z_i. The linear model with the response variable Z_i^2 and the predictors $S_i, U_{1i}, V_{1i}, V_{2i}$ and $\log(\tilde{T}_i)$ is also used to estimate $E(Z_i^2|\Omega_i)$.

The second two-phase sampling scenario allows Z_i to be missing for cases as well as for non-cases. Let $\vartheta_i = P(\xi_i = 1|\Omega_i, \delta_i = 1)$ and $\alpha_i = P(\xi_i = 1|\Omega_i, \delta_i = 0)$. Allowing differentiation in the sampling probabilities for cases and non-cases, ϑ_i and α_i are modeled with separate logistic regression models using the predictors $S_i, U_{1i}, V_{1i}, V_{2i}$. Our simulation experiment considers the average sampling probabilities $p_1 = 0.5$ for the cases, and $p_0 = 0.10$,0.25 and 0.50 for the non-cases. As with the first set-up, we use linear models with the predictors $S_i, U_{1i}, V_{1i}, V_{2i}$ and $\log(\tilde{T}_i)$ to estimate $E(Z_i|\Omega_i, \delta_i = 1)$ and $E(Z_i^2|\Omega_i, \delta_i = 1)$ based on the observations for which Z is not missing and with observed values of Z_i. Similarly, linear models with the predictors $S_i, U_{1i}, V_{1i}, V_{2i}$ and $\log(\tilde{T}_i)$ are used to estimate $E(Z_i|\Omega_i, \delta_i = 0)$ and $E(Z_i^2|\Omega_i, \delta_i = 0)$ based on the observations that are non-cases and with observed values of Z_i.

Tables 1 ,2 and 3 present our simulation results for $n = 600, 750, 1000$, and for the average sampling probabilities $p_1 = 1.0 and 0.5$ for the cases, and $p_0 = 0.1, 0.25 and 0.50$ for the non-cases. The weight function $W_i(t) = 1$ is used in the simulations. Each entry of Table 1 to Table 3 is based on 1000 simulation runs. Table 1 summarizes the bias (Bias), the empirical standard error (SSE), the average of the estimated standard error (ESE), and the empirical coverage probability (CP) of 95% confidence intervals of the AIPW estimator for θ. Table 1 shows that the AIPW estimator for θ performs well under two-phase sampling with the combinations of the average sampling probabilities $p_1 = 1.0$ and $p_0 = 0.1$, 0.25 and 0.50. The biases are small for the sample sizes $n = 600, 750$ and 1000. The averages of the estimated standard errors are very close to the empirical standard errors and the coverage probabilities are very close to the 0.95 nominal level, indicating appropriateness of the proposed estimator for the variance of $\hat{\theta}$.

Table 1: Bias, empirical standard error (SSE), average of the estimated standard error (ESE) and empirical coverage probabiity (CP) of 95% confidence intervals for the AIPW estimator of θ under model (17) with $\rho = 0.50$ and about 80% censoring percentage based on 1000 simulations, where p_1 is the sampling probability for the cases and p_0 is the sampling probability for the non-cases.

Size	Select. P.		β_1				β_2				γ			
n	p_1	p_0	Bias	SSE	ESE	CP	Bias	SSE	ESE	CP	Bias	SSE	ESE	CP
600	1.0	0.10	-0.0034	0.0568	0.0633	0.958	-0.0032	0.0340	0.0372	0.962	-0.0069	0.1185	0.1654	0.955
	1.0	0.25	-0.0005	0.0568	0.0545	0.948	-0.0001	0.0317	0.0314	0.952	-0.0041	0.0763	0.0746	0.951
	1.0	0.50	-0.0003	0.0529	0.0540	0.949	-0.0013	0.0315	0.0311	0.953	-0.0013	0.0601	0.0598	0.951
	0.5	0.10	-0.0013	0.0615	0.0610	0.950	-0.0013	0.0334	0.0346	0.948	-0.0051	0.1223	0.1456	0.951
	0.5	0.25	-0.0012	0.0535	0.0546	0.958	-0.0019	0.0318	0.0315	0.948	-0.0018	0.0854	0.0904	0.950
	0.5	0.50	0.0010	0.0532	0.0542	0.945	-0.0008	0.0322	0.0313	0.945	-0.0017	0.0749	0.0793	0.963
750	1.0	0.10	-0.0051	0.0486	0.0512	0.943	-0.0021	0.0292	0.0293	0.957	-0.0022	0.0921	0.1122	0.941
	1.0	0.25	-0.0007	0.0494	0.0486	0.951	-0.0015	0.0290	0.0280	0.954	0.0018	0.0659	0.0656	0.944
	1.0	0.50	-0.0009	0.0484	0.0483	0.951	-0.0019	0.0275	0.0278	0.946	0.0002	0.0534	0.0531	0.956
	0.5	0.10	0.0007	0.0496	0.0509	0.953	-0.0012	0.0284	0.0296	0.947	-0.0030	0.0997	0.1162	0.947
	0.5	0.25	-0.0015	0.0485	0.0486	0.960	0.0001	0.0281	0.0281	0.951	0.0008	0.0757	0.0784	0.945
	0.5	0.50	0.0003	0.0469	0.0483	0.949	-0.0011	0.0279	0.0279	0.947	-0.0040	0.0664	0.0690	0.948
1000	1.0	0.10	-0.0069	0.0420	0.0430	0.950	-0.0018	0.0237	0.0248	0.945	0.0012	0.0767	0.0854	0.943
	1.0	0. 25	-0.0004	0.0425	0.0419	0.951	-0.0004	0.0240	0.0242	0.949	-0.0001	0.0538	0.0557	0.946
	1.0	0.50	0.0001	0.0403	0.0417	0.942	-0.0010	0.0242	0.0241	0.954	-0.0015	0.0455	0.0460	0.950
	0.5	0.10	-0.0001	0.0443	0.0430	0.952	-0.0013	0.0245	0.0249	0.961	0.0011	0.0816	0.0917	0.954
	0.5	0. 25	-0.0008	0.0414	0.0420	0.945	-0.0009	0.0237	0.0242	0.945	-0.0005	0.0662	0.0663	0.945
	0.5	0.50	-0.0021	0.0410	0.0418	0.951	0.0002	0.0247	0.0241	0.950	-0.0052	0.0560	0.0586	0.951

We also compare the performance of the proposed AIPW estimator with the IPW estimator described in Section 2.3 and the complete-case (CC) estimator obtained by deleting subjects with missing values of Z_i. We present the estimation results for the full cohort where all the values of Z_i are fully observed, which is denoted by Full.

Table 2 compares the bias of these estimators for estimating θ. It shows that the biases of both the IPW and AIPW estimators for θ are very small at a level comparable to the Full estimator,as if all the values of the covariate Z_i were observed. The complete-case estimator (CC) yields larger biases.

Table 2: Comparison of Bias for the AIPW,IPW, CC and Full estimators of θ under model (17) with $\rho = 0.50$ and about 80% censoring percentage based on 1000 simulations, where p_1 is the sampling probability for the cases and p_0 is the sampling probability for the non-cases.

Size	Select. P.		Bias(β_1)				Bias(β_2)				Bias(γ)			
n	p_1	p_0	Full	AIPW	IPW	CC	Full	AIPW	IPW	CC	Full	AIPW	IPW	CC
600	1.0	0.10	-0.0008	-0.0034	-0.0160	-0.2192	-0.0014	-0.0032	-0.0108	-0.1923	-0.0014	-0.0069	-0.0166	-0.1951
	1.0	0.25	-0.0002	-0.0005	-0.0027	-0.1546	0.0006	-0.0001	-0.0015	-0.1467	-0.0001	-0.0041	-0.0063	-0.1485
	1.0	0.50	-0.0001	-0.0003	-0.0013	-0.0790	-0.0010	-0.0013	-0.0016	-0.0752	0.0000	-0.0013	-0.0013	-0.0297
	0.5	0.10	-0.0002	-0.0013	-0.0072	-0.1092	0.0006	-0.0013	-0.0064	-0.0946	-0.0001	-0.0051	-0.0125	-0.2298
	0.5	0.25	-0.0008	-0.0012	-0.0032	-0.0558	-0.0014	-0.0019	-0.0035	-0.0466	-0.0014	-0.0018	-0.0047	-0.1500
	0.5	0.50	0.0010	0.0010	0.0000	0.0001	-0.0007	-0.0008	-0.0009	-0.0008	-0.0005	-0.0017	-0.0017	-0.0001
750	1.0	0.10	-0.0038	-0.0051	-0.0137	-0.2107	-0.0009	-0.0021	-0.0078	-0.1914	-0.0002	0.0022	-0.0089	-0.1963
	1.0	0.25	-0.0001	-0.0007	-0.0034	-0.1594	-0.0012	-0.0015	-0.0022	-0.1487	0.0027	0.0018	-0.0001	-0.1378
	1.0	0.50	-0.0008	-0.0009	-0.0014	-0.0782	-0.0020	0.0019	0.0014	-0.0692	0.0001	0.0002	0.0004	0.0324
	0.5	0.10	0.0011	0.0007	-0.0068	-0.1117	-0.0002	-0.0012	-0.0059	-0.0928	-0.0021	-0.0030	-0.0094	-0.2343
	0.5	0.25	-0.0012	-0.0015	-0.0022	-0.0533	-0.0005	0.0000	-0.0014	-0.0467	0.0019	0.0008	-0.0016	-0.1448
	0.5	0.50	0.0004	0.0003	0.0013	0.0002	-0.0010	-0.0011	-0.0011	-0.0027	-0.0020	-0.0040	-0.0044	-0.0060
1000	1.0	0.10	0.0001	-0.0009	-0.0051	-0.1986	-0.0012	-0.0018	-0.0054	-0.1896	0.0020	0.0012	-0.0032	-0.1845
	1.0	0.25	0.0005	0.0002	-0.0023	-0.1602	-0.0001	-0.0004	-0.0016	-0.1477	-0.0002	-0.0009	-0.0009	-0.1434
	1.0	0.50	0.0001	0.0001	0.0001	-0.0771	-0.0010	-0.0010	-0.0013	-0.0749	0.0009	0.0015	0.0017	0.0343
	0.5	0.10	0.0006	-0.0001	-0.0065	-0.1184	-0.0005	-0.0013	-0.0054	-0.0980	0.0023	0.0011	-0.0049	-0.2357
	0.5	0.25	-0.0003	-0.0008	-0.0016	-0.0451	-0.0005	-0.0009	-0.0018	-0.0467	0.0005	-0.0005	-0.0010	-0.1458
	0.5	0.50	-0.0021	-0.0021	-0.0015	-0.0017	0.0002	0.0002	0.0005	0.0016	-0.0033	-0.0052	-0.0053	-0.0034

Table 3: Relative efficiency (REE) of the AIPW, IPW, CC estimators compared to the Full estimator for θ under model (17) with $\rho = 0.50$ and about 80% censoring percentage based on 1000 simulations, where p_1 is the sampling probability for the cases and p_0 is the sampling probability for the non-cases.

Size	Select. P.		REE(β_1)			REE(β_2)			REE(γ)		
n	p_1	p_0	AIPW	IPW	CC	AIPW	IPW	CC	AIPW	IPW	CC
600	1.0	0.10	0.9348	0.5869	0.2169	0.9201	0.6287	0.2265	0.4551	0.4642	0.2289
	1.0	0. 25	0.9957	0.8662	0,3418	0.9876	0.8904	0.3374	0.7266	0.7028	0.3429
	1.0	0.50	0.9969	0.9598	0.5564	0.9979	0.9714	0.5679	0.9058	0.9014	0.5624
	0.5	0.10	0.9202	0.6133	0.2120	0.9382	0.6713	0.2048	0.4536	0.4454	0.2159
	0.5	0. 25	0.9923	0.8241	0.3689	0.9858	0.8735	0.3975	0.6315	0.6158	0.3874
	0.5	0.50	0.9973	0.9225	0.6887	0.9925	0.9468	0.6936	0.7229	0.7174	0.6904
750	1.0	0.10	0.9721	0.6332	0.2216	0.9635	0.6954	0.2238	0.5278	0.4942	0.2250
	1.0	0.25	0.9886	0.8888	0.3451	0.9951	0.9113	0.3538	0.7110	0.6959	0.3269
	1.0	0.50	0.9982	0.9761	0.5689	0.9972	0.99862	0.5527	0.9107	0.9108	0.5721
	0.5	0.10	0.9557	0.6355	0.2101	0.9499	0.6924	0.2046	0.4769	0.4509	0.2174
	0.5	0.25	0.9910	0.8299	0.3819	0.9938	0.8956	0.3777	0.6287	0.6092	0.3800
	0.5	0.50	0.9961	0.9436	0.6835	0.9971	0.9453	0.7070	0.7197	0.7194	0.6919
1000	1.0	0.10	0.9652	0.6884	0.2219	0.9817	0.7525	0.2206	0.5558	0.5301	0.2335
	1.0	0.25	0.9948	0.8901	0.3450	0.9924	0.9256	0.3469	0.7722	0.7502	0.3529
	1.0	0.50	0.9966	0.9761	0.5567	0.9988	0.9815	0.5479	0.9174	0.9099	0.5714
	0.5	0.10	0.9753	0.6754	0.2224	0.9613	0.7174	0.2187	0.4947	0.4592	0.2222
	0.5	0.25	0.9927	0.8854	0.3792	0.9919	0.8881	0.3744	0.6345	0.6241	0.3747
	0.5	0.50	0.9999	0.9364	0.7159	0.9984	0.9651	0.7242	0.7276	0.7187	0.7140

Table 3 compares the relative efficiency (REE) of the AIPW, IPW and CC estimators relative to the Full estimators, where REE for each of the estimators is defined as SSE of the Full estimator divided by SSE of the corresponding estimator. It shows that the relative efficiency of the AIPW-R50 estimator for θ is larger than the relative efficiency of the IPW estimator, which is in turn larger than that of the complete-case estimator, in each case. The efficiency of the AIPW estimator gains the most over the IPW estimator when the sampling probability for the non-cases is small, e.g., $p_0 = 0.1$. This is because the AIPW estimator can more efficiently utilize information on the failure time and other fully observable covariates for individuals with missing covaiate(s).

Figure 1 presents the comparison of the estimators for the nonparametric component $A_1(t) = \int_0^t \alpha_1(s)\, ds$ for $n = 600$ and with average sampling probabilities $p_1 = 0.5$ for the cases and $p_0 = 0.1$ for the non-cases. Figure 1(a) plots the biases of the estimators for $A_1(t)$

for $0 < t \leq 1.5$ for each of the estimators, AIPW-R50, AIPW-R75, AIPW-R95, IPW, CC and Full. Figure 1(b) plots the relative efficiencies of these estimators. The coverage probabilities of the 95% pointwise confidence intervals for $A_1(t)$ for each t using the AIPW estimators are given in Figure 1(c).The coverage probability of the IPW estimator is not presented because the estimation of its standard error is much more complicated due to lack of orthogonality, that is, one has to take into consideration the estimation variance for the sampling probability models.

