UNIFIED APPROACH TO PARTIALLY LINEAR MODEL AND COX PROPORTIONAL HAZARDS MODEL WITH MISSING COVARIATES

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

STATISTICS

UNIVERSITY OF REGINA

By

Wei Tang

Regina, Saskatchewan

May, 2014

UNIVERSITY OF REGINA

FACULTY OF GRADUATE STUDIES AND RESEARCH

SUPERVISORY AND EXAMINING COMMITTEE

Wei Tang, candidate for the degree of Doctor of Philosophy in Statistics**,** has presented a thesis titled, ***Unified Approach to Partially Linear Model and Cox Proportional Hazards Model with Missing Covariates,*** in an oral examination held on March 26, 2014. The following committee members have found the thesis acceptable in form and content, and that the candidate demonstrated satisfactory knowledge of the subject material.

External Examiner:	*Dr. Wenyu Jiang, Queen's University
Supervisor:	Dr. Yang Zhao, Department of Mathematics & Statistics
Committee Member:	Dr. Chun-Hua Guo, Department of Mathematics & Statistics
Committee Member:	Dr. Taehan Bae, Department of Mathematics & Statistics
Committee Member:	Dr. Boting Yang, Department of Computer Science
Chair of Defense:	Dr. Dongyan Blachford, Faculty of Graduate Studies & Research

*via tele-conference

Abstract

In regression analysis the problem of missing covariate data is common in various fields of applications. Many methods have been developed to deal with this problem in the past three decades. These methods are workable under most missing data scenarios. However, when missing covariate data appear in a general missing data pattern, many methods become too complicated in computation.

In this thesis, we extend the unified approach of Chen and Chen (2000) and Zhao et al. (2013) to deal with partially linear model (Engle et al., 1986) and Cox proportional hazards model (Cox, 1972) with missing covariates. The unified approach possesses some superior characteristics in dealing with regression models with missing data. First, the unified approach requires less extra assumptions to be applied than many other methods, which may need additional modeling for variables with missing values. Second, this extension of the unified approach can deal with both the simple monotone missing data pattern and the general missing data pattern under missing completely at random and missing at random settings. Third, no iteration is needed in computing the proposed estimate. In general, compared to other methods, the unified approach is conceptually and computationally simple.

This thesis describes the proposed estimators separately for the partially linear model

and the Cox proportional hazards model with missing covariates. The asymptotic properties of the estimators are investigated or justified. Simulations are conducted under different settings to examine the performance of the proposed estimators for these two models.

Acknowledgments

I would like to express my sincere gratitude to my supervisor, Dr. Yang Zhao, who offered me an opportunity to pursue a PhD degree at the University of Regina. Her inspiring guidance and support have been invaluable throughout my time in academic study. Without her help, it would be impossible for me to continue my research work and finish this thesis.

I wish to also express heartfelt thanks to Dr. Chun-Hua Guo, Dr. Taehan Bae and Dr. Boting Yang for serving on my thesis examining committee.

I would like to thank all the faculty members, staff and graduate students in the Department of Mathematics and Statistics for their support and friendships while I have pursued this degree.

Finally, I would like to thank my wife, my son, my parents and my siblings back in China, for loving me and supporting me unconditionally.

Contents

Chapter 1

Introduction

Missing data can occur for various reasons in a variety of fields. In regression analysis, the problem of missing covariate data arises frequently in such fields as clinical studies and medical research. It can be problematic if not dealt with properly. The initial naive method is to perform a complete-case analysis which is proven to be biased sometimes if missingness is not completely at random and inefficient because incomplete observations are discarded. Statisticians have been working on improving the efficiency of the estimator by incorporating the incomplete observations into analysis. In the past three decades, many methods have been proposed to improve the efficiency of the parameter estimators in regression analysis with missing data. A comprehensive overview of methods for regression models was provided by Little and Rubin (2002). More recently, Ibrahim et al. (2005) reviewed four common approaches for making inference in generalized linear models with missing covariate data. Methods proposed for the simple monotone missing data pattern may not be easily extended to deal with the general missing data patterns. Methods developed for the general missing data pattern are often computationally or constructionally complex (Robins et al., 1994). The unified approach (Chen and Chen, 2000) can be extended to deal with the general missing data pattern without a signigicant increase in complexity.

Chen and Chen (2000) proposed a unified approach for the generalized estimating equation under double-sampling design. Zhao et al. (2013) extended this unified approach to augment the data with nonmonotone missing data pattern. In this thesis, we further extend the method (Zhao et al., 2013) to both the partially linear model and the Cox proportional hazards model with missing covariates and call it a unified inverse probability weighted (UIPW) estimator. Chen and Chen (2000) showed that the unified estimator corresponds to one of the estimator classes proposed by Robins et al. (1994) under the missing completely at random mechanism. Compared to the estimator classes of Robins et al. (1994), the unified estimator is simple in construction and less intensive in computation. Even without correct specification of the surrogate model, it will not lose efficiency. Its efficiency gain depends on how much information it can extract from auxiliary data through the surrogate model. The UIPW estimator inherits these strengths from the original unified approach and adds the ability to deal with missing covariates in the general missing data patterns for the partially linear model and the Cox proportional hazards model.

This thesis includes two regression models and discusses them separately in different chapters. Chapter 1 gives a succinct introduction to the basic idea of the unified approach shared by these two models and a general overview of the unified approach for the generalized estimating equations. In Chapter 2, we define the UIPW estimator for the partially linear model with missing covariates. Some corollaries are proven to assert the convergence of the estimator. In Chapter 3, we describe the UIPW estimator for the Cox proportional hazards model with missing covariates. Justifications are provided for the asymptotic behaviors of the estimator. In Chapter 4, we discuss the further research on the unified approach.

1.1 Missing data pattern

The missing data pattern describes which values are observed and which values are missing in the data matrix. Univariate nonresponse, multivariate two patterns, monotone pattern and general pattern are examples of these patterns (Little and Rubin, 2002). In this thesis, we investigate the partially linear model and the Cox proportional hazards model with data missing in simple monotone pattern and general pattern.

Let $X' = (W', U')'$ denote a vector of random variables, where $X \in R^p, W \in R^k$ and $U \in R^{p-k}$. The simple monotone missing data pattern is that the elements of $W = (x_1, ..., x_k)'$ being fully observed and the elements of $U = (x_{k+1}, ..., x_p)'$ being observed or missing together. Data matrix 1 holds the data of the n observations of random vector X and its missing pattern is simple monotonic. The symbol hyphen - represents the missing value in the data matrix. The first k rows and the last $p - k$ rows in data matrix 1 are n replications of W and U respectively. This is the simplest monotonic missing data pattern.

Many statistical methods are developed to deal with the data in this missingness pattern. However, in practice, simple monotone missing data pattern is rare. A general missing data pattern appears more complicated than the simple monotone missing data pattern.

Data matrix 1.

$$\begin{pmatrix} x_{11} & \dots & x_{1m} & x_{1(m+1)} & \dots & x_{1n} \\ \dots & \dots & \dots & \dots & \dots. & \dots \\ x_{k1} & \dots & x_{km} & x_{k(m+1)} & \dots. & x_{kn} \\ x_{(k+1)1} & \dots & x_{(k+1)m} & - & - & - \\ \dots & \dots & \dots & - & - & - \\ x_{p1} & \dots & x_{pm} & - & - & - \end{pmatrix}$$

A general missing data pattern shown in data matrix 2 occurs when missing values are haphazardly dispersed throughout the data matrix. This is a typical example of the general missing data pattern. The methods of analysis proposed for the simple monotone missing data pattern may fail to deal with the general missing pattern or face challenging complexity in computation. Special methods have to be developed for the general missing data pattern.

Data matrix 2.

$$\begin{pmatrix} x_{11} & \dots & x_{1m} & - & \dots & x_{1n} \\ - & \dots & \dots & - & \dots. & \dots \\ x_{k1} & - & x_{km} & x_{k(m+1)} & \dots. & - \\ x_{(k+1)1} & \dots & x_{(k+1)m} & - & - & x_{(k+1)n} \\ \dots & \dots & - & \dots & \dots. & \dots \\ x_{p1} & \dots & x_{pm} & x_{p(m+1)} & \dots & x_{pn} \end{pmatrix}$$

The missing data mechanism describes whether the missingness of variables is related to the values in the data set (Little and Rubin, 2002). The missing data mechanism is called missing complete at random (MCAR), if the missingness does not depend on the missing or observed values of observations. The missing data mechanism is called missing at random (MAR), if the missingness depends only on the observed values but not the missing values. The missing data mechanism is called not missing at random (NMAR), if the missingness depends on the unobserved values given the observed data. In this study, we will focus on extending the UIPW method to deal with the partially linear model and the Cox proportional hazards model under the MCAR and the MAR settings.

1.2 Basic idea of unified approach

In regression model, a relationship between a response Y and a vector of explanatory variables X is expressed in term of the conditional mean of Y given X. In a double-sampling desgin, a sample of units is taken strictly to obtain auxiliary information, and then a subsample where the variables of interest are observed. Auxiliary information on $(\tilde{Y}, \tilde{X})$ is collected on a primary sample of subjects. The measurements on (Y, X) are collected only on a validation subsample of the primary sample due to the practical limitations. It is common to employ double-sampling design to improve the estimate of interest for (Y, X) by using the auxiliary information $(\tilde{Y}, \tilde{X})$.

Chen and Chen (2000) proposed a unified approach to deal with the generalized estimating equations under the MCAR setting. Estimating equations based on the validation subsample considered in Chen and Chen (2000) are

$$0 = \sum_{i \in V} S_i(\beta) \equiv \sum_{i \in V} P_j[Y_i - g(X_i; \beta)]$$

and

$$0 = \sum_{i \in V} \tilde{S}_i(\gamma) \equiv \sum_{i \in V} \tilde{P}_j[\tilde{Y}_i - h(\tilde{X}_i; \gamma)],$$

where V is the index set of the validation subsample, $g(X; \beta) = E[Y|X; \beta]$, $h(\tilde{X}; \gamma) = E[\tilde{Y}|\tilde{X}; \gamma]$, P and $\tilde{P}$ are functions of X and $\tilde{X}$ respectively, and β and γ are vectors of parameters. Let $\hat{\beta}$ and $\hat{\gamma}$ denote the solutions to these two set estimating equations, respectively. A more efficient estimate $\bar{\gamma}$ for γ can be obtained from the primary sample with N observations by solving the following estimating equation

$$0 = \sum_{i=1}^{N} \bar{S}_i(\gamma) \equiv \sum_{i=1}^{N} \bar{P}_j[\tilde{Y}_i - h(\bar{X}_i; \gamma)].$$

Given the general regularity conditions in Appendix 5.1, $\hat{\beta}$ is consistent for β^* and asymptotically normal. $\hat{\gamma}$ and $\bar{\gamma}$ are consistent for γ^* and asymptotically normal. The joint distribution of $\sqrt{n}(\hat{\beta}' - \beta^{*\prime}, \hat{\gamma}' - \gamma^{*\prime})'$ is asymptotically normal as well. By the multivariate normal distribution theory, the conditional distribution of $\sqrt{n}(\hat{\beta}-\beta^*)$ given $\sqrt{n}(\hat{\gamma}-\gamma^*)$ is asymptotically normal with mean $\sqrt{n}D_{00}^{-1}C_{01}C_{11}^{-1}D_{11}(\hat{\gamma}-\gamma^*)$, where D_{00}, C_{01}, C_{11} and D_{11} are given below. The unified estimator proposed by Chen and Chen (2000) for the generalized estimating equation with double-sampling has the form

$$\bar{\beta} = \hat{\beta} - \hat{D}_{00}^{-1}\hat{C}_{01}\hat{C}_{11}^{-1}\hat{D}_{11}(\hat{\gamma}-\bar{\gamma}),$$

where $\hat{D}_{00}, \hat{D}_{11}, \hat{C}_{01}, \hat{C}_{11}$ are estimates of D_{00}, D_{11}, C_{01} and C_{11}, with

$$\begin{aligned}
D_{00} &\equiv E[\partial S(\beta^*)/\partial\beta], \\
D_{11} &\equiv E[\partial\tilde{S}(\gamma^*)/\partial\gamma], \\
C_{00} &\equiv E[S(\beta^*)S^{'}(\beta^*)], \\
C_{01} &\equiv E[S(\beta^*)\tilde{S}^{'}(\gamma^*)], \\
C_{10} &\equiv E[\tilde{S}(\gamma^*)S^{'}(\beta^*)], \text{ and} \\
C_{11} &\equiv E[\tilde{S}(\gamma^*)\tilde{S}^{'}(\gamma^*)].
\end{aligned}$$

This approach is originally designed for the generalized estimating equations under double-sampling. It can be extended to deal with missing covariates in regression models. When $\tilde{Y}=Y$ and $\tilde{X}$ consists some elements of X, the model for $(\tilde{Y},\tilde{X})$ can still be treated as a surrogate model. This is a simple example of missing covariates data in regression model.

The basic idea of unified approach is to employ the auxiliary information $(\tilde{Y},\tilde{X})$ to obtain the estimates $\hat{\gamma}$ and $\bar{\gamma}$ from the validation sample and the primary sample respectively, and then improve the estimate of β by utilizing $\hat{\gamma}$ and $\bar{\gamma}$ through the correlation between $\hat{\beta}$

and $\hat{\gamma}$.

The unified approach works well with fully observed auxiliary variables $(\tilde{Y}, \tilde{X})$, but fails when auxiliary variables have missing values. Zhao et al. (2013) extended this approach to deal with the general missing pattern. Let $\tilde{X}_i^{(l)}$ be a subvector of $\tilde{X}_i, l = 1, ..., q$. The elements of $\tilde{X}_i^{(l)}$ has the same missing pattern, which means they are observed or missing together. A sequence of estimating equations is considered to deal with data with general missing pattern

$$0 = \sum_{i \in V} \tilde{S}_i^{(l)}(\gamma_l) \equiv \sum_{i \in V} \tilde{P}_i^{(l)}[\tilde{Y}_i^{(l)} - h(\tilde{X}_i^{(l)}; \gamma_l)], l = 1, ..., q, \text{ and}$$
$$0 = \sum_{i=1}^{N_l} \bar{S}_i^{(l)}(\gamma_l) \equiv \sum_{i=1}^{N_l} \bar{P}_i^{(l)}[\bar{Y}_i^{(l)} - h(\bar{X}_i^{(l)}; \gamma_l)], l = 1, ..., q.$$

Estimates $\hat{\gamma}_l$ and $\bar{\gamma}_l$ are solutions to the lth estimating equations for the validation sample and the lth primary sample respectively. The improved estimate takes the form

$$\bar{\beta} = \hat{\beta} - \hat{D}_{00}^{-1} \hat{C}_{01} \hat{C}_{11}^{-1} \hat{D}_{11} (\hat{\gamma} - \bar{\gamma}),$$

where $\hat{\gamma} = (\hat{\gamma}_1', ..., \hat{\gamma}_q')'$ and $\bar{\gamma} = (\bar{\gamma}_1', ..., \bar{\gamma}_q')'$, $\hat{D}_{00}$, $\hat{C}_{01}$, $\hat{C}_{11}$ and $\hat{D}_{11}$ are estimates of D_{00}, C_{01}, C_{11} and D_{11}, respectively. Denote $\tilde{S}_{Qi}(\gamma) = (\tilde{S}_{ji}'(\gamma_1), ..., \tilde{S}_{ji}'(\gamma_q))', i = 1, ...n$. We note that $\tilde{S}_{Qi}(\gamma), i = 1, ..., n$ has the same distribution as $\tilde{S}_Q(\gamma)$. Matrices C_{01}, C_{11} and D_{11} are defined as

$$C_{01} \equiv E[S(\beta^*) \tilde{S}_Q'(\gamma^*)],$$
$$C_{11} \equiv E[\tilde{S}_Q(\gamma^*) \tilde{S}_Q'(\gamma^*)], \text{ and}$$
$$D_{11} \equiv E[\partial \tilde{S}_Q(\gamma^*)/\partial \gamma].$$

1.3 Objectives

In this thesis, we investigate whether the UIPW estimator works for the partially linear model and the Cox proportional hazards model with missing covariates and how much efficiency it gains compared to the complete-case analysis or the weighted complete-case analysis. These two regression models are different in such areas as applications, assumptions and asymptotic properties, however they also share some common points. First, there is no intercept in these two regression models. The intercept is absorbed by the nonparametric function in the partially linear model and the baseline hazard function in the Cox proportional hazards model respectively. Second, terms in the estimating equations for these two models are not independent. In the generalized estimating equation, the estimating equation is the sum of independent random variables. However for the partially linear model and the Cox proportional hazards model, the assumption of independence is violated. Terms in the estimating equations for both of these two models are correlated. Traditional methods may fail to work for them. Special techniques are required to deal with them in obtaining the asymptotic properties of the estimator. The assumptions, proofs and simulations for the UIPW estimator have to be investigated and examined for these two models.

In this thesis, we investigate the asymptotic properties of the UIPW estimator for the partially linear model and justify that for the Cox proportional hazards model. Simulations are designed to examine the performance of the UIPW estimators for both of these two models with missing covariates under different settings.

Chapter 2

Partially linear model with missing covariates

Consider the partially linear model (Härdle et al., 2000),

$$Y_i = X_i^{'}\beta + \nu(T_i) + \epsilon_i, i = 1, ..., n, \tag{2.1}$$

where $X_i = (x_{i1}, ..., x_{ip})^{'}$ is a vector of explanatory variables for the parametric component, $T_i = (t_i)$ is a scalar of explanatory variable for the nonparametric component, ϵ_i is random error with mean zero and finite variance $\sigma^2 = E[\epsilon_i^2]$. We assume that the vectors of random variables (X_i, T_i, ϵ_i) are independently and identically distributed (i.i.d.), which are replicates of (X, T, ϵ). $\beta = (\beta_1, ..., \beta_p)^{'}$ is a vector of unknown parameters and $\nu(\cdot)$ is an unknown function. Random error ϵ is independent of (X, T). Denote $Y = X^{'}\beta + \nu(T) + \epsilon$.

This model was first introduced by Engle et al. (1986) in their weather and electricity consumption study in US cities. The partially linear model gives a great flexibility in dealing with some practical problems compared to either the simple linear model or the nonparametric model. When $\beta = 0$, it reduces to a nonparametric model. When $\nu(\cdot) = 0$, it is a linear model. This flexibility of the partially linear model is a compromise between the simple linear model and the nonparametric model. On one hand, the goodness of fit of the linear model is improved by introducing the nonparametric component, while on the other hand, the linear component mitigates the curse of the dimensionality problem which is common in the nonparametric model in high dimension regression.

Following the introduction to the partially linear model by Engle et al. (1986), a substantial literature has emerged. Heckman (1986), Rice (1986), Green and Silverman (1994) and Eubank et al. (1998) worked on the partially linear model using the spline smoothing technique. The piecewise polynomial smoothing method was adopted by Chen (1988). Speckman (1988) studied the asymptotic behaviors of the estimators based on the kernel smoothing functions. Robinson (1988) constructed a $\sqrt{n}$-consistent estimator by a Nadaraya-Waston kernel estimator. Härdle et al. (2000) summarized the major results for the partially linear model.

The partially linear model has been extensively investigated for a variety of applications. However most approaches to studying the partially linear model are confined to fully observed data. In practice, missing data may happen either by deliberate choice due to practical limitations or by failure to collect complete data for various reasons. The study of the partially linear model with missing data has received increasing attention recently. Four main approaches have been described in the literature for the generalized linear model with missing covariates, which are maximum likelihood, multiple imputation, fully Bayesian and weighted estimating equation methods. Ibrahim et al. (2005) provided a comprehensive survey of these approaches. Liang et al. (2004) developed an estimator for the partially linear models with missing covariates by combining the methods of Robins et al. (1994) and Wang et al. (1998). This estimator is more efficient than the complete-case based estimator with Horvitz-Thompson weights. However, the finite-sample performance of the estimator is unsatisfactory because of the technical problems, the confidence region is based on a normal approximation. Liang and Qin (2008) introduced an empirical likelihood-based inference to avoid the normal approximation used in the latter estimator (Liang et al., 2004) for the partially linear model with missing covariates. Wang et al. (2004) developed an empirical likelihood method to make inference for the mean of a response when the response is missing at random. Wang (2009) developed a model calibration approach and an inverse probability weighted estimator to incorporate the incomplete observations into estimation for the monotone missing data pattern. The estimator has the same asymptotic variance as the one given in Liang et al. (2004) but it is less complicated in computation.

In this thesis, we extend the unified inverse probability weighted (UIPW) estimator class proposed by Chen and Chen (2000) and Zhao et al. (2013) to the partially linear model with missing covariates. The UIPW estimator is designed to improve the estimation efficiency of the complete-case analysis based on the generalized estimating equations under double-sampling. We investigate the asymptotic properties of the estimator and conduct simulation studies to examine the performance of the estimator for the partially linear model. In practice, the parametric component is of main interest and the nonparametric component is treated as a nuisance function in estimation. In this study, the estimation of the nonparametric component is not of interest either for the same reason, even though it is estimable and may obtain improvement of efficiency by using the unified approach.

The rest of this chapter is organized as follows. In Section 2.1, we introduce the notation for missing covariates. In Section 2.2, the UIPW estimators and their variances are described for the partially linear model with missing covariates under the MCAR and the MAR settings. The asymptotic behaviors are provided. In Section 2.3, simulations are conducted to evaluate the performance of the proposed estimator. Section 2.4 concludes this chapter and presents future research of interest.

2.1 Notations

In this study, we consider covariate data missing in the general pattern as shown in Figure 2.1 under the MCAR and the MAR missing mechanisms. Note that a complete-case analysis is necessary to be performed for using the unified approach, or it fails to work. We sort the data such that the first m observations are complete-cases. We define a covariate group as the combination of the elements of a covariate vector. For example, if $X = (x_1, x_2)'$ and $T = (t)$ in model (2.1), then (x_1), (x_2), (t), $(x_1, x_2)'$, $(x_1, t)'$ and $(x_2, t)'$ are six possible covariate groups. Assume that there are q covariate groups for covariates X and T in model (2.1). Let W_l denote the vector of covariates for the parametric component in the lth covariate group, $l = 1, 2, ..., q$. W_l consists of partial elements of X and U_l consists of the remaining elements by taking W_l out from covariate X. Let $W_l = (x_{l_1}, x_{l_2}, ..., x_{l_{p_l}})'$ and $U_l = (x_1, .., x_{l_1 - 1}, x_{l_1 + 1}, ..., x_{l_2 - 1}, x_{l_2 + 1}, ..., x_{l_{p_l} - 1}, x_{l_{p_l} + 1}, ..., x_p)'$.

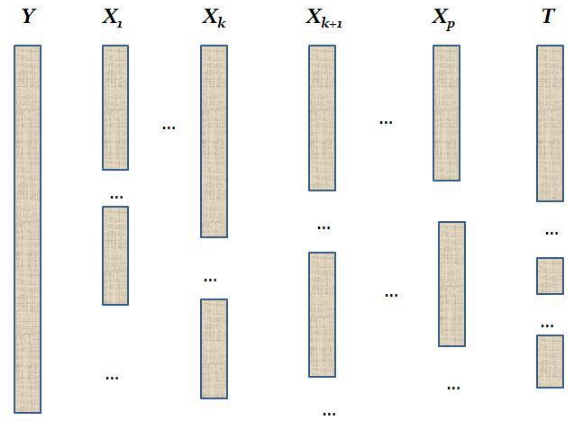

Figure 2.1: General missing data pattern

Under the MAR setting, two kinds of missingness are considered. Firstly, the selection probability function only depends on response Y but not X and T. Under this setting, both X and T are allowed to be missing. Assume the missingness of X and T is independent and T is missing completely at random. Secondly, the selection probability depends on the response Y and covariate T. Under this scenario, only covariate X is allowed to have missing values. The procedures for dealing data with the latter setting are similar to those for dealing data in the first one. We only introduce the procedures for dealing data in the first setting.

Let missing indicator variable $M_{li} = 1$ if the lth covariate group is observed for the ith observation and 0 otherwise, the complete-case indicator $M_{0i} = 1$ if $M_{1i} = M_{2i} = \ldots M_{qi} = 1$ and 0 otherwise. Denote the data selection probabilities as

$$\pi_{li} = P(M_{li} = 1 | Y_i),$$

$$\pi_{0i} = P(M_{0i} = 1|Y_i) \text{ and}$$

$$\pi_{lj,i} = P(M_{ji} = 1, M_{li} = 1|Y_i),$$

for $i = 1, ..., n$. If the missing data mechanism is MCAR, we have

$$\pi_{li} = P(M_{li} = 1|Y_i) = P(M_{li} = 1) = \pi_l,$$

$$\pi_{0i} = P(M_{0i} = 1|Y_i) = P(M_{0i} = 1) = \pi_0 \text{ and}$$

$$\pi_{lj,i} = P(M_{ji} = 1, M_{li} = 1|Y_i) = P(M_{ji} = 1, M_{li} = 1) = \pi_{lj},$$

for $i = 1, ..., n$.

Let $\Delta_0 = P(M_0 = 1|Y)$ and $A_l = \{i : M_{li} = 1\}$ denote the selection probability and the index set respectively. We define an indicator R_l for the covariate of the nonparametric component and $R_l = 1$ if the lth covariate group includes the covariate T and 0 otherwise.

The lth group of surrogate data, $(Y_i, W_{li}, T_i, M_{li}), i \in A_l$, includes the observed values of (Y, W_l, T, M_l). Let n_l be the number of observations in the lth group of surrogate data and $n_l > m$, the number of the observations in the complete-case data. The model suggested for the lth group of surrogate data is called the lth surrogate model.

