The Pennsylvania State University
The Graduate School
Eberly College of Science

SEMIPARAMETRIC ANALYSIS OF FAILURE TIME DATA IN THE PRESENCE OF DEPENDENT CENSORING

A Dissertation in
Statistics
by
Youngjoo Cho

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2015

The dissertation of Youngjoo Cho was reviewed and approved* by the following:

Debashis Ghosh
Professor of Statistics and Public Health Sciences
Dissertation Advisor, Chair of Committee

David Hunter
Professor of Statistics

Runze Li
Verne M. Willaman Professor of Statistics

Christopher Zorn
Liberal Arts Research Professor of Political Science and Sociology

Aleksandra Slavkovic
Associate Professor of Statistics
Associate Head for Graduate Studies

*Signatures are on file in the Graduate School.

Abstract

Survival analysis is a well-established field in statistics. It has many applications to biology, economics, industrial engineering, and so on. Independent censoring is one of the crucial assumptions in survival analysis. However, this is impractical in many medical studies. For example, in many medical studies, disease occurrence and dependent censoring exist simultaneously, where the presence of dependent censoring leads to the difficulty in analyzing covariate effects on disease outcomes. Such a data structure has been termed 'semicompeting risks data'.

Much research have been performed on modeling semicompeting risks data. One approach to handle the dependent censoring is to use semiparametric accelerated failure time (AFT) model. This dissertation focuses on addressing restrictions on previous methodology of the semiparametric AFT model and provide solutions.

In the first part of the dissertation, we propose a new weighted estimator for the AFT model under dependent censoring. One of the advantages in our approach is that these weights are optimal among all the linear combinations of the previously mentioned two estimators. To calculate these weights, a novel resampling-based scheme is employed. Attendant asymptotic statistical results for the estimator are established. In addition, simulation studies, as well as an application to real data, show the gains in efficiency for our estimator.

Goodness of fit procedures are essential tools for assessing model adequacy in statistics. While many authors have proposed goodness of fit tests for U-statistics of order 1, little has been developed for higher order U-statistics. In the second part of the dissertation, we develop a general theory and approach to goodness of fit techniques based on U-processes for the AFT model. Many of the examples will focus on U-statistics of order 2. We propose goodness of fit tests for U-statistics of order 2 by using theoretical results from U-process theory. For numerical summary of hypothesis testing, a generalization of resampling approach adapted from goodness of fit tests based on U-statistics of order 1 is developed. Simulation studies are used to illustrate the proposed methods.

In many medical studies, estimation of treatment effects is often of primary

scientific interest. As mentioned before, standard methods for evaluating the treatment effect in survival analysis typically require the assumption of independent censoring. In semicompeting risks framework, estimating treatment effect for the disease occurrence is difficult due to the dependent censoring. The approach to use semiparametric AFT model to adjust the dependent censoring requires an artificial censoring technique. However, when covariates are continuous and have large variability, this can lead to excessive artificial censoring resulting in numerically unstable estimates. In the third part of the dissertation, we propose a strategy for weighted estimation of treatment effect that adjusts for covariates. Weights are based on propensity score modeling of the treatment conditional on confounder variables. This novel application of propensity scores avoids excess artificial censoring caused by continuous covariates. Simulation studies and an application to data from the Radiation Therapy Oncology Group (RTOG) are used to illustrate the methodology.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to express gratitude to my advisor, Dr. Debashis Ghosh, who has led me until this time by brilliant and keen ideas and a lot of support. Without his support and patience, I could not do my research. From him, I learned not only knowledge of statistics but also intuition and broad view of every research problem, and socializing with people.

I also would like to give thanks to faculty, staffs and students at the Department of Statistics of the Penn State University. They have provided invaluable support which have made me acquire precious knowledge, attitude as a researcher and experiences.

Finally, I would like to thank my family, friends, and people who have helped me in various ways. Their love and encouragement provide strength to proceed step by step. Without their love and help, I could not write this dissertation.