Figure 1(a) shows that the estimation biases of both the IPW and AIPW-R50, AIPW-R75, AIPW-R95 estimator for $A_1(t)$ are very small, comparable to the estimator as if all the values of the covariate Z_i were observed.The bias of the complete case estimation of $A_1(t)$ is much larger.Figure 1b shows that the relative efficiency of AIPW-R50, AIPW-R75 and AIPW-R95 is significantly greater than that of IPW estimation and complete case estimation.The relative efficiency of AIPW-R50 is slightly smaller than that of AIPW-R75 and AIPW-R95. Fig 1c shows that the point wise coverage probability for $A_1(t)$ for AIPW-R50, AIPW-R75 and AIPW-R95 estimators are very close to the 0.95 nominal level.

The simulation results on relative efficiency, for $n = 600$ with $p_1 = 0.5$ and $p_0 = 0.1$, for estimating γ is 0.4536 for AIPW-R50, 0.4454 for IPW and 0.2159 for complete case estimator. The relative efficiencies for estimating $\beta_1 and \beta_2$ are 0.9202 and 0.9382 for AIPW, 0.6133 and 0.6713 for IPW estimator, respectively.The relative efficiency of estimating $A_1(t)$ for AIPW is close to 0.99 and is less than 0.80 for IPW. It shows that the advantage of using AIPW estimator over the IPW estimator is greater for estimating the effects of those covariates that are fully observed and less for the covariates with missing values.

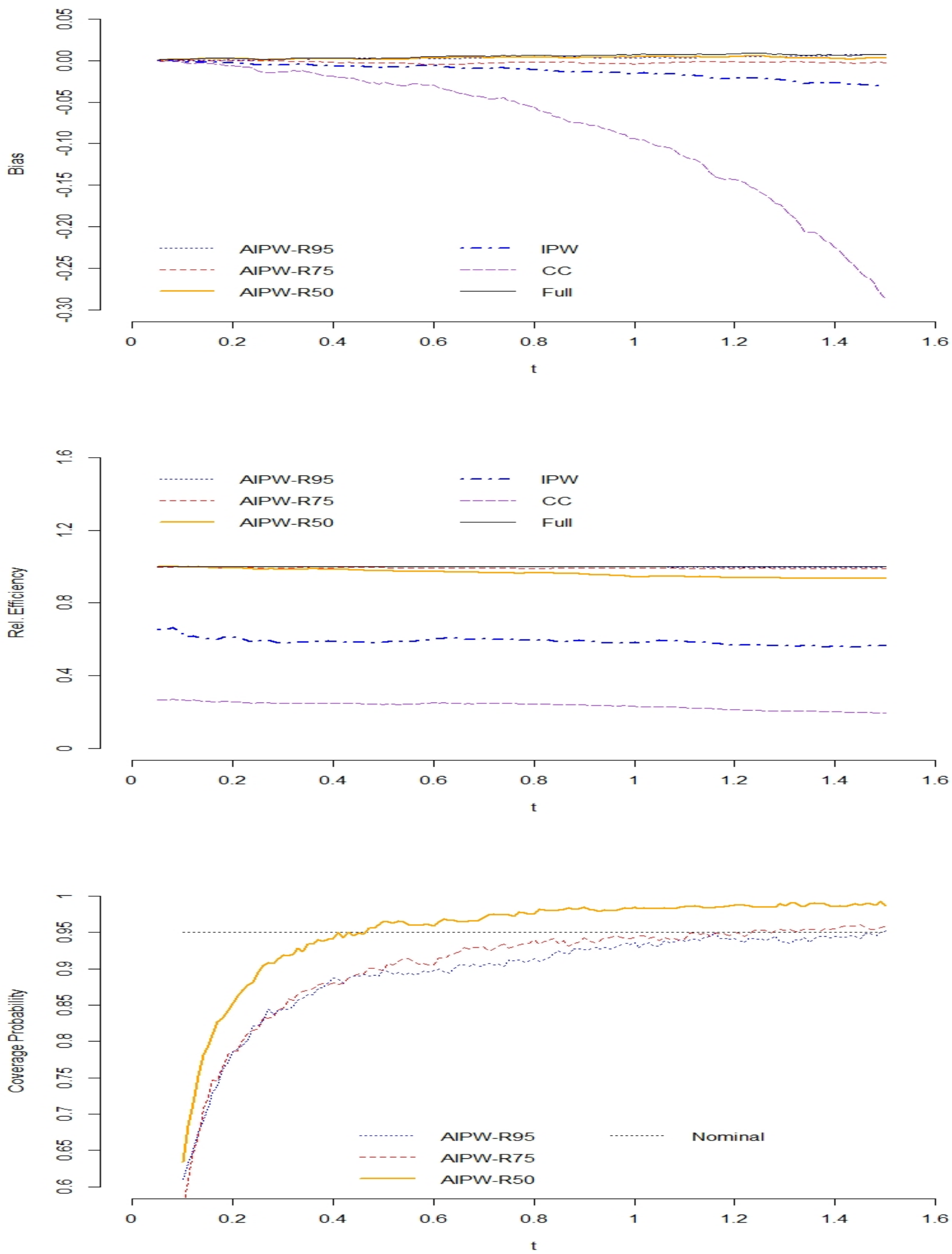

Figure 1: Comparision of the AIPW-R50, AIPW-R75, AIPW-R95, IPW, CC and Full estimators for the cumulative coefficient $A_1(t)$ under model (17) based on 1000 simulations with $n = 600$ and with sampling probabilities 0.5 for the cases and 0.10 for the non cases: (a) the plots of the biases of the estimates: (b) the plots of the relative efficiencies of the estimators: (c) the coverage probabilities of the 95% pointwise confidence intervals for $A_1(t)$for each t using the AIPW estimators

CHAPTER 5

5 Data Application

The RV144 was a preventive HIV vaccine efficacy trial randomized 16,394 HIV-1 negative volunteers to receive vaccine (n = 8198) and placebo (n= 8196) and monitered them for 42 months for occurrence of the primary study endpoint of HIV-1 infection. We apply the proposed AIPW estimation procedure to the vaccine group,which included 5035 men and 3163 women. They were assigned to received four vaccinations at weeks 0, 4, 12 and 24. Blood samples were collected from all participants at the week 26 visit for potentially measuring biomarkers of immune response to the vaccine. Participants were monitored for 42 months after the week 26 visit for occurrence of the primary endpoint of HIV-1 infection, with 43 observed HIV-1 infections among the 8198 vaccine recipients.

The tested vaccine contained three specific HIV-1 gp120 sequences −92TH023 in the ALVAC canarypox vector prime component administered at week 0, 4, 12, 24 and A244 and MN in the AIDSVAX protein boost component administered at week 12 and 24. The 92TH023 and A244 sequences are CRF01_AE HIV-1s whereas the MN sequence is a subtype B HIV-1. Since the CRF01_AE vaccine-insert sequences were genetically much closer to the population of HIV-1 sequences that trial participants were exposed to than MN, we expect they were more likely to have induced protective immune responses. Moreover, the A244 vaccine-insert sequence is of special interest because various analyses have supported that it may have the best induced protective V2 antibodies (e.g., Alam et al., 2013).

Accordingly, our analysis focuses on the A244 vaccine-insert sequence and on post Week 26 HIV-1 infection endpoints of vaccine recipients. The observed failure time $\tilde{T}_i$ is the time from week 26 visit to the HIV-1 infection diagnosis (the failure time) or the time to right censoring (study dropout, administrative censoring at 42 months, etc.).

Many papers (e.g., Haynes et al., 2012; Yates et al., 2014; Zolla-Pazner et al., 2014) reported analyses supporting that vaccine recipients with higher week 26 levels of antibodies binding to the V1V2 portion of the HIV-1 envelope protein had a significantly lower rate of HIV-1 infection over the subsequent 36 months. This observation was made for antibodies measured to each of several specific V1V2 sequences, including the A244 vaccine-insert sequence. In general, vaccines induce immune responses to the vaccine-insert sequences but not necessarily to sequences not inside the vaccine construct, suchthat vaccines tend to confer greater efficacy to prevent infection or disease with sequences matching the vaccine-insert sequences than with sequences mismatching the vaccine-insert sequences. Therefore, it is of interest to refine the analysis of the association of V1V2 antibodies with HIV-1 infection to account for the genetic type of the HIV-1 infection defined in terms of the V1V2 sequence. In particular, the theory is that if A244-induced V1V2 antibodies are a cause of vaccine efficacy, then we would expect to see that the association of V1V2 antibodies with HIV-1 infection would be stronger against infection with V1V2 sequences close to the A244 sequence than against infection with V1V2 sequences far from the A244 sequence.

Our analysis studies the same week 26 biomarker measuring level of IgG antibodies to the V1V2 portion of A244 that was previously studied by Yang et al. (2016); we label this week 26 biomarker IgG-A244V1V2. This biomarker wasmeasured for 34 of 41 HIV-1 infected vaccine recipients with HIV-1 V1V2 sequence data and for a random sample of 205 vaccine recipient controls. The observed biomarker was standardized to have mean 0 and variance 1 for the analysis.

Let R_i denote the biomarker IgG-A244V1V2 for subject i. The amount of exposure to HIV-1 includes the phase-one covariate baseline behavirol risk score in the model with three levels(low, medium and high). We code this risk score by two dummy indicator variables U_1 and U_2 , where $U_1 = 1$ if subject i is in the Low group and $U_2 = 1$ if subject i is in the Medium group. The biomarker $R_i$is the phase-two covariate, which can be missing for both case and non-case subjects.

We consider the following semiparametric additive hazards regression model

$$h(t|X,Z) = \alpha_0(t) + \alpha_1(t)U_1 + \alpha_2(t)U_2 + \gamma Z, \quad 0 \le t \le 3.5, \tag{18}$$

where $X = (1, U_1, U_2)^T$ and $Z = R_i$. Let ξ_i be the indicator that R_i is observed in phase-two sampling. In this data , R_i is measured in 79% (34 out of 43) and 2.5% (205 of 8155)of the non-cases. The phase -two variable, R_i, has mean (μ)= 9.2462 and standard deviation(s)=1.5991.We have standardized this covariate. The correlation between the covariates vaccine number and the biomarker $A244V1V2$ is 0.8538. So,We consider num_vacc as the auxiliary variable denoted by S. Two logistic regression models are used to model the sampling probabilities for the cases and for the non-cases separately. The estimated sampling probabilities $\hat{\vartheta}_i$ for the cases are given by $\text{logit}(\hat{\vartheta}_i) = -5.6680 + 1.9357 S_i - 0.1430 U_{1i} + 2.3336 U_{2i}$ with respective standard errors 2.4488, 0.6597, 1.1213, 1.9943. The estimated sampling probabilities $\hat{\alpha}_i$ for the non-cases are given by $\text{logit}(\hat{\alpha}_i) = -6.3159 + 0.7283 S_i - 0.2012 U_{1i} - 0.1117 U_{2i}$ with standard errors 0.7129, 0.1784, 0.1737, 0.1895. The weights q_i are estimated by $\hat{q}_i = \delta_i \xi_i / \hat{\vartheta}_i + (1-\delta_i)\xi_i/\hat{\alpha}_i$.

Let $Q_i = (1, S_i, U_{1i}, U_{2i}, \log(\tilde{T}_i))$. The linear models with the predictors Q_i are used to estimate $E(Z_i|\Omega_i, \delta_i = 1)$ and $E(Z_i^2|\Omega_i, \delta_i = 1)$ based on the observations that are cases and with observed Z_i's. The term $E(Z_i|\Omega_i, \delta_i = 1)$ is estimated by $\hat{\eta}^T Q_i$ where $\hat{\eta} = (-9.2865, 2.4023, -0.85600, -0.5167, 0.2414)^T$ with respective standard errors 1.1776, 0.2992, 0.2781, 0.2686, 0.2437, and $E(Z_i^2|\Omega_i, \delta_i = 1)$ is estimated by $\hat{\nu}^T Q_i$ where $\hat{\nu} = (33.0345, -8.1843, 0.6707, 0.1602, -0.2040)^T$ with respective standard errors 1.1516, 0.2926, 0.2720, 0.2626, 0.2384. Similarly, the linear models with the predictors Q_i are used to estimate $E(Z_i|\Omega_i, \delta_i = 0)$ and $E(Z_i^2|\Omega_i, \delta_i = 0)$ based on the observations that are non-cases and with observed Z_i's. The term $E(Z_i|\Omega_i, \delta_i = 0)$ is estimated by $\hat{\varphi}^T Q_i$ where $\hat{\varphi} = (-7.3815, 2.5245, 0.0358, -0.0281, -2.0291)^T$ with respective standard errors 1.97486, 0.10066, 0.08417, 0.09056, 1.58167, and $E(V_i^2|\Omega_i, \delta_i = 0)$ is estimated by $\hat{\kappa}^T Q_i$ where $\hat{\kappa} = (42.2632, -10.7899, 0.1296, -0.1157, 0.8496)^T$ with respective stand errors 7.4488, 0.3797, 0.3175, 0.3416, 5.9658.

Our method with $W_i(t) = 1$ gives the estimated effect of the immune response $\hat{\gamma} = -0.0001819378$ with standard error of 0.0001653135, yielding p-value 0.2711 for testing $\gamma = 0$, confirming that the immune response does not have significant effect on the hazard of HIV-1 infection. The estimates of the cumulative coefficients $A_0(t) = \int_0^t \alpha_0(s)\,ds$, $A_1(t) = \int_0^t \alpha_1(s)\,ds$ and $A_2(t) = \int_0^t \alpha_2(s)\,ds$ are plotted in Figure 2 along with their 95% pointwise confidence bands. Figure 2(b) and 2(c) indicate that estimated cumulative coefficients are decreasing.