The lth surrogate model is a partially linear model if $R_l = 1$ and a simple linear model if $R_l = 0$.

2.2 Estimators and asymptotic properties

For model (2.1), Speckman (1988) and Robinson (1988) proposed a least squares estimator of the parameter β and a kernel-based estimator of the nonparametric function $\nu(\cdot)$

based on the complete-case data, which are

$$\hat{\beta} = \left(\frac{1}{m}\sum_{i=1}^{m}\tilde{X}_i\tilde{X}_i'\right)^{-1}\left(\frac{1}{m}\sum_{i=1}^{m}\tilde{X}_i\tilde{Y}_i\right), \text{ and} \tag{2.2}$$

$$\hat{\nu}(t) = \sum_{i=1}^{m} K_0(t, T_i)(Y_i - X_i'\hat{\beta}), \tag{2.3}$$

where $\tilde{X}_i = X_i - \sum_{j=1}^{m} K_0(T_i, T_j)X_j$, $\tilde{Y}_i = Y_i - \sum_{j=1}^{m} K_0(T_i, T_j)Y_j$ and the weight $K_0(s, T_i)$ is defined as

$$K_0(s, T_i) = k\left(\frac{s - T_i}{b_0}\right)\Big/\sum_{j=1}^{m} k\left(\frac{s - T_j}{b_0}\right),$$

and k is a kernel function and b_0 is a bandwidth. The variance of $\sqrt{m}(\hat{\beta} - \beta^*)$ is estimated by

$$\widehat{var}(\sqrt{m}(\hat{\beta} - \beta^*)) = \hat{D}_{00}^{-1}\hat{C}_{00}\hat{D}_{00}^{-1}, \tag{2.4}$$

where

$$\hat{D}_{00} = \frac{1}{m}\sum_{i=1}^{m}\tilde{X}_i\tilde{X}_i', \text{ and}$$

$$\hat{C}_{00} = \frac{1}{m}\sum_{i=1}^{m}\tilde{X}_i(\tilde{Y}_i - \tilde{X}_i'\hat{\beta})(\tilde{Y}_i - \tilde{X}_i'\hat{\beta})'\tilde{X}_i'.$$

The above least squares estimator $\hat{\beta}$ is consistent for β^*, the true value of β (Robinson, 1988; Speckman, 1988).

We define q surrogate models corresponding to the q groups of surrogate data. For the lth group of surrogate data, if $R_l = 1$, a partially linear model is

$$Y_i = W_{li}'\gamma_l + \mu_l(T_i) + \eta_{li}, i \in A_l, \tag{2.5}$$

where $W_{li} \in R^{p_l}$ and T_i are the observed values of covariates for parametric and nonparametric components for observation i respectively, $\mu_l(\cdot)$ is an unknown function, $E(\eta_{li}) = 0$

and $E(\eta_{li}^2) = \tau_l^2$. If $R_l = 0$, we get a linear model

$$Y_i = W_{li}'\gamma_l + \eta_{li}, i \in A_l. \tag{2.6}$$

Under the MCAR setting, the complete-case data are a random subsample of the observations. The estimate based on the complete-case analysis possesses the same statistical properties as the one with fully observed data. We define the kernel smooth function in estimation as

$$K_l(s, T_i) = k\left(\frac{s - T_i}{b_l}\right) \Big/ \sum_{j \in A_l} k\left(\frac{s - T_j}{b_l}\right), i \in A_l. \tag{2.7}$$

Estimator of γ_l for the lth surrogate model based on the complete cases under the MCAR setting is given as

$$\hat{\gamma}_l = \left(\frac{1}{m}\sum_{i=1}^{m} \tilde{W}_{li}\tilde{W}_{li}'\right)^{-1}\left(\frac{1}{m}\sum_{i=1}^{m} \tilde{W}_{li}\tilde{Y}_i\right), \tag{2.8}$$

if $R_l = 1$, or

$$\hat{\gamma}_l = \left(\frac{1}{m}\sum_{i=1}^{m} W_{li}W_{li}'\right)^{-1}\left(\frac{1}{m}\sum_{i=1}^{m} W_{li}Y_i\right), \tag{2.9}$$

if $R_l = 0$, where $\tilde{W}_i = W_i - \sum_{j=1}^{m} K_0(T_i, T_j)W_j$.

For the lth covariate group, the number of observations n_l is greater than the number of the complete-case observations m. Estimate (2.8) or (2.9) for the lth surrogate model only uses the complete-case observations. The efficiency of the estimator $\hat{\gamma}_l$ can be improved by employing all available data in the lth group of surrogate data. If $R_l = 1$, the lth surrogate model is a partially linear model, then the least squares estimate of γ_l is

$$\bar{\gamma}_l = \left(\frac{1}{n_l}\sum_{i \in A_l} \bar{W}_{li}\bar{W}_{li}'\right)^{-1}\left(\frac{1}{n_l}\sum_{i \in A_l} \bar{W}_{li}\bar{Y}_i\right), i \in A_l \tag{2.10}$$

where $\bar{W}_{li} = W_{li} - \sum_{j \in A_l} K_l(T_i, T_j) W_{lj}$ and $\bar{Y}_i = Y_i - \sum_{j \in A_l} K_l(T_i, T_j) Y_j$. If $R_l = 0$, the lth surrogate model is a linear model, the least squares estimate of γ is

$$\bar{\gamma}_l = \left(\frac{1}{n_l} \sum_{i \in A_l} W_{li} W_{li}^{'} \right)^{-1} \left(\frac{1}{n_l} \sum_{i \in A_l} W_{li} Y_i \right), i \in A_l. \tag{2.11}$$

Let $\ddot{X} = X - E[X|T]$, $\ddot{W}_l = W_l - E[W_l|T]$, $\ddot{Y} = Y - E[Y|T]$, $S(W, Y, \gamma) = (S_1^{'}(W_1, Y, \gamma_1), ..., S_q^{'}(W_q, Y, \gamma_q))'$, with $S_l(W_l, Y, \gamma_l) = \ddot{W}_l(\ddot{Y} - \ddot{W}_l^{'} \gamma_l)$ if $R_l = 1$ and $S_l(W_l, Y, \gamma_l) = W_l(Y - W_l^{'} \gamma_l)$ if $R_l = 0$. Denote $\mathbb{S}(W) = diag(\mathbb{S}_1(W_1), ..., \mathbb{S}_q(W_q))$, with $\mathbb{S}_l(W_l) = \ddot{W}_l \ddot{W}_l^{'}$ if $R_l = 1$ and $\mathbb{S}_l(W_l) = W_l W_l^{'}$ if $R_l = 0$, $l = 1, ..., q$. Denote $\hat{\gamma} = (\hat{\gamma}_1^{'}, ..., \hat{\gamma}_q^{'})'$, $\bar{\gamma} = (\bar{\gamma}_1^{'}, ..., \bar{\gamma}_q^{'})'$ and $\gamma^* = (\gamma_1^{*'}, ..., \gamma_q^{*'})'$. The estimators $\hat{\gamma}$ and $\bar{\gamma}$ are consistent for γ^*. To utilize the estimators for the surrogate models to improve the estimation efficiency of $\hat{\beta}$, we propose the estimator

$$\bar{\beta} = \hat{\beta} - \hat{D}_{00}^{-1} \hat{C}_{01} \hat{C}_{11}^{-1} \hat{D}_{11} (\hat{\gamma} - \bar{\gamma}), \tag{2.12}$$

where $\hat{D}_{00}$, $\hat{C}_{01}$, $\hat{C}_{11}$ and $\hat{D}_{11}$ are estimates of D_{00}, C_{01}, C_{11} and D_{11}, which are defined as

$$\begin{aligned} D_{00} &\equiv E[\ddot{X}\ddot{X}^{'}], \\ C_{01} &\equiv E[\ddot{X}(\ddot{Y} - \ddot{X}^{'}\beta^*)S^{'}(W, Y, \gamma^*)], \\ C_{11} &\equiv E[S(W, Y, \gamma^*)S^{'}(W, Y, \gamma^*)], \text{ and} \\ D_{11} &\equiv E[\mathbb{S}(W)]. \end{aligned}$$

Matrix $\hat{D}_{00}$ is the same as that in the variance formula (2.4). Matrices $\hat{C}_{01}$, $\hat{C}_{11}$ and $\hat{D}_{11}$

are estimated by

$$\hat{C}_{01} = \frac{1}{m}\sum_{i=1}^{m} \tilde{X}_i(\tilde{Y}_i - \tilde{X}_i'\hat{\beta})\hat{S}'(W_i, Y_i, \hat{\gamma}),$$

$$\hat{C}_{11} = \frac{1}{m}\sum_{i=1}^{m} \hat{S}(W_i, Y_i, \hat{\gamma})\hat{S}'(W_i, Y_i, \hat{\gamma}), \text{ and}$$

$$\hat{D}_{11} = \frac{1}{m}\sum_{i=1}^{m} \hat{\mathbb{S}}(W_i),$$

where $\hat{S}(W_i, Y_i, \hat{\gamma}) = (\hat{S}_1'(W_{1i}, Y_i, \hat{\gamma}_1), ..., \hat{S}_q'(W_{qi}, Y_i, \hat{\gamma}_q))'$, with $\hat{S}_l(W_{li}, Y_i, \hat{\gamma}_l) = \tilde{W}_{li}(\tilde{Y}_i - \tilde{W}_{li}'\hat{\gamma}_l)$ if $R_l = 1$, and $\hat{S}_l(W_{li}, Y_i, \hat{\gamma}_l) = W_{li}(Y_i - W_{li}'\hat{\gamma}_l)$ if $R_l = 0$; $\hat{\mathbb{S}}(W_i) = diag(\hat{\mathbb{S}}_1(W_{1i}), \hat{\mathbb{S}}_2(W_{2i}), ..., \hat{\mathbb{S}}_q(W_{qi}))$, with $\hat{\mathbb{S}}_l(W_{li}) = \tilde{W}_{li}\tilde{W}_{li}'$ if $R_l = 1$, and $\hat{\mathbb{S}}_l(W_{1l}) = W_{li}W_{li}'$ if $R_l = 0$, $l = 1, ..., q$.

To deal with the missing data under MAR mechanism, we will use the inverse probability weighted estimating equations. Let us denote the probability for the ith observation being a complete-case as $\pi_{0i} = P(M_{0i} = 1|Y)$. When $\pi_{0i}, i = 1, ..., q$ are known, the inverse selection probability weighted estimator based on the complete cases is

$$\hat{\beta}_\pi = \left(\frac{1}{n}\sum_{i=1}^{n} \frac{M_{0i}}{\pi_{0i}}\tilde{X}_i\tilde{X}_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} \frac{M_{0i}}{\pi_{0i}}\tilde{X}_i\tilde{Y}_i\right), \quad (2.13)$$

where $\tilde{X}_i = X_i - \sum_{j=1}^{n} \frac{M_{0j}}{\pi_{0j}} K_0(T_i, T_j)X_j$ and $\tilde{Y}_i = Y_i - \sum_{j=1}^{n} K_0(T_i, T_j)Y_j$.

The variance of $\sqrt{n}(\hat{\beta}_\pi - \beta^*)$ can be estimated by

$$\widehat{var}(\sqrt{n}(\hat{\beta}_\pi - \beta^*)) = \hat{D}_{\pi 00}^{-1}\hat{C}_{\pi 00}\hat{D}_{\pi 00}^{-1} \quad (2.14)$$

where

$$\hat{D}_{\pi 00} = \frac{1}{n}\sum_{i=1}^{n} \frac{M_{0i}}{\pi_{0i}}\tilde{X}_i\tilde{X}_i'$$

and

$$\hat{C}_{\pi 00} = \frac{1}{n}\sum_{i=1}^{n}\frac{M_{0i}}{\pi_{0i}^2}\tilde{X}_i(\tilde{Y}_i - \tilde{X}_i'\hat{\beta})(\tilde{Y}_i - \tilde{X}_i'\hat{\beta})'\tilde{X}_i'.$$

To estimate $\gamma_{\pi l}$ based on the complete-case data for the lth surrogate model, we use

$$\hat{\gamma}_{\pi l} = \left(\frac{1}{n}\sum_{i=1}^{n}\frac{M_{0i}}{\pi_{0i}}\tilde{W}_{li}\tilde{W}_{li}'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{M_{0i}}{\pi_{0i}}\tilde{W}_{li}\tilde{Y}_i\right), \tag{2.15}$$

if $R_l = 1$, and

$$\hat{\gamma}_{\pi l} = \left(\frac{1}{n}\sum_{i=1}^{n}\frac{M_{0i}}{\pi_{0i}}W_{li}W_{li}'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{M_{0i}}{\pi_{0i}}W_{li}Y_i\right). \tag{2.16}$$

if $R_l = 0$, where $\tilde{W}_i = W_i - \sum_{j=1}^{n}\frac{M_{0i}}{\pi_{0i}}K_0(T_i, T_j)W_j$.

To improve the efficiency of estimator $\hat{\gamma}$, we use all the available data for the surrogate models. For the lth group of surrogate data, the probability of the lth covariate group being observed for the ith observation is $\pi_{li} = P(M_{li} = 1|Y)$. To use the lth group of surrogate data in estimation, the weight $1/\pi_{li}$ is employed for the ith observation. If $R_l = 1$, the improved estimate of γ_l using all the lth group of surrogate data for model (2.5) is

$$\bar{\gamma}_{\pi l} = \left(\frac{1}{n}\sum_{i=1}^{n}\frac{M_{li}}{\pi_{li}}\bar{W}_{li}\bar{W}_{li}'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{M_{li}}{\pi_{li}}\bar{W}_{li}\bar{Y}_i\right), \tag{2.17}$$

where $\bar{W}_{li} = W_{li} - \sum_{j=1}^{n}\frac{M_{lj}}{\pi_{li}}K_l(T_i, T_j)W_{lj}$ and $\bar{Y}_i = Y_i - \sum_{j=1}^{n}M_{lj}K_l(T_i, T_j)Y_j$. Note that the kernel smooth function $K_l(\cdot,\cdot)$ in (2.7) has to be used in computation for $\bar{\gamma}_{\pi l}$. If $R_l = 0$, then the improved estimator for model (2.6) by using all the lth group of surrogate data is

$$\bar{\gamma}_{\pi l} = \left(\frac{1}{n}\sum_{i=1}^{n}\frac{M_{li}}{\pi_{li}}W_{li}W_{li}'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{M_{li}}{\pi_{li}}W_{li}Y_i\right). \tag{2.18}$$

Let $S_\pi(W, Y, \gamma) = (S_{\pi 1}'(W_1, Y, \gamma_1), ..., S_{\pi q}'(W_q, Y, \gamma_q))'$, with $S_{\pi l}(W_l, Y, \gamma_l) = \ddot{W}_l\frac{M_0}{\Delta_0}$ $(\ddot{Y} - \ddot{W}_l'\gamma_l)$ if $R_l = 1$ and $S_{\pi l}(W_l, Y, \gamma_l) = W_l\frac{M_0}{\Delta_0}$ $(Y - W_l'\gamma_l)$ if $R_l = 0$. Let $\mathbb{S}_\pi(W) =$

$diag(\mathbb{S}_{\pi 1}(W_1), ..., \mathbb{S}_{\pi q}(W_q))$, with $\mathbb{S}_{\pi l}(W_l) = \ddot{W}_l \frac{M_0}{\Delta_0} \ddot{W}_l'$ if $R_l = 1$ and $\mathbb{S}_{\pi l}(W_l) = W_l \frac{M_0}{\Delta_0} W_l'$ if $R_l = 0$. $\hat{\gamma}_\pi = (\hat{\gamma}_{\pi 1}', ..., \hat{\gamma}_{\pi q}')'$ and $\bar{\gamma}_\pi = (\bar{\gamma}_{\pi 1}', ..., \bar{\gamma}_{\pi q}')'$. The UIPW estimator of β is given as

$$\bar{\beta}_\pi = \hat{\beta}_\pi - \hat{D}_{\pi 00}^{-1} \hat{C}_{\pi 01} \hat{C}_{\pi 11}^{-1} \hat{D}_{\pi 11} (\hat{\gamma}_\pi - \bar{\gamma}_\pi) \tag{2.19}$$

where $\hat{D}_{\pi 00}$, $\hat{C}_{\pi 01}$, $\hat{C}_{\pi 11}$ and $\hat{D}_{\pi 11}$ are estimates of $D_{\pi 00}$, $C_{\pi 01}$, $C_{\pi 11}$ and $D_{\pi 11}$ respectively, which are defined as

$$\begin{aligned}
D_{\pi 00} &\equiv E[\ddot{X} \frac{M_0}{\Delta_0} \ddot{X}'], \\
C_{\pi 01} &\equiv E[\ddot{X} \frac{M_0}{\Delta_0} (\ddot{Y} - \ddot{X}' \beta^*) S_\pi'(W, Y, \gamma^*)], \\
C_{\pi 11} &\equiv E[S_\pi(W, Y, \gamma^*) S_\pi'(W, Y, \gamma^*)], \text{ and} \\
D_{\pi 11} &\equiv E[\mathbb{S}_\pi(W)].
\end{aligned}$$

Matrix $D_{\pi 00}$ is estimated the same as in (2.14). Matrices $C_{\pi 01}$, $C_{\pi 11}$ and $D_{\pi 11}$ are estimated by

$$\begin{aligned}
\hat{C}_{\pi 01} =& \frac{1}{n} \sum_{i=1}^n \frac{M_{0i}}{\pi_{0i}} \tilde{X}_i (\tilde{Y}_i - \tilde{X}_i' \hat{\beta}) \hat{S}_\pi'(W_i, Y_i, \hat{\gamma}), \\
\hat{C}_{\pi 11} =& \frac{1}{n} \sum_{i=1}^n \hat{S}_\pi(W_i, Y_i, \hat{\gamma}) \hat{S}_\pi'(W_i, Y_i, \hat{\gamma}), \text{ and} \\
\hat{D}_{\pi 11} =& \frac{1}{n} \sum_{i=1}^n \hat{\mathbb{S}}_\pi(W_i),
\end{aligned}$$

respectively, where $\hat{S}_\pi(W_i, Y_i, \hat{\gamma}) = (\hat{S}_{\pi 1}'(W_{1i}, Y_i, \gamma_1), ..., \hat{S}_{\pi q}'(W_{qi}, Y_i, \gamma_q))'$, with $\hat{S}_{\pi l}(W_{li}, Y_i, \hat{\gamma}_l) = \tilde{W}_{li} \frac{M_{0i}}{\pi_{0i}} (\tilde{Y}_i - \tilde{W}_{li}' \hat{\gamma}_l)$ if $R_l = 1$, and $\hat{S}_{\pi l}(W_{li}, Y_i, \hat{\gamma}_l) = W_{li} \frac{M_{0i}}{\pi_{0i}} (Y_i - W_{li}' \hat{\gamma}_l)$ if $R_l = 0$; $\hat{\mathbb{S}}_\pi(W_i) = diag(\hat{\mathbb{S}}_{\pi 1i}(W_{1i}), \hat{\mathbb{S}}_{\pi 2i}(W_{2i}), ..., \hat{\mathbb{S}}_{\pi qi}(W_{qi}))$, with $\hat{\mathbb{S}}_{\pi l}(W_{li}) = \tilde{W}_{li} \frac{M_{0i}}{\pi_{0i}} \tilde{W}_{li}'$ if $R_l = 1$ and $\hat{\mathbb{S}}_{\pi l}(W_{1l}) = W_{li} \frac{M_{0i}}{\pi_{0i}} W_{li}'$ if $R_l = 0$, $l = 1, ..., q$.

The selection probability may be known or unknown but need to be parametrically estimable. The estimator using estimated selection probabilities can be more efficient than that using the known selection probabilities (Robins et al., 1995). If the selection probabilities π_{li} can be parametrically modeled as $\pi_i(\alpha_l) = P(M_{li} = 1|Y_i, \alpha_l), l = 0, 1, ..., q$, where α_ls are vectors of parameters for the selection probabilities. The estimator $\hat{\alpha}_l$ is the solution to the estimating equations $\sum_i^n H_{li}(\alpha_l) = 0, l = 0, 1, ..., q$. The estimators $\hat{\beta}_{\hat{\pi}}$, $\hat{\gamma}_{\hat{\pi}}$ and $\bar{\gamma}_{\hat{\pi}}$ are obtained by replacing the selection probabilities π_{0i} and π_{li} in formulae (2.13), (2.15), (2.16), (2.17) and (2.18) with their estimates $\pi_{0i}(\hat{\alpha}_0)$ and $\pi_{li}(\hat{\alpha}_l), i = 1, ..., n$ and $l = 1, ..., q$. Let $\alpha = (\alpha_1^{'}, ..., \alpha_q^{'})^{'}$ and $H_i(\alpha) = (H_{1i}^{'}(\alpha_1), ..., H_{qi}^{'}(\alpha_q))^{'}$. Following Robins et al. (1995), we denote the residual for the ith subject from the least squares regression of A_i on B_i, $i = 1, ..., n$, as

$$\hat{R}es(A_i, B_i) = A_i - \left(\sum_{i=1}^n A_i B_i^{'}\right)\left(\sum_{i=1}^n B_i B_i^{'}\right)^{-1} B_i.$$

The UIPW estimator of β with the estimated selection probabilities is given as

$$\bar{\beta}_{\hat{\pi}} = \hat{\beta}_{\hat{\pi}} - \hat{D}_{\hat{\pi}00}^{-1} \hat{C}_{\hat{\pi}01} \hat{C}_{\hat{\pi}11}^{-1} \hat{D}_{\hat{\pi}11} (\hat{\gamma}_{\hat{\pi}} - \bar{\gamma}_{\hat{\pi}}), \tag{2.20}$$

where $\hat{D}_{\hat{\pi}00}$, $\hat{C}_{\hat{\pi}01}$, $\hat{C}_{\hat{\pi}11}$ and $\hat{D}_{\hat{\pi}11}$ are estimated as follows

$$\begin{aligned}
\hat{D}_{\hat{\pi}00} =& \frac{1}{n} \sum_{i=1}^n \frac{M_{0i}}{\pi_{0i}(\hat{\alpha}_0)} \tilde{X}_i \tilde{X}_i^{'}, \\
\hat{C}_{\hat{\pi}01} =& \frac{1}{n} \sum_{i=1}^n \hat{R}es\left(\frac{M_{0i}}{\pi_{0i}(\hat{\alpha}_0)} \tilde{X}_i(\tilde{Y}_i - \tilde{X}_i^{'}\beta), H_{0i}(\hat{\alpha}_0))\right) \hat{R}es^{'}\left(\hat{S}_{\hat{\pi}}(W_i, Y_i, \hat{\gamma}), H_{0i}(\hat{\alpha}_0)\right), \\
\hat{C}_{\hat{\pi}11} =& \frac{1}{n} \sum_{i=1}^n \hat{R}es\left(\hat{S}_{\hat{\pi}}(W_i, Y_i, \hat{\gamma}), H_{0i}(\hat{\alpha}_0)\right) \hat{R}es^{'}\left(\hat{S}_{\hat{\pi}}(W_i, Y_i, \hat{\gamma}), H_{0i}(\hat{\alpha}_0)\right), \text{and} \\
\hat{D}_{\hat{\pi}11} =& \frac{1}{n} \sum_{i=1}^n \hat{\mathbb{S}}_{\hat{\pi}}(W_i),
\end{aligned}$$

where $\hat{S}_{\hat{\pi}}(W_i, Y_i, \hat{\gamma}) = (\hat{S}'_{\hat{\pi}1}(W_{1i}, Y_i, \gamma_1), ..., \hat{S}'_{\hat{\pi}q}(W_{qi}, Y_i, \gamma_q))'$, $\hat{S}_{\hat{\pi}l}(W_{li}, Y_i, \hat{\gamma}_l) = \tilde{W}_{li}\frac{M_{0i}}{\pi_{0i}(\hat{\alpha}_0)}$ $(\tilde{Y}_i - \tilde{W}'_{li}\hat{\gamma}_l)$ if $R_l = 1$ and $\hat{S}_{\hat{\pi}l}(W_{li}, Y_i, \hat{\gamma}_l) = W_{li}\frac{M_{0i}}{\pi_{0i}(\hat{\alpha}_0)}(Y_i - W'_{li}\hat{\gamma}_l)$ if $R_l = 0$; $\hat{\mathbb{S}}_{\hat{\pi}}(W_i) = diag(\hat{\mathbb{S}}_{\hat{\pi}1i}(W_{1i}), \hat{\mathbb{S}}_{\hat{\pi}2i}(W_{2i}), ..., \hat{\mathbb{S}}_{\hat{\pi}qi}(W_{qi}))$, $\hat{\mathbb{S}}_{\hat{\pi}l}(W_{li}) = \tilde{W}_{li}\ \frac{M_{0i}}{\pi_{0i}(\hat{\alpha}_0)}\tilde{W}'_{li}$ if $R_l = 1$ and $\hat{\mathbb{S}}_{\hat{\pi}l}(W_{1l})$ $= W_{li}\frac{M_{0i}}{\pi_{0i}(\hat{\alpha}_0)}W'_{li}$ if $R_l = 0, l = 1, ..., q$.

We do not propose a nonparametric model for the surrogate data with X missing but T observed. We know that even if a partially linear model is misspecified as a nonparametric regression model, the estimator for the model is still consistent. However its convergence rate is not the same as its partially linear counterpart. The convergence rate of the nonparametric estimator is usually of lower order $m^{-r}, 0 < r < 1/2$ (Härdle, 1994; Härdle et al., 2000), where m is the number of the complete-case observations. The estimator for the parametric component in the partially linear model tends to approach the true value typically at speed $1/\sqrt{m}$ with appropriate kernel bandwidth and other parameters required by convergence criteria. Because of the different convergence rate for the parametric and the nonparametric estimators, the joint distribution of the estimator for the parametric component $\sqrt{m}(\hat{\beta} - \beta^*)$ and the nonparametric component $\sqrt{m}(\hat{\nu}(t) - \nu^*(t))$ might not converge together in distribution. The unified approach might not be applicable for the case with covariates X missing completely but T observed. A detailed investigation of these issues is interesting, but is beyond this thesis.