Chapter 1 Introduction

1.1 The Problem

Survival analysis is a well-developed field in statistics. Not only does it have many applications in statistics, but also its theory and techniques are widely used in biology, industrial engineering, economics, and so on. For example, in medical studies, researchers are interested in studying death, which is the event of interest. Due to the cost and time restriction of the study, the outcomes are not observed on all subjects and hence are censored. *Censoring* is a key feature in survival analysis. Due to this censoring, information of subject's lifetime and event is only known to have occurred in certain time interval. Most of the data in survival analysis is right-censored, which implies that the event can be observed only if it occurs before some fixed time point or during the study (Klein and Moeschberger, 2003, Chapter 3, p.64).

One of the basic assumptions in survival analysis is an independent censoring assumption. To explain this assumption, we will introduce some notations and a definition. Let T be the failure time of an individual and $\mathbf{Z}$ be covariates corresponding to the individual. Define the failure rate $\lambda(t; \mathbf{Z})$ as

$$\lambda(t; \mathbf{Z}) = \lim_{h \to 0} \frac{P(t \le T < t + h | T \ge t, \mathbf{Z})}{h}. \tag{1.1}$$

The independent censoring assumption states that the failure rates for individuals under existence of censoring at each time is the same as the failure rates for individuals without censoring. It implies that when we select an individual from people who neither are censored nor have an event of interest, the failure rate is

$\lambda(t; \mathbf{Z})$. In other words,

$$\lim_{h \to 0} \frac{P(t \leq T < t + h | \mathbf{Z}, T \geq t)}{h} = \lim_{h \to 0} \frac{P(t \leq T < t + h | \mathbf{Z}, T \geq t, Y(t) = 1)}{h}. \quad (1.2)$$

where $Y(t) = 1$ means that an individual is in the risk set, i.e, the individual neither is censored nor has the event. Generally, a censoring is independent if for every individual, the probability of censoring at time t only depends on covariates $\mathbf{Z}$ and does not depend on failure times (Kalbfleisch and Prentice, 2002, pp.12-13). This independent censoring assumption is a very crucial assumption in the modeling in survival analysis. If this assumption does not hold, i.e, when failure and censoring are statistically dependent, the joint distribution of failure and censoring is not identifiable (Tsiatis, 1975) so that statistical inference is not possible. Many models in survival analysis are based on this assumption.

However, in many cases, the assumption that people are independently censored may not hold. Typically, this occurs when we model multiple failure times. According to Ghosh (2000, Chapter 1), multiple events can be divided into two categories: The first is events that are distinguishable. For example, as we will see in the second example, time to HIV RNA level more than baseline quantity and time to withdrawal of study are two different events (Albrecht et al. 2001). The second one is recurrent events, which indicate that one event can occur several times for one individual. Tumor recurrence from oncology studies (Byar, 1980) belongs to this category. More specifically, censoring of subjects in the study is not statistically independent of failure in the modeling of two distinct events or recurrent events in the presence of a terminal event, which can be death or withdrawal of the study. The following two examples demonstrate this situation.

Example 1 (Bone Marrow Transplantation for Leukemia) : Bone marrow transplants are a typical treatment for acute leukemia. However, after the transplantation, patients may encounter various events. These events are:

1. Developing acute or chronic graft-versus-host disease (GVHD)
2. Returning the platelet count back to normal levels (platelet recovery)
3. Returning level of granulocytes back to normal levels