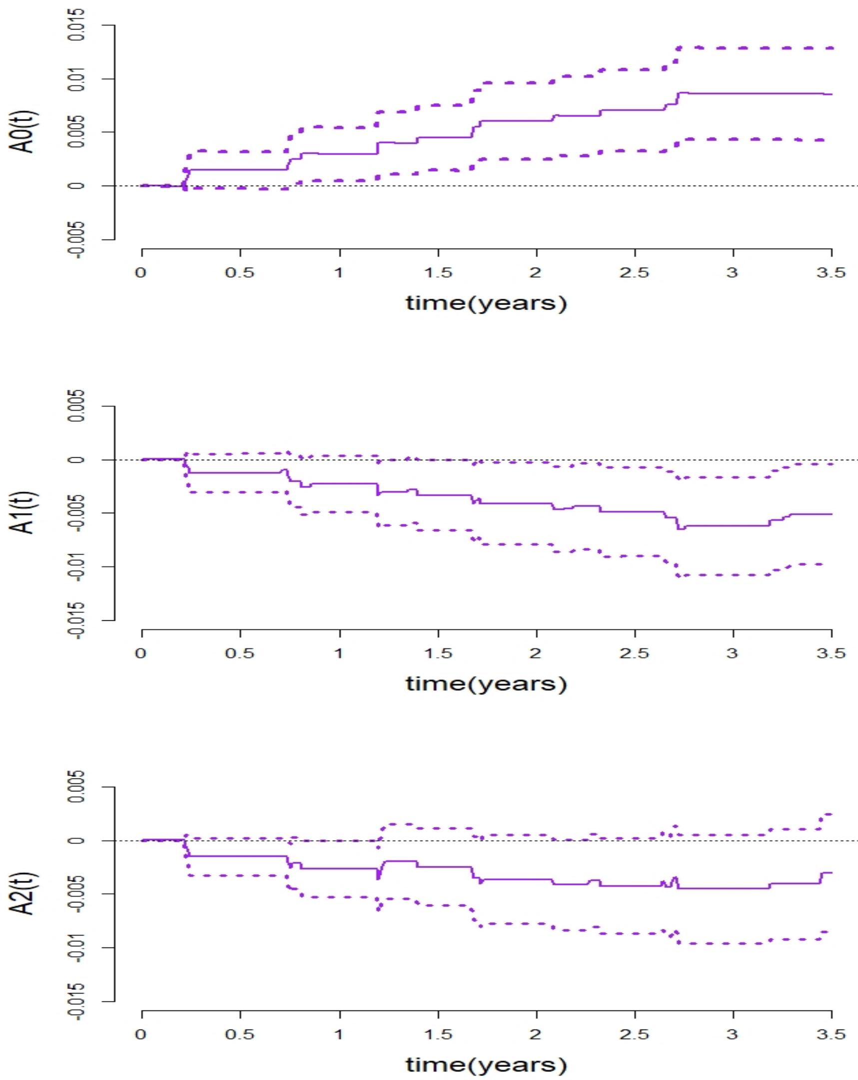

Figure 2: The AIPW estimation of the cumulative coefficients under model (18) using vaccine number as auxiliary variable: (a) the plot of $\hat{A}_0(t)$ for the cumulative baseline function with 95% pointwise confidence bands; (b) the plot of $\hat{A}_1(t)$ for the cumulative effect of risk score of low group with 95% pointwise confidence bands; (c) the plot of $\hat{A}_2(t)$ for the cumulative effect of risk score of medium group with 95% pointwise confidence bands

References

Aalen, O.O. (1980). A model for nonparametric regression analysis of counting processes. *Lecture Notes in Statistics-2: Mathematical Statistics and Probability Theory*, 1–25. Springer Verlag, New York.

Barlow, W.E. (1994). Robust variance estimation for the case-cohort design. *Biometrics* **50**, 1064–1072.

Borgan, Ø., Langholz, B., Samuelsen, S.O., Goldstein, L. and Pogoda, J. (2000). Exposure stratified case-cohort designs. *Lifetime Data Analysis* **6**, 39–58.

Breslow, N. E. and Lumley, T. (2013). Semiparametric models and two-phase samples: Applications to Cox regression. *Banerjee, M., Bunea, F., Huang, J., Koltchinskii, V., and Maathuis, M. H., eds., From Probability to Statistics and Back: High-Dimensional Models and Processes – A Festschrift in Honor of Jon A. Wellner, (Beachwood, Ohio, USA: Institute of Mathematical Statistics)* **9**, 65–77.

Breslow, N., Lumley, T., Ballantyne, C., Chambless, L. and Kulich, M. (2009a). Improved Horvitz-Thompson estimation of model parameters from two-phase stratified samples: Applications in epidemiology. *Statistics in Biosciences* **1**, 32–49.

Breslow, N., Lumley, T., Ballantyne, C., Chambless, L. and Kulich, M. (2009b). Using the whole cohort in the analysis of case-cohort data. *American Journal of Epidemiology* **169**, 1398–1405.

Breslow, N. and Wellner, J. (2007). Weighted likelihood for semiparametric models and two-phase stratified samples, with application to Cox regression. *Scandinavian Journal of Statistics* **34**, 86–102.

Chen, K. (2001). Generalized case-cohort sampling. *Journal of the Royal Statistical Society*, Ser. B **63**, 791–809.

Chen, K. and Lo, S.H. (1999). Case-cohort and case-control analysis with Cox's model. *Biometrika* **86**, 755–764.

Cheng, S.C., Wei, L.J. and Ying , Z. (1995). Analysis of transformation models with censored data. *Biometrika* **82**, 835–845.

Flynn, N. M., Forthal, D. N., Harro, C. D., Judson, F. N., Mayer, K. H., Para, M. F., and the rgp120 HIV Vaccine Study Group. (2005). Placebo-controlled trial of a recombinant glycoprotein 120 vaccine to prevent HIV infection. *Journal of Infectious Diseases* **191**, 654–665.

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**, 875–891.

Gilbert PB, Peterson ML, Follmann D, Hudgens MG, Francis DP, Gurwith M, Heyward WL, Jobes DV, Popovic V, Self SG, Sinangil F, Burke D, Berman PW. (2005) Correlation between immunologic responses to a recombinant glycoprotein 120 vaccine and incidence of HIV-1 infection in a phase 3 HIV-1 preventive vaccine trial. *Journal of Infectious Diseases*. **191**, 666–677.

Gottschalk, P. and Dunn, J. (2005). The five-parameter logistic: a characterization and comparison with the four-parameter logistic. *Analytical Biochemistry* **343**, 54–65.

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

Huber, P. J. (1981). *Robust Statistics.* Wiley, New York.

Huffer, F.W. and McKeague, I.W. (1991). Weighted least squares estimation for Aalen's additive risk model. *Journal of the American Statistical Association* **86**, 114–129.

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

Kalbfleisch, J.D. and Lawless, J.F. (1988). Likelihood analysis of multi-state models for disease incidence and mortality. *Statistics in Medicine* **7**, 149–160.

Kang, S., Cai, J. and Chambless, L. (2013). Marginal additive hazards model for case-cohort studies with multiple disease outcomes: an application to the Atherosclerosis Risk in Communities (ARIC) study. *Biostatistics* **14**, 28–41.

Kong, L. and Cai, J. (2009). Case-cohort analysis with accelerated failure time model. *Biometrics* **65**, 135–142.

Kulich, M. and Lin, D.Y. (2000). Additive hazard regressions for case-cohort studies. *Biometrika* **87**, 73–87.

Kulich, M. and Lin, D.Y. (2004). Improving the efficiency of relative-risk estimation in case-cohort studies. *Journal of the American Statistical Association* **99**, 832–844.

Li, Z., Gilbert, P.B., Nan, B. (2008). Weighted likelihood method for grouped survival data in case-cohort studies with application to HIV vaccine trials. *B*iometrics **64**, 1247–1255.

Lin, D. Y. and Ying, Z. (1993). Cox regression with incomplete covariate measurements. *Journal of the American Statistical Association* **88**, 1341–1349.

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

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

McKeague, I.W. and Sasieni, P.D. (1994). A partly parametric additive risk model. *Biometrika* **81**, 501–514.

Murphy, S. A., Rossini, A. J., and van der Vaart, A. W. (1997). Maximum Likelihood Estimation in the Proportional Odds Model. *Journal of the American Statistical Association* **92**, 968–976.

Nan, B. and Wellner, J. A. (2013). A general semiparametric Z-estimation approach for case-cohort studies. *Statistica Sinica* **23**, 1155-1180.

Plotkin, S.A. and Gilbert, P.B. (2012). Nomenclature for immune correlates of protection after vaccination. *Clin Infect Dis* **54**, 1615–1617.

PrÆstgaard, J. and Wellner, J. A. (1993). Exchangeably weighted bootstraps of the general empirical process. *Annals of Probability* **21**, 2053–2086.

Prentice, R.L. (1986). A Case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika* **73**, 1–11.

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

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

Saegusa, T. and Wellner, J. A. (2013) Weighted likelihood estimation under two-phase sampling. Ânnals of Statistics **41**, 269–295.

Samuelsen, S. O., Ånested, H. and Skrondal, A. (2007). Stratified case-cohort analysis of general cohort sampling designs. *Scandinavian Journal of Statistics* **34**, 103–119.

Self, S.G. and Prentice, R.L. (1988). Asymptotic distribution theory and efficiency results for case-cohort studies. *The Annals of Statistics* **16**, 64–81.

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

Tsiatis, A. A. (2006). *Semiparametric Theory and Missing Data*. Springer, New York.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, New York.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes. Springer Series in Statistics*. Springer-Verlag, New York.

Appendix A: Proof of the Theorems

First we present Lemma 1 and Lemma 2, which are the key results for proving Theorem 2 and Theorem 3. The proofs of two lemmas are given just before the proofs of the theorems.

Lemma 1. *Let* $R_i(t), t \in [0,\tau]$, $i = 1,\ldots,n$, *be* q*-dimensional iid processes whose sample paths are of bounded variation. Assume that* $(E\{\|V[R_i;s,t]\|^2\})^{1/2} \le C(t-s)^{\alpha}$, *for* $s,t \in [0,\tau]$, *where* $\alpha > 0$ *and* $C > 0$ *are constants, and* $\|\cdot\|$ *is the Euclidean norm. Then* $n^{-1}\sum_{i=1}^n R_i(t) \xrightarrow{P} ER_1(t)$, *uniformly in* $t \in [0,\tau]$ *and* $\mathcal{R}(t) = n^{-1/2}\sum_{i=1}^n (R_i(t) - ER_1(t))$ *converges weakly to a mean-zero Gaussian process* $\mathcal{R}_0(t), t \in [0,\tau]$, *with continuous paths.*

The asymptotic results for the M-estimators are established in Theorem 5.7 and Theorem 5.21 in van der Vaart (1998). Under Condition A, the limits of the M-estimators $\hat{\psi}$, and $\hat{\varphi}_1$ and $\hat{\varphi}_2$ exist with $\hat{\psi} \xrightarrow{P} \psi^*$, $\hat{\varphi}_1 \xrightarrow{P} \varphi_1^*$ and $\hat{\varphi}_2 \xrightarrow{P} \varphi_2^*$ as $n \to \infty$, and the following expressions hold

$$n^{1/2}(\hat{\psi} - \psi^*) = n^{-1/2}\sum_{i=1}^n \psi_i + o_p(1) \tag{19}$$

$$n^{1/2}(\hat{\varphi}_1 - \varphi_1^*) = n^{-1/2}\sum_{i=1}^n \phi_{1,i} + o_p(1) \tag{20}$$

$$n^{1/2}(\hat{\varphi}_2 - \varphi_2^*) = n^{-1/2}\sum_{i=1}^n \phi_{2,i} + o_p(1), \tag{21}$$

where $(\psi_i, \phi_{1,i}, \phi_{2,i})$ are mean zero independent identically distributed (iid) random vectors. In addition, if $P(\xi_i = 1|\Omega_i) = \pi(\Omega_i, \psi)$ is correctly specified, then $\psi^* = \psi_0$, where ψ_0 is the true value of ψ if $\pi(\Omega_i, \psi)$ is the correct model for $P(\xi_i = 1|\Omega_i)$. If both $E(Z_i(t)|\Omega_i) = \mu_1(\Omega_i, \varphi_1)$ and $E\{Z_i(t)Z_i^T(t)|\Omega_i\} = \mu_2(\Omega_i, \varphi_2)$ are modelled correctly, then $\varphi_1^* = \varphi_{10}$ and $\varphi_2^* = \varphi_{20}$, where φ_{10} and φ_{20} are the true values of φ_1 and φ_2 when $\mu_1(\Omega_i, \varphi_1)$ and $\mu_2(\Omega_i, \varphi_2)$ are the correct models for $E(Z_i(t)|\Omega_i)$ and $E(Z_i(t)Z_i^T(t)|\Omega_i)$, respectively.

Let $q_i^* = \xi_i/\pi(\Omega_i, \psi^*)$ and let $E^*\{Z_i(t)|\Omega_i\} = \mu_1(\Omega_i, \varphi_1^*)$ and $E^*\{Z_i(t)Z_i^T(t)|\Omega_i\} = \mu_2(\Omega_i, \varphi_2^*)$. Let $E^*(X_i(t)|\Omega_i)$ and $E^*(X_i(t)X_i^T(t)|\Omega_i)$ correspond to $E(X_i(t)|\Omega_i)$ and $E(X_i(t)X_i^T(t)|\Omega_i)$ defined in (7) and (8) with $E(Z_i(t)|\Omega_i)$ replaced by $E^*\{Z_i(t)|\Omega_i\}$ and $E(Z_i(t)Z_i^T(t)|\Omega_i)$

replaced by $E^*\{Z_i(t)Z_i^T(t)|\Omega_i\}$, respectively. Define $E^*_{ux}(t)$, $E^*_{xx}(t)$ and $E^*_{xn}(t)$ similar to $E_{ux}(t)$, $E_{xx}(t)$ and $E_{xn}(t)$ given in (6) with q_i, $E(X_i(t)|\Omega_i)$ and $E(X_i(t)X_i^T(t)|\Omega_i)$ replaced by q_i^*, $E^*(X_i(t)|\Omega_i)$ and $E^*(X_i(t)X_i^T(t)|\Omega_i)$, respectively.

In addition, we define

$$\begin{aligned}
E^*_{uz}(t) &= n^{-1}\textstyle\sum_{i=1}^n W_i(t)Y_i(t)U_i(t)\{q_i^* Z_i^T(t) + (1-q_i^*)E^*(Z_i^T(t)|\Omega_i)\}\\
E^*_{zz}(t) &= n^{-1}\textstyle\sum_{i=1}^n W_i(t)Y_i(t)\{q_i^* Z_i(t)Z_i^T(t) + (1-q_i^*)E^*(Z_i(t)Z_i^T(t)|\Omega_i)\}\\
E^*_{vz}(t) &= n^{-1}\textstyle\sum_{i=1}^n W_i(t)Y_i(t)V_i(t)\{q_i^* Z_i^T(t) + (1-q_i^*)E^*(Z_i^T(t)|\Omega_i)\}\\
E^*_{zn}(t) &= n^{-1}\textstyle\sum_{i=1}^n \int_0^t W_i(s)Y_i(s)\{q_i^* Z_i(s) + (1-q_i^*)E(Z_i(s)|\Omega_i)\}\,dN_i(s)\\
E_{vu}(t) &= n^{-1}\textstyle\sum_{i=1}^n W_i(t)Y_i(t)V_i(t)U_i^T(t)\\
E_{uu}(t) &= n^{-1}\textstyle\sum_{i=1}^n W_i(t)Y_i(t)U_i(t)U_i^T(t)\\
E_{un}(t) &= n^{-1}\textstyle\sum_{i=1}^n \int_0^t W_i(s)Y_i(s)U_i(s)\,dN_i(s).
\end{aligned}$$

Let $\hat{E}_{vz}(t)$, $\hat{E}_{zz}(t)$, $\hat{E}_{uz}(t)$ and $\hat{E}_{zn}(t)$ be the counterparts of $E^*_{vz}(t)$, $E^*_{zz}(t)$, $E^*_{uz}(t)$ and $E^*_{zn}(t)$, obtained by replacing q_i^* with $\hat{q}_i = \xi_i/\pi(\Omega_i,\hat{\psi})$, and by replacing $E^*(Z_i(t)|\Omega_i)$ and $E^*(Z_i(t)Z_i^T(t)|\Omega_i)$ with $\mu_1(\Omega_i,\hat{\varphi}_1)$ and $\mu_2(\Omega_i,\hat{\varphi}_2)$, respectively. Because $U_i(t)$ and $V_i(t)$ are observable, we let $\hat{E}_{uu}(t) = E^*_{uu}(t) = E_{uu}(t)$, $\hat{E}_{vu}(t) = E^*_{vu}(t) = E_{vu}(t)$ and $\hat{E}_{un}(t) = E^*_{un}(t) = E_{un}(t)$.