The estimators $\hat{\beta}$ and $\hat{\nu}$ are consistent and asymptotic normal under some assumptions (Speckman, 1988; Robinson, 1988). The asymptotic behaviors of estimator $\hat{\beta}$ have been investigated using different smoothing techniques by Heckman (1986), Rice (1986), Speckman (1988), Robinson (1988), etc. Silverman (1984) showed that the Spline smoothing is asymptotically equivalent to the kernel smoothing. Under the assumptions in Appendix 5.2, Robinson (1988) constructed a $\sqrt{m}$-consistent estimator by kernel smoothing for random design case. The joint normal distribution of $\sqrt{m}(\hat{\beta}-\beta^*)$ and $\sqrt{m}(\hat{\gamma}-\gamma^*)$ is sufficient to apply the unified approach for the partially linear model in this study.

Robinson (1988) showed the estimator $\sqrt{m}(\hat{\beta}-\beta^*)$ given in (2.2) for model (2.1) is consistent and asymptotically normal with mean 0 and variance $\sigma^2 E^{-1}[\{X-E(X|T)\}\{X-E(X|T)\}^{'}]$. In the following part, we use some results from Robinson (1988) to show the asymptotic properties of the estimators for the surrogate models. The proofs of Corollary 2.2.1 - 2.2.3 are provided in Appendix 5.3.

Corollary 2.2.1. *Under the assumptions given in Appendix 5.2, estimator* $\hat{\gamma}_l$ (2.8) *is consistent for* γ_l^* *and* $\sqrt{m}(\hat{\gamma}_l-\gamma_l^*)$ *is asymptotically normal with mean* 0 *and variance* $\sigma^2\Phi_l^{-1}+\Phi_l^{-1}var(\ddot{W}_l\ddot{U}_l^{'}\beta_{l2})\Phi_l^{-1}$, *where* $\Phi_l = E[\ddot{W}_l\ddot{W}_l^{'}]$, β_{l2} *is the parameters corresponding to the covariates* W_l *in the* l*th covariate group .*

Corollary 2.2.2. *Under the assumptions given in Appendix 5.2 along with the positive definiteness of* $E[XX^{'}] = \Sigma$ *and* $E[W_l\nu(T)] = \Psi < \infty$, *estimator* $\hat{\gamma}_l$ (2.9) *is consistent for* γ_l^* *and* $\sqrt{m}(\hat{\gamma}_l-\gamma_l^*)$ *is asymptotically normal with mean* 0 *and variance* $\sigma^2\Sigma_l^{-1}+\Sigma_l^{-1}var(W_lU_l^{'}\beta_{l2}+W_l\nu(T))\Sigma_l^{-1}$, *where* $E[W_lW_l^{'}] = \Sigma_l$.

Corollary 2.2.3. *Under assumptions given in Appendix 5.2 and Corollary 2.2.1, 2.2.2, the joint distribution of* $\sqrt{m}(\hat{\beta}-\beta^*)$ *and* $\sqrt{m}(\hat{\gamma}_l-\gamma_l^*), l = 1, ..., q$ *is asymptotically normal* *with mean* 0 *and variance*

$$\begin{pmatrix} \sigma^2\Phi^{-1} & \Phi^{-1}\sigma^2 E[\ddot{X}\ddot{W}_l']\Phi_l^{-1} \\ \Phi_l^{-1}\sigma^2 E[\ddot{W}_l\ddot{X}']\Phi^{-1} & \Phi_l^{-1}\sigma^2 + \Phi_l^{-1}var(\ddot{W}_l\ddot{U}\beta_{l2})\Phi_l^{-1} \end{pmatrix},$$

if $R_l = 1$*, or variance*

$$\begin{pmatrix} \sigma^2\Phi^{-1} & \Phi^{-1}\sigma^2 E[\ddot{X}W_l']\Sigma_l^{-1} \\ \Sigma_l^{-1}\sigma^2 E[W_l\ddot{X}']\Phi^{-1} & \Sigma_l^{-1}\sigma^2 + \Sigma_l^{-1}var(\ddot{W}_l\ddot{U}\beta_{l2} + W\nu(T))\Sigma_l^{-1} \end{pmatrix},$$

if $R_l = 0$*.*

We do not provide a proof of the joint normality of $\sqrt{m}(\hat{\beta} - \beta^*)$ and $\sqrt{m}(\hat{\gamma} - \gamma^*)$, however extending proof of the Corollary 2.2.3 is part of my future research. We presumptively assume that the joint distribution of $\sqrt{m}(\hat{\beta} - \beta^*)$ and $\sqrt{m}(\hat{\gamma} - \gamma^*)$ is normal without explicit proof in the following parts. As a result, the asymptotic normality of $\bar{\beta}$ is asserted because of the asymptotic normality of $\hat{\beta}$, $\hat{\gamma}$ and $\bar{\gamma}$ and by Slutsky's theorem. The consistency of $\bar{\beta}$ is sustained due to the consistencies of $\hat{\beta}$, $\hat{\gamma}$ and $\bar{\gamma}$ and the regularity conditions in Appendix 5.2. We conjecture that estimators $\hat{\beta}_\pi$ and $\hat{\gamma}_\pi$ have similar asymptotic properties as $\hat{\beta}$ and $\hat{\gamma}$.

Under the MCAR setting, $\pi_{01} = ...\pi_{0n} = \pi_0$, $\pi_{l1} = ...\pi_{ln} = \pi_l$, and $\pi_{lj,1} = ...\pi_{lj,n} = \pi_{lj}$, the proposed estimator $\bar{\beta}$ in (2.12) is consistent for β^* and $\sqrt{m}(\bar{\beta} - \beta^*)$ is asymptotically normal with mean 0 and variance estimated by

$$\widehat{var}(\sqrt{m}(\bar{\beta} - \beta^*)) = \hat{D}_{00}^{-1}\hat{C}_{00}\hat{D}_{00}^{-1} - \hat{D}_{00}^{-1}\hat{C}_{01}(I - \pi_0\hat{C}_{11}^{-1}(\hat{C}_{11} \circ V))\hat{C}_{11}^{-1}\hat{C}_{01}'\hat{D}_{00}^{-1}$$

where symbol $\circ$ denote Hadamard product of matrices and

$$V = \begin{pmatrix} \pi_{11}/(\pi_1\pi_1) & ... & \pi_{1q}/(\pi_1\pi_q) \\ ... & ... & ... \\ \pi_{q1}/(\pi_q\pi_1) & ... & \pi_{qq}/(\pi_q\pi_q) \end{pmatrix}.$$

Under the MAR setting, the proposed estimator $\bar{\beta}$ in (2.19) is consistent for β^* and $\sqrt{m}(\bar{\beta} - \beta^*)$ is asymptotically normal with mean 0 and variance estimated by

$$\begin{aligned}\widehat{var}(\sqrt{n}(\bar{\beta}_\pi - \beta^*)) =& \hat{D}_{\pi 00}^{-1}\hat{C}_{\pi 00}\hat{D}_{\pi 00}^{-1} \\ & -\hat{D}_{\pi 00}^{-1}\hat{C}_{\pi 01}\hat{C}_{\pi 11}^{-1}\Big(\hat{C}_{\pi 11} - \hat{C}_{\pi 22} + \hat{C}_{\pi 12} + \hat{C}_{\pi 12}'\Big)\hat{C}_{\pi 11}^{-1}\hat{C}_{\pi 01}'\hat{D}_{\pi 00}^{-1} \\ & +\hat{D}_{\pi 00}^{-1}\Big(\hat{C}_{\pi 01}\hat{C}_{\pi 11}^{-1}\hat{C}_{\pi 02}' + \hat{C}_{\pi 02}\hat{C}_{\pi 11}^{-1}\hat{C}_{\pi 01}'\Big)\hat{D}_{\pi 00}^{-1}, \end{aligned} \tag{2.21}$$

where $\hat{C}_{\pi 00}$ is the same as that in (2.13). Estimates $\hat{C}_{02}$, $\hat{C}_{\pi 12}$ and $\hat{C}_{\pi 22}$ are as follows

$$\begin{aligned}\hat{C}_{\pi 02} =& \frac{1}{n}\sum_{i=1}^{n}\frac{M_{0i}}{\pi_{0i}}\tilde{X}_i(\tilde{Y}_i - \tilde{X}_i'\hat{\beta})\bar{S}_\pi'(W_i, Y_i, \bar{\gamma}), \\ \hat{C}_{\pi 12} =& \frac{1}{n}\sum_{i=1}^{n}\hat{S}_\pi(W_i, Y_i, \hat{\gamma})\bar{S}_\pi'(W_i, Y_i, \bar{\gamma}) \text{ and} \\ \hat{C}_{\pi 22} =& \frac{1}{n}\sum_{i=1}^{n}\bar{S}_\pi(W_i, Y_i, \bar{\gamma})\bar{S}_\pi'(W_i, Y_i, \bar{\gamma}),\end{aligned}$$

where $\bar{S}_\pi(W_i, Y_i, \hat{\gamma}) = (\bar{S}_{\pi 1}'(W_{1i}, Y_i, \gamma_1), ..., \bar{S}_{\pi q}'(W_{qi}, Y_i, \gamma_q))'$ with $\bar{S}_{\pi l}(W_{li}, Y_i, \bar{\gamma}_l) = \bar{W}_{li}$ $\frac{M_{li}}{\pi_{li}}(\bar{Y}_i - \bar{W}_{li}'\bar{\gamma}_l)$ if $R_l = 1$ and $\bar{S}_{\pi l}(W_{li}, Y_i, \bar{\gamma}_l) = W_{li}\frac{M_{li}}{\pi_{li}}(Y_i - W_{li}'\bar{\gamma}_l)$ if $R_l = 0, l = 1, ..., q$.

If the selection probability π is estimated, the variance of $\sqrt{n}(\bar{\beta}_{\hat{\pi}} - \beta^*)$ can be estimated by

$$\begin{aligned}\widehat{var}(\sqrt{n}(\bar{\beta}_{\hat{\pi}} - \beta^*)) =& \hat{D}_{\hat{\pi} 00}^{-1}\hat{C}_{\hat{\pi} 00}\hat{D}_{\hat{\pi} 00}^{-1} \\ & -\hat{D}_{\hat{\pi} 00}^{-1}\hat{C}_{\hat{\pi} 01}\hat{C}_{\hat{\pi} 11}^{-1}\Big(\hat{C}_{\hat{\pi} 11} - \hat{C}_{\hat{\pi} 22} + \hat{C}_{\hat{\pi} 12} + \hat{C}_{\hat{\pi} 12}'\Big)\hat{C}_{\hat{\pi} 11}^{-1}\hat{C}_{\hat{\pi} 01}'\hat{D}_{\hat{\pi} 00}^{-1}\end{aligned}$$

$$+\hat{D}_{\hat{\pi}00}^{-1}\bigg(\hat{C}_{\hat{\pi}01}\hat{C}_{\hat{\pi}11}^{-1}\hat{C}_{\hat{\pi}02}' + \hat{C}_{\hat{\pi}02}\hat{C}_{\hat{\pi}11}^{-1}\hat{C}_{\hat{\pi}01}'\bigg)\hat{D}_{\hat{\pi}00}^{-1}, \tag{2.22}$$

where $\hat{D}_{\hat{\pi}00}$, $\hat{C}_{\hat{\pi}01}$, $\hat{C}_{\hat{\pi}11}$ and $\hat{D}_{\hat{\pi}11}$ are the same terms as in estimator $\bar{\beta}_{\hat{\pi}}$ in (2.20), $\hat{C}_{\hat{\pi}00}$, $\hat{C}_{\hat{\pi}02}$, $\hat{C}_{\hat{\pi}12}$ and $\hat{C}_{\hat{\pi}22}$ are as follows:

$$\hat{C}_{\hat{\pi}00} = \frac{1}{n}\sum_{i=1}^{n}\hat{R}es\bigg(\frac{M_{0i}}{\pi_{0i}(\hat{\alpha}_0)}\tilde{X}_i(\tilde{Y}_i - \tilde{X}_i'\hat{\beta}), H_{0i}(\hat{\alpha}_0)\bigg)\hat{R}es'\bigg(\frac{M_{0i}}{\pi_{0i}(\hat{\alpha}_0)}\tilde{X}_i(\tilde{Y}_i - \tilde{X}_i'\hat{\beta}), H_{0i}(\hat{\alpha}_0)\bigg),$$
$$\hat{C}_{\hat{\pi}02} = \frac{1}{n}\sum_{i=1}^{n}\hat{R}es\bigg(\frac{M_{0i}}{\pi_{0i}(\hat{\alpha}_0)}\tilde{X}_i(\tilde{Y}_i - \tilde{X}_i'\hat{\beta}), H_{0i}(\hat{\alpha}_0)\bigg)\hat{R}es'\bigg(\bar{S}_{\hat{\pi}}(W_i, Y_i, \bar{\gamma}), H_i(\hat{\alpha})\bigg),$$
$$\hat{C}_{\hat{\pi}12} = \frac{1}{n}\sum_{i=1}^{n}\hat{R}es\bigg(\hat{S}_{\hat{\pi}}(W_i, Y_i, \hat{\gamma}), H_{0i}(\hat{\alpha}_0)\bigg)\hat{R}es'\bigg(\bar{S}_{\hat{\pi}}(W_i, Y_i, \bar{\gamma}), H_i(\hat{\alpha})\bigg) \text{ and}$$
$$\hat{C}_{\hat{\pi}22} = \frac{1}{n}\sum_{i=1}^{n}\hat{R}es\bigg(\bar{S}_{\hat{\pi}}(W_i, Y_i, \bar{\gamma}), H_i(\hat{\alpha})\bigg)\hat{R}es'\bigg(\bar{S}_{\hat{\pi}}(W_i, Y_i, \bar{\gamma}), H_i(\hat{\alpha})\bigg),$$

where $\bar{S}_{\hat{\pi}}(W_i, Y_i, \hat{\gamma}) = (\bar{S}_{\hat{\pi}1}'(W_{1i}, Y_i, \gamma_1), ..., \bar{S}_{\hat{\pi}q}'(W_{qi}, Y_i, \gamma_q))'$ with $\bar{S}_{\hat{\pi}l}(W_{li}, Y_i, \bar{\gamma}_l) = \bar{W}_{li} \frac{M_{li}}{\pi_{li}(\hat{\alpha}_l)}(\bar{Y}_i - \bar{W}_{li}'\bar{\gamma}_l)$ if $R_l = 1$, and $\bar{S}_{\hat{\pi}l}(W_{li}, Y_i, \bar{\gamma}_l) = W_{li}\frac{M_{li}}{\pi_{li}(\hat{\alpha}_l)}(Y_i - W_{li}'\bar{\gamma}_l)$ if $R_l = 0$, $l = 1, ..., q$.

2.3 Simulation study

Three simulation studies are conducted to investigate the performance of the UIPW estimator for the partially linear model. Simulation study one is for the MCAR setting. Simulation study two and three are for the MAR settings.

Simulation results are reported in Table 2.1- 2.3. Here: $bias(\hat{\beta})$ and $bias(\bar{\beta})$ are the averages of $\hat{\beta} - \beta$ and $\bar{\beta} - \beta$ respectively; $s.d.(\hat{\beta})$ and $s.d.(\bar{\beta})$ are the sample standard deviations of the estimators $\hat{\beta}$ and $\bar{\beta}$ respectively; $s.e.(\hat{\beta})$ and $s.e.(\bar{\beta})$ are the averages of estimated standard error of the $\hat{\beta}$ and $\bar{\beta}$ respectively; CP of $\hat{\beta}$ and CP of $\bar{\beta}$ are the coverage probabilities of the 95% confidence interval of β for estimators $\hat{\beta}$ and $\bar{\beta}$ respectively; $MSE(\hat{\beta})$ and $MES(\bar{\beta})$ are the averages of the squared differences $\hat{\beta} - \beta$ and $\bar{\beta} - \beta$; ARE is the asymptotic relative efficiency and $ARE = (s.d.(\hat{\beta})/s.d.(\bar{\beta}))^2$. We generate 1000 datasets for each case.

2.3.1 Simulation one (MCAR setting)

We consider the model $Y = X\beta + sin(T) + \epsilon$, where $X = (x_1, x_2)$ and $\beta = (1, 0.75)'$. Covariates x_1 and x_2 are independent random variables, $x_1 \sim U(0, 1)$ and $x_2 \sim N(0, 1)$. Covariate t and random error ϵ follow $N(0, 1)$. Sample size $n = 200$ and 500 are chosen. The selection probabilities for covariate x_1 and covariate x_2 are 0.4 and 0.8 respectively. In this simulation, the normal kernel function $K(u) = exp(-u^2/2h^2)$ is applied and the kernel bandwidth $h = 1/\sqrt[5]{n}$ is used.

Simulation results are presented in Table 2.1. We see that bias($\bar{\beta}$)'s, $s.e.(\bar{\beta})$'s, $s.d.(\bar{\beta})$'s and MSE($\bar{\beta}$)'s are noticeably smaller than bias($\hat{\beta}$)'s, $s.e.(\hat{\beta})$'s, $s.d.(\hat{\beta})$'s and MSE($\hat{\beta}$)'s respectively. Comparing the efficiency gain between $\bar{\beta}_1$ and $\bar{\beta}_2$ in Table 2.1, the estimates of β_2 gain more efficiency than that of β_1. This is because there are more available data for β_2 to compensate the loss of information due to discarding the incomplete cases in estimation. The $ARE(\beta_1)$'s are around 1.15 and the $ARE(\beta_2)$'s are around 2.4. The CPs of $\hat{\beta}$ and $\bar{\beta}$ are close to the 95% nominal coverage rate.

Table 2.1: Simulation one (MCAR). $Y = X\beta + sin(T) + \epsilon$,$X = (x_1, x_2)', x_1 \sim U(0,1)$, x_2, t and $\epsilon \sim N(0,1)$. Selection probabilities for x_1 and x_2 are 0.4 and 0.8 respectively.

	n=200		n=500	
	$\beta_1 = 1$	$\beta_2 = 0.75$	$\beta_1 = 1$	$\beta_2 = 0.75$
bias($\hat{\beta}$)	-0.0082	0.0027	0.0082	-0.0014
$s.d.(\hat{\beta})$	0.4462	0.1373	0.2841	0.0826
$s.e.(\hat{\beta})$	0.4298	0.1245	0.2734	0.0788
CP of $\hat{\beta}$	0.9330	0.9210	0.9420	0.9380
MSE($\hat{\beta}$)	0.1989	0.0188	0.0807	0.0068
bias($\bar{\beta}$)	0.0032	0.0056	0.0084	-0.0004
$s.d.(\bar{\beta})$	0.4251	0.0892	0.2642	0.0553
$s.e.(\bar{\beta})$	0.4095	0.0920	0.2578	0.0549
CP of $\bar{\beta}$	0.9340	0.9500	0.9450	0.9500
MSE($\bar{\beta}$)	0.1806	0.0080	0.0698	0.0031
ARE	1.1013	2.3698	1.1566	2.2286

2.3.2 Simulation two (MAR setting)

We consider model $Y = X\beta + g(T) + \epsilon$, where $X = (x_1, x_2)$, $g(T) = 0$, $\beta = (0, 0.5)$, and $x_1, x_2, t \sim U(0,1)$ and random error $\epsilon \sim N(0,1)$. The selection probability of x_1 equals to $Pr(\delta = 1|Y, T) = 1/(1 + exp(0.5 - 0.5Y))$, which generates 60% of x_1 with missing values. The selection probability for x_2 is 0.8. The quartic kernel, $K(u) = 15/16(1-u^2)^2 I_{(|u|\leq 1)}$, is used in this example. The kernel bandwidth is $h = 1/\sqrt[5]{n}$.

The selection probability is known or estimated. The upper half table in Table 2.2 is the results for the known selection probabilities. The lower half table in Table 2.2 is the results for the estimated selection probabilities. It is noticed that the estimators with the estimated selection probabilities are slightly more efficient than that with the known selection probabilities.

Similar results as in Table 2.1 are seen in Table 2.2. We note that bias($\bar{\beta}$)'s, $s.e.(\bar{\beta})$'s, $s.d.(\bar{\beta})$'s and MSE($\bar{\beta}$)'s are remarkably smaller than bias($\hat{\beta}$)'s, $s.e.(\hat{\beta})$'s and MSE($\hat{\beta}$)'s respectively. The CPs of $\hat{\beta}$ and $\bar{\beta}$ are close to the 95% nominal coverage rate. The $ARE(\beta_1)$'s are around 1.2 and the $ARE(\beta_2)$'s are around=3.5.

Table 2.2: Simulation two (MAR). $Y = X\beta + g(T) + \epsilon$, where $g(T) = 0$. x_1,x_2 and $t \sim U(0,1)$, $\epsilon \sim N(0,1)$, $\epsilon \sim N(0,1)$. Selection probability for x_1 is $1/(1+exp(0.5-0.5Y))$. This generates 60% of observations with missing x_1 values. Selection probability for x_2 is 0.8. 32% of observations are complete-cases.

	n=200		n=500	
	$\beta_1 = 0$	$\beta_2 = 0.5$	$\beta_1 = 0$	$\beta_2 = 0.5$
bias($\hat{\beta}$)	0.0133	0.0491	0.0091	0.0225
$s.d.(\hat{\beta})$	0.5228	0.5342	0.3346	0.3510
$s.e.(\hat{\beta})$	0.4870	0.4839	0.3150	0.3138
CP of $\hat{\beta}$	0.9220	0.9290	0.9290	0.9190
MSE($\hat{\beta}$)	0.2732	0.2875	0.1119	0.1236
bias($\bar{\beta}$)	0.0095	0.0059	0.0156	0.0063
$s.d.(\bar{\beta})$	0.4751	0.2939	0.3052	0.1856
$s.e.(\bar{\beta})$	0.4538	0.2900	0.2873	0.1781
CP of $\bar{\beta}$	0.9340	0.9330	0.9220	0.9420
MSE($\bar{\beta}$)	0.2256	0.0863	0.0933	0.0344
ARE	1.2108	3.3028	1.2023	3.5765

Table 2.2 – *Continued from previous page*

	n=200		n=500	
	$\beta_1 = 0$	$\beta_2 = 0.5$	$\beta_1 = 0$	$\beta_2 = 0.5$
bias($\hat{\beta}$)	$0.010\,8$	$0.032\,7$	$0.004\,0$	$0.033\,4$
$s.d.(\hat{\beta})$	$0.492\,0$	$0.530\,4$	$0.331\,4$	$0.335\,9$
$s.e.(\hat{\beta})$	$0.472\,7$	$0.471\,1$	$0.325\,0$	$0.323\,7$
CP of $\hat{\beta}$	$0.926\,0$	$0.912\,0$	$0.948\,0$	$0.932\,0$
MSE($\hat{\beta}$)	$0.241\,9$	$0.282\,1$	$0.109\,7$	$0.113\,8$
bias($\bar{\beta}$)	$0.000\,4$	$0.014\,4$	$0.010\,8$	$0.002\,6$
$s.d.(\bar{\beta})$	$0.464\,7$	$0.305\,7$	$0.300\,3$	$0.181\,1$
$s.e.(\bar{\beta})$	$0.444\,1$	$0.293\,6$	$0.295\,5$	$0.176\,4$
CP of $\bar{\beta}$	$0.927\,0$	$0.920\,0$	$0.942\,0$	$0.937\,0$
MSE($\bar{\beta}$)	$0.215\,7$	$0.093\,6$	$0.090\,2$	$0.032\,8$
ARE	$1.120\,8$	$3.010\,2$	$1.217\,2$	$3.441\,6$

2.3.3 Simulation three (MAR setting)

We consider the similar setting as simulation study two. In this simulation, we set $x_1, x_2 \sim U(0,1), t \sim U(-1,1)$, $\beta_1 = \beta_2 = 0, g(T) = 0$. The kernel function and bandwidth are chosen the same one as in simulation two. The selection probability for x_1

is $Pr(\delta = 1|Y, T) = \phi(0.5Y + 0.5sign(T)T^2)$, where Φ is the cumulative function of the standard normal distribution, and the selection probability for x_2 is a constant 0.8.

The simulation results are similar as that of simulation 2. We note that the CPs of $\hat{\beta}$ are slightly deviant from the 95% nominal coverage rates for size=200. The reason may be the sample size is not adequate for asymptotic properties. When sample size increases to 500, the 95% confidence coverage probabilities stay close to the nominal rates. The $ARE(\beta_1)$'s are around 1.20 and the $ARE(\beta_2)$'s are around 4.8.