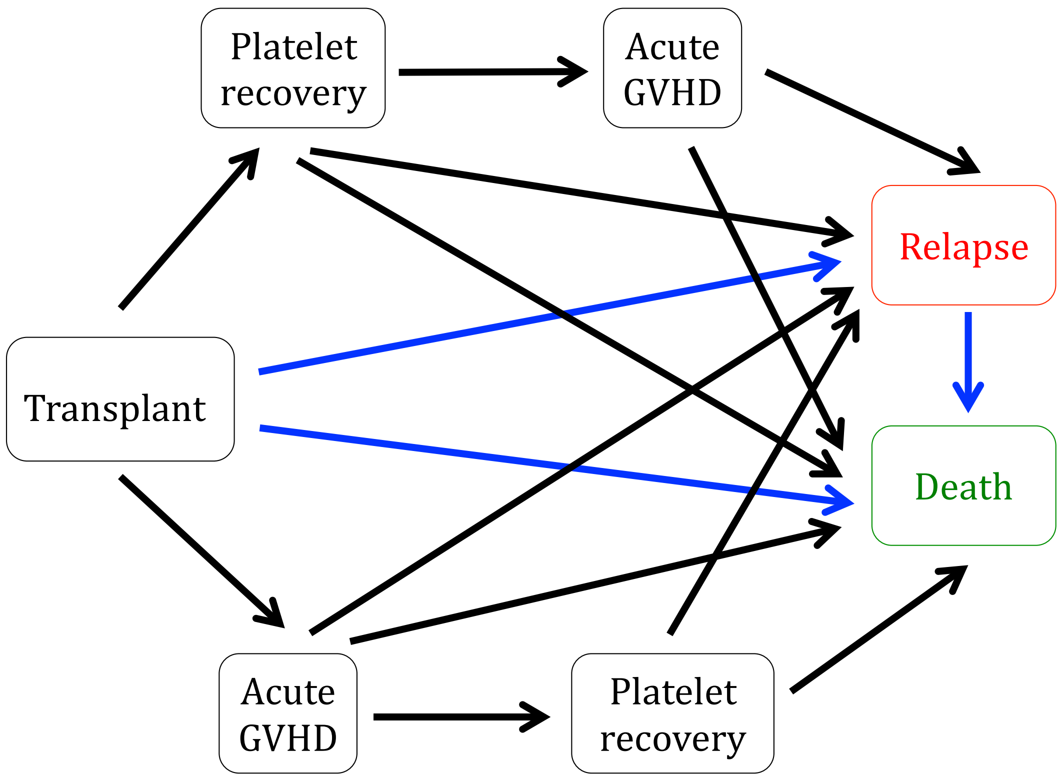
4. Development of infections

Platelet recovery (recovery of platelet count recovers to greater than or equal to $40 \times 10^9/l$) and acute GVHD typically occur within first 100 days. Patients may go through some or all of these events after transplant, and they may die or relapse after experiencing them. Moreover, relapse of leukemia and death from transplant could happen directly. It is also possible that patients may not experience some or all of stages. Figure 1.1 shows the process which the patients may experience after bone marrow transplant. In this study, there were 137 patients. 99 of them had acute myeloctic leukemia (AML) and the rest of patients had acute lymphoblastic leukemia. They received treatment in four hospitals: 76 at The Ohio State University Hospitals (OSU) in Columbus; 21 at Hahnemann University (HU) in Philadelphia; 23 at St. Vincent's Hospital (SVH) in Sydney, Australia; and 17 at Alfred Hospital (AH) in Melbourne. Transplants considered in this study were ones in these hospitals from March 1, 1984 to June 30, 1989. 42 patients experienced relapse and 41 patients died during remission. This dataset also contains various covariates which may be associated with death or relapse: risk groups according to their health status at the time of transplantation, time to acute GVHD, time to chronic GVHD, age of patient and donor, sex of patient and donor, waiting time from diagnosis to transplantation, etc. Details about this dataset can be found in Copelan et al. (1991) and Chapter 1 of Klein and Moeschberger (2003), pp. 3-6.

Example 2 (AIDS Study) : In AIDS Clinical Trial Group Study 364 (Albrecht et al. 2001), patients who have human immunodeficiency virus (HIV) RNA level greater than or equal to 500 copies per milliliters are considered. The first virologic failure is defined as the first time that HIV RNA level is greater than or equal to 2000 copies per milliliters. Treatments of interest are nelfinavir(NFV), efavirenz(EFV) or combination of NFV and EFV in combination with two nucleoside therapies. Total number of patients is 194, and they were randomly assigned to one of these treatments. The main interest of the study is to analyze treatment effect to time of the first virologic failure.