Let

$$\begin{aligned}
e^*_{ux}(t) &= E[W_i(t)Y_i(t)u_i(t)\{q_i^* X_i^T(t) + (1-q_i^*)E^*(X_i^T(t)|\Omega_i)\}],\\
e^*_{xx}(t) &= E[W_i(t)Y_i(t)\{q_i^* X_i(t)X_i^T(t) + (1-q_i^*)E^*(X_i(t)X_i^T(t)|\Omega_i)\}],\\
e^*_{xn}(t) &= E[\int_0^t W_i(s)Y_i(s)\{q_i^* X_i(s) + (1-q_i^*)E^*(X_i^T(s)|\Omega_i)\}\,dN_i(s)].
\end{aligned}$$

If $P(\xi_i = 1|\Omega_i) = \pi(\Omega_i,\psi)$, and/or both $E(Z_i(t)|\Omega_i) = \mu_1(\Omega_i,\varphi_1)$ and $E(Z_i(t)Z_i^T(t)|\Omega_i)$ $= \mu_2(\Omega_i,\varphi_2)$ are modelled correctly, then under MAR, $e^*_{ux}(t) = e_{ux}(t)$, $e^*_{xx}(t) = e_{xx}(t)$, and $e^*_{xn}(t) = e_{xn}(t)$. Under Conditions (A.1), (A.2), (A.4) and (A.7), by the properties of functions with bounded variation (cf. Folland (1999)), the processes, say $R_i(t)$, within each sum of $E^*_{uz}(t)$, $E^*_{zz}(t)$, $E^*_{vz}(t)$, $E^*_{zn}(t)$, $E_{vu}(t)$, $E_{uu}(t)$, $E_{un}(t)$ have bounded variations, and $(E\{\|V[R_i; s,t]\|^2\})^{1/2} \leq C(t-s)^{\alpha}$, for $s,t \in [0,\tau]$, where $\alpha > 0$ and $C > 0$ are constants. It follows by Lemma 1 that if $P(\xi_i = 1|\Omega_i) = \pi(\Omega_i, \psi)$, and/or both $E(Z_i(t)|\Omega_i) = \mu_1(\Omega_i, \varphi_1)$ and $E(Z_i(t)Z_i^T(t)|\Omega_i) = \mu_2(\Omega_i, \varphi_2)$ are modelled correctly, then $E_{uu}(t) \xrightarrow{P} e_{uu}(t)$, $E_{un}(t) \xrightarrow{P} e_{un}(t)$, $E^*_{ux}(t) \xrightarrow{P} e_{ux}(t)$, $E^*_{xx}(t) \xrightarrow{P} e_{xx}(t)$, and $E^*_{xn}(t) \xrightarrow{P} e_{xn}(t)$, uniformly in $t \in [0,\tau]$, and $n^{1/2}\{E_{uu}(t) - e_{uu}(t)\}$, $n^{1/2}\{E^*_{ux}(t) - e_{ux}(t)\}$, $n^{1/2}\{E^*_{xx}(t) - e_{xx}(t)\}$ and $n^{1/2}\{E^*_{xn}(t) - e_{xn}(t)\}$ converge weakly to zero-mean Gaussian processes on $[0,\tau]$, respectively. By the first order expansions of $\hat{E}_{ux}(t)$, $\hat{E}_{xx}(t)$ and $\hat{E}_{xn}(t)$ around ψ^*, φ_1^* and φ_2^*, we also have $\hat{E}_{ux}(t) \xrightarrow{P} e_{ux}(t)$, $\hat{E}_{xx}(t) \xrightarrow{P} e_{xx}(t)$, and $\hat{E}_{xn}(t) \xrightarrow{P} e_{xn}(t)$ uniformly in $t \in [0,\tau]$ as $n \to \infty$.

Let $\pi'(\Omega_i, \psi)$ be the m dimensional column vector for the derivative of $\pi(\Omega_i, \psi)$ with respect to ψ. Let $\mu_1'(\Omega_i, \varphi_1)$ be the rk_1 dimensional column vector consisting the derivative of $\mu_1(\Omega_i, \varphi_1)$ with respect to φ_1, the elements from $(j-1)k_1 + 1$ to jk_1 are the derivatives of the jth component of $\mu_1(\Omega_i, \varphi_1)$ with respect to φ_1 for $j = 1, \ldots, r$. Here I_r is the $r \times r$ identity matrix and $\otimes$ is the Kronecker product of matrices. Let $\mu_2'(\Omega_i, \varphi_2)$ be the $rk_2 \times r$ matrix consisting the derivative of $\mu_2(\Omega_i, \varphi_2)$ with respect to φ_2, the elements on the lth column with rows from $(j-1)k_2 + 1$ to jk_2 are the derivatives of the (j,l)th element of $\mu_2(\Omega_i, \varphi_2)$ with respect to φ_2 for $j, l = 1, \ldots, r$.

By the matrix expressions for $E(X_i(t)|\Omega_i)$ and $E(X_i(t)X_i^T(t)|\Omega_i)$ defined in 7 and 8 and by (19)-(21), the following lemma implies that each of the terms $\sqrt{n}\{\hat{E}_{xn}(t) - E^*_{xn}(t)\}$, $\sqrt{n}\{\hat{E}_{xu}(t) - E^*_{xu}(t)\}$ and $\sqrt{n}\{\hat{E}_{xx}(t) - E^*_{xx}(t)\}$ can be approximated by the sum of iid random processes.

Lemma 2. *Suppose that Condition A holds.*

(a) $n^{1/2}\{\hat{E}_{uz}(t) - E^*_{uz}(t)\} = g_{uz,\pi}(t)\hat{\Phi}_{1,n} + g_{uz,\mu_1}(t)\hat{\Psi}_n + o_p(1)$;

(b) $n^{1/2}\{\hat{E}_{vz}(t) - E^*_{vz}(t)\} = g_{vz,\pi}(t)\hat{\Phi}_{1,n} + g_{vz,\mu_1}(t)\hat{\Psi}_n + o_p(1)$;

(c) $n^{1/2}\{\hat{E}_{zz}(t) - E^*_{zz}(t)\} = g_{zz,\pi}(t)\hat{\Phi}_{2,n} + g_{zz,\mu_2}(t)\hat{\Psi}_n + o_p(1)$;

(d) $n^{1/2}\{\hat{E}_{zn}(t) - E^*_{zn}(t)\}^T = g_{zn,\pi}(t)\hat{\Phi}_{1,n} + g_{zn,\mu_1}(t)\hat{\Psi}_n + o_p(1)$,

where I_r *is the* $r \times r$ *identity matrix and* $\otimes$ *is the Kronecker product of matrices,* $\hat{\Phi}_{1,n} = n^{1/2}I_r \otimes (\hat{\varphi}_1 - \varphi_1^*)$, $\hat{\Phi}_{2,n} = n^{1/2}I_r \otimes (\hat{\varphi}_2 - \varphi_2^*)$, $\hat{\Psi}_n = n^{1/2}I_r \otimes (\hat{\psi} - \psi^*)$, *and*

$$
\begin{aligned}
g_{uz,\pi}(t) &= E\{(1-q_i^*)W_i(t)Y_i(t)U_i(t)(\mu_1'(\Omega_i,\varphi_1^*))^T\} \\
g_{uz,\mu_1}(t) &= -E\{\xi_i W_i(t)Y_i(t)[U_i(t)\{Z_i(t)-\mu_1(\Omega_i,\varphi_1^*)\}^T] \otimes (\pi'(\Omega_i,\psi^*))^T\pi^{-2}(\Omega_i,\psi^*)\} \\
g_{vz,\pi}(t) &= E\{(1-q_i^*)W_i(t)Y_i(t)V_i(t)(\mu_1'(\Omega_i,\varphi_1^*))^T\} \\
g_{vz,\mu_1}(t) &= -E\{\xi_i W_i(t)Y_i(t)[V_i(t)\{Z_i(t)-\mu_1(\Omega_i,\varphi_1^*)\}^T] \otimes (\pi'(\Omega_i,\psi^*))^T\pi^{-2}(\Omega_i,\psi^*)\} \\
g_{zz,\pi}(t) &= E\{(1-q_i^*)W_i(t)Y_i(t)(\mu_2'(\Omega_i,\varphi_2^*))^T\} \\
g_{zz,\mu_2}(t) &= -E\{\xi_i W_i(t)Y_i(t)\{Z_i(t)Z_i^T(t)-\mu_2(\Omega_i,\varphi_2^*)\} \otimes (\pi'(\Omega_i,\psi^*))^T\pi^{-2}(\Omega_i,\psi^*)\} \\
g_{zn,\pi}(t) &= E\{\textstyle\int_0^t(1-q_i^*)W_i(s)Y_i(s)(\mu_1'(\Omega_i,\varphi_1^*))^T\,dN_i(s)\} \\
g_{zn,\mu_1}(t) &= -E\{\textstyle\int_0^t \xi_i W_i(s)Y_i(s)\{Z_i(t)-\mu_1(\Omega_i,\varphi_1^*)\}^T \otimes (\pi'(\Omega_i,\psi^*))^T\pi^{-2}(\Omega_i,\psi^*)\,dN_i(s)\}.
\end{aligned}
$$

Further, if $P(\xi_i = 1|\Omega_i) = \pi(\Omega_i,\psi)$ is correctly specified, then $g_{uz,\pi}(t) = 0$, $g_{vz,\pi}(t) = 0$, $g_{zz,\pi}(t) = 0$, $g_{zn,\pi}(t) = 0$; If $E(Z_i(t)|\Omega_i) = \mu_1(\Omega_i,\varphi_1)$ and $E(Z_i(t)Z_i^T(t)|\Omega_i) = \mu_2(\Omega_i,\varphi_2)$ are both modelled correctly, then $g_{uz,\mu_1}(t) = 0$, $g_{vz,\mu_1}(t) = 0$, $g_{zz,\mu_2}(t) = 0$, $g_{zn,\mu_1}(t) = 0$. To help readers follow-up with the notations, we note that the covariates $U(\cdot)$, $V(\cdot)$ and $Z(\cdot)$ are p, q and r dimensional vectors, respectively, and the parameters ψ, φ_1 and φ_2 are m, k_1 and k_2 dimensional vectors, respectively. The dimensions of the matrices defined in Lemma 2 are as follows: $g_{uz,\pi}(t)$ is an p by rk_1 matrix, $g_{vz,\pi}(t)$ is an q by rk_1 matrix, $g_{zz,\pi}(t)$ is an r by rk_2 matrix, $g_{zn,\pi}(t)$ is an 1 by rk_1 matrix, $I_r \otimes (\hat{\varphi}_1 - \varphi_1^*)$ is an rk_1 by r matrix, and $I_r \otimes (\hat{\varphi}_2 - \varphi_2^*)$ is an rk_2 by r matrix. Also, $g_{uz,\mu_1}(t)$ is an p by rm matrix, $g_{vz,\mu_1}(t)$ is an q by rm matrix,

$g_{zz,\mu_2}(t)$ is an r by rm matrix, $g_{zn,\mu_1}(t)$ is an 1 by rm matrix, and $I_r \otimes (\hat{\psi} - \psi^*)$ is an rm by r matrix.

Proof of Lemma 1.

We prove the lemma for $q = 1$ since the vector processes converge (in probability or weakly) if each component processes converge (in probability or weakly). Consider the processes $\{R_i(t), t \in [0,\tau]\}$, $i = 1, \ldots, n$, as a random sample from a probability distribution $\mathcal{P}$ on a measurable space $(\mathcal{X}, \mathcal{A})$. Let $\mathcal{F}$ be the class of coordinate projections $f_t : \mathcal{X} \longrightarrow R$, where $f_t(R_i) = R_i(t)$, for $t \in [0,\tau]$. The $L_r(\mathcal{P})$-norm of f_t is given by $\|f_t\|_{\mathcal{P},r} = (\mathcal{P}|f_t|^r)^{1/r} = (E|f_t(R_i)|^r)^{1/r}$. By the Jordan Decomposition (cf. Folland (1999)), $R_i(t) - R_i(0) = R_i^+(t) - R_i^-(t)$, where $R_i^+(t) = V^+[R_i; 0, t] = \sup\{\sum_{i=1}^n [R_i(t_k) - R_i(t_{k-1})]^+ : \Gamma \text{ is a partition of } [0,t]\}$ and $R_i^-(t) = V^-[R_i; 0, t] = \sup\{\sum_{i=1}^n [R_i(t_k) - R_i(t_{k-1})]^- : \Gamma \text{ is a partition of } [0,t]\}$ correspond to the positive variation and the negative variation of $R_i(\cdot)$ on $[0,t]$, respectively. Here $x^+ = \max\{x, 0\}$ and $x^- = -\min\{x, 0\}$. Further, the total variation of $R_i(\cdot)$ on $[0,t]$ is $V[R_i; 0, t] = R_i^+(t) + R_i^-(t)$. The condition $(E\{\|V[R_i; s, t]\|^2\})^{1/2} \leq C(t-s)^\alpha$ implies that $(E\{(R_i^\pm(t) - R_i^\pm(s))^2\})^{1/2} \leq C(t-s)^\alpha$ for $s < t$.

Let $0 = t_0 < t_1 < \cdots < t_K = \tau$ be a partition of $[0,\tau]$ such that $\max_{1 \leq j \leq K}(t_j - t_{j-1}) < \epsilon$. Then for any $f_t \in \mathcal{F}$, there is a bracket $[l_j, u_j]$ such that $l_j \leq f_t \leq u_j$, where $l_j = R_i^+(t_{j-1}) - R_i^-(t_j)$ and $u_j = R_i^+(t_j) - R_i^-(t_{j-1})$ with $t \in [t_{j-1}, t_j]$. The $L_2(\mathcal{P})$ bracket size of $[l_j, u_j]$ is $\|u_j - l_j\|_{P,2} \leq \|R_i^+(t_j) - R_i^+(t_{j-1})\|_{P,2} + \|R_i^-(t_j) - R_i^-(t_{j-1})\|_{P,2} \leq 2C(t_j - t_{j-1})^\alpha \leq 2C\epsilon^\alpha$. Hence, the bracketing number $N_{[\,]}(\epsilon^\alpha, \mathcal{F}, L_2(\mathcal{P})) \leq \kappa(1/\epsilon)$ for every $\epsilon > 0$ for some constant κ. Thus, $N_{[\,]}(\epsilon, \mathcal{F}, L_2(\mathcal{P})) \leq \kappa(1/\epsilon)^{1/\alpha}$ for every $\epsilon > 0$. So the bracketing integral $J_{[\,]}(1, \mathcal{F}, L_2(\mathcal{P})) = \int_0^1 \sqrt{\log N_{[\,]}(\epsilon, \mathcal{F}, L_2(\mathcal{P}))}\, d\epsilon \leq \int_0^1 \sqrt{\log\{\kappa(1/\epsilon)^{1/\alpha}\}}\, d\epsilon = \int_0^\infty \sqrt{\log\kappa + u/\alpha}\, \exp(-u)\, du < \infty$. By Glivenko-Cantelli Theorem and Donsker Theorem (Theorem 19.4 and Theorem 19.5 of van der Vaart; 1998), $n^{-1}\sum_{i=1}^n R_i(t) \xrightarrow{P} ER_1(t)$, uniformly in $t \in [0,\tau]$ and the empirical process $\{n^{-1/2}\sum_{i=1}^n R_i(t), t \in [0,\tau]\}$ converges weakly to a mean-zero Gaussian process which can be constructed to have continuous paths by Theorem 18.14 and Lemma 18.15 (van der Vaart; 1998). □

Proof of Lemma 2.