Table 2.3: Simulation three (MAR). $Y = X\beta + g(T) + \epsilon$, where $g(T) = 0$. $x_1, x_2 \sim U(0,1), T \sim U(-1,1)$. Selection probability for x_1 is $\phi(0.5Y + 0.5sign(T)T^2)$. This generates 50% of observations with missing x_1 values. 44% of observations are fully observed.

	n=200		n=500	
	$\beta_1 = 0$	$\beta_2 = 0.5$	$\beta_1 = 0$	$\beta_2 = 0.5$
bias($\hat{\beta}$)	0.0366	-0.0015	0.0157	-0.0014
$s.d.(\hat{\beta})$	0.5744	0.6156	0.3733	0.3825
$s.e.(\hat{\beta})$	0.4971	0.4983	0.3435	0.3484
CP of $\hat{\beta}$	0.9200	0.9020	0.9390	0.9390
MSE($\hat{\beta}$)	0.3310	0.3786	0.1395	0.1462
bias($\bar{\beta}$)	0.0377	0.0044	0.0184	0.0037

Table 2.3 – *Continued from previous page*

	n=200		n=500	
	$\beta_1 = 0$	$\beta_2 = 0.5$	$\beta_1 = 0$	$\beta_2 = 0.5$
$s.d.(\bar{\beta})$	$0.541\,5$	$0.298\,2$	$0.342\,1$	$0.173\,1$
$s.e.(\bar{\beta})$	$0.467\,5$	$0.280\,3$	$0.315\,6$	$0.174\,4$
CP of $\bar{\beta}$	$0.920\,0$	$0.920\,0$	$0.934\,0$	$0.955\,0$
MSE($\bar{\beta}$)	$0.294\,4$	$0.088\,9$	$0.117\,2$	$0.029\,9$
ARE	$1.125\,2$	$4.260\,9$	$1.191\,0$	$4.883\,6$

2.4 Discussion

In this chapter, the performance and asymptotic properties of the UIPW estimator for the partially linear model are investigated under the MCAR and the MAR settings. The simulation results are consistent with the theory for the UIPW estimator expectedly. In constructing the estimator for the MAR setting with the selection probability depending on Y, we choose a linear model serving as a surrogate model where T is missing but X is observed. This choice increases the difficulty in obtaining the asymptotic properties of the estimator. Extra assumptions have to be made to guarantee the convergence of the estimator. The data with missing X but observed T are discarded in the estimation. These data may contain information which can be used to improve the efficiency of the proposed estimator. The future work may include finding a method to incorporate these data into our estimator.

In this study, the error terms of the partially linear models are assumed to be homoscedastic. Heteroscedastic errors for the partially linear model have been investigated in the literature (Schick, 1996; Härdle et al., 2000). For missing data in the partially linear model with heteroscedastic errors, the properties of the UIPW estimator needs further investigation. If the error terms are autocorrelated, the model becomes a partially linear model with serially correlated errors (You and Chen, 2007a; 2007b). It is a meaningful extension of the UIPW estimator to the partially linear model with autocorrelated errors.

Although the estimator of the parametric component in the partially linear model is of interest in this study and the nonparametric component is considered as a nuisance component in estimation, the estimator of the nonparametric component sometimes is also valuable. We are intrigued by the following questions: How much efficiency of the estimator for the nonparametric component is gained when using unified approach? Whether the improvement of the parametric component necessarily causes the improvement of the nonparametric component? And when the UIPW estimator works better for the nonparametric component?

Fan and Lin (2006) provided a general overview of the partially linear model with longitudinal data. It is also an interesting matter of future investigation on how the UIPW estimator performs for the partially linear model with missing longitudinal covariate data.

Chapter 3

Cox proportional hazards model with missing covariates

The Cox proportional hazards model was first introduced by Cox (1972), which assumes that the hazard rate of an individual with covariate Z has the proportional form

$$\lambda(t|Z) = \lambda_0(t)exp(\beta^{'}Z), t \geq 0, \tag{3.1}$$

where t denotes the time to an event, $\lambda_0(t)$ denotes an unknown and unspecified nonnegative function, vectors $Z \in R^p$ and $\beta \in R^p$ are the associated covariates and the unknown regression parameters respectively. Cox (1972) proposed a maximum likelihood method to obtain a consistent estimator of β for this proportional hazards model. In Cox (1975), he used a generalized partial likelihood for the proportional hazards model. The partial likelihood is an alternative choice when the full likelihood is too difficult to be computed or there are too many nuisance parameters in the likelihood function.

Cox (1972 and 1975) conjectured that the partial likelihood shares the asymptotic behaviors of a full likelihood and made reasonable assumptions for the partial likelihood. Tsiatis (1981) proved the asymptotic properties of the estimator $\hat{\beta}$ using partial likelihood. Naes (1982) simplified the proof by introducing discrete time martingale theory. Andersen and Gill (1982) used the counting process theory to show the asymptotic properties of $\hat{\beta}$.

In practice, some covariates are not always observed, either by design or by happenstance. A complete-case analysis can be performed after discarding the incomplete observations, however it is well known that the estimator obtained by using the complete-case observations may be biased and is not efficient. Horvitz-Thompson estimator (Horvitz and Thompson, 1952) removed the bias of the estimator based on the complete-case observations by using the inverse probability weights in estimation, but it still failed to improve efficiency due to failure to utilize all the available data in the analysis.

How to incorporate the incomplete data into the analysis has been a challenge to statisticians. Many methods have been developed to incorporate incomplete observations into the analysis. These methods are divided into four categories: maximum likelihood, multiple imputation, fully Bayesian and weighted estimating equations (Ibrahim et al., 2005).

Herring and Ibrahim (2001) demonstrated a likelihood-based estimator to handle the Cox model with missing covariates when the missing data mechanism is MAR. Multiple imputation is a common way to handle missing data in practice because complete data methods can be applied to the imputed data directly. White and Royston (2009) imputed the missing binary or Normal covariates by using a logistic or a linear regression model.

Wang and Chen (2001) extended the idea of Robins et al. (1994) to form an augmented estimator for the Cox model with missing covariates. Qi et al. (2005) extended the estimator of Wang and Chen (2001) through introducing the nonparametric method to estimate selection probability in estimation. Qi et al. (2010) compared the multiple imputation estimators with the augmented inverse probability weighted estimators. Their study indicated that multiple imputation may yield a consistent and more efficient estimator than the complete-case analysis if imputation models are correctly specified, however the choices of the imputation models are critical to the consistency of the estimator.

Weighted estimating equations are widely used in the Cox proportional hazards model with missing data. Pugh et al. (1994) proposed an inverse probability weighted estimating equations for the Cox regression with missing covariates under the MAR assumption. Robins et al. (1994) introduced the augmented inverse probability weighted estimating equations for parametric models. The aforementioned method was furthered to the Cox proportional hazards model by Wang and Chen (2001). The augmented inverse probability weighted estimators are robust. Correct specification of the conditional distribution of the missing data given observed data or the missingness of the data are essential for the consistency of the estimator. The augmented inverse probability weighted estimating equations produce an efficient estimator by incorporating information from the incomplete observations into the analysis, however, the estimating equations can be complicated due to the estimation of the conditional expectation given the surrogate data.

The initial idea of the unified approach proposed by Chen and Chen (2000) demonstrates its simplicity in dealing with MCAR data in a simple monotone missing data pattern, however it fails to handle data with general missingness pattern and the MAR missing mechanism. Zhao et al. (2013) extended the work of Chen and Chen (2000) to the general missing data pattern and the MAR missing data mechanism for parametric regression models. In this chapter, we will extend this idea to the Cox proportional hazards model with missing covariate data. The extended unified approach inherits the conceptual simplicity and the computational feasibility of the original method. These strengths are superior to other methods in handling the Cox proportional hazards model with missing covariates.

This chapter includes the following sections. Section 3.1 defines notations and missing data patterns used in this chapter. In Section 3.2, the UIPW estimator is described using the idea of the unified approach and the asymptotic properties of the UIPW estimator are provided and justified. In Section 3.3, numerical simulations are used to examine the performance of the UIPW estimator under different settings. Section 3.4 provides some conclusions and a discussion about further research.

3.1 Notation

First, we consider the Cox model (3.1) with missing covariates in a simple monotone pattern shown in figure 3.1. In this data pattern, the first k elements of covariate vector Z are fully observed; the last $p - k$ elements of covariate vector Z have missing values; the first m observations are complete cases and the last $n - m$ observations have missing values.

When the data are MAR and in the simple monotone missing data pattern, the unified approach proposed by Chen and Chen (2000) can be easily employed to handle it. Adopting the Chen and Chen (2000)'s idea, the first m observations are employed to perform the complete-case analysis to obtain the estimator $\hat{\beta}$ for the Cox regression model specified by the hazard function (3.1). A working model with the observed covariate data $(z_{1i}, ..., z_{ki})^{'}, i = 1, ..., n$, which omits covariates with missing values from the full model (3.1), serves as the surrogate model and its hazard function is as follows

$$\tilde{\lambda}(t|W) = \tilde{\lambda}_0(t)exp(\gamma^{'}W),$$

where $W = (z_1, ..., z_k)^{'}$ and $\gamma \in R^k$. For this surrogate model, two estimators $\hat{\gamma}$ and $\bar{\gamma}$ are obtained by employing the first m observations and all n observations in estimation respectively.

Estimators $\hat{\gamma}$ and $\bar{\gamma}$ are consistent for γ^*. Estimator $\bar{\gamma}$ is more efficient, because estimator $\bar{\gamma}$ is from a larger sample with size n, but estimator $\hat{\gamma}$ is from the complete cases with size m, where $n > m$. Statistically, it is possible for the estimator $\bar{\gamma}$ to drag the $\hat{\beta}$ close to the value β^* through the correlation between $\hat{\beta}$ and $\hat{\gamma}$. Roughly speaking, this is how the unified approach works in improving the efficiency of the estimate of β.

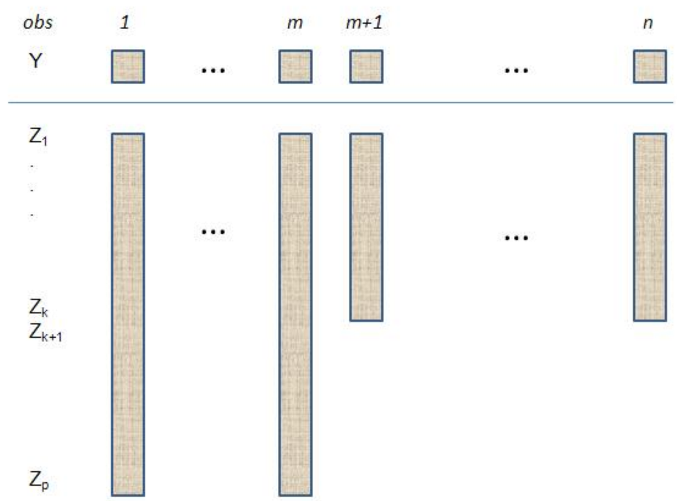

Figure 3.1: Single missing data pattern

When the data have a general missing data pattern or the missing data mechanism is not MCAR , the original unified approach cannot be applied. To improve the applicability of the unified approach, we extend the UIPW estimator (Chen and Chen, 2000; Zhao et al., 2013) to the Cox proportional hazards model with missing covariates in general missing data pattern.

In the following discussion, we consider the Cox model (3.1) with the observed covariates Z in the general missing data pattern shown in figure 3.2.

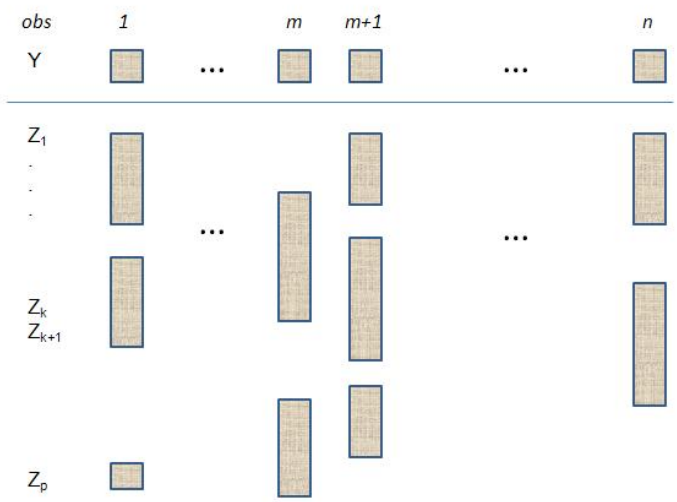

Figure 3.2: General missing data pattern

A complete-case analysis is compulsory for using the unified approach. We rearrange the observations such that the first m subjects form the complete-case data. There are q distinct covariate groups. A covariate group is a combination of the elements of covariate vector Z, which have the same missing pattern within the group. Denote the kth covariate group as $W_k = \{z_{k_1}, z_{k_2}, ..., z_{k_{p_k}}\}'$, $k = 1, ..., q$.

The missing indicator variable $M_{ki} = 1$ if the kth covariate group W_k is observed for subject i, namely W_{ki} is observed, and 0 otherwise, $i = 1, ..., n$ and $k = 1, ..., q$. If $M_{1i} = M_{2i} = ... = M_{qi} = 1$ for subject i, then we let $M_{0i} = 1$ and 0 otherwise.

We let

$$\pi_{ji} = P(M_{ji} = 1|X_i, \delta_i),$$

$$\pi_{0i} = P(M_{0i} = 1|X_i, \delta_i) \text{ and}$$

$$\pi_{kj,i} = P(M_{ji} = 1, M_{ki} = 1 | X_i, \delta_i),$$

where X_i is the event time of subject i and δ_i is the failure indicator.

The kth surrogate data contain observed values for the kth covariate group W_k. We denote the index set for kth surrogate data as $A_k = \{i : M_{ki} = 1\}$. There are n_k observations in the kth surrogate data and $n_k > m$, the number of complete-case observations.

3.2 Estimators

In this study, we focus on cases with the general missing data pattern the MAR missing mechanism. The procedure for the case with the simple monotone missing pattern and MCAR missing mechanism is omitted, because it can be treated as a special case of the general missing data pattern.

Given the observed data in Figure 3.2, the parameters of the Cox proportional hazards model (3.1) can be estimated by performing the complete-case analysis with inverse probability weights, but this method discards the incomplete observations in the estimation.

To improve the efficiency, we will extend the UIPW estimator to deal with the Cox proportional hazards model with the general missing data pattern. The UIPW estimator incorporates the incomplete observations in the estimation. It is more efficient than the one based on the complete-case analysis or the complete-case based analysis with simple inverse probability weights (SIPW).

In the Cox proportional hazards model, we let T,C and $X = min(T, C)$ be the random variables representing the failure, censored, and observed times for a subject. The failure indicator, $\delta = I(T \leq C)$, is 1 if the subject experiences failure and 0 otherwise. Let $Z(t)$ denote a set of time-dependent covariates. We assume that T and C are independent for given $Z(t)$. Define the counting process $N(t) = \delta I(X \leq t)$ and the at-risk process $Y(t) = I(X > t)$ corresponding to (X, δ). Let M be the missing-data indicator, and $M = 1$ if the corresponding element of $Z(t)$ is observed and 0 otherwise. Suppose that $(X_i, \delta_i, Z_i(t), M_i), i = 1, ..., n$, are i.i.d. replicates of $(X, \delta, Z(t), M)$. For the ith individual let the values of $Z(t)$ be $Z_i(t) = (z_{i1}(t), ..., z_{ip}(t))'$. The n individuals have lifetimes T_1,...,T_n and censored times C_1,...,C_n. An observed time $X_i = min(T_i, C_i)$ is either the lifetime or a censored time.

Counting process $N_i(t) = 0$ means that the ith individual has not experienced the failure before time t and $N_i(t) = 1$ indicates that it fails at time $t, T_i \leq t$. Failure indicator $\delta_i(t) = 1$ means that the individual is uncensored at time $t, C_i > t$ and $\delta_i(t) = 0$ means that it is censored before time $t, C_i \leq t$. We note that no tie is assumed to occur in this process. If considering ties, a small change should be made to the likelihood function. Here, we will skip this part about how to deal with the tied data. A short justification will be given in later part for dealing with ties.

In Cox (1972 and 1975), covariates were assumed to be time invariant, namely constants in the process. This is a restrictive condition in practice. Naes (1982) relaxed this restriction on the covariates to adapt more general practical applications and introduced the martingale idea in proving the asymptotic properties. Andersen and Gill (1982) extended the study using the counting process. The Cox proportional hazards model with time dependent covariates has the form

$$\lambda(t|Z) = \lambda_0(t) exp(\beta' Z(t)).$$

For the observed data in Figure 3.2, the first m observations $Z_i'(u) \in R^p, i = 1, ..., m$ are fully observed, then the likelihood function based on the complete cases has the form

$$L(\beta, t) = \prod_{i=1}^{m} \prod_{0 \le u \le t} \left[\frac{Y_i(u) exp(\beta' Z_i(u))}{\sum_{j=1}^{n} Y_j(u) exp(\beta' Z_j(u))} \right]^{\triangle N_i(u)},$$

and its score function is

$$\begin{aligned} U(\beta, t) &= \sum_{i=1}^{m} U_i(\beta, t) \\ &= \sum_{i=1}^{m} \int_0^t \left[Z_i(u) - \frac{S^{(1)}(\beta, u)}{S^{(0)}(\beta, u)} \right] dN_i(u), \end{aligned}$$

where for $l = 0, 1, 2$,

$$S^{(l)}(\beta, u) = \frac{1}{m} \sum_{j=1}^{m} Y_j(u) Z_j^{\otimes l}(u) exp(\beta' Z_j(u)),$$

with $Z_j^{\otimes 0}(u) = 1$, $Z_j^{\otimes 1}(u) = Z$, and $Z_j^{\otimes 2}(u) = Z_j(u) Z_j'(u)$ and $t = sup\{u : Pr(Y(u) = 1) > 0\}$. The solution $\hat{\beta}$ to the score equation $U(\beta, t) = 0$ is the estimator based on the complete-case observations. Under regularity conditions given in Appendix 5.4, Andersen and Gill (1982) showed that the estimator $\hat{\beta}$ is consistent for β^*, the true value of β, and asymptotically normal.

To reduce the bias, a weighted estimating equation using the inverse probabilities as the Horvitz-Thompson weights can be applied. For known selection probabilities, the weighted estimating function takes the form

$$U_\pi(\beta,t) = \sum_{i=1}^{n} \frac{M_{0i}}{\pi_{0i}} U_{\pi i}(\beta,t)$$
$$= \sum_{i=1}^{n} \frac{M_{0i}}{\pi_{0i}} \int_0^t \left[Z_i(u) - \frac{S_\pi^{(1)}(\beta,u)}{S_\pi^{(0)}(\beta,u)} \right] dN_i(u), \tag{3.2}$$

where for $l = 0, 1, 2,$

$$S_\pi^{(l)}(\beta,u) = \frac{1}{n} \sum_{j=1}^{n} \frac{M_{0i}}{\pi_{0i}} Y_j(u) Z_j^{\otimes l}(u) exp(\beta^{'} Z_j(u)),$$

and M_{0i} is the indicator for the subject i being a complete-case and π_{0i} is the probability of the subject i being a complete-case. The estimator $\hat{\beta}_\pi$ is the solution to the estimating equation $U_\pi(\beta,t) = 0$ and consistent for β^*.

To incorporate the incomplete data into the analysis, given the observed data in Figure 3.2, we propose a surrogate Cox proportional hazards model for each of the corresponding q surrogate data sets. The kth hazard function for t given covariates W_k, suggested for the kth surrogate data, is given as

$$\tilde{\lambda}(t|W_k) = \tilde{\lambda}_{0k}(t) exp(\gamma_k^{'} W_k), k = 1, ..., q,$$

where $W_k \in R^{p_k}$ is the kth covariate group. We use the kth surrogate data to compute two estimators $\hat{\gamma}_k$ and $\bar{\gamma}_k$ for γ_k, where $\hat{\gamma}_k$ is obtained from the complete-case analysis and $\bar{\gamma}$ is based on all the observed data in the kth surrogate data set.

The partial likelihood $\tilde{L}_k(\gamma_k, t)$ for the kth surrogate model based on the complete cases, is given as

$$\tilde{L}_k(\gamma_k, t) = \prod_{i=1}^{m} \prod_{0 \le u \le t} \left[\frac{Y_i(u) exp(\gamma_k^{'} W_{ki}(u))}{\sum_{j=1}^{n} Y_j(u) exp(\gamma_k^{'} W_{kj}(u))} \right]^{\triangle N_i(u)}, k = 1, ..., q,$$

and its score function is

$$\tilde{U}_k(\gamma_k, t) = \sum_{i=1}^{m} \tilde{U}_{ki}(\gamma_k, t)$$
$$= \sum_{i=1}^{m} \int_0^t \left[W_{ki}(u) - \frac{\tilde{S}_k^{(1)}(\gamma_k, u)}{\tilde{S}_k^{(0)}(\gamma_k, u)} \right] dN_i(u),$$

where for $l = 0, 1, 2,$

$$\tilde{S}_k^{(l)}(\gamma_k, u) = \frac{1}{m} \sum_{j=1}^{m} Y_j(u) W_{kj}^{\otimes l}(u) exp(\gamma_k^{'} W_{kj}(u)),$$

with $W_{kj}^{\otimes 0}(u) = 1$, $W_{kj}^{\otimes 1}(u) = W_{kj}$, and $W_{kj}^{\otimes 2}(u) = W_{kj}(u) W_j^{'}(u)$. The estimator $\hat{\gamma}_k$ is the solution to the estimating equation $\tilde{U}_k(\gamma_k, t) = 0$, and consistent for γ_k^*, $k = 1, ..., q$.

The weighted score function for the kth surrogate model is

$$\tilde{U}_{\pi k}(\gamma_k, t) = \sum_{i=1}^{n} \frac{M_{0i}}{\pi_{0i}} \tilde{U}_{\pi k i}(\gamma_k, t)$$
$$= \sum_{i=1}^{n} \int_0^t \frac{M_{0i}}{\pi_{0i}} \left[W_{ki}(u) - \frac{\tilde{S}_{\pi k}^{(1)}(\gamma_k, u)}{\tilde{S}_{\pi k}^{(0)}(\gamma_k, u)} \right] dN_i(u), \tag{3.3}$$

where for $l = 0, 1, 2,$

$$\tilde{S}_{\pi k}^{(l)}(\gamma_k, u) = \frac{1}{n} \sum_{j=1}^{n} \frac{M_{0i}}{\pi_{0i}} Y_j(u) W_{kj}^{\otimes l}(u) exp(\gamma_k^{'} W_{kj}(u)).$$

The estimators $\hat{\gamma}_{\pi k}$ is the solution to the estimating equation $\tilde{U}_{\pi k}(\gamma_k, t) = 0$ and consistent for $\gamma_k^*, k = 1, ..., q$.

The number of observations in the kth surrogate data is n_k, which is greater than the number of observations in the complete cases m, namely $n_k > m$. The above equations employ only the complete cases in the analyses. To utilize all the n_k observations in the kth surrogate data set to estimate γ_k for the kth surrogate model, the weighted estimating equation takes form:

$$\bar{U}_{\pi k}(\gamma_k, t) = \sum_{i=1}^{n} \frac{M_{ki}}{\pi_{ki}} \bar{U}_{\pi ki}(\gamma_k, t)$$
$$= \sum_{i=1}^{n} \int_0^t \frac{M_{ki}}{\pi_{ki}} \left[W_{ki}(u) - \frac{\bar{S}_{\pi k}^{(1)}(\gamma_k, u)}{\bar{S}_{\pi k}^{(0)}(\gamma_k, u)} \right] dN_i(u), \tag{3.4}$$

where for $l = 0, 1, 2$

$$\bar{S}_{\pi k}^{(l)}(\gamma_k, u) = \frac{1}{n} \sum_{j=1}^{n} \frac{M_{ki}}{\pi_{ki}} Y_j(u) W_{kj}^{\otimes l}(u) exp(\gamma_k' W_{kj}(u)).$$

Solutions to the estimating equation $\bar{U}_{\pi k}(\gamma_k, t) = 0$ is the estimator $\bar{\gamma}_{\pi k}, k = 1, ..., q$.

The q surrogate models, suggested for the q surrogate data sets, serve as working models in the analyses. These surrogate models are generated from the full model by omitting partial covariates. Although the surrogate model can be chosen arbitrarily, we choose the model with omitted variables in this study, which may lessen the assumptions for the specified models. The effects of omitting covariates in the Cox proportional hazards model have been investigated by statisticians and can be found in the literature (Bretagnolle and Huber-Carol, 1988; Morgan et al., 1986; Struthers and Kalbfleisch, 1986; Lagakos and Schoenfeld, 1984; O'Neill, 1986). Under the assumptions in Appendix (5.4), it has been shown that, even when the true model $\lambda(t|Z(t)) = \lambda_0(t)exp(\beta' Z(t))$ is misspecified as $\tilde{\lambda}(t|W_k(t)) = \tilde{\lambda}_{0k}(t)exp(\gamma_k' W_k(t))$, which omits partial covariates from the full model, the asymptotic properties of the estimator $\hat{\gamma}_k$ still hold.

Suppose that the assumptions in Appendix (5.4) given by Andersen and Gill (1982) hold for both the full model and the surrogate models. The estimators $\hat{\beta}$, $\hat{\gamma}_k$ and $\bar{\gamma}_k, k = 1, ..., q$, are the solutions to the estimating equations (3.2), (3.3) and (3.4) respectively. The asymptotic properties, consistency and normality, of the estimators $\hat{\beta}$, $\hat{\gamma}$ and $\bar{\gamma}$ are sustained under the assumptions, however the asymptotic joint normality of $(\hat{\beta}'$, $\hat{\gamma}_1'$,...,$\hat{\gamma}_q')'$ remains unproven. It is reasonable to conjecture that it is asymptotically normal. If extra conditions are given such that the joint random sequence $(U(\hat{\beta},t)', \tilde{U}(\hat{\gamma}_1,t)', ..., \tilde{U}(\hat{\gamma}_q,t)')'$ satisfies Lindeberg-Feller Multivariate Central Limit Theorem, then the asymptotic joint normality of $(\hat{\beta}'$, $\hat{\gamma}_1'$,...,$\hat{\gamma}_q')'$ is obtained. We will skip this step and assume the asymptotic joint normality of $\hat{\beta}$, $\hat{\gamma}_1$,...,$\hat{\gamma}_q$ holds. Similar results are presumptively held for $\hat{\beta}_\pi$, $\hat{\gamma}_{\pi 1}$,...,$\hat{\gamma}_{\pi q}$.