In the bone marrow transplant example, since researchers are interested in occurring death or relapse as failure, the independent censoring mechanism works in the modeling. However, if our interest is the relapse, then this mechanism is not valid because the death of people might be due to deterioration of their health

Figure 1.1: Processes after a Bone Marrow Transplant (Source: Klein and Moeschberger, 2003, p.4.)



status from relapse or side effect of acute or chronic GVHD. In this dataset, 40 people who died had experienced relapse. Thus, it is reasonable to consider that relapse and death are closely related. In the AIDS study, researchers analyzed the dataset by assuming that patients were independently censored by withdrawal, i.e, they imposed independent censoring assumption on the withdrawal and the data is analyzed by this framework (Albrecht et al. 2001; Peng and Fine, 2006). However, the first virologic failure and withdrawal of study are closely related. Moreover, the side effect from treatment may cause patient dropouts from the study. The following analysis supports this hypothesis. In this study, 101 patients experienced virologic failure and the other 93 were censored, and 83 patients among them were censored in administrative time C and 10 patients dropped out from the study before the administrative time. 9 of these 10 people received combination of NFV and EFV, so it may be possible that treatment affects the dropout (Peng and Fine, 2006). In these examples, the event which censored failure of interest is death or withdrawal, but the event of interest and censoring (death or withdrawal of study) are not statistically independent because the failure of interest affects the censoring. The kind of censoring is called 'dependent censoring'. Modeling failure of interest under existence of dependent censoring is very complicated because many typical models in the survival analysis are based on the independent censoring assumption.

1.2 Background

As can be seen in the previous section, modeling time to failure of interest under the existence of death or withdrawal of study is related to many difficult issues. The naive way of modeling is to ignore this dependent censoring and use traditional models by treating this possibly informative dropout as independent censoring. However, clearly, this way of modeling is problematic. In Chapter 3, we will see that this naive way of modeling causes the issue of bias when estimating parameters of the model.

Now we will introduce some literatures for this dependent censoring problem. Several researchers have proposed models of failure time in the presence of dependent censoring. They concentrated their effort on making identifiable models of failure time under the existence of dependent censoring. For example, Emoto and Matthews (1990) proposed a bivariate Weibull model when failure and censoring are dependent. In addition, Robins and Rotnitzky (1992) proposed a model by imposing weight from characteristics of the surrogate marker data. However, these are too restrictive in that they can be only applied under certain conditions.

Due to limitations of making identifiable models, researchers have started to focus on joint models of the failure time and the dependent censoring. Day et al. (1997) proposed using a biomarker for prediction of disease into frailty models. By adapting the idea of Day et al. (1997), Fine et al. (2001) proposed semicompeting risks data structure which contains both a single nonterminal event and terminal event. One of the fundamental estimators in survival analysis is an estimator for survival function. Fine et al. (2001) proposed an estimator for the bivariate survival function for Clayton copula model on semicompeting risks data. Wang (2003) extended an approach of Fine et al. (2001) to general copula models and general parameterization of models used in Day et al. (1997).

These models in the previous paragraph have provided a good solution by constructing a model incorporating both a nonterminal event and a terminal event, which are a generalization of disease occurrence and withdrawal of the study. However, these models have limited applications because they did not include covariates in the model. Lin et al. (1996) proposed a bivariate accelerated failure time (AFT) model including covariates in the semicompeting risks data structure (although they did not use term 'semicompeting risks data' explicitly for the data in which they were interested, the structure they considered is equivalent to semicompeting risks data). They also used the artificial censoring technique to adjust the dependent censoring. Peng and Fine (2006) proposed another procedure, which also employed the artificial censoring but their artificial censoring is different from Lin et al. (1996). Ding et al. (2009) proposed a regression model whose response variable is an increasing function of time instead of using the log function as Lin et al. (1996).

It is natural to extend models in the semicompeting risks data to one containing recurrent events under the existence of the terminal event. In this case, a main interest is to model time to each recurrent event. Ghosh and Lin (2003) proposed the procedure using bivariate AFT model for both recurrent events and death, which is a generalization of one in Lin et al. (1996). Ghosh (2010) and Hsieh et al. (2011) extended the approach of Peng and Fine (2006) to a case of recurrent events with the dependent censoring.