Part (a). Consider the decomposition

$$n^{1/2}\{\hat{E}_{uz}(t)-E^*_{uz}(t)\} = n^{-1/2}\sum_{i=1}^n(1-q_i^*)W_i(t)Y_i(t)U_i(t)\{\mu_1(\Omega_i,\hat{\varphi}_1)-\mu_1(\Omega_i,\varphi_1^*)\}^T + n^{-1/2}\sum_{i=1}^n(\hat{q}_i-q_i^*)W_i(t)Y_i(t)U_i(t)\{Z_i(t)-\mu_1(\Omega_i,\hat{\varphi}_1)\}^T \quad (22)$$

By the first order Taylor expansion of $\mu_1(\Omega_i,\hat{\varphi}_1)-\mu_1(\Omega_i,\varphi_1^*)$, the first term in (22) is

$$n^{-1}\sum_{i=1}^n(1-q_i^*)W_i(t)Y_i(t)U_i(t)(\mu_1'(\Omega_i,\varphi_1^*))^T\{n^{1/2}I_r\otimes(\hat{\varphi}_1-\varphi_1^*)\}+o_p(1), \quad (23)$$

By the Glivenko-Cantelli theorem (Theorem 19.4 of van der Vaart, 1998),

$$n^{-1}\sum_{i=1}^n(1-q_i^*)W_i(t)Y_i(t)U_i(t)(\mu_1'(\Omega_i,\varphi_1^*))^T\xrightarrow{P}g_{uz,\pi}(t),$$

uniformly in $t\in[0,\tau]$, where $g_{uz,\pi}(t)=E\{(1-q_i^*)W_i(t)Y_i(t)U_i(t)(\mu_1'(\Omega_i,\varphi_1^*))^T\}$. Hence the first term in (22) is $g_{uz,\pi}(t)\hat{\Phi}_{1,n}+o_p(1)$.

The second term in (22) is

$$\begin{aligned}
&n^{-1/2}\sum_{i=1}^n(\pi^{-1}(\Omega_i,\hat{\psi})-\pi^{-1}(\Omega_i,\psi^*))\xi_iW_i(t)Y_i(t)U_i(t)\{Z_i(t)-\mu_1(\Omega_i,\varphi_1^*)\}^T\\
&\quad -n^{-1/2}\sum_{i=1}^n(\pi^{-1}(\Omega_i,\hat{\psi})-\pi^{-1}(\Omega_i,\psi^*))\xi_iW_i(t)Y_i(t)U_i(t)\{\mu_1(\Omega_i,\hat{\varphi}_1)-\mu_1(\Omega_i,\varphi_1^*)\}^T\\
&= -n^{-1}\sum_{i=1}^n\xi_iW_i(t)Y_i(t)U_i(t)\{Z_i(t)-\mu_1(\Omega_i,\varphi_1^*)\}^T(\pi'(\Omega_i,\psi^*))^T\pi^{-2}(\Omega_i,\psi^*)n^{1/2}(\hat{\psi}-\psi^*)\\
&\quad +o_p(1)\\
&= -n^{-1}\sum_{i=1}^n\xi_iW_i(t)Y_i(t)[U_i(t)\{Z_i(t)-\mu_1(\Omega_i,\varphi_1^*)\}^T]\otimes(\pi'(\Omega_i,\psi^*))^T\pi^{-2}(\Omega_i,\psi^*)\\
&\quad \{n^{1/2}I_q\otimes(\hat{\psi}-\psi^*)\}+o_p(1)
\end{aligned} \quad (24)$$

uniformly in $t\in[0,\tau]$. Here the second term is $o_p(1)$ by the first order Taylor expansion, the

Glivenko-Cantelli theorem and the fact that $\hat{\psi} - \psi^* = O_p(n^{-1/2})$, $\hat{\varphi}_1 - \varphi_1^* = O_p(n^{-1/2})$. The details are omitted. Also by the Glivenko-Cantelli theorem,

$$-n^{-1}\sum_{i=1}^{n}\xi_i W_i(t)Y_i(t)[U_i(t)\{Z_i(t)-\mu_1(\Omega_i,\varphi_1^*)\}^T]\otimes(\pi'(\Omega_i,\psi^*))^T\pi^{-2}(\Omega_i,\psi^*)\xrightarrow{P} g_{uz,\mu_1}(t),$$

uniformly in $t \in [0, \tau]$.

It follows from (22), (23) and (24) that the last r columns of $n^{1/2}\{\hat{E}_{ux}(t) - E^*_{ux}(t)\}$ equals

$$n^{1/2}\{\hat{E}_{uz}(t) - E^*_{uz}(t)\} = g_{uz,\pi}(t)\hat{\Phi}_{1,n} + g_{uz,\mu_1}(t)\hat{\Psi}_n + o_p(1). \tag{25}$$

Part (b). The proof is same as for part (a) only to replace $U_i(t)$ by $V_i(t)$.

Part (c). Consider the decomposition

$$\begin{aligned} n^{1/2}\{\hat{E}_{zz}(t) - E^*_{zz}(t)\} &= n^{-1/2}\sum_{i=1}^{n}(1-q_i^*)W_i(t)Y_i(t)\{\mu_2(\Omega_i,\hat{\varphi}_2)-\mu_2(\Omega_i,\varphi_2^*)\} \\ &\quad + n^{-1/2}\sum_{i=1}^{n}(\hat{q}_i - q_i^*)W_i(t)Y_i(t)\{Z_i(t)Z_i^T(t)-\mu_2(\Omega_i,\hat{\varphi}_2)\} \end{aligned} \tag{26}$$

By the first order Taylor expansion of $\mu_2(\Omega_i, \hat{\varphi}_2) - \mu_2(\Omega_i, \varphi_2^*)$, the first term in (26) is

$$n^{-1}\sum_{i=1}^{n}(1-q_i^*)W_i(t)Y_i(t)(\mu_2'(\Omega_i,\varphi_2^*))^T\{n^{1/2}I_r\otimes(\hat{\varphi}_2-\varphi_2^*)\} + o_p(1). \tag{27}$$

By the Glivenko-Cantelli theorem (Theorem 19.4 of van der Vaart, 1998),

$$n^{-1}\sum_{i=1}^{n}(1-q_i^*)W_i(t)Y_i(t)(\mu_2'(\Omega_i,\varphi_2^*))^T\xrightarrow{P} g_{zz,\pi}(t),$$

uniformly in $t \in [0, \tau]$, where $g_{zz,\pi}(t) = E\{(1-q_i^*)W_i(t)Y_i(t)(\mu_2'(\Omega_i, \varphi_2^*))^T\}$. Hence the first term in (26) is $g_{zz,\pi}(t)\hat{\Phi}_{2,n} + o_p(1)$.

The second term in (26) is

$$
\begin{aligned}
& n^{-1/2}\sum_{i=1}^{n}(\pi^{-1}(\Omega_i,\hat{\psi})-\pi^{-1}(\Omega_i,\psi^*))\xi_i W_i(t)Y_i(t)\{Z_i(t)Z_i^T(t)-\mu_2(\Omega_i,\varphi_2^*)\} \\
& \quad -n^{-1/2}\sum_{i=1}^{n}(\pi^{-1}(\Omega_i,\hat{\psi})-\pi^{-1}(\Omega_i,\psi^*))\xi_i W_i(t)Y_i(t)\{\mu_2(\Omega_i,\hat{\varphi}_2)-\mu_2(\Omega_i,\varphi_2^*)\} \\
& = -n^{-1}\sum_{i=1}^{n}\xi_i W_i(t)Y_i(t)\{Z_i(t)Z_i^T(t)-\mu_2(\Omega_i,\varphi_2^*)\}(\pi'(\Omega_i,\psi^*))^T\pi^{-2}(\Omega_i,\psi^*)n^{1/2}(\hat{\psi}-\psi^*) \\
& \quad +o_p(1) \\
& = -n^{-1}\sum_{i=1}^{n}\xi_i W_i(t)Y_i(t)\{Z_i(t)Z_i^T(t)-\mu_2(\Omega_i,\varphi_2^*)\}\otimes(\pi'(\Omega_i,\psi^*))^T\pi^{-2}(\Omega_i,\psi^*) \\
& \quad \{n^{1/2}I_r\otimes(\hat{\psi}-\psi^*)\}+o_p(1) \qquad (28)
\end{aligned}
$$

uniformly in $t\in[0,\tau]$. Here the second term is $o_p(1)$ by the first order Taylor expansion, the Glivenko-Cantelli theorem and the fact that $\hat{\psi}-\psi^*=O_p(n^{-1/2})$, $\hat{\varphi}_1-\varphi_1^*=O_p(n^{-1/2})$. The details are omitted. Also by the Glivenko-Cantelli theorem,

$$
-n^{-1}\sum_{i=1}^{n}\xi_i W_i(t)Y_i(t)\{Z_i(t)Z_i^T(t)-\mu_2(\Omega_i,\varphi_2^*)\}\otimes(\pi'(\Omega_i,\psi^*))^T\pi^{-2}(\Omega_i,\psi^*)\xrightarrow{P}g_{zz,\mu_2}(t),
$$

uniformly in $t\in[0,\tau]$.

It follows from (26), (27) and (28) that the last $r\times r$ diagonal block of $n^{1/2}\{\hat{E}_{xx}(t)-E_{xx}^*(t)\}$ equals

$$
n^{1/2}\{\hat{E}_{zz}(t)-E_{zz}^*(t)\}=g_{zz,\pi}(t)\hat{\Phi}_{2,n}+g_{zz,\mu_2}(t)\hat{\Psi}_n+o_p(1). \qquad (29)
$$

Part (d). Similar to (22),

$$
\begin{aligned}
& n^{1/2}\{\hat{E}_{zn}(t)-E_{zn}^*(t)\}^T \\
& = n^{-1/2}\sum_{i=1}^{n}\int_0^t(1-q_i^*)W_i(s)Y_i(s)\{\hat{E}(Z_i(s)\Omega_i)-E^*(Z_i(s)|\Omega_i)\}^T\,dN_i(s) \\
& \quad +n^{-1/2}\sum_{i=1}^{n}\int_0^t(\hat{q}_i-q_i^*)W_i(s)Y_i(s)\{Z_i(s)-\hat{E}(Z_i(s)|\Omega_i)\}^T\,dN_i(s). \qquad (30)
\end{aligned}
$$

The rest of the proof is similar to the proof for part (a) by letting $U_i(\cdot) = 1$ and replacing the terms in the summations with their integrations with respect to $dN_i(s)$.

Further, if $\pi(\Omega_i, \psi)$ is the correct model for $P(\xi_i = 1|\Omega_i)$, then $\psi^* = \psi_0$, where ψ_0 is the true value of ψ. Thus $\pi(\Omega_i, \psi^*) = P(\xi_i = 1|\Omega_i)$ and $E(q_i^*|\Omega_i) = E(\xi_i/\pi(\Omega_i, \psi^*)|\Omega_i) = 1$. Hence $g_{uz,\pi}(t) = 0$ by the double expectation property. Similarly, $g_{vz,\pi}(t) = 0$, $g_{zz,\pi}(t) = 0$, $g_{zn,\pi}(t) = 0$. If $\mu_1(\Omega_i, \varphi_1)$ is the correct model for $E(Z_i(t)|\Omega_i)$ and $\mu_2(\Omega_i, \varphi_2)$ is the correct model for $E\{Z_i(t)Z_i^T(t)|\Omega_i\}$, then $\varphi_1^* = \varphi_{10}$ and $\varphi_2^* = \varphi_{20}$, where φ_{10} and φ_{20} are the true values of φ_1 and φ_2, respectively. In this case, $\mu_1(\Omega_i, \varphi_1^*) = E(Z_i(t)|\Omega_i)$ and $\mu_2(\Omega_i, \varphi_2^*) = E(Z_i(t)Z_i^T(t)|\Omega_i)$. Thus $E\{Z_i(t) - \mu_1(\Omega_i, \varphi_1^*)|\Omega_i\} = 0$ and $E\{Z_i(t)Z_i^T(t) - \mu_2(\Omega_i, \varphi_2^*)|\Omega_i\} = 0$. Under MAR and by the double expectation property, we have $g_{uz,\mu_1}(t) = 0$, $g_{vz,\mu_1}(t) = 0$, $g_{zz,\mu_2}(t) = 0$, and $g_{zn,\mu_1}(t) = 0$.

Proof of Theorem 1.

The following notations are introduced for the simplicity of the expressions in the proof for Theorem 1. Let $X(t) = [Y_1(t)X_1(t), \ldots, Y_n(t)X_n(t)]^T$, $U(t) = [Y_1(t)U_1(t), \ldots, Y_n(t) U_n(t)]^T$, $N(t) = [N_1(t), N_2(t), \ldots, N_n(t)]^T$ and $\lambda(t) = [\lambda_1(t), \lambda_2(t), \ldots, \lambda_n(t)]^T$. Let $W(t) = diag\{W_i(t)\}$, $H_q = diag\{q_i\}$ and $H_{1-q} = diag\{(1-q_i)\}$ be $n \times n$ diagonal weight matrices. Let $\tilde{\Omega} = (\Omega_1, \ldots, \Omega_n)$, $E[X(t)|\tilde{\Omega}] = V_x(t)$ and $E[X^T(t)W(t)H_{1-q}X(t)|\,\tilde{\Omega}] = V_{xx}(t)$. Let $X^*(t) = H_qX(t) + H_{1-q}V_x(t)$, $U^{\sim}(t) = \{U^T(t)W(t)U(t)\}^{-1}U^T(t)W(t)$ and $H(t) = W(t)\{I - U(t)U^{\sim}(t)\}$.