Chen and Chen (2000) proposed an estimator for β based on the generalized estimating equations with the observed data in simple monotone missing data pattern shown in Figure 3.1 and MCAR mechanism, which incorporated the incomplete observations into the analysis. Due to the simple monotone missing data pattern restriction, the original approach can only incorporate one extra estimator $\hat{\gamma}_1$ from the surrogate model into the unified estimator $\bar{\beta}$. The form of the proposed estimator is

$$\bar{\beta} = \hat{\beta} - \hat{D}_1^{-1}\hat{C}_{12}\hat{C}_{22}^{-1}\hat{D}_2(\hat{\gamma}_1 - \bar{\gamma}_1),$$

where $\hat{D}_1, \hat{D}_2, \hat{C}_{12}, \hat{C}_{22}$ are estimates of D_1, D_2, C_{12} and C_{22} respectively, with $D_1 \equiv E[\partial S(\beta^*)/\partial\beta]$, $D_2 \equiv E[\partial\tilde{S}(\gamma_1^*)/\partial\gamma_1]$ and $C \equiv E\left[\begin{pmatrix} S(\beta^*) \\ \tilde{S}(\gamma_1^*) \end{pmatrix} \left(S'(\beta^*), \tilde{S}'(\gamma_1^*)\right)\right]$, and $S(\beta)$ and $\tilde{S}(\gamma_1)$ are the generalized estimating equations for β and γ_1 respectively.

Zhao et al. (2013) not only extended the unified approach (Chen and Chen, 2000) from the simple monotone missing data pattern to general missing data pattern, but also relaxed the MCAR restriction to the MAR for parametric regression models.

We replace the estimating equations $S(\beta)$ and $\tilde{S}(\gamma)$ in (Zhao et al., 2013) with the score functions $U_\pi(\beta,t)$ and $\tilde{U}_\pi(\gamma,t)$ respectively, where $\tilde{U}_\pi(\gamma,t) = (\tilde{U}'_{\pi 1}(\gamma_1,t),...,\tilde{U}'_{\pi q}(\gamma_q,t))'$. The diagonal matrix $D = (D_{\pi 00}, D_{\pi 11})$, with $D_{\pi 00} = E[\partial U_\pi(\beta,t)/\partial\beta]$ and $D_{\pi 11} = E[\partial\tilde{U}_\pi(\gamma,t)/\partial\gamma]$. Let $C = \begin{pmatrix} C_{\pi 00} & C_{\pi 01} \\ C_{\pi 10} & C_{\pi 11} \end{pmatrix}$ with $C_{\pi 00} = E[U_\pi(\beta,t)U'_\pi(\beta,t)]$, $C_{\pi 01} = E[U_\pi(\beta,t)\tilde{U}'_\pi(\gamma,t)]$, and $C_{\pi 11} = E[\tilde{U}_\pi(\gamma,t)\tilde{U}'_\pi(\gamma,t)]$.

Denote that $\hat{\gamma}_\pi = (\hat{\gamma}'_{\pi 1},...,\hat{\gamma}'_{\pi q})'$, $\bar{\gamma}_\pi = (\bar{\gamma}'_{\pi 1},...,\bar{\gamma}'_{\pi q})'$, and $\gamma^* = (\gamma_1^{*'},...,\gamma_q^{*'})'$. Under the given regularity conditions, $\hat{\beta}_\pi$ and $\hat{\gamma}_\pi$ are consistent for β^* and γ^*, and $\sqrt{n}(\hat{\beta}'_\pi - \beta^{*'}, \hat{\gamma}'_\pi - \gamma^{*'})'$ converges to an asymptotic normal distributions with mean 0 and variance $D^{-1}CD^{-1}$. The conditional distribution of $\sqrt{n}(\hat{\beta}_\pi - \beta^*)$ given $\sqrt{n}(\hat{\gamma}_\pi - \gamma^*)$ is asymptotic normal with mean $D^{-1}_{\pi 00}C_{\pi 01}C^{-1}_{\pi 11}D_{\pi 11}(\hat{\gamma}_\pi - \gamma^*)$. The UIPW estimator for the Cox proportional hazards model with missing covariates takes the form

$$\bar{\beta}_\pi = \hat{\beta}_\pi - \hat{D}^{-1}_{\pi 00}\hat{C}_{\pi 01}\hat{C}^{-1}_{\pi 11}\hat{D}_{\pi 11}(\hat{\gamma}_\pi - \bar{\gamma}_\pi).$$

The proposed estimator $\bar{\beta}_\pi$ is consistent and asymptotic normal. The consistency of $\bar{\beta}_\pi$ is obtained by the uniform law of large numbers and the consistencies of $\hat{\beta}_\pi$, $\hat{\gamma}_\pi$ and $\bar{\gamma}_\pi$. Its asymptotic normality is obtained by asymptotic normality of $\hat{\beta}_\pi$, $\hat{\gamma}_\pi$ and $\bar{\gamma}_\pi$, and by Slutsky's theorem. It is easy to show that the $\bar{\beta}_\pi$ is asymptotically more efficient than $\hat{\beta}_\pi$ in term of variance.

The variance $var(\sqrt{n}(\bar{\beta}_\pi - \beta^*))$ can be estimated by

$$\begin{aligned}\widehat{var}(\sqrt{n}(\bar{\beta}_\pi - \beta^*)) =& \hat{D}^{-1}_{\pi 00}\hat{C}_{\pi 00}\hat{D}^{-1}_{\pi 00} \\ & -\hat{D}^{-1}_{\pi 00}\hat{C}_{\pi 01}\hat{C}^{-1}_{\pi 11}\left(\hat{C}_{\pi 11} - \hat{C}_{\pi 22} + \hat{C}_{\pi 12} + \hat{C}'_{\pi 12}\right)\hat{C}^{-1}_{\pi 11}\hat{C}'_{\pi 01}\hat{D}^{-1}_{\pi 00} \\ & +\hat{D}^{-1}_{\pi 00}\left(\hat{C}_{\pi 01}\hat{C}^{-1}_{\pi 11}\hat{C}'_{\pi 02} + \hat{C}_{\pi 02}\hat{C}^{-1}_{\pi 11}\hat{C}'_{\pi 01}\right)\hat{D}^{-1}_{\pi 00},\end{aligned}$$

where $\hat{D}_{\pi 00}$, $\hat{C}_{\pi 00}$, $\hat{C}_{\pi 01}$, $\hat{C}_{\pi 11}$, $\hat{C}_{\pi 22}$, $\hat{C}_{\pi 12}$ and $\hat{C}_{\pi 02}$ are estimates of $D_{\pi 00}$, $C_{\pi 00}$, $C_{\pi 01}$, $C_{\pi 11}$, $C_{\pi 22}$, $C_{\pi 12}$ and $C_{\pi 02}$ respectively.

We adopt the robust variance formulae (Lin and Wei, 1989) for the Cox proportional hazards model in the computation of the estimators. Matrices $\hat{C}_{00}$, $\hat{C}_{11}$ and $\hat{C}_{22}$ can be derived explicitly from the formulae given by Lin and Wei (1989). The matrices $\hat{C}_{01}$, $\hat{C}_{02}$ and $\hat{C}_{12}$ are defined in the following part. Denote $\bar{Z}(\hat{\beta}_\pi, t) = \frac{S_\pi^{(1)}(\hat{\beta}_\pi, X_i)}{S_\pi^{(0)}(\hat{\beta}_\pi, X_i)}$ and $\tilde{Z}_k(\hat{\gamma}_{\pi k}, t) = \frac{\tilde{S}_{\pi k}^{(1)}(\hat{\gamma}_{\pi k}, t)}{\tilde{S}_{\pi k}^{(0)}(\hat{\gamma}_{\pi k}, t)}, k = 1, ..., q$. The score residuals $L_{\pi i}^{(0)}(\hat{\beta}_\pi)$, $L_{\pi i}^{(1)}(\hat{\gamma}_\pi) = (L_{\pi 1i}^{(1)'}(\hat{\gamma}_{\pi 1}), ..., L_{\pi qi}^{(1)'}(\hat{\gamma}_{\pi q}))'$ and $L_{\pi i}^{(2)}(\bar{\gamma}_\pi) = (L_{\pi 1i}^{(2)'}(\bar{\gamma}_{\pi l}), ..., L_{\pi qi}^{(2)'}(\bar{\gamma}_{\pi q}))'$ are defined as follows:

$$\begin{aligned} L_{\pi i}^{(0)}(\hat{\beta}_\pi) =& \frac{M_{0i}}{\pi_{0i}}\Bigg[\delta_i\bigg(Z_i(X_i) - \bar{Z}(\hat{\beta}_\pi, X_i)\bigg) - \\ & \sum_{j=1}^{n} \frac{\delta_j Y_i(X_j) exp(\hat{\beta}_\pi' Z_i(X_j))}{\pi_{0j} S_\pi^{(0)}(\hat{\beta}_\pi, X_j)}\bigg(Z_i(X_j) - \bar{Z}(\hat{\beta}_\pi, X_j)\bigg)\Bigg], \\ L_{\pi ki}^{(1)}(\hat{\gamma}_{\pi k}) =& \frac{M_{0i}}{\pi_{0i}}\Bigg[\delta_i\bigg(W_{ki}(X_i) - \tilde{Z}_k(\hat{\gamma}_{\pi k}, X_i)\bigg) - \\ & \sum_{j=1}^{n} \frac{\delta_j Y_i(X_j) exp(\hat{\gamma}_{\pi k}' W_{ki}(X_j))}{\pi_{0j} S_\pi^{(0)}(\hat{\gamma}_{\pi k}, X_j)}\bigg(W_{ki}(X_j) - \tilde{Z}_k(\hat{\gamma}_{\pi k}, X_j)\bigg)\Bigg], k = 1, ..., q, \\ L_{\pi ki}^{(2)}(\bar{\gamma}_{\pi k}) =& \frac{M_{ki}}{\pi_{ki}}\Bigg[\delta_i\bigg(W_{ki}(X_i) - \tilde{Z}_k(\bar{\gamma}_{\pi k}, X_i)\bigg) - \\ & \sum_{j=1}^{n} \frac{\delta_j Y_i(X_j) exp(\bar{\gamma}_{\pi k}' W_{ki}(X_j))}{\pi_{kj} S_\pi^{(0)}(\bar{\gamma}_{\pi k}, X_j)}\bigg(W_{ki}(X_j) - \tilde{Z}_k(\bar{\gamma}_{\pi k}, X_j)\bigg)\Bigg], k = 1, ..., q. \end{aligned}$$

Use the above notations, we can compute the matrices of variance and covariance as follows:

$$\begin{aligned} \hat{D}_{\pi 00} &= \frac{1}{n}\sum_{i=1}^{n} \frac{M_{0i}\delta_i}{\pi_{0i}}\left(\frac{S_\pi^{(2)}(\hat{\beta}_\pi, X_i)}{S_\pi^{(0)}(\hat{\beta}_\pi, X_i)} - \frac{S_\pi^{(1)}(\hat{\beta}_\pi, X_i) \otimes S_\pi^{(1)}(\hat{\beta}_\pi, X_i)}{(S_\pi^{(0)}(\hat{\beta}_\pi, X_i))^2}\right), \\ \hat{D}_{\pi 11} &= \frac{1}{n}\sum_{i=1}^{n} \frac{M_{0i}\delta_i}{\pi_{0i}}\left(\frac{\tilde{S}_\pi^{(2)}(\hat{\beta}_\pi, X_i)}{\tilde{S}_\pi^{(0)}(\hat{\beta}_\pi, X_i)} - \frac{\tilde{S}_\pi^{(1)}(\hat{\beta}_\pi, X_i) \otimes \tilde{S}_\pi^{(1)}(\hat{\beta}_\pi, X_i)}{(\tilde{S}_\pi^{(0)}(\hat{\beta}_\pi, X_i))^2}\right), \end{aligned}$$

$$\hat{C}_{\pi 00} = \frac{1}{n}\sum_{i=1}^{n} L_{\pi i}^{(0)}(\hat{\beta}_\pi) L_{\pi i}^{(0)'}(\hat{\beta}_\pi),$$

$$\hat{C}_{\pi 01} = \frac{1}{n}\sum_{i=1}^{n} L_{\pi i}^{(0)}(\hat{\beta}_\pi) L_{\pi i}^{(1)'}(\hat{\gamma}_\pi),$$

$$\hat{C}_{\pi 02} = \frac{1}{n}\sum_{i=1}^{n} L_{\pi i}^{(0)}(\hat{\beta}_\pi) L_{\pi i}^{(2)'}(\bar{\gamma}_\pi),$$

$$\hat{C}_{\pi 11} = \frac{1}{n}\sum_{i=1}^{n} L_{\pi i}^{(1)}(\hat{\gamma}_\pi) L_{\pi i}^{(1)'}(\hat{\gamma}_\pi),$$

$$\hat{C}_{\pi 12} = \frac{1}{n}\sum_{i=1}^{n} L_{\pi i}^{(1)}(\hat{\gamma}_\pi) L_{\pi i}^{(2)'}(\bar{\gamma}_\pi) \text{ and}$$

$$\hat{C}_{\pi 22} = \frac{1}{n}\sum_{i=1}^{n} L_{\pi i}^{(2)}(\bar{\gamma}_\pi) L_{\pi i}^{(2)'}(\bar{\gamma}_\pi).$$

The selection probabilities are assumed to be known in the above discussion. The method can be extended to deal with the cases where the missing data probabilities are unknown but parametrically estimable. Robins et al. (1995) showed the estimator with the estimated selection probabilities are in general more efficient than that using the true selection probabilities.

Assume the selection probabilities for the complete cases and the kth surrogate data are $\pi_i(\alpha_0) = P(M_{0i} = 1|X_i, \delta_i, \alpha_0)$ and $\pi_i(\alpha_k) = P(M_{ki} = 1|X_i, \delta_i, \alpha_k), k = 1, ..., q$, respectively. Denote the corresponding estimating equations as $\sum_i^n H_{0i}(\alpha_0) = 0$ and $\sum_i^n H_{ki}(\alpha_k) = 0, k = 1, ..., q$. The estimates $\hat{\alpha}_0$ and $\hat{\alpha}_k, k = 1, ..., q$ are solutions to $\sum_i^n H_{0i}(\alpha_0) = 0$ and $\sum_i^n H_{ki}(\alpha_k) = 0$ respectively. Denote $\alpha = (\alpha_1^{'}, ..., \alpha_q^{'})'$ and $H_i(\alpha) = (H_{1i}^{'}(\alpha_1), ..., H_{qi}^{'}(\alpha_q))'$. Replacing the selection probabilities π_{ki} in the likelihood functions

(3.2), (3.3) and (3.4) with the estimated selection probabilities $\pi_{ki}(\hat{\alpha}_k), k = 0, ..., q$, we obtain the estimates $\hat{\beta}_{\hat{\pi}}$, $\hat{\gamma}_{\hat{\pi}}$ and $\bar{\gamma}_{\hat{\pi}}$. Following Robins et al. (1995), we denote the residual for subject i from the least squares regression of A_i on B_i, $i = 1, ..., n$, as

$$\hat{R}es(A_i, B_i) = A_i - \bigg(\sum_{i=1}^{n} A_i B_i^{'}\bigg)\bigg(\sum_{i=1}^{n} B_i B_i^{'}\bigg)^{-1} B_i.$$

The UIPW estimator with the estimated selection probabilities is given as

$$\bar{\beta}_{\hat{\pi}} = \hat{\beta}_{\hat{\pi}} - \hat{D}_{\hat{\pi}00}^{-1}\hat{C}_{\hat{\pi}01}\hat{C}_{\hat{\pi}11}^{-1}\hat{D}_{\hat{\pi}11}(\hat{\gamma}_{\hat{\pi}} - \bar{\gamma}_{\hat{\pi}}),$$

where

$$\begin{aligned}
\hat{D}_{\hat{\pi}00} =& \frac{1}{n}\sum_{i=1}^{n}\frac{M_{0i}\delta_i}{\pi_{0i}(\hat{\alpha}_0)}\bigg(\frac{S_{\hat{\pi}}^{(2)}(\hat{\beta}_{\hat{\pi}}, X_i)}{S_{\hat{\pi}}^{(0)}(\hat{\beta}_{\hat{\pi}}, X_i)} - \frac{S_{\hat{\pi}}^{(1)}(\hat{\beta}_{\hat{\pi}}, X_i) \otimes S_{\hat{\pi}}^{(1)}(\hat{\beta}_{\hat{\pi}}, X_i)}{(S_{\hat{\pi}}^{(0)}(\hat{\beta}_{\hat{\pi}}, X_i))^2}\bigg),\\
\hat{D}_{\hat{\pi}11} =& \frac{1}{n}\sum_{i=1}^{n}\frac{M_{0i}\delta_i}{\hat{\pi}_{0i}}\bigg(\frac{\tilde{S}_{\hat{\pi}}^{(2)}(\hat{\beta}_{\hat{\pi}}, X_i)}{\tilde{S}_{\hat{\pi}}^{(0)}(\hat{\beta}_{\hat{\pi}}, X_i)} - \frac{\tilde{S}_{\hat{\pi}}^{(1)}(\hat{\beta}_{\hat{\pi}}, X_i) \otimes \tilde{S}_{\hat{\pi}}^{(1)}(\hat{\beta}_{\hat{\pi}}, X_i)}{(\tilde{S}_{\hat{\pi}}^{(0)}(\hat{\beta}_{\hat{\pi}}, X_i))^2}\bigg),\\
\hat{C}_{\hat{\pi}01} =& \frac{1}{n}\sum_{i=1}^{n}\hat{R}es\bigg(L_{\hat{\pi}i}^{(0)}(\hat{\beta}_{\hat{\pi}}), H_{0i}(\hat{\alpha}_0)\bigg)\hat{R}es^{'}\bigg(L_{\hat{\pi}i}^{(1)}(\hat{\gamma}_{\hat{\pi}}), H_{0i}(\hat{\alpha}_0)\bigg),
\end{aligned}$$

and $S_{\hat{\pi}}^{(2)}(\hat{\beta}_{\hat{\pi}}, X_i)$, $S_{\hat{\pi}}^{(1)}(\hat{\beta}_{\hat{\pi}}, X_i)$, $S_{\hat{\pi}}^{(0)}(\hat{\beta}_{\hat{\pi}}, X_i)$, $\tilde{S}_{\hat{\pi}}^{(2)}(\hat{\beta}_{\hat{\pi}}, X_i)$, $\tilde{S}_{\hat{\pi}}^{(1)}(\hat{\beta}_{\hat{\pi}}, X_i)$, $\tilde{S}_{\hat{\pi}}^{(0)}(\hat{\beta}_{\hat{\pi}}, X_i)$, $L_{\hat{\pi}i}^{(0)}(\hat{\beta}_{\hat{\pi}})$, and $L_{\hat{\pi}i}^{(1)}(\hat{\gamma}_{\hat{\pi}})$ are obtained by replacing $\pi_{ki}, k = 0, ..., q$, with $\pi_{ki}(\hat{\alpha}_k)$ in the formulae for $S_{\pi}^{(2)}(\hat{\beta}_{\pi}, X_i)$, $S_{\pi}^{(1)}(\hat{\beta}_{\pi}, X_i)$, $S_{\pi}^{(0)}(\hat{\beta}_{\pi}, X_i)$, $\tilde{S}_{\pi}^{(2)}(\hat{\beta}_{\pi}, X_i)$, $\tilde{S}_{\pi}^{(1)}(\hat{\beta}_{\pi}, X_i)$, $\tilde{S}_{\pi}^{(0)}(\hat{\beta}_{\pi}, X_i)$, $L_{\pi i}^{(0)}(\hat{\beta}_{\pi})$, and $L_{\pi i}^{(1)}(\hat{\gamma}_{\pi})$ respectively.

The variance of $\sqrt{n}(\bar{\beta}_{\hat{\pi}} - \beta^*)$ is estimated by

$$\begin{aligned}
\widehat{var}(\sqrt{n}(\bar{\beta}_{\hat{\pi}} - \beta^*)) =& \hat{D}_{\hat{\pi}00}^{-1}\hat{C}_{\hat{\pi}00}\hat{D}_{\hat{\pi}00}^{-1}\\
&- \hat{D}_{\hat{\pi}00}^{-1}\hat{C}_{\hat{\pi}01}\hat{C}_{\hat{\pi}11}^{-1}\bigg(\hat{C}_{\hat{\pi}11} - \hat{C}_{\hat{\pi}22} + \hat{C}_{\hat{\pi}12} + \hat{C}_{\hat{\pi}12}^{'}\bigg)\hat{C}_{\hat{\pi}11}^{-1}\hat{C}_{\hat{\pi}01}^{'}\hat{D}_{\hat{\pi}00}^{-1}\\
&+ \hat{D}_{\hat{\pi}00}^{-1}\bigg(\hat{C}_{\hat{\pi}01}\hat{C}_{\hat{\pi}11}^{-1}\hat{C}_{\hat{\pi}02}^{'} + \hat{C}_{\hat{\pi}02}\hat{C}_{\hat{\pi}11}^{-1}\hat{C}_{\hat{\pi}01}^{'}\bigg)\hat{D}_{\hat{\pi}00}^{-1},
\end{aligned}$$

where

$$\hat{C}_{\hat{\pi}00} = \frac{1}{n}\sum_{i=1}^{n}\hat{Res}\Big(L_{\hat{\pi}i}^{(0)}(\hat{\beta}_{\hat{\pi}}), H_{0i}(\hat{\alpha}_0)\Big)\hat{Res}'\Big(L_{\hat{\pi}i}^{(0)}(\hat{\beta}_{\hat{\pi}}), H_{0i}(\hat{\alpha}_0)\Big),$$
$$\hat{C}_{\hat{\pi}02} = \frac{1}{n}\sum_{i=1}^{n}\hat{Res}\Big(L_{\hat{\pi}i}^{(0)}(\hat{\beta}_{\hat{\pi}}), H_{0i}(\hat{\alpha}_0)\Big)\hat{Res}'\Big(L_{\hat{\pi}i}^{(2)}(\bar{\gamma}_{\hat{\pi}}), H_{i}(\hat{\alpha})\Big),$$
$$\hat{C}_{\hat{\pi}11} = \frac{1}{n}\sum_{i=1}^{n}\hat{Res}\Big(L_{\hat{\pi}i}^{(1)}(\hat{\gamma}_{\hat{\pi}}), H_{0i}(\hat{\alpha}_0)\Big)\hat{Res}'\Big(L_{\hat{\pi}i}^{(1)}(\hat{\gamma}_{\hat{\pi}}), H_{0i}(\hat{\alpha}_0)\Big),$$
$$\hat{C}_{\hat{\pi}12} = \frac{1}{n}\sum_{i=1}^{n}\hat{Res}\Big(L_{\hat{\pi}i}^{(1)}(\hat{\gamma}_{\hat{\pi}}), H_{0i}(\hat{\alpha}_0)\Big)\hat{Res}'\Big(L_{\hat{\pi}i}^{(2)}(\bar{\gamma}_{\hat{\pi}}), H_{i}(\hat{\alpha})\Big) \text{ and}$$
$$\hat{C}_{\hat{\pi}22} = \frac{1}{n}\sum_{i=1}^{n}\hat{Res}\Big(L_{\hat{\pi}i}^{(2)}(\bar{\gamma}_{\hat{\pi}}), H_{i}(\hat{\alpha})\Big)\hat{Res}'\Big(L_{\hat{\pi}i}^{(2)}(\bar{\gamma}_{\hat{\pi}}), H_{i}(\hat{\alpha})\Big).$$

If the missing data mechanism is MCAR, $\pi_{01} = ...\pi_{0n} = \pi_0$, $\pi_{l1} = ...\pi_{ln} = \pi_l$ and $\pi_{lj,1} = ...\pi_{lj,n} = \pi_{lj}$, for all $k, j = 1, ..., q$, then the variance formula for $\bar{\beta}_\pi$ can be simplified as

$$\widehat{var}(\sqrt{n}(\bar{\beta}_\pi - \beta^*)) = D_{00}^{-1}C_{00}D_{00}^{-1} - D_{00}^{-1}C_{01}(I - \pi_0 C_{11}^{-1}(C_{11} \circ V))C_{11}^{-1}C_{01}^{'}D_{00}^{-1},$$

where symbol $\circ$ denote Hadamard product of matrix and

$$V = \begin{pmatrix} \pi_{11}/(\pi_1\pi_1) & ... & \pi_{1q}/(\pi_1\pi_q) \\ ... & ... & ... \\ \pi_{q1}/(\pi_q\pi_1) & ... & \pi_{qq}/(\pi_q\pi_q) \end{pmatrix}.$$

This formula may be further simplified. For example, if the missing data pattern is the same as the one shown in Figure 3.1 and the data are MCAR and the number of missing data pattern $q = 1$, then the variance of $\bar{\beta}$ can be estimated by

$$\widehat{var}(\sqrt{n}(\bar{\beta}_\pi - \beta^*)) = D_{00}^{-1}C_{00}D_{00}^{-1} - (1-\rho)D_{00}^{-1}C_{01}C_{11}^{-1}C_{01}^{'}D_{00}^{-1},$$

where $\rho = m/n$. m is the number of the complete-case observations and n is the sample size. This formula takes the exactly same form as that in Chen and Chen (2000).

To deal with the data with ties we use Breslow method (Breslow, 1974) to approximate the likelihood function. If the number of failures at each distinct survival time is relatively small compared to the number of subjects at risk, then Breslow's approximation is very close to the exact partial likelihood, otherwise, the approximation can be poor.