1.3 Outline of Dissertation

Estimators from Lin et al. (1996) and Peng and Fine (2006) have their own characteristics. For example, as we will see in detail in Chapter 3, the estimator from Peng and Fine (2006) is based on U-statistics of order 2 while the estimator of Lin et al. (1996) is based on U-statistics of order 1. In the view of efficiency, neither of them is better than the other with respect to standard error. If we extend the scope of estimators based on these two procedures, it is desirable to obtain the estimator which has better efficiency than ones by Lin et al. (1996) and Peng and Fine (2006).

In the first part of this dissertation, we propose a weighted estimator of combining Lin et al. (1996) and Peng and Fine (2006) for semicompeting risks data. In these estimations, our goal is to obtain the most efficient one by combining estimators from Lin et al. (1996) and Peng and Fine (2006) by inputting weights on each estimator with respect to standard error. The theory of the weighted estimator is based on the argument of Wei et al. (1989).

Goodness of fit is an essential part in statistics. In survival analysis, most estimation procedures are U-statistics of order 1 (Lin et al. 1993; Lin et al. 1996). However, there are also U-statistics of order 2. A good example of U-statistic of order 2 is the Wilcoxon test statistic for linear regression (Jin et al. 2001). Moreover, the estimating function proposed by Peng and Fine (2006) is also U-statistics of order 2. In Chapter 5, we propose the goodness of fit approach for U-statistic of order 2.

In Lin et al. (1996) and Peng and Fine (2006), they used a technique called an artificial censoring. The idea of artificial censoring to match uncensored times between treatment group and control group by censoring uncensored quantities (Lin et al. 1996). Details about the artificial censoring is explained in Chapter 3. Artificial censoring is a key technique to adjust dependence between time to the event of interest and time to the dependent censoring. Artificial censoring is an increasing function of covariates. Hence when covariates have large variability, the magnitude of the artificial censoring is large. However, a large artificial censoring implies that a lot of uncensored observations will be censored so that in the extreme case, there are few uncensored events after artificial censoring adjustment. This is problematic in the observational studies. The bone marrow transplant study is a good example for this case. Unlike the randomized studies, the existence of the confounder variables requires to use all variables in model building.

In many medical studies, the key interest of researchers is to examine the treatment effect for the time to disease progression or the time to death. If the distribution of confounders is equal between the treatment group and the control group, then it is possible to establish models using the treatment effect only. In randomized studies, by randomization, the distribution of confounders is equal between control group and treatment group. In the observational studies, the usual way to make the distributions of the confounders be equal is to use the propensity score. This covariate adjustment is useful for reducing the excessive artificial censoring. In Chapter 6, we develop treatment effect estimation by using covariate adjustment based on the propensity score.

The order of this dissertation is as follows. In Chapter 2, we are going to discuss the underlying models: the AFT model for a single event and recurrent events, and the linear transformation model under the independent censoring assumption. In Chapter 3, previous research on dependent censoring will be discussed. In Chapter 4, we propose the weighted estimation for the bivariate AFT model in the presence of the dependent censoring. In Chapter 5, a goodness of fit procedure for U-statistic of order 2 for the AFT model is proposed. In Chapter 6, we show the method of covariate adjustment based on the propensity score for reducing artificial censoring. In Chapter 7, the future work for modeling semicompeting risks data and recurrent events in the presence of dependent censoring will be discussed.

Chapter 2 Basic Models

2.1 Introduction

In survival analysis, perhaps the most famous model is the Cox proportional hazard model. However, this model has several problems. First, this model assumes the proportionality of hazard, which may not be applicable to many real datasets. Moreover, it is difficult to find a relationship between regression coefficients and survival time. For these reasons, alternatives of the Cox proportional hazard model have been developed. The most well-known alternatives of this Cox model are an accelerated failure time model and a linear transformation model. In this section, we will briefly discuss basic of these models. We assume independent censoring in this chapter.