We have estimation equation (4)

$$\sum_{i=1}^{n} \Big[q_iU_i(t)W_i(t)\{dN_i(t) - \lambda_i(t)\,dt\} + (1-q_i)E\{U_i(t)W_i(t)(dN_i(t) \\ -\lambda_i(t)\,dt)|\Omega_i\} \Big] \quad = 0$$

where $\lambda_i(t)\,dt = U_i(t)\alpha_i(t)\,dt + X_i(t)\theta\,dt$

With some algebra, this equation can be simplified to

$$U^T(t)W(t)H_q\,[dN(t) - U(t)dA(t) - X(t)\theta\,dt] \\ = -U^T(t)W(t)H_{1-q}\,[dN(t) - U(t)dA(t) - V_x(t))\theta\,dt]$$

$$U^T(t)W(t)[H_q + H_{1-q}]dN(t) - U^T(t)W(t)[H_q + H_{1-q}]U(t)\,dA(t) \\ = U^T(t)W(t)[H_qX(t) + H_{1-q}V_x(t)]\,\theta dt$$

Where

$$V_x(t) = E[X(t)|\tilde{\Omega}]$$

Since

$$H_q + H_{1-q} = I$$

$$U^T(t)W(t)dN(t) - U^T(t)W(t)U(t)\,dA(t) = U^T(t)W(t)X^*(t)\,\theta dt$$
$$U^T(t)W(t)U(t)dA(t) = U^T(t)W(t)dN(t) - U^T(t)W(t)X^*(t)\,\theta dt$$
$$dA(t) = U^{\sim}(t)[dN(t) - X^*(t)\theta\,dt] \tag{31}$$

We have estimation equation (5)

$$\sum_{i=0}^{n}\int_0^\tau \Big[q_iY_i(t)X_I(t)W_i(t)\{dN_i(t) - \lambda_i(t)\,dt\} + (1-q_i)E\{X_i(t)W_i(t)(dN_i(t) - \lambda_i(t)\,dt)|\Omega_i\}\Big] = 0$$

where $\lambda_i(t)\,dt = U_i(t)\alpha_i(t)\,dt + X_i(t)\theta\,dt$

It can be simplified to

$$\int_0^\tau \Big\{X^T(t)W(t)H_q\,[dN(t) - U(t)dA(t) - X(t)\theta\,dt]\Big\}$$
$$= -\sum_{i=1}^{n}\int_0^\tau \Big\{(1-q_i)\,E[X_i(t)W_i(t)\{dN_i(t) - U_i(t)dA(t) - X_i(t)\theta\,dt\}|\Omega_i]\Big\}$$

$$\int_0^\tau \Big\{X^T(t)W(t)H_q\,[dN(t) - U(t)dA(t) - X(t)\theta\,dt]\Big\}$$
$$= -\int_0^\tau V_x^T(t)W(t)H_{1-q}dN(t)$$
$$+\int_0^\tau V_x^T(t)W(t)H_{1-q}U(t)dA(t) + \int_0^\tau V_{xx}(t)\theta\,dt$$

$$\int_0^\tau \{X^T(t)H_q + V_x^T(t)H_{1-q}\}W(t)dN(t) - \int_0^\tau \{X^T(t)H_q + V_x^T(t)H_{1-q}\}W(t)U(t)dA(t)$$
$$= \int_0^\tau \{X^T(t)W(t)H_qX(t) + V_{xx}(t)\}\theta\,dt$$

$$\int_0^\tau (X^*(t))^T W(t) dN(t) - \int_0^\tau (X^*(t))^T W(t) U(t) dA(t)$$
$$= \int_0^\tau \{X^T(t) W(t) H_q X(t) + V_{xx}(t)\} \theta \, dt \tag{32}$$

substituting the value of $dA(t)$ from (31) into (32)

$$\int_0^\tau (X^*(t))^T W(t) dN(t) - \int_0^\tau (X^*(t))^T W(t) U(t) U^{\sim}(t) dN(t)$$
$$+ \int_0^\tau (X^*(t))^T W(t) U(t) U^{\sim}(t) X^*(t) \theta \, dt = \int_0^\tau \{X^T(t) W(t) H_q X(t) + V_{xx}(t)\} \theta \, dt$$

$$\int_0^\tau (X^*(t))^T W(t) [I - U(t) U^{\sim}(t)] dN(t) = \int_0^\tau \{X^T(t) W(t) H_q X(t) + V_{xx}(t)$$
$$- (X^*(t))^T W(t) U(t) U^{\sim}(t) X^*(t)\} \theta \, dt$$

$$\int_0^\tau (X^*(t))^T H(t) dN(t) = \int_0^\tau \{X^T(t) W(t) H_q X(t) + V_{xx}(t)$$
$$- (X^*(t))^T W(t) U(t) U^{\sim}(t) X^*(t)\} \theta \, dt$$

$$\tilde{\theta} = \left\{ \int_0^\tau [V_{xx}(t) + X^T(t) W(t) H_q X(t) \right.$$
$$\left. - (X^*(t))^T W(t) U(t) U^{\sim}(t) X^*(t)] \, dt \right\}^{-1} \int_0^\tau (X^*(t))^T H(t) \, dN(t) \tag{33}$$

$$\tilde{A}(t) = \int_0^t U^{\sim}(s) \{dN(s) - X^*(s) \tilde{\theta} \, ds\} \tag{34}$$

Further, recall the definitions given in (6), simple matrix operations show that

$$
\begin{aligned}
&n^{-1}U^T(t)W(t)U(t) = E_{uu}(t),\\
&n^{-1}U^T(t)W(t)X^*(t) = n^{-1}U^T(t)W(t)\{H_qX(t) + H_{1-q}V_x(t)\} = E_{ux}(t),\\
&n^{-1}\{X^T(t)W(t)H_qX(t) + V_{xx}(t)\} = E_{xx}(t),\\
&n^{-1}\int_0^t U^T(s)W(s)\,dN(s) = E_{un}(t) \qquad (35)\\
&n^{-1}\int_0^t (W(s)X^*(s))^T dN(s) = n^{-1}\int_0^t W(s)\{H_qX(s) + H_{1-q}V_x(s)\}^T dN(s) = E_{xn}(t).
\end{aligned}
$$

We note that

$$U^{\sim}(t) = \{U^T(t)W(t)U(t)\}^{-1}U^T(t)W(t) = E_{uu}^{-1}(t)n^{-1}U^T(t)W(t) \qquad (36)$$

$$H(t) = W(t)[I - U(t)U^{\sim}(t)] = W(t) - W(t)U(t)E_{uu}^{-1}(t)n^{-1}U^T(t)W(t) \qquad (37)$$

It follows that

$$
\begin{aligned}
&(X^*(t))^T H(t)dN(t),\\
&= (X^*(t))^T W(t)dN(t) - (X^*(t))^T W(t)U(t)E_{uu}^{-1}(t)n^{-1}U^T(t)W(t),\\
&= n\,dE_{xn}(t) - n\,E_{ux}^T(t)E_{uu}^{-1}(t)dE_{un}(t) \qquad (38)
\end{aligned}
$$

Similarly

$$
\begin{aligned}
&V_{xx}(t) + X^T(t)W(t)H_qX(t) - (X^*(t))^T W(t)U(t)U^{\sim}(t)X^*(t)\\
&= n\,E_{xx}(t) - n\,(E_{ux}(t))^T E_{uu}^{-1}(t)E_{ux}(t)
\end{aligned}
\qquad (39)
$$

Substituting the values from (39) and (38) to (33)

$$\tilde{\theta} = \left\{ \int_0^\tau \{E_{xx}(t) - E_{ux}^T(t)E_{uu}^{-1}(t)E_{ux}(t)\}dt \right\}^{-1} \left\{ \int_0^\tau [dE_{xn}(t) - E_{ux}^T(t)E_{uu}^{-1}(t)dE_{un}(t)] \right\}, \tag{40}$$

Substituting the values from (36) and (35) to (34)

$$\tilde{A}(t) = \int_0^t E_{uu}^{-1}(s)dE_{un}(s) - \int_0^t E_{uu}^{-1}(s)E_{ux}(s)\tilde{\theta}\, ds, \tag{41}$$

This completes the proof of Theorem (1) □

Proof of Theorem 2.

Part (a).

Let $\hat{A} = \int_0^\tau \left\{ \hat{E}_{xx}(t) - \hat{E}_{xu}(t)(E_{uu}(t))^{-1}\hat{E}_{ux}^T(t) \right\} dt$
and $\hat{R} = \hat{E}_{xn}(\tau) - \int_0^\tau \hat{E}_{xu}(t)(E_{uu}(t))^{-1}\, dE_{un}(t)$.

By the arguments preceding Lemma 2 in the Appendix, we have $\hat{A} \xrightarrow{P} A$ and $\hat{R} \xrightarrow{P} R$ if $P(\xi_i = 1|\Omega_i) = \pi(\Omega_i, \psi)$, and/or both $E(Z_i(t)|\Omega_i) = \mu_1(\Omega_i, \varphi_1)$ and $E\{Z_i(t)\ Z_i^T(t)|\Omega_i\} = \mu_2(\Omega_i, \varphi_2)$ are modelled correctly, where

$$R = \int_0^\tau \{de_{xn}(t) - e_{xu}(t)e_{uu}^{-1}(t)\, de_{un}(t)\} = \int_0^\tau \{e_{xx}(t) - e_{xu}(t)e_{uu}^{-1}(t)e_{xu}^T(t)\}\, dt\, \theta_0 = A\theta_0,$$

since $e_{xn}(t) = \int_0^t \{e_{xu}(s)\alpha_0(s) + e_{xx}(s)\theta_0\}\, ds$ and $e_{un}(t) = \int_0^t \{e_{uu}(s)\alpha_0(s) + e_{ux}(s)\theta_0\}\, ds$.
It follows that $\hat{\theta} = \hat{A}^{-1}\hat{R} \xrightarrow{P} A^{-1}R = A^{-1}A\theta_0 = \theta_0$.

Part (b).

Let $A^* = \int_0^\tau \left\{ E_{xx}^*(t) - E_{xu}^*(t)(E_{uu}(t))^{-1}(E_{xu}^*(t))^T \right\} dt$ and $R^* = E_{xn}(\tau) - \int_0^\tau E_{xu}^*(t)(E_{uu}(t))^{-1}$ $dE_{un}(t)$. Let $\theta^* = A^{*-1}R^*$ and write $\sqrt{n}(\hat{\theta} - \theta_0) = \sqrt{n}(\theta^* - \theta_0) + \sqrt{n}(\hat{\theta} - \theta^*)$.

We first derive the iid decomposition for $\sqrt{n}(\theta^* - \theta_0)$. With similar arguments in part (a), we have $A^* \xrightarrow{P} A$ and $R^* \xrightarrow{P} R = A\theta_0$.

$$\begin{aligned} \sqrt{n}(\theta^* - \theta_0) &= \sqrt{n}\left[A^{*-1}R^* - A^{-1}A\,\gamma_0\right] \\ &= \sqrt{n}(A^{*-1} - A^{-1})R^* + \sqrt{n}A^{-1}\left[R^* - A\,\theta_0\right] \\ &= -\sqrt{n}A^{-1}(A^* - A)\theta^* + \sqrt{n}A^{-1}\left[R^* - \int_0^\tau \{de_{xn}(t) - e_{xu}(t)e_{uu}^{-1}(t)\, de_{un}(t)\}\right]. \end{aligned} \tag{42}$$

Consider the decomposition

$$\begin{aligned} A^* - A = \int_0^\tau & [E_{xx}^*(t) - E_{xu}^*(t)(E_{uu}(t))^{-1}(E_{xu}^*(t))^T - \{e_{xx}(t) - e_{xu}(t)e_{uu}^{-1}(t)e_{xu}^T(t)\}]\, dt \\ = \int_0^\tau & [\{E_{xx}^*(t) - e_{xx}(t)\} - \{E_{xu}^*(t) - e_{xu}(t)\}(E_{uu}(t))^{-1}(E_{xu}^*(t))^T \\ & + e_{xu}(t)e_{uu}^{-1}(t)\{E_{uu}(t) - e_{uu}(t)\}(E_{uu}(t))^{-1}(E_{xu}^*(t))^T \\ & - e_{xu}(t)e_{uu}^{-1}(t)\{(E_{xu}^*(t))^T - e_{xu}^T(t)\}]\, dt. \end{aligned}$$

By the consistency of θ^*, the uniform convergence of $E_{uu}(t)\xrightarrow{P}e_{uu}(t)$ and $E^*_{xu}(t)\xrightarrow{P}e_{xu}(t)$ and by the weak convergence of $n^{1/2}\{E^*_{xx}(t)-e_{xx}(t)\}$, $n^{1/2}\{E^*_{xu}(t)-e_{xu}(t)\}$, and $n^{1/2}\{E_{uu}(t)-e_{uu}(t)\}$, applying Lemma A.1 of Lin and Ying (2001), the first term of (42) equals

$$-A^{-1}\sqrt{n}\int_0^\tau [\{E^*_{xx}(t)-e_{xx}(t)\}-\{E^*_{xu}(t)-e_{xu}(t)\}e^{-1}_{uu}(t)e^T_{xu}(t)+e_{xu}(t)e^{-1}_{uu}(t)$$
$$\{E_{uu}(t)-e_{uu}(t)\}e^{-1}_{uu}(t)e^T_{xu}(t)-e_{xu}(t)e^{-1}_{uu}(t)\{(E^*_{xu}(t))^T-e^T_{xu}(t)\}]\,dt\,\theta_0+o_p(1). \tag{43}$$

Similarly, the second term of (42) equals

$$\begin{aligned}
&= A^{-1}\sqrt{n}\int_0^\tau [dE^*_{xn}(t)-E^*_{xu}(t)(E_{uu}(t))^{-1}dE_{un}(t)-\{de_{xn}(t)-e_{xu}(t)e^{-1}_{uu}(t)de_{un}(t)\}]\\
&= A^{-1}\sqrt{n}\int_0^\tau \big[d\{E^*_{xn}(t)-e_{xn}(t)\}-\{E^*_{xu}(t)-e_{xu}(t)\}(E_{uu}(t))^{-1}\,dE_{un}(t)\\
&\quad +e_{xu}(t)e^{-1}_{uu}(t)\{E_{uu}(t)-e_{uu}(t)\}(E_{uu}(t))^{-1}\,dE_{un}(t)-e_{xu}(t)e^{-1}_{uu}(t)\,d\{E_{un}(t)-e_{un}(t)\}\big]\\
&= A^{-1}\sqrt{n}\int_0^\tau \big[d\{E^*_{xn}(t)-e_{xn}(t)\}-\{E^*_{xu}(t)-e_{xu}(t)\}e^{-1}_{uu}(t)\,de_{un}(t)\\
&\quad +e_{xu}(t)e^{-1}_{uu}(t)\{E_{uu}(t)-e_{uu}(t)\}e^{-1}_{uu}(t)\,de_{un}(t)-e_{xu}(t)e^{-1}_{uu}(t)\,d\{E_{un}(t)-e_{un}(t)\}\big]\\
&\quad +o_p(1).
\end{aligned} \tag{44}$$

Combining (43) and (44) yields

$$\begin{aligned}
&\sqrt{n}(\theta^*-\theta_0)\\
&= -A^{-1}\sqrt{n}\int_0^\tau \Big[\{E^*_{xx}(t)-e_{xx}(t)\}-\{E^*_{xu}(t)-e_{xu}(t)\}e^{-1}_{uu}(t)e^T_{xu}(t)+e_{xu}(t)e^{-1}_{uu}(t)\\
&\quad \{E_{uu}(t)-e_{uu}(t)\}e^{-1}_{uu}(t)e^T_{xu}(t)-e_{xu}(t)e^{-1}_{uu}(t)\{(E^*_{xu}(t))^T-e^T_{xu}(t)\}\Big]\,dt\,\theta_0\\
&\quad +A^{-1}\sqrt{n}\int_0^\tau \Big[d\{E^*_{xn}(t)-e_{xn}(t)\}-\{E^*_{xu}(t)-e_{xu}(t)\}e^{-1}_{uu}(t)\,de_{un}(t)\\
&\quad +e_{xu}(t)e^{-1}_{uu}(t)\{E_{uu}(t)-e_{uu}(t)\}e^{-1}_{uu}(t)\,de_{un}(t)-e_{xu}(t)e^{-1}_{uu}(t)\,d\{E_{un}(t)-e_{un}(t)\}\Big]\\
&\quad +o_p(1).
\end{aligned} \tag{45}$$