3.3 Simulation study

We conduct simulations to investigate the behaviors of the proposed UIPW estimator. These simulations are divided into three categories. Simulation scenario 1 is designed to investigate the performance of the UIPW estimator under different settings, which include combinations of several missing data mechanisms, parameter values, covariate distributions and correlation coefficients between covariates. Simulation scenario 2 compares the UIPW estimator with the estimator suggested by Qi et al. (2005). We consider both the simple monotone missing data pattern and the general missing data pattern. Simulation scenario 3 simulates data with the time-dependent covariates. This setting is designed to investigate the properties of the UIPW estimator with the time-dependent covariates.

Simulation results are reported in Tables 3.1-3.5. In the tables, $bias(\hat{\beta})$, $s.d.(\hat{\beta})$, $s.e.(\hat{\beta})$, CP of $\hat{\beta}$ and $MSE(\hat{\beta})$ are for the estimates based on the complete-case analysis by using the SIPW method; $bias(\bar{\beta})$, $s.d.(\bar{\beta})$, $s.e.(\bar{\beta})$, CP of $\bar{\beta}$ and $MSE(\bar{\beta})$ are for the UIPW estimator. $bias(\hat{\beta})$ and $bias(\bar{\beta})$ are the averages of $\hat{\beta} - \beta_0$ and $\bar{\beta} - \beta_0$ respectively; $s.d.(\hat{\beta})$

and $s.d.(\bar{\beta})$ are the standard deviation of $\hat{\beta}$ and $\bar{\beta}$ respectively; $s.e.(\hat{\beta})$ and $s.e.(\bar{\beta})$ are the averages of the square roots of the estimated variances of $\hat{\beta}$ and $\bar{\beta}$ respectively; CP of $\hat{\beta}$ and CP of $\bar{\beta}$ are the coverage probabilities of the 95% confidence intervals of β for the estimators $\hat{\beta}$ and $\bar{\beta}$ respectively; $MSE(\hat{\beta})$ and $MSE(\bar{\beta})$ are the averages of the squared differences $\hat{\beta} - \beta$ and $\bar{\beta} - \beta$; ARE is the asymptotic relative efficiency of $\bar{\beta}$ compared to $\hat{\beta}$.

3.3.1 Simulation scenario 1

These simulations are designed to show the performance of the the estimators under the MAR and the MCAR settings with different parameter values, covariate distributions and correlation coefficients between the covariates. We consider a Cox proportional hazards model with the proportional hazard function $\lambda(t|Z) = \lambda_0(t)exp(\beta_0^{'}Z)$, where covariates z_1 and z_2 are either independent or correlated with correlation coefficients $r = 0.3$, $r = 0.4$ or $r = 0.5$.

The missing data mechanisms are MCAR or MAR. The parameters $(\beta_{01}, \beta_{02})^{'}$ take three pairs of values $(1,1)^{'}$, $(1,4)^{'}$ and $(1,9)^{'}$. We set the baseline hazard rate $\lambda_0 = 0.4$. Censored times are generated from an exponential distribution with mean $\mu = 0.2$.

Simulation results for the MCAR and the MAR setting are presented in Tables 3.1 and 3.2 respectively.

3.3.1.1 MCAR setting

Under the MCAR setting, we consider two cases. The first case is that $z_1 \sim U(0,1)$ and $z_2 \sim N(1,1)$ are independent. The second case is that $z_1 \sim N(r, r^2)$ and $z_2 \sim N(r,1)$ are correlated with correlation coefficient $r = 0.3, 0.4$ or 0.5. The missingnesses of z_1 and z_2 are independent with the selection probabilities 0.4 and 0.9 respectively. This yields 36% of cohort members to be fully observed. The censored time has an exponential distribution with mean $\mu = 0.2$, which censors $35\% \sim 45\%$ and $45\% \sim 55\%$ of cohort members for independent and correlated covariates setting respectively. The sample size is 250 and the simulation results are based on 1000 replications.

In Table 3.1, we see that almost all the biases, the standard deviations and the theoretical standard errors of $\bar{\beta}$ are noticeably smaller than that of $\hat{\beta}$ for all the parameter values of β_0. The 95% CPs of $\bar{\beta}$ are close to the nominal value. We see that in most cases AREs > 1, which indicates that the UIPW estimators are more efficient than the SIPW estimators. The AREs of β_1 at parameter values $(1,4)'$ and $(1,9)'$ are numerically equal to 1. This suggests that in these cases there is no improvement in $\bar{\beta}$ compared to $\hat{\beta}$. The reason for this is that in these cases 36% of the observed z_1 are employed in the complete-case analysis, but only 4% observations are used to compensate the loss of information for z_1 in the UIPW estimators.

For the setting with correlated covariates, as correlation coefficient increases the censored rates and the parameters in the distributions for z_1 and z_2 change as well. Simulation results for different correlation coefficients are not directly comparable. If we just consider the $AREs$, no significant difference is found among these simulation results.

As the value of parameter β_0 changes from $(1,1)'$ to $(1,4)'$ and $(1,9)'$, noticeable increases of the biases, $s.d.$'s and $s.e.$'s for both the SIPW estimator and the UIPW estimator are observed. This is partially because β_2 is overwhelmingly larger than β_1, as a result, the term $z_2\beta_2$ becomes the dominate term in the hazard function. It is believed that if these two terms make similar contributions to regression model, the performances of the UIPW estimator for β_1 and β_2 are supposed to be similar, and vice versa. The example with the parameter value $\beta_0 = (1,1)'$ always show good performances in all settings.

Table 3.1: Simulation results under the MCAR setting.

	β_1=1	$\beta_2 = 1$	$\beta_1 = 1$	$\beta_2 = 4$	$\beta_1 = 1$	$\beta_2 = 9$
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
			Independent covariates			
bias($\hat{\beta}$)	0.013 3	0.018 2	0.030 7	0.090 3	0.053 0	0.233 9
$s.d.(\hat{\beta})$	0.508 4	0.167 3	0.484 6	0.433 4	0.493 0	1.018 3
$s.e.(\hat{\beta})$	0.466 5	0.164 0	0.427 3	0.396 2	0.445 2	0.902 5
CP of $\hat{\beta}$	0.928 0	0.943 0	0.925 0	0.935 0	0.923 0	0.925 0
MSE($\hat{\beta}$)	0.258 4	0.028 3	0.235 6	0.195 8	0.245 6	1.090 5
bias($\bar{\beta}$)	0.019 1	0.012 8	0.029 9	0.054 2	0.050 5	0.137 3
$s.d.(\bar{\beta})$	0.493 8	0.115 4	0.486 1	0.290 5	0.493 3	0.657 4

Continued on next page

	β_1=1	$\beta_2 = 1$	$\beta_1 = 1$	$\beta_2 = 4$	$\beta_1 = 1$	$\beta_2 = 9$
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
$s.e.(\bar{\beta})$	0.4535	0.1194	0.4228	0.2947	0.4427	0.7008
CP of $\bar{\beta}$	0.9300	0.9410	0.9190	0.9450	0.9240	0.9490
MSE($\bar{\beta}$)	0.2439	0.0135	0.2370	0.0873	0.2457	0.4506
ARE	1.0602	2.1004	0.9939	2.2249	0.9986	2.3988
Correlated covariates (r=0.3)						
bias($\hat{\beta}$)	0.0389	0.0428	0.0109	0.1097	0.0235	0.3155
$s.d.(\hat{\beta})$	0.5984	0.2158	0.5678	0.4745	0.5976	1.2533
$s.e.(\hat{\beta})$	0.5604	0.1941	0.5209	0.4542	0.5414	1.0657
CP of $\hat{\beta}$	0.9360	0.9270	0.9290	0.9250	0.9250	0.9180
MSE($\hat{\beta}$)	0.3592	0.0484	0.3222	0.2370	0.3574	1.6687
bias($\bar{\beta}$)	0.0363	0.0246	0.0109	0.0474	0.0226	0.1286
$s.d.(\bar{\beta})$	0.5767	0.1394	0.5738	0.3065	0.6000	0.7527
$s.e.(\bar{\beta})$	0.5529	0.1454	0.5150	0.3409	0.5370	0.8470
CP of $\bar{\beta}$	0.9350	0.9430	0.9200	0.9500	0.9230	0.9620
MSE($\bar{\beta}$)	0.3335	0.0200	0.3290	0.0961	0.3601	0.5825
ARE	1.0768	2.3968	0.9793	2.3970	0.9923	2.7726
Correlated covariates (r=0.4)						
bias($\hat{\beta}$)	0.0497	0.0373	0.0293	0.1198	0.0262	0.3039

Continued on next page

	β_1=1	$\beta_2 = 1$	$\beta_1 = 1$	$\beta_2 = 4$	$\beta_1 = 1$	$\beta_2 = 9$
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
$s.d.(\hat{\beta})$	0.469 6	0.216 1	0.442 8	0.505 1	0.632 4	1.167 0
$s.e.(\hat{\beta})$	0.428 2	0.193 1	0.400 2	0.447 8	0.538 3	1.064 5
CP of $\hat{\beta}$	0.922 0	0.931 0	0.912 0	0.918 0	0.911 0	0.928 0
MSE($\hat{\beta}$)	0.222 7	0.048 0	0.196 7	0.269 2	0.400 2	1.453 0
bias($\bar{\beta}$)	0.042 4	0.014 2	0.028 3	0.047 1	0.024 8	0.151 8
$s.d.(\bar{\beta})$	0.456 0	0.138 9	0.443 8	0.319 9	0.634 3	0.752 3
$s.e.(\bar{\beta})$	0.429 6	0.150 1	0.395 9	0.341 4	0.533 7	0.830 7
CP of $\bar{\beta}$	0.938 0	0.961 0	0.917 0	0.948 0	0.909 0	0.955 0
MSE($\bar{\beta}$)	0.209 5	0.019 5	0.197 6	0.104 4	0.402 5	0.588 4
ARE	1.060 3	2.420 2	0.995 5	2.493 8	0.994 1	2.406 6
Correlated covariates (r=0.5)						
bias($\hat{\beta}$)	0.035 2	0.029 7	0.037 0	0.114 4	0.024 8	0.245 4
$s.d.(\hat{\beta})$	0.369 6	0.210 0	0.361 7	0.477 7	0.387 5	1.121 4
$s.e.(\hat{\beta})$	0.350 3	0.196 9	0.328 6	0.440 0	0.343 8	1.000 1
CP of $\hat{\beta}$	0.934 0	0.925 0	0.923 0	0.929 0	0.923 0	0.923 0
MSE($\hat{\beta}$)	0.137 7	0.044 9	0.132 0	0.241 1	0.150 6	1.316 5
bias($\bar{\beta}$)	0.029 1	0.010 8	0.035 7	0.065 6	0.024 6	0.097 8
$s.d.(\bar{\beta})$	0.361 0	0.142 8	0.361 8	0.325 8	0.388 6	0.731 7

Continued on next page

Table 3.1 – *Continued from previous page*

	β_1=1	$\beta_2 = 1$	$\beta_1 = 1$	$\beta_2 = 4$	$\beta_1 = 1$	$\beta_2 = 9$
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
$s.e.(\bar{\beta})$	0.354 0	0.157 3	0.325 0	0.336 2	0.341 1	0.803 8
CP of $\bar{\beta}$	0.935 0	0.959 0	0.913 0	0.949 0	0.921 0	0.964 0
MSE($\bar{\beta}$)	0.131 0	0.020 5	0.132 0	0.110 4	0.151 5	0.544 4
ARE	1.048 4	2.162 9	0.999 2	2.149 6	0.994 1	2.348 9

3.3.1.2 MAR setting

In this section, we investigate the performance of the UIPW estimators under the MAR setting. The setting is the same as the one for MCAR except for the selection probabilities. The selection probabilities for z_1 is associated with the event time x but neither z_1 nor z_2. The selection probabilities for z_1 and z_2 given X are $\pi(X, \delta) = \Phi(-0.05log(X))$ and 0.6 respectively, where Φ is the standard normal cumulative distribution function and X is the event time of the subject. There are 40% of the values of the covariate z_1 with missing values and 36% of the subjects are fully observed. The censoring function is the same as the one in the MCAR setting. The sample size is 250 and simulation results are based on 1000 replications.

Similar results in Table 3.1 are seen as that in Table 3.2 in regards to biases, $s.d.$'s and $s.e.$'s. Compared to the results in MCAR, there are noticeable increases of biases, $s.d.$'s and $s.e.$'s for the MAR cases with the same settings. Especially, for the setting with $\beta_0 = (1,4)', (1,9)'$, the CPs of $\hat{\beta}$ and $\bar{\beta}$ are slightly deviant from the 95% nominal coverage rates. However, the CPs of $\bar{\beta}$ are closer to the 95% nominal coverage than the CPs of $\hat{\beta}$ in the simulation results.

Table 3.2: Simulation results under the MAR setting.

	β_1=1	$\beta_2 = 1$	$\beta_1 = 1$	$\beta_2 = 4$	$\beta_1 = 1$	$\beta_2 = 9$
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
Independent covariates						
bias($\hat{\beta}$)	0.015 8	0.025 3	0.060 3	0.168 1	0.005 6	0.341 7
$s.d.(\hat{\beta})$	0.624 7	0.217 1	0.626 3	0.576 0	0.632 7	1.179 7
$s.e.(\hat{\beta})$	0.591 7	0.198 9	0.569 9	0.492 4	0.564 8	1.103 9
CP of $\hat{\beta}$	0.945 0	0.932 0	0.936 0	0.910 0	0.926 0	0.940 0
MSE($\hat{\beta}$)	0.390 1	0.047 7	0.395 5	0.359 6	0.400 0	1.507 0
bias($\bar{\beta}$)	0.019 5	0.023 5	0.056 1	0.117 6	0.009 9	0.234 4
$s.d.(\bar{\beta})$	0.551 4	0.168 7	0.617 0	0.443 7	0.636 7	0.974 0
$s.e.(\bar{\beta})$	0.521 1	0.161 8	0.556 3	0.429 4	0.560 1	0.999 0
CP of $\bar{\beta}$	0.942 0	0.938 0	0.930 0	0.923 0	0.925 0	0.942 0
MSE($\bar{\beta}$)	0.304 1	0.029 0	0.383 4	0.210 5	0.405 0	1.002 6

Continued on next page

Table 3.2 – *Continued from previous page*

	β_1=1	$\beta_2 = 1$	$\beta_1 = 1$	$\beta_2 = 4$	$\beta_1 = 1$	$\beta_2 = 9$
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
ARE	1.283 4	1.654 6	1.030 4	1.684 9	0.987 6	1.467 0
Correlated covariates (r=0.3)						
bias($\hat{\beta}$)	0.017 8	0.035 7	0.035 2	0.110 7	0.013 5	0.260 1
$s.d.(\hat{\beta})$	0.653 8	0.223 3	0.589 3	0.510 4	0.595 1	1.097 9
$s.e.(\hat{\beta})$	0.576 1	0.200 6	0.518 6	0.450 2	0.502 2	0.979 1
CP of $\hat{\beta}$	0.905 0	0.933 0	0.919 0	0.928 0	0.911 0	0.927 0
MSE($\hat{\beta}$)	0.427 3	0.051 1	0.348 2	0.272 5	0.353 9	1.271 9
bias($\bar{\beta}$)	−0.003 5	0.023 8	0.023 6	0.080 4	0.010 3	0.174 5
$s.d.(\bar{\beta})$	0.563 6	0.169 4	0.582 7	0.393 4	0.590 8	0.904 9
$s.e.(\bar{\beta})$	0.512 6	0.165 0	0.505 3	0.387 7	0.497 6	0.888 9
CP of $\bar{\beta}$	0.911 0	0.937 0	0.919 0	0.944 0	0.915 0	0.936 0
MSE($\bar{\beta}$)	0.317 3	0.029 2	0.339 7	0.161 1	0.348 8	0.848 4
ARE	1.345 7	1.737 4	1.023 1	1.683 2	1.014 6	1.472 3
Correlated covariates (r=0.4)						
bias($\hat{\beta}$)	0.045 7	0.047 9	0.018 2	0.115 6	0.051 1	0.224 1
$s.d.(\hat{\beta})$	0.486 2	0.220 3	0.434 5	0.498 3	0.444 7	1.019 2
$s.e.(\hat{\beta})$	0.440 2	0.201 4	0.396 9	0.441 0	0.384 5	0.940 0
CP of $\hat{\beta}$	0.927 0	0.924 0	0.923 0	0.926 0	0.904 0	0.921 0

Continued on next page

	β_1=1	$\beta_2 = 1$	$\beta_1 = 1$	$\beta_2 = 4$	$\beta_1 = 1$	$\beta_2 = 9$
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
MSE($\hat{\beta}$)	0.238 3	0.050 8	0.188 9	0.261 4	0.200 2	1.087 9
bias($\bar{\beta}$)	0.012 2	0.028 3	−0.005 6	0.071 7	0.045 8	0.207 3
$s.d.(\bar{\beta})$	0.409 0	0.169 7	0.427 6	0.364 4	0.448 0	0.881 8
$s.e.(\bar{\beta})$	0.397 3	0.169 6	0.387 3	0.384 3	0.380 7	0.846 9
CP of $\bar{\beta}$	0.944 0	0.935 0	0.920 0	0.950 0	0.905 0	0.929 0
MSE($\bar{\beta}$)	0.167 3	0.029 6	0.182 7	0.137 8	0.202 6	0.819 7
ARE	1.413 0	1.685 3	1.032 2	1.870 0	0.985 5	1.336 0
			Correlated covariates (r=0.5)			
bias($\hat{\beta}$)	0.031 2	0.028 1	0.018 3	0.096 3	0.017 6	0.216 1
$s.d.(\hat{\beta})$	0.394 3	0.209 7	0.366 7	0.472 9	0.373 7	1.044 3
$s.e.(\hat{\beta})$	0.361 5	0.201 2	0.328 8	0.428 9	0.316 7	0.915 3
CP of $\hat{\beta}$	0.922 0	0.955 0	0.923 9	0.936 9	0.917 0	0.930 0
MSE($\hat{\beta}$)	0.156 3	0.044 7	0.134 7	0.232 7	0.139 8	1.136 1
bias($\bar{\beta}$)	0.002 3	0.012 7	0.008 6	0.058 2	0.011 7	0.135 0
$s.d.(\bar{\beta})$	0.345 0	0.164 8	0.360 4	0.382 0	0.376 3	0.860 4
$s.e.(\bar{\beta})$	0.326 7	0.171 4	0.320 4	0.374 3	0.313 6	0.834 6
CP of $\bar{\beta}$	0.936 0	0.954 0	0.921 9	0.944 9	0.900 0	0.932 0
MSE($\bar{\beta}$)	0.118 9	0.027 3	0.129 9	0.149 2	0.141 6	0.757 7

Continued on next page

Table 3.2 – *Continued from previous page*

	β_1=1	$\beta_2 = 1$	$\beta_1 = 1$	$\beta_2 = 4$	$\beta_1 = 1$	$\beta_2 = 9$
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
ARE	1.306 4	1.619 0	1.035 3	1.532 7	0.986 6	1.473 1

3.3.2 Simulation scenario 2

In this section, we use simulation study to compare the fully augmented weighted estimator (FAWE) (Qi et al., 2005) with the UIPW estimator. Two settings are conducted. The first one is the same one as example 2 in Qi et al. (2005). In the second one we attach a selection probability to z_2 in example 2 to allow general missing data pattern. The second setting is not supported by the original method in Qi et al. (2005).

Simulation sample sizes n are chosen to be 250, 500 and 1000. The baseline hazard function takes the form $\lambda(t|Z) = \lambda_0(t)exp(\beta' Z)$, where $\lambda_0(t) = 1$, $\beta_0 = (-ln(2), ln(2))'$ and $z_1, z_2 \sim Bernoulli(0.5)$. The selection probability of z_1 is $\pi(X, \delta) = 1/(1+exp(1.5-2.5\delta - X))$. The variable z_2 is fully observed in the first setting and has a selection probability 0.8 in the second setting. In this study, 65% of the cohort members have missing z_1 in the first setting, and 28% of observations are complete-cases in the second setting. The censored time has an exponential distribution with mean $\mu = 0.5z_1 + 0.1$. This generates

20% uncensored cases.

In the simulation results for the first setting, we see that the performances of the UIPW estimator is similar to that of the FAWE (Qi et al., 2005). No significant differences are noticed for the biases, $s.e.$'s, $s.d.$'s and CPs of these two estimators. Regarding the AREs of the UIPW estimator and the FAWE, the FAWE is slightly more efficient than the UIPW estimator in term of asymptotic variance.

The FAWE fails to deal with the data with general missing data pattern generated in the second setting, but the UIPW estimator still gains efficiency under this setting. This is one of the advantages of the UIPW estimator over FAWE.

The results indicate that the UIPW estimator with the estimated selection probabilities is slightly more efficient than that with the true selection probabilities.

Table 3.3: Simulation results for scenario 2.

	n=250		n=500		n=1000	
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
			z_2 fully observed			
bias($\hat{\beta}$)	-0.0673	0.0296	-0.0342	0.0210	-0.0079	0.0083
$s.d.(\hat{\beta})$	0.5660	0.4427	0.3607	0.2918	0.2536	0.2024
$s.e.(\hat{\beta})$	0.5066	0.4002	0.3519	0.2777	0.2467	0.1946
CP of $\hat{\beta}$	0.9190	0.9230	0.9550	0.9470	0.9420	0.9350

Continued on next page

Table 3.3 – *Continued from previous page*

	n=250		n=500		n=1000	
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
MSE($\hat{\beta}$)	0.324 6	0.196 6	0.131 2	0.085 5	0.064 3	0.041 0
bias($\bar{\beta}$)	−0.070 1	0.025 0	−0.034 1	0.010 6	−0.008 2	0.004 4
$s.d.(\bar{\beta})$	0.567 0	0.317 2	0.361 5	0.214 2	0.253 8	0.148 5
$s.e.(\bar{\beta})$	0.506 4	0.310 7	0.351 9	0.212 0	0.246 7	0.146 9
CP of $\bar{\beta}$	0.920 0	0.952 0	0.956 0	0.947 0	0.941 0	0.941 0
MSE($\bar{\beta}$)	0.326 1	0.101 1	0.131 7	0.045 9	0.064 4	0.022 0
ARE	0.996 6	1.948 1	0.996 1	1.856 7	0.997 9	1.857 8
z_2 with selection probability 0.8						
bias($\hat{\beta}$)	−0.081 8	0.044 8	−0.060 6	0.013 5	−0.026 6	0.014 0
$s.d.(\hat{\beta})$	0.635 4	0.467 6	0.418 1	0.327 4	0.296 3	0.222 9
$s.e.(\hat{\beta})$	0.574 7	0.450 9	0.396 9	0.311 9	0.276 4	0.217 9
CP of $\hat{\beta}$	0.928 0	0.945 0	0.938 0	0.933 0	0.934 0	0.943 0
MSE($\hat{\beta}$)	0.410 1	0.220 4	0.178 3	0.107 3	0.088 4	0.049 9
bias($\bar{\beta}$)	−0.064 2	0.030 7	−0.055 6	0.008 7	−0.024 8	0.006 6
$s.d.(\bar{\beta})$	0.544 1	0.353 8	0.374 1	0.240 1	0.264 4	0.163 9
$s.e.(\bar{\beta})$	0.538 5	0.356 5	0.363 1	0.240 2	0.250 8	0.165 2
CP of $\bar{\beta}$	0.943 0	0.948 0	0.940 0	0.949 0	0.944 0	0.950 0
MSE($\bar{\beta}$)	0.299 9	0.126 0	0.142 9	0.057 6	0.070 4	0.026 9

Continued on next page

Table 3.3 – *Continued from previous page*

	n=250		n=500		n=1000	
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
ARE	1.363 7	1.746 9	1.249 0	1.859 9	1.256 1	1.850 1
z_1 with estimated selection probability and z_2 with selection probability 0.8						
bias($\hat{\beta}$)	−0.077 6	0.059 8	−0.028 9	0.028 1	−0.018 4	0.016 3
$s.d.(\hat{\beta})$	0.635 5	0.488 1	0.401 2	0.324 4	0.272 5	0.230 3
$s.e.(\hat{\beta})$	0.570 4	0.451 7	0.392 6	0.311 4	0.276 3	0.218 1
CP of $\hat{\beta}$	0.927 0	0.933 0	0.948 0	0.936 0	0.951 0	0.942 0
MSE($\hat{\beta}$)	0.409 5	0.241 6	0.161 7	0.105 9	0.074 5	0.053 2
bias($\bar{\beta}$)	−0.072 4	0.043 1	−0.023 3	0.022 2	−0.021 1	0.004 8
$s.d.(\bar{\beta})$	0.567 9	0.347 6	0.362 2	0.238 3	0.244 7	0.162 2
$s.e.(\bar{\beta})$	0.531 4	0.357 8	0.358 4	0.238 4	0.250 8	0.165 4
CP of $\bar{\beta}$	0.933 0	0.960 0	0.942 0	0.948 0	0.956 0	0.957 0
MSE($\bar{\beta}$)	0.327 4	0.122 5	0.131 6	0.057 2	0.060 2	0.026 3
ARE	1.252 3	1.972 1	1.227 2	1.853 9	1.240 3	2.016 8

3.3.3 Time-varying covariates

This simulation is specifically designed to examine the performance of the UIPW estimator with time-dependent covariates.

Due to the extensive application in clinical trials, this example will simulate the processes for patients who receive treatments in clinics. We will use the algorithm introduced by MacKenzie and Abrahamowicz (2002) and validated by Sylvestre and Abrahamowicz (2008) to simulate treatment history in clinics.