2.2 Accelerated Failure Time Model

2.2.1 Single Event Case

The accelerated Failure Time Model (AFT) is a linear model for the logarithm of the failure time. Let T be time to an event of interest and C be time to independent censoring. Define $\mathbf{Z}$ to be $p \times 1$ a vector of covariates. Note that due to the existence of censoring, we can only observe $\tilde{T} = T \wedge C$, where $a \wedge b$ means minimum of a and b. The censoring indicator is $\Delta = I(T \leq C)$ where $I(\cdot)$ denotes an indicator function. The observed data consists of n independent and identically distributed copies $\{\tilde{T}_i, \Delta_i, \mathbf{Z}_i\}, i = 1, ..., n$. The model is

$$\log(T) = \mathbf{Z}^T \boldsymbol{\eta}_0 + \epsilon, \tag{2.1}$$

where $\boldsymbol{\eta}_0$ is $p \times 1$ vector of regression coefficients and ϵ is an error term. The ϵs are independent and identically distributed and do not depend on $\mathbf{Z}$. Expression (2.1) implies that AFT model is a linear model, so the relationship between failure time and covariates is direct. Interpretation of coefficients is more natural than that of the Cox proportional hazard model. Note that the equivalent formulation of the expression (2.1) in terms of a survival function of T given $\mathbf{Z}$ is

$$S(t|\mathbf{Z}) = S_0(te^{-\mathbf{Z}^T \boldsymbol{\eta}_0}),$$

where $S_0(\cdot)$ is a survival function of T when $\mathbf{Z} = 0$. Thus in this model, a person's survival time is accelerated or decelerated according to $\mathbf{Z}^T \boldsymbol{\eta}_0$. The key element in statistical inference for $\boldsymbol{\eta}_0$ is whether we impose an assumption of parametric or nonparametric distribution on the error term. In the parametric setting, distributions for ϵ_i usually include the Weibull distribution, the logistic distribution, the normal distribution or the generalized gamma distribution. However, this parametric assumption is sometimes too restrictive for particular datasets. For this reason, a semiparametric model is a preferable alternative. For the semiparametric AFT model, we usually consider rank estimation methods to estimate parameters and perform statistical inference in the AFT model. A range of literature has discussed this type of the rank-based estimation (Wei and Gail, 1983; Louis, 1981; Tsiatis, 1990; Jin et al. 2003). This rank-based estimation is a generalization of linear rank tests for right censored data (Prentice, 1978). Let us consider a single covariate Z. By using the counting process notation, linear rank tests can be expressed as

$$U(W_n) = \sum_{i=1}^{n} \int_0^\infty W_n \left[Z_i - \frac{\sum_{j=1}^n I(\tilde{T}_j \geq t) Z_j}{\sum_{j=1}^n I(\tilde{T}_j \geq t)} \right] dN_i(t)$$

where $N_i(t) = I(\tilde{T}_i \leq t, \Delta_i = 1)$ and W_n is a weight function. In this case, this function can be interpreted as the sum of differences across the event time of observed covariates for those who have the event and the average of covariate values for those who are at risk at the event time. If we replace $\tilde{T}_i$ to $\tilde{T}_i(\eta_0) = \tilde{T}_i e^{-\eta_0 Z_i}$, this function is a test statistic for the hypothesis $H_0 : \eta = \eta_0$. If $\eta = \eta_0$, this function takes value 0, which implies that we can use this rank statistic as an estimating function (Tsiatis, 1990). If we extend this function to a case of multiple covariates, we can obtain the estimating function for $\boldsymbol{\eta}_0$, which is given by

$$\mathbf{S}_n(\boldsymbol{\eta}) = \sum_{i=1}^{n} w_i(\boldsymbol{\eta})\Delta_i \left[\mathbf{Z}_i - \frac{\sum_{j=1}^{n} \mathbf{Z}_j I\{\tilde{T}_j(\boldsymbol{\eta}) \geq \tilde{T}_i(\boldsymbol{\eta})\}}{\sum_{j=1}^{n} I\{\tilde{T}_j(\boldsymbol{\eta}) \geq \tilde{T}_i(\boldsymbol{\eta})\}}\right], \tag{2.2}$$

where $w(\cdot)$ is nonnegative weight functions and $\tilde{T}_i(\boldsymbol{\eta}) = \tilde{T}_i e^{-\mathbf{Z}_i^T \boldsymbol{\eta}}$. We obtain a solution by solving $\mathbf{S}_n(\boldsymbol{\eta}) = 0$. Let us denote this solution as $\hat{\boldsymbol{\eta}}$. Under regularity conditions stated in Ying (1993), $\hat{\boldsymbol{\eta}}$ is strongly consistent and asymptotically normal. Since the function in (2.2) is zero-crossing function, i.e., the function crosses the point zero, the solution always exists theoretically.