By $de_{xn}(t) = \{e_{xu}(t)\alpha_0(t) + e_{xx}(t)\theta_0\}\, dt$ and $de_{un}(t) = \{e_{uu}(t)\alpha_0(t) + e_{ux}(t)\theta_0\}\, dt$, ignoring the $o_p(1)$ term, $\sqrt{n}(\theta^* - \theta_0)$ equals

$$
\begin{aligned}
& A^{-1}\sqrt{n}\int_0^\tau \Big[d\{E^*_{xn}(t) - e_{xn}(t)\} - \{E^*_{xu}(t) - e_{xu}(t)\}\alpha_0(t)\, dt - \{E^*_{xx}(t) - e_{xx}(t)\}\theta_0\, dt\Big] \\
& -A^{-1}\sqrt{n}\int_0^\tau e_{xu}(t)e_{uu}^{-1}(t)\Big[d\{E_{un}(t) - e_{un}(t)\} - \{E_{uu}(t) - e_{uu}(t)\}\alpha_0(t)\, dt \\
& \qquad\qquad -\{E^*_{ux}(t) - e_{ux}(t)\}\theta_0\, dt\Big] \\
& = A^{-1}\sqrt{n}\int_0^\tau \{dE^*_{xn}(t) - E^*_{xu}(t)\alpha_0(t)\, dt - E^*_{xx}(t)\theta_0\, dt\} \\
& -A^{-1}\sqrt{n}\int_0^\tau e_{xu}(t)e_{uu}^{-1}(t)\{dE_{un}(t) - E_{uu}(t)\alpha_0(t)\, dt - E^*_{ux}(t)\theta_0\, dt\}. \qquad (46)
\end{aligned}
$$

It is easy to see from here that, ignoring the $o_p(1)$ term, $\sqrt{n}(\theta^* - \theta_0)$ is the sum of independent identically distributed random variables with mean zero. The first integral in (46) equals $n^{-1}\sum_{i=1}^n \eta_{x,i}$, where

$$
\begin{aligned}
\eta_{x,i} \;=\; & \int_0^\tau W_i(t)Y_i(t)\big[\{q_i^* X_i(t) + (1 - q_i^*)E^*(X_i(t)|\Omega_i)\}\, dN_i(t) \\
& -[q_i^* X_i(t)X_i^T(t) + (1 - q_i^*)E^*\{X_i(t)X_i^T(t)|\Omega_i\}]\,\theta_0\, dt \qquad (47) \\
& -\{q_i^* X_i(t) + (1 - q_i^*)E^*(X_i(t)|\Omega_i)\}U_i^T(t)\alpha_0(t)\, dt\big].
\end{aligned}
$$

The ssecond integral in (46) equals $n^{-1}\sum_{i=1}^n \eta_{u,i}$, where

$$
\begin{aligned}
\eta_{z,i} \;=\; & \int_0^\tau e_{xu}(t)e_{uu}^{-1}(t)W_i(t)Y_i(t)\big[U_i(t)\, dN_i(t) \qquad (48) \\
& -U_i(t)\{q_i^* X_i^T(t) + (1 - q_i^*)E^*(X_i^T(t)|\Omega_i)\}\,\theta_0\, dt - U_i(t)U_i^T(t)dA_0(t)\big].
\end{aligned}
$$

It follows by (46) that

$$
\sqrt{n}(\theta^* - \theta_0) = A^{-1}n^{-1/2}\sum_{i=1}^n(\eta_{x,i} - \eta_{u,i}) + o_p(1). \qquad (49)
$$

Next we consider the iid decomposition for $\sqrt{n}(\hat{\theta} - \theta^*)$. Note that

$$\sqrt{n}(\hat{\theta} - \theta^*) \;\; = \;\; -\sqrt{n}A^{*-1}(\hat{A} - A^*)\hat{\theta} + \sqrt{n}A^{*-1}\big(\hat{R} - R^*\big). \tag{50}$$

Consider the decomposition

$$\begin{aligned}\hat{A} - A^* \qquad &= - \int_0^\tau \Big[\{\hat{E}_{xu}(t) - E^*_{xu}(t)\}(E_{uu}(t))^{-1}(\hat{E}_{xu}(t))^T \\ &\quad + E^*_{xu}(t)(E_{uu}(t))^{-1}\{\hat{E}_{ux}(t) - E^*_{ux}(t)\} \\ &\quad - \{\hat{E}_{xx}(t) - E^*_{xx}(t)\} \Big] \, dt.\end{aligned} \tag{51}$$

By Lemma 2 and the weak convergence of $n^{1/2}\{E_{zz}(t) - e_{zz}(t)\}$, $n^{1/2}\{E^*_{zx}(t) - e_{zx}(t)\}$, $n^{1/2}\{E^*_{xx}(t) - e_{xx}(t)\}$ and $n^{1/2}\{E^*_{xn}(t) - e_{xn}(t)\}$, we have the weak convergence of $n^{1/2}\{\hat{E}_{zx}(t) - E^*_{zx}(t)\}$, $n^{1/2}\{\hat{E}_{xx}(t) - E^*_{xx}(t)\}$, and $n^{1/2}\{\hat{E}_{xn}(t) - E^*_{xn}(t)\}$. By $\hat{\gamma} \xrightarrow{P} \gamma_0$, the uniform convergence of $\hat{E}_{xx}(t) \xrightarrow{P} e_{xx}(t)$ and $\hat{E}_{zx}(t) \xrightarrow{P} e_{zx}(t)$, $E^*_{xx}(t) \xrightarrow{P} e_{xx}(t)$ and $E^*_{zx}(t) \xrightarrow{P} e_{zx}(t)$, and applying Lemma A.1 of Lin and Ying (2001), the first term of (50) equals

$$\begin{aligned} &-A^{-1}\sqrt{n} \int_0^\tau \Big[\{\hat{E}_{xx}(t) - E^*_{xx}(t)\}\{\hat{E}_{xu}(t) - E^*_{xu}(t)\}e^{-1}_{uu}(t)e^T_{ux}(t) \\ &-e_{xu}(t)e^{-1}_{uu}(t)\{(\hat{E}_{ux}(t)) - (E^*_{ux}(t))\} \Big] \, dt \, \theta_0 + o_p(1), \end{aligned} \tag{52}$$

the second term of (50) equals

$$-A^{-1}\sqrt{n} \int_0^\tau \Big[\{\hat{E}_{xu}(t) - E^*_{xu}(t)\}e^{-1}_{uu}(t) \, de_{un}(t) - d\{\hat{E}_{xn}(t) - E^*_{xn}(t)\} \Big] + o_p(1). \tag{53}$$

Combining (52) and (53) yields

$$\begin{aligned}
&\sqrt{n}(\hat{\theta} - \theta^*) \\
&= A^{-1}\sqrt{n}\int_0^\tau \Big[d\{\hat{E}_{xn}(t) - E^*_{xn}(t)\} - \{\hat{E}_{xu}(t) - E^*_{xu}(t)\}\alpha_0(t)\,dt \\
&-\{\hat{E}_{xx}(t) - E^*_{xx}(t)\}\,\theta_0\,dt\Big] + A^{-1}\sqrt{n}\int_0^\tau e_{xu}(t)e_{uu}^{-1}\{\hat{E}_{ux}(t) - E^*_{ux}(t)\}\,\theta_0\,dt + o_p(1).
\end{aligned} \tag{54}$$

Next, we derive the explicit expressions for the sum of iid approximation of $\sqrt{n}(\hat{\theta} - \theta^*)$. Let $\theta_0(t) = (\beta_0^T, \gamma_0^T)^T$, where $\beta_0(t)$ and $\gamma_0(t)$ are the regression coefficients for $V(t)$ and $Z(t)$, respectively. Because $U_i(t)$ and $V_i(t)$ are observable, $\hat{E}_{uu}(t) = E^*_{uu}(t) = E_{uu}(t)$, $\hat{E}_{vu}(t) = E^*_{vu}(t) = E_{vu}(t)$ and $\hat{E}_{un}(t) = E^*_{un}(t) = E_{un}(t)$.

The first q elements of $n^{1/2}\{\hat{E}_{xn}(t) - E^*_{xn}(t)\}$ are zero. By Lemma 2 (d) and (19)–(21), the last r elements equal to

$$\begin{aligned}
&n^{1/2}\{\hat{E}_{zn}(t) - E^*_{zn}(t)\} \\
&= n^{-1/2}\sum_{i=1}^n \left(\{g_{zn,\pi}(t)(I_r \otimes \phi_{1,i})\}^T + \{g_{zn,\mu_1}(t)(I_r \otimes \psi_i)\}^T\right) + o_p(1).
\end{aligned} \tag{55}$$

The first q elements of $n^{1/2}\{\hat{E}_{xu}(t) - E^*_{xu}(t)\}\alpha_0(t)$ are zero. By Lemma 2 (a) and (19)–(21), the last r elements equal to

$$\begin{aligned}
&n^{1/2}\{\hat{E}_{zu}(t) - E^*_{zu}(t)\}\alpha_0(t) \\
&= \left\{g_{uz,\pi}(t)\hat{\Phi}_{1,n} + g_{uz,\mu_1}(t)\hat{\Psi}_n\right\}^T\alpha_0(t) + o_p(1) \\
&= n^{-1/2}\sum_{i=1}^n \left\{g_{uz,\pi}(t)(I_r \otimes \phi_{1,i}) + g_{uz,\mu_1}(t)(I_r \otimes \psi_i)\right\}^T\alpha_0(t) + o_p(1).
\end{aligned} \tag{56}$$

The first q columns of $n^{1/2}\{\hat{E}_{ux}(t) - E^*_{ux}(t)\}$ are zero. By Lemma 2 (a) and (19)–(21),

$$
\begin{aligned}
&n^{1/2}\{\hat{E}_{ux}(t) - E^*_{ux}(t)\}\theta_0 = n^{1/2}\{\hat{E}_{uz}(t) - E^*_{uz}(t)\}\gamma_0 \\
&= \{g_{uz,\pi}(t)\hat{\Phi}_{1,n} + g_{uz,\mu_1}(t)\hat{\Psi}_n\}\gamma_0 + o_p(1) \\
&= n^{-1/2}\sum_{i=1}^{n}\big(g_{uz,\pi}(t)\{I_r \otimes \phi_{1,i}\} + g_{uz,\mu_1}(t)\{I_r \otimes \psi_i\}\big)\gamma_0 + o_p(1).
\end{aligned} \tag{57}
$$

By Lemma 2 (b) and (19)–(21), the first q components of $n^{1/2}\{\hat{E}_{xx}(t) - E^*_{xx}(t)\}\theta_0$ is

$$
\begin{aligned}
&n^{1/2}\{\hat{E}_{vz}(t) - E^*_{vz}(t)\}\gamma_0 \\
&= \big(g_{vz,\pi}(t)\hat{\Phi}_{1,n} + g_{vz,\mu_1}(t)\hat{\Psi}_n\big)\gamma_0 + o_p(1) \\
&= n^{-1/2}\sum_{i=1}^{n}\big(g_{vz,\pi}(t)\{I_r \otimes \phi_{1,i}\} + g_{vz,\mu_1}(t)\{I_r \otimes \psi_i\}\big)\gamma_0 + o_p(1).
\end{aligned} \tag{58}
$$

By Lemma 2 (b) and (c), and (19)–(21), the last r components of $n^{1/2}\{\hat{E}_{xx}(t) - E^*_{xx}(t)\}\theta_0$ is

$$
n^{1/2}\{\hat{E}_{vz}(t) - E^*_{vz}(t)\}^T\beta_0 + n^{1/2}\{\hat{E}_{zz}(t) - E^*_{zz}(t)\}\gamma_0, \tag{59}
$$

where

$$
\begin{aligned}
&n^{1/2}\{\hat{E}_{vz}(t) - E^*_{vz}(t)\}^T\beta_0 \\
&= \big\{g_{vz,\pi}(t)\hat{\Phi}_{1,n} + g_{vz,\mu_1}(t)\hat{\Psi}_n\big\}^T\beta_0 + o_p(1) \\
&= n^{-1/2}\sum_{i=1}^{n}\big\{g_{vz,\pi}(t)(I_r \otimes \phi_{1,i}) + g_{vz,\mu_1}(t)(I_r \otimes \psi_i)\big\}^T\beta_0 + o_p(1),
\end{aligned} \tag{60}
$$

and

$$
\begin{aligned}
&n^{1/2}\{\hat{E}_{zz}(t) - E^*_{zz}(t)\}\gamma_0 \\
&= \big\{g_{zz,\pi}(t)\hat{\Phi}_{2,n} + g_{zz,\mu_2}(t)\hat{\Psi}_n\big\}^T\gamma_0 + o_p(1) \\
&= n^{-1/2}\sum_{i=1}^{n}\big\{g_{zz,\pi}(t)(I_r \otimes \phi_{2,i}) + g_{zz,\mu_2}(t)(I_r \otimes \psi_i)\big\}\gamma_0 + o_p(1).
\end{aligned} \tag{61}
$$

Plugging the expressions (55)–(61) into (54), we have

$$\sqrt{n}(\hat{\theta}-\theta^*) = A^{-1}n^{-1/2}\sum_{i=1}^{n}(\varepsilon_{\Phi,i}+\varepsilon_{\Psi,i}) + o_p(1), \tag{62}$$

where

$$\begin{aligned}
\varepsilon_{\Phi,i} \;=\; & -\int_0^\tau \Bigg[\begin{pmatrix} 0_{q\times 1} \\ d\{g_{zn,\pi}(t)(I_r\otimes\phi_{1,i})\}^T - \{g_{uz,\pi}(t)(I_r\otimes\phi_{1,i})\}^T\alpha_0(t)\,dt \end{pmatrix} \\
& - \begin{pmatrix} g_{vz,\pi}(t)(I_r\otimes\phi_{1,i})\gamma_0 \\ \{g_{vz,\pi}(t)(I_r\otimes\phi_{1,i})\}^T\beta_0(t) + g_{zz,\pi}(t)(I_r\otimes\phi_{1,i})\gamma_0 \end{pmatrix} dt\Bigg] \\
& + \int_0^\tau e_{xu}(t)e_{uu}^{-1}(t)g_{uz,\pi}(t)(I_r\otimes\phi_{1,i})\gamma_0\,dt,
\end{aligned} \tag{63}$$

$$\begin{aligned}
\varepsilon_{\Psi,i} \;=\; & -\int_0^\tau \Bigg[\begin{pmatrix} 0_{q\times 1} \\ d\{g_{zn,\mu_1}(t)(I_r\otimes\psi_i)\}^T - \{g_{uz,\mu_1}(t)(I_r\otimes\psi_i)\}^T\alpha_0(t)\,dt \end{pmatrix} \\
& - \begin{pmatrix} g_{vz,\mu_1}(t)(I_r\otimes\psi_i)\gamma_0 \\ \{g_{vz,\mu_1}(t)(I_r\otimes\psi_i)\}^T\beta_0 + g_{zz,\mu_2}(t)(I_r\otimes\psi_i)\gamma_0 \end{pmatrix} dt\Bigg] \\
& + \int_0^\tau e_{xu}(t)e_{uu}^{-1}(t)g_{uz,\mu_1}(t)(I_r\otimes\psi_i)\gamma_0\,dt.
\end{aligned} \tag{64}$$

Because $\phi_{1,i}$, $\phi_{2,i}$ and ψ_i are iid with mean zero, it follows from (63) and (64) that $\varepsilon_{\Phi,i}$ and $\varepsilon_{\Psi,i}$ are iid random vectors with mean zero. Combining (49) and (62), we have

$$\sqrt{n}(\hat{\theta}-\theta_0) = A^{-1}n^{-1/2}\sum_{i=1}^{n}(\eta_{x,i}-\eta_{u,i}+\varepsilon_{\Phi,i}+\varepsilon_{\Psi,i}) + o_p(1). \tag{65}$$

It follows that $\sqrt{n}(\hat{\theta}-\theta_0)\xrightarrow{\mathcal{D}}N(0, A^{-1}\Sigma A^{-1})$, where

$$\Sigma = E\{(\eta_{x,i}-\eta_{u,i}+\varepsilon_{\Phi,i}+\varepsilon_{\Psi,i})(\eta_{x,i}-\eta_{u,i}+\varepsilon_{\Phi,i}+\varepsilon_{\Psi,i})^T\}. \tag{66}$$

Part (c).