In most clinical applications, the covariate data change in time as step functions. In clinical practice, it is common for the covariate data to be taken at intervals. For example, dose may be changed at a visit to a doctor according to the effect of the treatment with the dose the patient receives at last visit. The prescription dose is not subject to a change without doctor's prescription. It is a constant between at least two visits. In general, this simplification to the Cox proportional hazards model is adequate without loss of practical meaning clinically and statistically (Therneau and Foundation, 1999).

Table 3.4 is a sample of the simulated data. Field ID is a unique number assigned to each patient who receives treatment in the clinic. Fields Start and Stop denote the numbers of days when a set of dose values for a patient starts and stops from the beginning. Field Status is the censored flag, Status=1 if the subject is censored and 0 otherwise. Field Life is the duration of the treatment for a patient in the clinic.

Table 3.4: Data sample for time-dependent covariates

ID	Status	Life	Start	Stop	x_1	x_2
3	0	143	0	31	2.3	0
3	0	143	31	73	1.4	9.61
3	0	143	73	101	2	0
3	1	143	101	143	0	8.69
4	0	37	0	20	0	0
4	0	37	20	37	1.2	8.59

For this example, we simulate processes for patients treated in clinics with 365 days follow-up. Patients receive two medicines measured in dose x_1 and x_2 and start their initial treatments at a date, which rounds uniform random variable on $[1, 365]$ to integer. Each dose takes the value which rounds uniform random variable on $[0, 10]$ to 1 decimal place. The duration of the dose is a random sample from Poisson distribution with $\lambda = 3$. Event time follows exponential distribution with $\lambda = 0.012$. Censor time follows uniform distribution on $[1, 870]$. The censored rate is about 10%.

The number of events obtained in the data returned by the function depends on both the distribution of the event and the censored times. Patients without an event before or on day 365 and who are not censored before day 365 are censored on day 365 by administrative censoring. The sample size n is 500 or 1000. The number of repetitions is 500 and $\beta_0 = (ln(1.04), ln(0.99))$. Variables z_1 and z_2 are missing completely at random with selection probability 0.6 and 0.5 respectively. Note that the selection policy adheres to the subject. If one subject is missing, all observations belonging to this subject are missing as well.

Table 3.5 shows the simulation results for the setting with time-dependent covariates. The CPs of $\bar{\beta}$ and $\hat{\beta}$ fall around the 95% nominal rate. The biases, $s.e.$'s and $s.d.$'s of $\bar{\beta}$ are noticeably smaller than that of $\hat{\beta}$. The AREs of β_1 and β_2 imply the great improvement of efficiency. The results indicate that the UIPW estimator $\bar{\beta}$ performs better than the complete-case estimator $\hat{\beta}$.

Table 3.5: Simulation results for time-dependent covariates

	n=500		n=1000	
	$\beta_1 = ln(1.04)$	$\beta_2 = ln(0.99)$	$\beta_1 = ln(1.04)$	$\beta_2 = ln(0.99)$
bias($\hat{\beta}$)	-0.0030	-0.0030	-0.0011	-0.0028
$s.d.(\hat{\beta})$	0.0485	0.0570	0.0315	0.0370
$s.e.(\hat{\beta})$	0.0459	0.0546	0.0336	0.0391
CP of $\hat{\beta}$	0.9340	0.9540	0.9700	0.9680
MSE($\hat{\beta}$)	0.0024	0.0033	0.0010	0.0014
bias($\bar{\beta}$)	-0.0024	-0.0023	-0.0001	-0.0031
$s.d.(\bar{\beta})$	0.0309	0.0428	0.0223	0.0279

Continued on next page

Table 3.5 – *Continued from previous page*

	$\beta_1 = ln(1.04)$	$\beta_2 = ln(0.99)$	$\beta_1 = ln(1.04)$	$\beta_2 = ln(0.99)$
$s.e.(\bar{\beta})$	0.034 4	0.044 5	0.024 5	0.031 0
CP of $\bar{\beta}$	0.962 0	0.930 0	0.958 0	0.966 0
MSE($\bar{\beta}$)	0.001 0	0.001 8	0.000 5	0.000 8
ARE	2.473 1	1.772 0	2.008 1	1.752 1

3.4 Discussion

In this chapter, we extend the UIPW estimator for the Cox proportional hazards model with missing covariates under MCAR and MAR settings. The simulation results indicate that the UIPW estimator is more efficient than the simple inverse probability weighted estimator.

Compared to other estimators, the UIPW estimator sometimes is less efficient but is more flexible in dealing with the general missing data pattern. This flexibility is one of its outstanding characteristics. The UIPW estimator, compromising efficiency and simplicity, is an effective and competitive method to deal with the Cox proportional hazards model with missing covariates in practice.

Chapter 4

Future research

Zhao et al. (2013) extended the unified approach to the parametric regression models with missing covariate data. In Chapters 2 and 3, we extend the unified approach estimator to the partially linear model and the Cox proportional hazards model with missing covariates. The unified estimators show excellent performances for these models, however their asymptotic behaviors are rarely scrutinized theoretically regarding the question about how much and under what conditions they outperform other estimators in terms of bias and asymptotic relative efficiency. Chen and Chen (2000) showed that if the selection probability is a constant, the unified estimator corresponds to one member of the class of Robins et al. (1994). It is reasonable to conjecture the same asymptotic behaviors of the UIPW estimator for data missing at random and in the general pattern as that for data missing completely at random and in the simple monotone pattern, however detailed investigations are needed.

The unified approach requires that the selection probability function is known or can be parametrically estimated using a parametric model. Wang et al. (1997) proposed a non-parametric kernel based estimator for the selection probability function. A kernel-assisted unified approach which incorporates the kernel estimator into unified approach is a good choice for the unknown probability function in estimation. Further investigations on the kernel-assisted unified approach estimator are valuable.

Chapter 5

Appendix

5.1 General regularity conditions

Let $\alpha = (\beta, \gamma)$, $\mathbf{u}(\alpha) = \{s(\beta), \tilde{s}(\gamma)\}$. The regularity conditions are listed below.

(a.) $\alpha^* = (\beta^*, \gamma^*)$ lies in the interior of a compact parameter space.

(b.) $E[\mathbf{u}(\alpha)] \neq 0$ if $\alpha \neq \alpha^*$.

(c.) There is a neighborhood $\mathbf{A}$ of α^* such that $E[sup_{\alpha\in\mathbf{A}}||\mathbf{u}(\alpha)||]$, $E[sup_{\alpha\in\mathbf{A}}||\partial\mathbf{u}(\alpha)/\partial\alpha||]$ and $E[sup_{\alpha\in\mathbf{A}}||\mathbf{u}(\alpha)\mathbf{u}^{'}(\alpha)||]$ are all finite, where $||M|| = (\sum_{ij} m_{ij}^2)^{1/2}$ for any matrix M with elements m_{ij}.

(d.) $E[\mathbf{u}(\alpha^*)\mathbf{u}^{'}(\alpha^*)]$ is finite and positive definite.

5.2 Assumptions of partially linear models

Robinson (1988) provided the following assumptions of partially linear models:

(i.) $(Y_i, X_i, T_i), i = 1, 2...$, are independent and are replications of (Y, X, T).

(ii.) model (2.1) is true.

(iii.) ϵ is independent of X, T.

(iv.) $E[\epsilon^2] = \sigma^2 < \infty$.

(v.) $E[|X|^4] < \infty$.

(vi.) T admits a pdf $f \in \mathscr{G}_\lambda^\infty$, for some $\lambda > 0$.

(vii.) $E[X|T = t] \in \mathscr{G}_\mu^2$, for some $\mu > 0$.

(viii.) $\nu(\cdot) \in \mathscr{G}_\upsilon^4$, for some $\upsilon > 0$.

(ix.) as $N \to \infty$, $Na^{2q}b^4 \to \infty$, $Na^{2min(\lambda+1,\mu)+2min(\lambda+1,\upsilon)}b^{-4} \to 0$, $a^{min(\lambda+1,2\lambda,\mu,\upsilon)}b^{-2} \to 0$, $b \to 0$.

(x.) $k \in \mathscr{K}_{max(l+m-1,l+n-1)}$, for the integers l, m, n such that $l - 1 < \lambda \leq l$, $m - 1 < \mu \leq m, n - 1 < \upsilon \leq n$.

(xi.) $\Phi = E[\{X - E(X|T)\}\{X - E(X|T)\}^{'}]$ is positive definite.

Here $\mathscr{K}_l, l \geq 1$, is a class of even functions $k : \mathscr{R} \to \mathscr{R}$ satisfying: $\int_{\mathscr{R}} u^i k(u) du = \delta_{i0}$ and $k(u) = O((1 + |u|^{l+1+\epsilon})^{-1})$, some $\epsilon > 0$. $\mathscr{G}_\mu^\alpha, \alpha > 0, \mu > 0$, is the class of functions $\nu : \mathscr{R}^q \to \mathscr{R}$ satisfying: ν is $(m-1)$-times partially differentiable, for $m-1 \leq \mu \leq m$ and all t; for some $\rho > 0, sup_{y \in_{t\rho}} |\nu(y) - \nu(t) - Q(y,t)|/|y-t|^\nu \leq h(t)$ for all t, where $\mathscr{L}_{t\rho} = y : |y - t| < \rho; Q = 0$ when $m = 1$; Q is a $(m - 1)$th-degree homogeneous polynomial in $y-t$ with coefficients the partial derivatives of ν at t of orders 1 through $m-1$ when $m>1$; and $\nu(t)$, its derivatives of order $m-1$ and less, and $h(t)$, have finite αth moments.

5.3 Some asymptotic properties

Corollary 2.2.1. *Under the assumptions given in Appendix 5.2, estimator $\hat{\gamma}_l$ (2.8) is consistent for γ_l^* and $\sqrt{m}(\hat{\gamma}_l-\gamma_l^*)$ is asymptotically normal with mean 0 and variance $\sigma^2\Phi_l^{-1}+\Phi_l^{-1}var(\ddot{W}_l\ddot{U}_l'\beta_{l2})\Phi_l^{-1}$, where $\Phi_l = E[\ddot{W}_l\ddot{W}_l']$, β_{l2} is the parameters corresponding to the covariates W_l in the lth covariate group .*

Proof. First, we show the consistency of $\hat{\gamma}_l$. Clearly, from the formula for estimator $\hat{\gamma}_l$ (2.8), we have:

$$\begin{aligned}
\hat{\gamma}_l &= \Big(\sum_{i=1}^m \tilde{W}_{li}\tilde{W}_{li}'\Big)^{-1}\Big(\sum_{i=1}^m \tilde{W}_{li}\tilde{Y}_i\Big)\\
&= \Big(\sum_{i=1}^m \tilde{W}_{li}\tilde{W}_{li}'\Big)^{-1}\Big(\sum_{i=1}^m \tilde{W}_{li}(I-K)Y_i\Big)\\
&= \Big(\sum_{i=1}^m \tilde{W}_{li}\tilde{W}_{li}'\Big)^{-1}\Big(\sum_{i=1}^m \tilde{W}_{li}(I-K)(X_i'\beta+\nu(T_i)+\epsilon_i)\Big).
\end{aligned}$$

Denote A as the matrix obtained by changing the order of the columns of identity matrix I such that the first p_l columns of $X'A$ is $W_l' = (W_{l1},...,W_{lp_l})'$. Denote that $X'A = (W_l', U_l')$ and $A^{-1}\beta = (\beta_{l1}', \beta_{l2}')'$, where $\beta_{l1}\in R^{p_l}$. Continuing the transformation by using $X'A = (W_l', U_l')$, we obtain

$$\begin{aligned}
\hat{\gamma}_l &= \Big(\sum_{i=1}^m \tilde{W}_{li}\tilde{W}_{li}'\Big)^{-1}\Big(\sum_{i=1}^m \tilde{W}_{li}(I-K)(X_i'\beta+\nu(T_i)+\epsilon_i)\Big)\\
&= \Big(\sum_{i=1}^m \tilde{W}_{li}\tilde{W}_{li}'\Big)^{-1}\Big(\sum_{i=1}^m \tilde{W}_{li}(\tilde{X}_i'\beta+\tilde{\nu}(T_i)+\tilde{\epsilon}_i)\Big)
\end{aligned}$$

$$
\begin{aligned}
&= \left(\sum_{i=1}^{m} \tilde{W}_{li}\tilde{W}_{li}'\right)^{-1}\left(\sum_{i=1}^{m} \tilde{W}_{li}(\tilde{X}_i' A A^{-1}\beta + \tilde{\nu}(T_i) + \tilde{\epsilon}_i)\right) \\
&= \left(\sum_{i=1}^{m} \tilde{W}_{li}\tilde{W}_{li}'\right)^{-1}\left(\sum_{i=1}^{m} \tilde{W}_{li}((\tilde{W}_{li}', \tilde{U}_{li}')(\beta_{l1}', \beta_{l2}')' + \tilde{\nu}(T_i) + \tilde{\epsilon}_i)\right) \\
&= \left(\sum_{i=1}^{m} \tilde{W}_{li}\tilde{W}_{li}'\right)^{-1}\left(\sum_{i=1}^{m} \tilde{W}_{li}(\tilde{W}_{li}'\beta_{l1} + \tilde{U}_{li}'\beta_{l2} + \tilde{\nu}(T_i) + \tilde{\epsilon}_i)\right) \\
&= \left(\sum_{i=1}^{m} \tilde{W}_{li}\tilde{W}_{li}'\right)^{-1}\left(\sum_{i=1}^{m} \tilde{W}_{li}\tilde{W}_{li}'\beta_{l1} + \tilde{W}_{li}\tilde{U}_{li}'\beta_{l2} + \tilde{W}_{li}\tilde{\nu}(T_i) + \tilde{W}_{li}\tilde{\epsilon}_i\right) \\
&= \beta_{l1} + \left(\sum_{i=1}^{m} \tilde{W}_{li}\tilde{W}_{li}'\right)^{-1}\sum_{i=1}^{m}\left(\tilde{W}_{li}\tilde{U}_{li}'\beta_{l2} + \tilde{W}_{li}\tilde{\nu}(T_i) + \tilde{W}_{li}\tilde{\epsilon}_i\right) \\
&= \beta_{l1} + \left(\sum_{i=1}^{m} \frac{\tilde{W}_{li}\tilde{W}_{li}'}{m}\right)^{-1}\sum_{i=1}^{m}\left(\frac{\tilde{W}_{li}\tilde{U}_{li}'}{m}\beta_{l2} + \frac{\tilde{W}_{li}\tilde{\nu}(T_i)}{m} + \frac{\tilde{W}_{li}\tilde{\epsilon}_i}{m}\right).
\end{aligned}
$$

Under assumptions given in Appendix 5.2, Robinson (1988) showed that $\sum_{i=1}^{m} \tilde{X}_i\tilde{\nu}(T_i)/\sqrt{m} \xrightarrow{p} 0$, $\sum_{i=1}^{m} \tilde{X}_i\tilde{\epsilon}_i/m \xrightarrow{p} 0$ and $\sum_{i=1}^{m} \tilde{X}_i\tilde{X}_i'/m \xrightarrow{p} \Phi$. Matrix $(W_{l1}, .., W_{ln_l})$ is a submatrix obtained by taking rows out from $(X_1, ..., X_n)$. For elementwise convergence, obviously, $\sum_{i=1}^{m} \tilde{W}_{li}\tilde{\nu}(T_i)/m \xrightarrow{p} 0$, $\sum_{i=1}^{m} \tilde{W}_{li}\tilde{\epsilon}_i/m \xrightarrow{p} 0$, $\sum_{i=1}^{m} \tilde{W}_{li}\tilde{W}_{li}'/m \xrightarrow{p} \Phi_l$, $\sum_{i=1}^{m} \tilde{W}_{li}\tilde{U}_{li}'/m \xrightarrow{p} \bar{\Phi}_l$ hold as well, where $\bar{\Phi}_l = E[\ddot{W}\ddot{U}]$. Φ_l and $\bar{\Phi}_l$ are the submatrix of Φ correspondent to the variance and covariance between covariates in W_l and U_l. The consistency of $\hat{\gamma}_l$ for partially linear surrogate model is sustained and $\hat{\gamma}_l \xrightarrow{p} \gamma_l^* = \beta_{l1} + \Phi_l^{-1}\bar{\Phi}_l\beta_{l2}$.

Consider the asymptotic normality of $\hat{\gamma}_l$. Following the same skills as Robinson (1988) to decompose the $\sqrt{m}(\hat{\gamma}_l - \gamma_l^*)$, we have

$$
\begin{aligned}
\sqrt{m}(\hat{\gamma}_l - \gamma_l^*) =& \sqrt{m}\left((\sum_{i=1}^{m} \tilde{W}_{li}\tilde{W}_{li}')^{-1}\sum_{i=1}^{m} \tilde{W}_{li}\tilde{Y}_i - \gamma_l^*\right) \\
=& \sqrt{m}\left(\sum_{i=1}^{m} \tilde{W}_{li}\tilde{W}_{li}'\right)^{-1}\sum_{i=1}^{m} \tilde{W}_{li}(\tilde{Y}_i - \tilde{W}_{li}'\gamma_l^*) \\
=& \sqrt{m}\left(\sum_{i=1}^{m} \tilde{W}_{li}\tilde{W}_{li}'\right)^{-1}\sum_{i=1}^{m} \tilde{W}_{li}\left(\tilde{Y}_i - \tilde{W}_{li}'(\Phi_l^{-1}\bar{\Phi}_l\beta_{l2} + \beta_{l1})\right)
\end{aligned}
$$

$$
\begin{aligned}
=&\sqrt{m}\bigg(\sum_{i=1}^{m}\tilde{W}_{li}\tilde{W}_{li}'\bigg)^{-1}\sum_{i=1}^{m}\tilde{W}_{li}\bigg(\tilde{W}_{li}\beta_{l1}+\tilde{U}_{li}'\beta_{l2}+\tilde{\nu}(T_i)+\tilde{\epsilon}_i-\\
&\tilde{W}_{li}'(\Phi_l^{-1}\bar{\Phi}_l\beta_{l2}+\beta_{l1})\bigg)\\
=&\sqrt{m}\bigg(\sum_{i=1}^{m}\tilde{W}_{li}\tilde{W}_{li}'\bigg)^{-1}\sum_{i=1}^{m}\tilde{W}_{li}\bigg(\tilde{U}_{li}'\beta_{l2}+\tilde{\nu}(T_i)+\tilde{\epsilon}_i-\tilde{W}_{li}'\Phi_l^{-1}\bar{\Phi}_l\beta_{l2}\bigg)\\
=&\bigg(\sum_{i=1}^{m}\frac{\tilde{W}_{li}\tilde{W}_{li}'}{m}\bigg)^{-1}\sum_{i=1}^{m}\tilde{W}_{li}\bigg(\frac{\tilde{U}_{li}'\beta_{l2}}{\sqrt{m}}+\frac{\tilde{\nu}(T_i)}{\sqrt{m}}+\frac{\tilde{\epsilon}_i}{\sqrt{m}}\bigg)-\sqrt{m}\Phi_l^{-1}\bar{\Phi}_l\beta_{l2}\\
=&\bigg(\sum_{i=1}^{m}\frac{\tilde{W}_{li}\tilde{W}_{li}'}{m}\bigg)^{-1}\sum_{i=1}^{m}\tilde{W}_{li}\bigg(\frac{\tilde{\nu}(T_i)}{\sqrt{m}}+\frac{\tilde{\epsilon}_i}{\sqrt{m}}\bigg)-\\
&\bigg(\sum_{i=1}^{m}\frac{\tilde{W}_{li}\tilde{W}_{li}'}{m}\bigg)^{-1}\sum_{i=1}^{m}\frac{\tilde{W}_{li}\tilde{U}_{li}'\beta_{l2}}{\sqrt{m}}-\sqrt{m}\Phi_l^{-1}\bar{\Phi}_l\beta_{l2}.
\end{aligned}
$$

The first term $\sum_{i=1}^{m}\tilde{W}_{li}\bigg(\tilde{\nu}(T_i)+\tilde{\epsilon}_i\bigg)/\sqrt{m}\approx\sum_{i=1}^{m}\ddot{W}_{li}\epsilon_i/\sqrt{m}$ and the second term $\sum_{i=1}^{m}\tilde{W}_{li}\tilde{U}_{li}'\beta_{l2}/\sqrt{m}\approx\sum_{i=1}^{m}\ddot{W}_{li}\ddot{U}_{li}'\beta_{l2}/\sqrt{m}$ (Robinson, 1988), where $\ddot{W}_{li}=W_{li}-E[W_l|T]$ and $\ddot{U}_{li}=U_{li}-E[U_l|T]$. So, we have

$$
\sqrt{m}(\hat{\gamma}_l-\gamma_l^*)\sim\Phi_l^{-1}\sum_{i=1}^{m}\frac{\ddot{W}_{li}\epsilon_i}{\sqrt{m}}-\Phi_l^{-1}\sum_{i=1}^{m}\frac{\ddot{W}_{li}\ddot{U}_{li}'\beta_{l2}}{\sqrt{m}}-\sqrt{m}\Phi_l^{-1}\bar{\Phi}_l\beta_{l2}.
$$

We note that the sum is made up of independent and identically distributed random variables $\Phi_l^{-1}\ddot{W}_{li}\epsilon_i/\sqrt{m}-\Phi_l^{-1}\ddot{W}_{li}\ddot{U}_{li}'\beta_{l2}/\sqrt{m}-\sqrt{m}\Phi_l^{-1}\bar{\Phi}_l\beta_{l2},,i=1,...,n$. According to the Central Limit Theorem, if the covariance matrix of the random variable is positive definite, then $\sqrt{m}(\hat{\gamma}_l-\gamma_l^*)$ will converge in distribution to a normal random variable with mean 0.

Because ϵ is independent of W_l and U_l, the variance of the random variable $\Phi^{-1}(\ddot{W}_{li}\epsilon_i+\ddot{W}_{li}\ddot{U}_{li}'\beta_{l2})$ can be written as

$$
\begin{aligned}
var(\sqrt{m}(\hat{\gamma}_l-\gamma^*))=&var\bigg(\Phi^{-1}(\ddot{W}_l\epsilon-\ddot{W}_l\ddot{U}_l'\beta_{l2})\bigg)\\
=&\sigma^2\Phi_l^{-1}+\Phi^{-1}var(\ddot{W}_l\ddot{U}_l'\beta_{l2})\Phi^{-1},l=1,...,q.
\end{aligned}
$$

Here, $\sigma^2\Phi_l^{-1}$ is positive definite, so is $var(\sqrt{m}(\hat{\gamma}_l - \gamma^*))$. This proves the corollary. $\square$

Corollary 2.2.2. *Under the assumptions given in Appendix 5.2 along with the positive definiteness of* $E[XX'] = \Sigma$ *and* $E[W_l\nu(T)] = \Psi < \infty$, *estimator* $\hat{\gamma}_l$ (2.9) *is consistent for* γ_l^* *and* $\sqrt{m}(\hat{\gamma}_l - \gamma_l^*)$ *is asymptotically normal with mean* 0 *and variance* $\sigma^2\Sigma_l^{-1} + \Sigma_l^{-1}var(W_lU_l'\beta_{l2} + W_l\nu(T))\Sigma_l^{-1}$, *where* $E[W_lW_l'] = \Sigma_l$.

Proof. From the formula for estimator $\hat{\gamma}_l$ (2.9), we know that

$$\hat{\gamma}_l = \left(\sum_{i=1}^{m} W_{li}W_{li}'\right)^{-1}\left(\sum_{i=1}^{m} W_{li}Y_i\right)$$
$$\hat{\gamma}_l = \left(\sum_{i=1}^{m} W_{li}W_{li}'\right)^{-1}\left(\sum_{i=1}^{m} W_{li}(X_i'\beta + \nu(T_i) + \epsilon_i)\right)$$
$$\hat{\gamma}_l = \left(\sum_{i=1}^{m} W_{li}W_{li}'\right)^{-1}\left(\sum_{i=1}^{m} W_{li}(W_{li}'\beta_{l1} + U_{li}'\beta_{l2} + \nu(T_i) + \epsilon_i)\right)$$
$$\hat{\gamma}_l = \beta_{l1} + \left(\sum_{i=1}^{m} W_{li}W_{li}'\right)^{-1}\sum_{i=1}^{m} W_{li}(U_{li}'\beta_{l2} + \nu(T_i) + \epsilon_i)$$
$$\hat{\gamma}_l = \beta_{l1} + \left(\sum_{i=1}^{m} \frac{W_{li}W_{li}'}{m}\right)^{-1}\sum_{i=1}^{m} \frac{W_{li}}{m}(U_{li}'\beta_{l2} + \nu(T_i) + \epsilon_i).$$

The assumptions $E[W_lW_l'] = \Sigma_l$, $E[W_lU_l] = \bar{\Sigma}_l$, $E[W_l\nu(T)] = \Psi$ and $E[W\epsilon] = 0$ imply that $W_{li}(U_{li}'\beta_{l2} + \nu(T_i) + \epsilon_i)/m$ converges to $\bar{\Sigma}_l\beta_{l2} + \Psi$ in probability, and so $\hat{\gamma}_l \xrightarrow{p} \gamma_l^* = \beta_{l1} + \Sigma_l^{-1}(\bar{\Sigma}_l\beta_{l2} + \Psi)$. The consistency of $\hat{\gamma}_l$ is sustained. Now, we show it is asymptotically normal. We further transform the above expression and obtain:

$$\begin{aligned}\sqrt{m}(\hat{\gamma}_l - \gamma_l^*) =& \sqrt{m}\left((\sum_{i=1}^{m} W_{li}W_{li}')^{-1}\sum_{i=1}^{m} W_{li}Y_i - \gamma^*\right)\\ =& \sqrt{m}\left(\sum_{i=1}^{m} W_{li}W_{li}'\right)^{-1}\sum_{i=1}^{m} W_{li}(Y_i - W_{li}'\gamma^*)\\ =& \sqrt{m}\left(\sum_{i=1}^{m} W_{li}W_{li}'\right)^{-1}\sum_{i=1}^{m} W_{li}\left(Y_i - W_{li}'(\beta_{l1} + \Sigma_l^{-1}\bar{\Sigma}_l\beta_{l2} + \Sigma_l^{-1}\Psi)\right)\end{aligned}$$

$$
\begin{aligned}
=&\sqrt{m}\Big(\sum_{i=1}^{m} W_{li}W_{li}'\Big)^{-1}\sum_{i=1}^{m} W_{li}\Big(W_{li}'\beta_{l1}+U_{li}'\beta_{l2}+\nu(T_i)+\epsilon_i-\\
&W_{li}'(\beta_{l1}+\Sigma_l^{-1}\bar{\Sigma}_l\beta_{l2}+\Sigma_l^{-1}\Psi)\Big)\\
=&\Big(\sum_{i=1}^{m}\frac{W_{li}W_{li}'}{m}\Big)^{-1}\sum_{i=1}^{m}\frac{W_{li}}{\sqrt{m}}(U_{li}'\beta_{l2}+\nu(T_i)+\epsilon_i)-\\
&\sqrt{m}(\Sigma_l^{-1}\bar{\Sigma}_l\beta_{l2}+\Sigma_l^{-1}\Psi)).
\end{aligned}
$$

The sum is made up of independent and identically distributed random variables $W_{li}(U_{li}'\beta_{l2}+\nu(T_i)+\epsilon_i-\sqrt{m}(\Sigma_l^{-1}\bar{\Sigma}_l\beta_{l2}+\Sigma_l^{-1}\Psi))/\sqrt{m}, i=1,...,n$. By using Central Limit Theorem, the sum converges to a normal random variable in distribution with mean 0 and variance $\sigma^2\Sigma_l^{-1}+\Sigma_l^{-1}var(W_lU_l'\beta_{l2}+W_l\nu(T))\Sigma_l^{-1}$. □

Corollary 2.2.3. *Under assumptions given in Appendix 5.2 and Corollary 2.2.1, 2.2.2, the joint distribution of* $\sqrt{m}(\hat{\beta}-\beta^*)$ *and* $\sqrt{m}(\hat{\gamma}_l-\gamma_l^*), l=1,...,q$ *is asymptotically normal with mean* 0 *and variance*

$$
\begin{pmatrix}
\sigma^2\Phi^{-1} & \Phi^{-1}\sigma^2E[\ddot{X}\ddot{W}_l']\Phi_l^{-1}\\
\Phi_l^{-1}\sigma^2E[\ddot{W}_l\ddot{X}']\Phi^{-1} & \Phi_l^{-1}\sigma^2+\Phi_l^{-1}var(\ddot{W}_l\ddot{U}\beta_{l2})\Phi_l^{-1}
\end{pmatrix},
$$

if $R_l=1$, *or variance*

$$
\begin{pmatrix}
\sigma^2\Phi^{-1} & \Phi^{-1}\sigma^2E[\ddot{X}W_l']\Sigma_l^{-1}\\
\Sigma_l^{-1}\sigma^2E[W_l\ddot{X}']\Phi^{-1} & \Sigma_l^{-1}\sigma^2+\Sigma_l^{-1}var(\ddot{W}_l\ddot{U}\beta_{l2}+W\nu(T))\Sigma_l^{-1}
\end{pmatrix},
$$

if $R_l=0$.

Proof. Firstly, we consider that the lth surrogate model is a partially linear model. We can show that, $\sqrt{m}(\hat{\beta}'-\beta^{*'},\hat{\gamma}_l'-\gamma_l^{*'})'$ can be approximated by the sum of independent and identically distributed random vectors, $i = 1, ..., n$,

$$\begin{pmatrix} \Phi^{-1}\ddot{X}_i\epsilon_i/\sqrt{m} \\ \frac{\Phi_l^{-1}\ddot{W}_{li}\epsilon_i}{\sqrt{m}} - \frac{\Phi_l^{-1}\ddot{W}_{li}\ddot{U}_{li}'\beta_{l2}}{\sqrt{m}} - \sqrt{m}\Phi_l^{-1}\bar{\Phi}_l\beta_{l2} \end{pmatrix},$$

The positive definiteness of variance of $\ddot{X}\epsilon$ and $\ddot{W}_l\epsilon - \ddot{W}_l\ddot{U}_l'\beta_{l2}$ is a necessary condition for the asymptotic joint normality of these two random sequences. The expectation of these two terms are zeros. Denote Var$_l$ as its variance, Var$_l[1,1] \in R^{p\times p}$, Var$_l[1,2] \in R^{p\times p_l}$ and Var$_l[2,2] \in R^{p_l\times p_l}$. Notice that ϵ is independent of X and T. We have Var$_l[1,1] = \sigma^2\Phi$, Var$_l[1,2] = E[\ddot{X}\epsilon\epsilon'\ddot{W}_l']$ and Var$_l[2,2] = \sigma^2\Phi_l + var(\ddot{W}_l\ddot{U}_l'\beta_{l2})$. Performing linear transformations on both sides of Var$_l$ to eliminate the elements $[1,2]$ and $[2,1]$. Finally, we have

$$\text{Var}_l[2,2] = \left(\sigma^2\Phi_l + var(\ddot{W}\ddot{U}'\beta_{l2}) - E(\ddot{W}\epsilon\epsilon'\ddot{X}')E^{-1}(\ddot{X}\epsilon\epsilon'\ddot{X}')E[\ddot{X}\epsilon\epsilon'\ddot{W}']\right).$$

Tripathi (1999) and Lavergne (2008) showed that for random matrices $A \in R^{n\times p}$, $B \in R^{n\times q}$, $E[||A||^2] < \infty$, $E[||B||^2] < \infty$ and $E[A'A]$ is nonsingular. Then $E[B'B] - E[B'A]E^{-1}[A'A]E[A'B] \geq 0$, with equality iff $B = AE^{-1}[A'A]E[A'B]$. Using the extended Cauchy inequality in our case, we have $E[\ddot{W}\epsilon\epsilon'\ddot{X}']E^{-1}[\ddot{X}\epsilon\epsilon'\ddot{X}']E[\ddot{X}\epsilon\epsilon'\ddot{W}'] \leq E[\ddot{W}\epsilon\epsilon'\ddot{W}']$. We use the above inequality for the Var$_1[2,2]$ and get

$$\begin{aligned} \text{Var}_l[2,2] =& \left(\sigma^2\Phi_l + var(\ddot{W}\ddot{U}'\beta_{l2}) - E[\ddot{W}\epsilon\epsilon'\ddot{W}']\Phi^{-1}E[\ddot{W}\epsilon\epsilon'\ddot{W}']/\sigma^2\right) \\ \geq& \left(\sigma^2\Phi_l + var(\ddot{W}\ddot{U}'\beta_{l2}) - \sigma^2E[\ddot{W}\epsilon\epsilon'\ddot{W}']\right) \\ =& var(\ddot{W}\ddot{U}'\beta_{l2}). \end{aligned}$$

By Theorem C1, given positive definite $E[\ddot{X}\ddot{X}']$, the $var(\ddot{W}\ddot{U}'\beta_{l2})$ is positive definite. This conclusion proves that $\text{Var}_l[2,2] > 0$ immediately. The corollary for partially linear surrogate is proven.

When the lth surrogate model is a linear model, the second term of $(\hat{\gamma}_l - \gamma_l^*)$ is changed to $W_{li}\{U_{li}'\beta_{l2} + \nu(T_i) + \epsilon_i - \sqrt{m}(\Sigma_l^{-1}\bar{\Sigma}_l\beta_{l2} + \Sigma_l^{-1}\Psi)\}/\sqrt{m}$. Similar transformations can be performed as the partial one, however extra condition has to be added to assert its positive definiteness. The estimate of parameter for the linear surrogate model might not converge expectedly if no extra conditions to restrict the behaviors of the nonparametric component in the partially linear true model. The positive definiteness of the variance of $W_lU_l\beta_{l2} + \nu(T)$ is a sufficient condition for the positive definiteness of Var_l. If the surrogate model is a partially linear model, no extra condition is necessary. □

Theorem C1. *A random row vector* $\boldsymbol{v} = (\boldsymbol{v}_1, \boldsymbol{v}_2), \boldsymbol{v}_1 \in R^{1\times k}, \boldsymbol{v}_2 \in R^{1\times(p-k)}$ *and* $E[\boldsymbol{v}] = 0$, *if* $E[\boldsymbol{v}'\boldsymbol{v}]$ *is positive definite, then for any* $\beta_1 \in R^k \neq 0$, *the matrix* $E[(\boldsymbol{v}_2'\boldsymbol{v}_1\beta_1 - E\boldsymbol{v}_2'\boldsymbol{v}_1\beta_1)(\boldsymbol{v}_2'\boldsymbol{v}_1\beta_1 - E\boldsymbol{v}_2'\boldsymbol{v}_1\beta_1)']$ *is positive definite.*

Proof. By definition of variance formula, we know matrix $E[(\boldsymbol{v}_2'\boldsymbol{v}_1\beta_1 - E\boldsymbol{v}_2'\boldsymbol{v}_1\beta_1)(\boldsymbol{v}_2'\boldsymbol{v}_1\beta_1 - E\boldsymbol{v}_2'\boldsymbol{v}_1\beta_1)']$ is the variance of random variable $\boldsymbol{v}_2'\boldsymbol{v}_1\beta_1$. It is supposed to be positive definite. We suppose this claim is false and prove its opposite (proof by contradiction).

By probability theory, if matrix $E[(\boldsymbol{v}_2'\boldsymbol{v}_1\beta_1 - E\boldsymbol{v}_2'\boldsymbol{v}_1\beta_1)(\boldsymbol{v}_2'\boldsymbol{v}_1\beta_1 - E\boldsymbol{v}_2'\boldsymbol{v}_1\beta_1)']$ is not positive definite, then there exists a row vector $a \in R^{1\times(p-k)}, a \neq 0$ and a number d such that $P(a\boldsymbol{v}_2'\boldsymbol{v}_1\beta_1 = d) = 1$ (Gross, 2003).

We construct a matrix $A = Diag(\beta_1', a)$, where both $\beta_1 \neq 0$ and $a \neq 0$. Matrix A has full row rank. Random variable $\boldsymbol{w} = A\boldsymbol{v}'$. Matrix $cov(\boldsymbol{w}, \boldsymbol{w})$ is positive definite as well by probability theory. From the definition of variance, we know that

$$\begin{aligned} cov(\boldsymbol{w}, \boldsymbol{w}) &= cov(A\boldsymbol{v}', A\boldsymbol{v}') \\ &= Acov(\boldsymbol{v}', \boldsymbol{v}')A' \\ &= AE\left[\begin{pmatrix} \boldsymbol{v}_1' \\ \boldsymbol{v}_2' \end{pmatrix} \begin{pmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 \end{pmatrix}\right]A' \\ &= AE\left[\begin{pmatrix} \boldsymbol{v}_1'\boldsymbol{v}_1 & \boldsymbol{v}_1'\boldsymbol{v}_2 \\ \boldsymbol{v}_2'\boldsymbol{v}_1 & \boldsymbol{v}_2'\boldsymbol{v}_2 \end{pmatrix}\right]A' \\ &= \begin{pmatrix} \beta_1' & 0 \\ 0 & a \end{pmatrix} E\left[\begin{pmatrix} \boldsymbol{v}_1'\boldsymbol{v}_1 & \boldsymbol{v}_1'\boldsymbol{v}_2 \\ \boldsymbol{v}_2'\boldsymbol{v}_1 & \boldsymbol{v}_2'\boldsymbol{v}_2 \end{pmatrix}\right] \begin{pmatrix} \beta_1 & 0 \\ 0 & a' \end{pmatrix} \\ &= E\left[\begin{pmatrix} \beta_1' & 0 \\ 0 & a \end{pmatrix} \begin{pmatrix} \boldsymbol{v}_1'\boldsymbol{v}_1 & \boldsymbol{v}_1'\boldsymbol{v}_2 \\ \boldsymbol{v}_2'\boldsymbol{v}_1 & \boldsymbol{v}_2'\boldsymbol{v}_2 \end{pmatrix}\right] \begin{pmatrix} \beta_1 & 0 \\ 0 & a' \end{pmatrix} \\ &= E\left[\begin{pmatrix} \beta_1'\boldsymbol{v}_1'\boldsymbol{v}_1\beta_1 & \beta_1'\boldsymbol{v}_1'\boldsymbol{v}_2a' \\ a\boldsymbol{v}_2'\boldsymbol{v}_1\beta_1 & a\boldsymbol{v}_2'\boldsymbol{v}_2a' \end{pmatrix}\right]. \end{aligned}$$

We note that $a\boldsymbol{v}_2'$ and $\boldsymbol{v}_1\beta_1$ are numbers. Denote $\boldsymbol{c}_1 = a\boldsymbol{v}_2'$ and $\boldsymbol{c}_2 = \boldsymbol{v}_1\beta_1$. Because we have $P(a\boldsymbol{v}_2'\boldsymbol{v}_1\beta_1 = d) = 1$, subsequently $\boldsymbol{c}_1\boldsymbol{c}_2 = d$ holds with probability 1. Denote $\mathcal{A} = \{v : a\boldsymbol{v}_2'\boldsymbol{v}_1\beta_1 = d\}$. We continue the transformation

$$cov(\boldsymbol{w}, \boldsymbol{w}) = E\left[\begin{pmatrix} \beta_1'\boldsymbol{v}_1'\boldsymbol{v}_1\beta_1 & \beta_1'\boldsymbol{v}_1'\boldsymbol{v}_2a' \\ a\boldsymbol{v}_2'\boldsymbol{v}_1\beta_1 & a\boldsymbol{v}_2'\boldsymbol{v}_2a' \end{pmatrix}\right]$$

$$= E_{\mathcal{A}}\left[\begin{pmatrix} \beta_1^{'}\boldsymbol{v}_1^{'}\boldsymbol{v}_1\beta_1 & \beta_1^{'}\boldsymbol{v}_1^{'}\boldsymbol{v}_2 a^{'} \\ a\boldsymbol{v}_2^{'}\boldsymbol{v}_1\beta_1 & a\boldsymbol{v}_2^{'}\boldsymbol{v}_2 a^{'} \end{pmatrix}\right]$$

$$= E_{\mathcal{A}}\left[\begin{pmatrix} \boldsymbol{c}_2^2 & d \\ d & \boldsymbol{c}_1^2 \end{pmatrix}\right].$$

We note that

$$d\begin{pmatrix} \boldsymbol{c}_2^2 & d \end{pmatrix} - \boldsymbol{c}_2^2\begin{pmatrix} d & \boldsymbol{c}_1^2 \end{pmatrix} = \begin{pmatrix} d\boldsymbol{c}_2^2 - \boldsymbol{c}_2^2 d & d^2 - \boldsymbol{c}_2^2\boldsymbol{c}_1^2 \end{pmatrix}$$
$$= \begin{pmatrix} 0 & 0 \end{pmatrix}.$$

The two rows in matrix $\begin{pmatrix} \boldsymbol{c}_2^2 & d \\ d & \boldsymbol{c}_1^2 \end{pmatrix}$ are linearly dependent. This matrix does not have full rank. Consequently, it is not positive definite either. This contradicts with positive definiteness of the matrix $cov(\boldsymbol{w}, \boldsymbol{w})$. □

5.4 Assumptions of Cox proportional hazards models

Andersen and Gill (1982) provided the following assumptions for the Cox proportional hazards models:

A (Finite interval). $\int_0^t \lambda_0(u)du < \infty$.

B (Asymptotic stability). There exists a neighborhood $\mathscr{B}$ of β_0 and scalar, vector and matrix functions $s^{(0)}$, $s^{(1)}$, and $s^{(2)}$ defined on $\mathscr{B} \times [0, t]$ such that for $j = 0, 1, 2$

$$sup_{u\in[0,t],\beta\in\mathscr{B}}||S^{(j)}(\beta, u) - s^{(j)}(\beta, u)|| \xrightarrow{p} 0.$$

C (Linderberg condition). There exists $\delta > 0$ such that

$$n^{-1/2} sup_{i,u} |Z_l(u)| Y_l(u) I\{\beta^{'} Z_l(u) > -\delta |Z_l(u)|\} \xrightarrow{p} 0.$$

D (Asymptotic regularity conditions). Let $\mathscr{B}, s^{(0)}, s^{(1)}$ and $s^{(2)}$ be as in Condition B and define $e = s^{(1)}/s^{(0)}$ and $v = s^{(2)}/s^{(0)} - e^{\otimes 2}$. For all $\beta \in \mathscr{B}, u \in [0, t]$:

$$s^{(1)}(\beta, u) = \frac{\partial}{\partial \beta} s^{(0)}(\beta, u), s^{(2)}(\beta, u) = \frac{\partial^2}{\partial \beta^2} s^{(0)}(\beta, u),$$

$s^{(0)}(\cdot, u)$, $s^{(1)}(\cdot, u)$ and $s^{(2)}(\cdot, u)$ are continuous functions of $\beta \in \mathscr{B}$, uniformly in $t \in [0, t]$, $s^{(0)}, s^{(1)}$ and $s^{(2)}$ are bounded on $\mathscr{B} \times [0, t]$; $s^{(0)}$ is bounded away from zero on $\mathscr{B} \times [0, t]$ and the matrix

$$\Sigma = \int_0^t v(\beta_0, u) s^{(0)}(\beta_0, u) \lambda_0(u) du$$

is positive definite.

Bibliography

Andersen, P. K. and R. D. Gill. 1982. *Cox's regression model for counting processes: a large sample study*, Annals of Statistics **10**, 1100–1121.

Breslow, N. 1974. *Covariance analysis of censored survival data*, Biometrics **30**, 89–99.

Bretagnolle, Jean and Catherine. Huber-Carol. 1988. *Effects of omitting covariates in Cox's model for survival data*, Scand. J. Statist. **15**, 125–138.

Chen, H. 1988. *Convergence rates for parametric components in a partly linear model*, Annals of Statistics **16**, 136–146.

Chen, YiHau and Hung Chen. 2000. *A unified approach to regression analysis under double-sampling design*, Journal of the Royal Statistical Society, Series B **62**, 449–460.

Cox, D. R. 1972. *Regression models and life tables*, Journal of the Royal Statistical Society, Series B **34**, 187–220.

______. 1975. *Partial likelihood*, Biometrika **62**, 269–276.

Engle, R., C. Granger, J. Rice, and A. Weiss. 1986. *Nonparametric estimates of the relation between weather and electricity sales*, Journal of the American Statistical Association **81**, 310–320.

Eubank, R.L., E.L. Kambour, J.T. Kim, K. Klipple, Reese C.S., and M. Schimek. 1998. *Estimation in partially linear models*, Computational Statistics & Data Analysis **29**, 27–34.

Fan, Jianqing and Runze Lin. 2006. *An overview on nonparametric and semiparametric techniques for longitudinal data, Frontiers in Statistics*, Imperial College Pres.

Green, P. and B. W. Silverman. 1994. *Nonparametric regression and generalized linear models: a roughness penalty approach*, Vol. 58, Chapman and Hall, New York.

Gross, Jügen. 2003. *Linear regression*, Springer.

Härdle, W. 1994. *Applied nonparametric regression*, Technical Report D-10178, Institute fur Statistik und Okonometrie, Spandauer Str 1, Berlin.

Härdle, W., J. Gao, and H. Liang. 2000. *Partially linear models*, Springer, New York.

Heckman, N.E. 1986. *Spline smoothing in partly linear models*, Journal of the Royal Statistical Society, Series B **48**, 224–248.

Herring, Amy H. and Joseph G. Ibrahim. 2001. *Likelihood-based methods for missing covariates in the Cox proportional hazards model*, Journal of the American Statistical Association **96**, 292–302.

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

Ibrahim, J.G., M.H. Chen, S.R. Lipsitz, and A.H. Herring. 2005. *Missing-data methods for generalized linear models: A comparative review*, Journal of the American Statistical Association **100**, 332–346.

Lagakos, S. and D. Schoenfeld. 1984. *Properties of proportional hazards score tests under misspecified regression models*, Biometrics **40**, 1037–1048.

Lavergne, Pascal. 2008. *A Cauchy-Schwarz inequality for expectation of matrices*, Dept. of Economics, Simon Fraser University, 8888 University Drive, Burnaby BC, V5A 1S6 Canada.

Liang, H. and Y. Qin. 2008. *Empirical likelihood-based inferences for partially linear models with missing covariates*, Aust. N. Z. J. Stat. **50**, 347–359.

Liang, H., S. Wang, J.M. Robins, and R. J. Carroll. 2004. *Estimation in partially linear models with missing covariates*, Journal of the American Statistical Association **99**, 357–367.

Lin, D. Y. and L. J. Wei. 1989. *The robust inference for the Cox proportional hazards model*, Journal of the American Statistical Association **84**, 1074–1078.

Little, Roderick J. A. and Donald B. Rubin. 2002. *Statistical analysis with missing data*, WILEY.

MacKenzie, Todd and Michal Abrahamowicz. 2002. *Marginal and hazard ratio specific random data generation: Applications to semi-parametric bootstrapping*, Statistics and Computing **12**, 245–252.

Morgan, Timothy M., S. W. Lagakos, and D. A. Schoenfeld. 1986. *Omitting covariates from the proportional hazards model*, Biometrics **42**, 993–995.

Naes, T. 1982. *The asymptotic distribution of the estimator for the regression parameter in Cox's regression model*, Scandinavian Journal of Statistics **9**, 107–115.

O'Neill, Terence J. 1986. *Inconsistency of the misspecified proportional hazards model*, Statist. Probab. Lett. **4**, 219–222.

Pugh, M., J. Robins, S. Lipsitz, and D. Harrington. 1994. *Inference in the Cox proportional hazards model with missing covariates*, Technical Report 758Z, Harvard School of Public Health,Department of Biostatistics, Boston, MA.

Qi, Lihong, C.Y. Wang, and Ross L. Prentice. 2005. *Weighted estimators for proportional hazards regression with missing covariates*, Journal of the American Statistical Association **100**, 1250–1263.

Qi, Lihong, Ying Fang Wang, and Yulei He. 2010. *A comparison of multiple imputation and fully augmented weighted estimators for Cox regression with missing covariates*, Statistics in Medicine **29**, 2592–2604.

Rice, J. 1986. *Convergence rates for partially splined models*, Statistics & Probability Letters **4**, 203–208.

Robins, James M., Andrea Rotnitzky, and Lue Ping Zhao. 1995. *Analysis of semiparametric regression models for repeated outcomes in the presence of missing data*, Journal of the American Statistical Association **90**, 106–121.

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

Robinson, P.M. 1988. *Root-n-consistent semiparametric regression*, Econometrica **56**, 931–954.

Schick, Anton. 1996. *Weighted least squares estimates in partly linear regression models*, Statistics & Probability Letters **27**, 281–287.

Silverman, B. W. 1984. *Spline smoothing: Ther equivalent variable kernel method*, The Annals of Statistics **12**, 898–916.

Speckman, P. 1988. *Kernel smoothing in partial linear models*, Journal of the Royal Statistical Society, Series B **50**, 413–436.

Struthers, C. A. and J. D. Kalbfleisch. 1986. *Misspecified proportional hazard models*, Biometrika **73**, 363–369.

Sylvestre, Marie-Pierre and Michal Abrahamowicz. 2008. *Comparison of algorithms to generate event times conditional on time-dependent covariates*, Statistics in medicine **27**, 2618–2634.

Therneau, Terry M. and Mayo Foundation. 1999. *A Package for Survival Analysis in S*.

Tripathi, Gautam. 1999. *A matrix extension of the Cauchy-Schwarz inequality*, Economics Letters **63(1)**, 1–3.

Tsiatis, A. A. 1981. *A large sample study of Cox's regression model*, The Annals of Statistics **9**, 93–108.

Wang, C. Y. and Hua Yun Chen. 2001. *Augmented inverse probability weighted estimator for Cox missing covariate regression*, Biometrics **57**, 414–419.

Wang, C. Y., Suojin Wang, Lue-Ping Zhao, and Shyh-Tyan Ou. 1997. *Weighted semiparametric estimation in regression analysis with missing covariate data*, Journal of the American Statistical Association **92**, 512–525.

Wang, C.Y., S. Wang, R.G. Gutierrez, and R.J. Carroll. 1998. *Local linear regression for generalized linear models witho missing data*, The Annals of Statistics **26**, 1028–1050.

Wang, Q. H. 2009. *Statistical estimation in partial linear models with covariate data missing at random*, Ann Inst Stat Math **61**, 47–84.

Wang, Q. H., O. Linton, and W. Härdle. 2004. *Semiparametric regression analysis with missing response at random*, Journal of the American Statistical Association **99**, 334–345.

White, IR and P Royston. 2009. *Imputing missing covariate values for the Cox model*, Statistics in Medicine **28(15)**, 1982–1998.

You, Jinhong and Gemai Chen. 2007a. *On inference for a semiparametric partially linear regression model with serially correlated errors*, The Canadian Journal of Statistics **35**, 515–531.

______. 2007b. *Semiparametric generalized least squares estimation in partially linear regression models with correlated errors*, Journal of Statistical Planning and Inference **137**, 117–132.

Zhao, Yang, Meng Liu, and Jerald F. Lawless. 2013. *Inference in parametric regression models with covariates missing in nonmonotone patterns*, Department of Mathematics & Statistics, University Of Regina, 3737 Wascana Parkway, Regina, SK, S4S 0A2.