By Lemma 2, if $P(\xi_i = 1|\Omega_i) = \pi(\Omega_i, \psi)$ is correctly specified, then $g_{uz,\pi}(t) = 0$, $g_{vz,\pi}(t) = 0$, $g_{zz,\pi}(t) = 0$, $g_{zn,\pi}(t) = 0$. It follows from (63) that $\varepsilon_{\Psi,i} = 0$. Similarly, by Lemma 2, if $E(Z_i(t)|\Omega_i) = \mu_1(\Omega_i, \varphi_1)$ and $E(Z_i(t)Z_i^T(t)|\Omega_i) = \mu_2(\Omega_i, \varphi_2)$ are both modelled correctly, then $g_{uz,\mu_1}(t) = 0$, $g_{vz,\mu_1}(t) = 0$, $g_{zz,\mu_2}(t) = 0$, $g_{zn,\mu_1}(t) = 0$. We have $\varepsilon_{\Phi,i} = 0$ by (64). $\square$

Proof of Theorem 3

Part (a).

By (12), $\hat{A}(t) = \int_0^t E_{uu}^{-1}(s)\, dE_{un}(s) - \int_0^t E_{uu}^{-1}(s)\hat{E}_{ux}(s)\, ds\, \hat{\theta}$. By the arguments in the proof of Theorem 2, if $P(\xi_i = 1|\Omega_i) = \pi(\Omega_i, \psi)$, and/or both $E(Z_i(t)|\Omega_i) = \mu_1(\Omega_i, \varphi_1)$ and $E(Z_i(t)Z_i^T(t)|\Omega_i) = \mu_2(\Omega_i, \varphi_2)$ are modelled correctly, then $\hat{E}_{xx}(t) \xrightarrow{P} e_{xx}(t)$, $\hat{E}_{xn}(t) \xrightarrow{P} e_{xn}(t)$ and $\hat{E}_{xz}(t) \xrightarrow{P} e_{xz}(t)$ uniformly in $t \in [0, \tau]$ as $n \to \infty$. By the consistency of $\hat{\theta}$ proved in Theorem 2, we have $\hat{A}(t) \xrightarrow{P} \int_0^t e_{uu}^{-1}(s)\, de_{un}(s) - \int_0^t e_{uu}^{-1}(s)e_{ux}(s)\, ds\, \theta_0 = A_0(t)$.

Part (b).

Let $A^*(t) = \int_0^t (E_{uu}(s))^{-1}\, dE_{un}(s) - \int_0^t (E_{uu}(s))^{-1} E_{ux}^*(s)\, ds\, \theta^*$. Write $\sqrt{n}(\hat{A}(t) - A_0(t)) = \sqrt{n}(A^*(t) - A_0(t)) + \sqrt{n}(\hat{A}(t) - A^*(t))$. The weak convergence of $\sqrt{n}(\hat{A}(t) - A_0(t))$ is proved through the iid decomposition for each of the two terms.

We consider the following decomposition:

$$
\begin{aligned}
&\sqrt{n}(A^*(t) - A_0(t)) \\
&= \sqrt{n}\left\{ \int_0^t (E_{uu}(s))^{-1}\, dE_{un}(s) - \int_0^t e_{uu}^{-1}(s)\, de_{un}(s) \right\} \qquad (67) \\
&\quad -\sqrt{n} \int_0^t \{(E_{uu}(s))^{-1} E_{ux}^*(s) - e_{uu}^{-1}(s)e_{ux}(s)\}\, ds\, \theta^* - \int_0^t e_{uu}^{-1}(s)e_{ux}(s)\, ds \sqrt{n}(\theta^* - \theta_0).
\end{aligned}
$$

By Lemma A.1 of Lin and Ying (2001), the first term of (67) equals

$$
\begin{aligned}
&\sqrt{n} \int_0^t [\{(E_{uu}(s))^{-1} - e_{uu}^{-1}(s)\}\, dE_{un}(s) + e_{uu}^{-1}(s)\, \{dE_{un}(s) - de_{un}(s)\}] \qquad (68) \\
&= -\sqrt{n} \int_0^t \{e_{uu}^{-1}(s)E_{uu}(s)e_{uu}^{-1}(s)\, de_{un}(s) - e_{uu}^{-1}(s)\, dE_{un}(s)\} + o_p(1).
\end{aligned}
$$

Similarly, the second term of (67) equals

$$
\begin{aligned}
&-\sqrt{n} \int_0^t \{(E_{uu}(s))^{-1} E_{ux}^*(s) - e_{uu}^{-1}(s)e_{ux}(s)\}\, ds\, \theta^* \qquad (69) \\
&= \sqrt{n} \int_0^t [e_{uu}^{-1}(s)E_{uu}(s)e_{uu}^{-1}(s)e_{ux}(s) - e_{uu}^{-1}(s)E_{ux}^*(s)]\, ds\, \theta_0 + o_p(1).
\end{aligned}
$$

The sum of the first two terms of (67) is

$$\sqrt{n}\Big\{-\int_0^t e_{uu}^{-1}(s)E_{uu}(s)\alpha_0(s)\,ds+\int_0^t e_{uu}^{-1}(s)\,dE_{un}(s)-\int_0^t e_{uu}^{-1}(s)E_{ux}^*(s)\,ds\,\theta_0\Big\}+o_p(1).$$
$$=\sqrt{n}\int_0^t e_{uu}^{-1}(s)\{dE_{un}(s)-E_{uu}(s)\alpha_0(s)\,ds-E_{ux}^*(s)\,ds\,\theta_0\}+o_p(1)$$
$$=n^{-1/2}\sum_{i=1}^n\zeta_i(t)+o_p(1), \tag{70}$$

where

$$\begin{aligned}\zeta_i(t) \;=\; & \int_0^t e_{uu}^{-1}(s)W_i(s)Y_i(s)\big[\{U_i(s)\,dN_i(s)\\ & -U_i(s)U_i^T(s)\,dA_0(s)\\ & -\{q_i^*X_i(t)+(1-q_i^*)E^*(X_i(t)|\Omega_i)\}U_i^T(t)\theta_0\,dt\big].\end{aligned} \tag{71}$$

Combining (67), (70) and (49), we obtain

$$\sqrt{n}(A^*(t)-A_0(t))=n^{-1/2}\sum_{i=1}^n\Big\{\zeta_i(t)-\int_0^t e_{uu}^{-1}(s)e_{ux}(s)\,ds\,A^{-1}(\eta_{x,i}-\eta_{u,i})\Big\}+o_p(1), \tag{72}$$

uniformly in $t\in[0,\tau]$.

Next, similar to (67), we have

$$\begin{aligned}\sqrt{n}(\hat{A}(t)-A^*(t)) \qquad & -\sqrt{n}\int_0^t\{(E_{uu}(s))^{-1}\hat{E}_{ux}(s)-(E_{uu}(s))^{-1}E_{ux}^*(s)\}\,ds\,\hat{\theta}\\ & -\int_0^t(E_{uu}(s))^{-1}E_{ux}^*(s)\,ds\sqrt{n}(\hat{\theta}-\theta^*).\end{aligned} \tag{73}$$

Using similar arguments of the asymptotic theory, the first term of (73) equals

$$\begin{aligned}
&\sqrt{n}(\hat{A}(t) - A^*(t)) \\
&= -\sqrt{n}\int_0^t e_{uu}^{-1}(s)\{\hat{E}_{ux}(s) - E_{ux}^*(s)\}\,\theta_0\,ds \\
&- \int_0^t e_{uu}^{-1}(s)e_{ux}(s)\,ds\,\sqrt{n}(\hat{\theta} - \theta^*) + o_p(1). \qquad (74)
\end{aligned}$$

By the expressions (55)–(62) and (74),

$$\sqrt{n}(\hat{A}(t) - A^*(t)) = n^{-1/2}\sum_{i=1}^{n}(\upsilon_{\Phi,i}(t) + \upsilon_{\Psi,i}(t)) + o_p(1), \qquad (75)$$

where

$$\begin{aligned}
\upsilon_{\Phi,i}(t) &= \int_0^t e_{uu}^{-1}(s)g_{uz,\pi}(s)(I_r \otimes \phi_{1,i})\gamma_0\,ds \\
&\quad - \int_0^t e_{uu}^{-1}(s)e_{ux}(s)\,ds\,A^{-1}\varepsilon_{\Phi,i}, \qquad (76)
\end{aligned}$$

$$\begin{aligned}
\upsilon_{\Psi,i}(t) &= \int_0^t e_{uu}^{-1}(s)g_{uz,\mu_1}(s)(I_r \otimes \psi_i)\gamma_0\,ds \\
&\quad - \int_0^t e_{uu}^{-1}(s)e_{ux}(s)\,ds\,A^{-1}\varepsilon_{\Psi,i}. \qquad (77)
\end{aligned}$$

From pact (b) of the proof of Theorem 2, $\varepsilon_{\Phi,i}$ and $\varepsilon_{\Psi,i}$ are iid with mean zero. Because $\phi_{1,i}$, $\phi_{2,i}$ and ψ_i are iid with mean zero, it follows from (76) and (77) that $\upsilon_{\Phi,i}(t)$ and $\upsilon_{\Psi,i}(t)$ are iid random vectors with mean zero.

It follows from (72) and (75) that

$$\begin{aligned}
\sqrt{n}(\hat{A}(t) - A_0(t)) &= n^{-1/2}\sum_{i=1}^{n}\left\{\zeta_i(t) - \int_0^t e_{uu}^{-1}(s)e_{ux}(s)\,ds\,A^{-1}(\eta_{x,i} - \eta_{u,i})\right\} \\
&\quad + n^{-1/2}\sum_{i=1}^{n}\left\{\upsilon_{\Phi,i}(t) + \upsilon_{\Psi,i}(t)\right\} + o_p(1), \qquad (78)
\end{aligned}$$

uniformly in $t \in [0, \tau]$. By Lemma 1, $\sqrt{n}(\hat{A}(t) - A_0(t))$ converges weakly to a zero-mean Gaussian process on $[0, \tau]$.

Part (c).

By Lemma 2, (19)–(21), (74)–(75), it is easy to see that when $P(\xi_i = 1|\Omega_i) = \pi(\Omega_i, \psi)$ is correctly specified, $\varepsilon_{\Phi,i} = 0$ and $\upsilon_{\Phi,i}(t) = 0$, and when $E(Z_i(t)|\Omega_i) = \mu_1(\Omega_i, \varphi_1)$ and $E(Z_i(t)Z_i^T(t)|\Omega_i) = \mu_2(\Omega_i, \varphi_2)$ are modelled correctly $\varepsilon_{\Psi,i} = 0$ and $\upsilon_{\Psi,i}(t) = 0$. $\square$

Appendix B: Regularity Conditions

Let $f(t)$ be a function $[a,b] \to R$. Given any finite partition $\Gamma = \{a = t_0 < \cdots < t_K = b\}$ of $[a,b]$, the variation of f over $[a,b]$ is $V[f;a,b] = \sup\{\sum_{k=1}^{K} |f(t_k) - f(t_{k-1})| : \Gamma \text{ is a partition of } [a,b]\}$. The function f has bounded variation on $[a,b]$ if $V[f;a,b] < \infty$. A vector f of functions has bounded variation if each component of f has bounded variation, and in this case, $V[f;a,b]$ is the vector of the variations of the component functions.

We assume the following regularity conditions throughout the paper:

Condition A.

A1. The processes $U_i(t)$, $X_i(t)$ and $W_i(t)$, $0 \le t \le \tau$, have bounded second moments, their sample paths are left continuous and of bounded variation. The variations of the processes $U_i(\cdot)$, $V_i(\cdot)$ and $W_i(\cdot)$ satisfy the conditions $(E\{\|V[U_i;s,t]\|^2\})^{1/2} \le C(t-s)^\alpha$, $(E\{\|V[V_i;s,t]\|^2\})^{1/2} \le C(t-s)^\alpha$, and $(E\{\|V[W_i;s,t]\|^2\})^{1/2} \le C(t-s)^\alpha$, for $s,t \in [0,\tau]$, where $\alpha > 0$ and $C > 0$ are constants, and $\|\cdot\|$ is the Euclidean norm.

A2. The $\alpha(t)$, $e_{uu}(t)$, $e_{xx}(t)$ and $e_{xu}(t)$ are twice differentiable on $[0,\tau]$, and $e_{uu}(t)$ is a nonsingular matrix, and $e_{uu}^{-1}(t)$ is bounded over $0 \le t \le \tau$.

A3. The matrix A is positive definite.

A4. $W_i(t)$ is a weight process depending only on phase-one variables, $W_i(t) \xrightarrow{P} w_i(t)$ uniformly in $t \in [0,\tau]$ and $1 \le i \le n$, and $w_i(t)$ is differentiable with uniformly bounded derivative.

A5. The censoring is independent in the sense that the censoring does not alter the risk of failure. This assumption is described by $E\{d\tilde{N}_i(t) | X_i[0,t], U_i[0,t], \tilde{T}_i \ge t\} = E\{dN_i^*(t) | X_i[0,t], U_i[0,t], T_i \ge t\}$, where $\tilde{N}_i(t) = I(\tilde{T}_i \le t)$, $N_i^*(t) = I(T_i \le t)$, and $X_i[0,t] = \{X_i(s), 0 \le s \le t\}$ and $U_i[0,t] = \{U_i(s), 0 \le s \le t\}$ are the covariate histories up to time t.

A6. The phase-two covariate Z_i is missing at random (MAR), i.e., $P(\xi_i = 1|Z_i, \Omega_i) = P(\xi_i = 1|\Omega_i)$.

A7. The function $\pi(\Omega_i, \psi)$ is twice differentiable with respect to ψ the compact set Θ_ψ, $\pi'_\psi(\Omega_i, \psi) = \partial\pi(\Omega_i, \psi)/\partial\psi$ is uniformly bounded, and there is a $\varepsilon > 0$ such that $\pi(\Omega_i, \psi) \geq \varepsilon > 0$ for all $i = 1, \ldots, n$.

A8. The functions $\mu_1(\Omega_i, \varphi_1)$ and $\mu_2(\Omega_i, \varphi_2)$ are twice differentiable with respect to φ_1 and φ_2 on the compact sets Θ_{φ_1} and Θ_{φ_2}, respectively.