Since the estimating function is a nonsmooth function, it is not possible to apply numerical algorithms for smooth functions. To obtain the solution for estimating equations, if there is a single covariate, a one-dimensional case root finding method such as the bisection method can be employed. If the dimension of covariate is greater than 1, the exact solution may not exist. Then we can define the estimator of $\boldsymbol{\eta}$, say $\hat{\boldsymbol{\eta}}$, as a minimizer of $||\mathbf{S}_n(\boldsymbol{\eta})||$ (Peng and Fine, 2006). For multiple covariates, we can use the linear programming method proposed by Jin et al. (2003) or the Nelder-Mead algorithm (Nelder and Mead, 1965).

2.2.2 Recurrent Events Case

For modeling in the recurrent event cases, Lin et al. (1998) proposed an extension of the AFT model from the single event case. For $i = 1, \ldots, n$ and $k = 1, 2, \ldots$, let T_{ik} be kth event time for ith subject. In this case, we do not assume any dependence structure between events for a subject. Let $N_i^*(t)$ be the number of events which ith subject has experienced until time t without censoring, which can be formulated as,

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

Then the mean function of $N_i^*(t)$ given covariates $\mathbf{Z}_i$ is

$$E\{N_i^*(t)|\mathbf{Z}_i\} = \mu_0(e^{\mathbf{Z}_i^T\boldsymbol{\eta}_0}t), \tag{2.3}$$

where $\boldsymbol{\eta}_0$ is a $p \times 1$ vector of unknown regression parameters. By transforming times as in the single event case, we can transform the recurrent event times and counting processes:

$$\tilde{T}_{ik}(\boldsymbol{\eta}) = T_{ik}e^{\mathbf{Z}_i^T\boldsymbol{\eta}}$$
$$\tilde{N}_i^*(t;\boldsymbol{\eta}) = \sum_{k=1}^{\infty} I\{\tilde{T}_{ik}(\boldsymbol{\eta}) \leq t\}.$$

Then (2.3) can be expressed as

$$E\{\tilde{N}_i^*(t;\boldsymbol{\eta_0})\} = \mu_0(t) \tag{2.4}$$

where $\mu_0(\cdot)$ is an unknown continuous function. (2.4) indicates that the effect of covariates affects the a number of recurrences across time by accelerating or decelerating the time of the events multiplied by factor $e^{\mathbf{Z}_i^T\boldsymbol{\eta}_0}$ with respect to those of when $\mathbf{Z} = 0$. (Lin et al. 1998). Due to the existence of censoring, we observe the censored version of $N^*(t)$. With the right-censored data, the observable counting processes are defined as

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

where $\tilde{C}_i(\boldsymbol{\eta}) = C_ie^{\mathbf{Z}_i^T\boldsymbol{\eta}}$. Let $Y_i(t;\boldsymbol{\eta}) = I(\tilde{C}_i(\boldsymbol{\eta}) \geq t) = I(C_i \geq te^{-\mathbf{Z}_i^T\boldsymbol{\eta}})$. Then define

$$M_i(t;\boldsymbol{\eta}) = \tilde{N}_i(t;\boldsymbol{\eta}) - \int_0^t Y_i(v;\boldsymbol{\eta})d\mu_0(v),$$

which is equivalent to

$$M_i(t;\boldsymbol{\eta}) = \int_0^t Y_i(v;\boldsymbol{\eta})d\{\tilde{N}_i^*(v;\boldsymbol{\eta}) - \mu_0(v)\}.$$

From (2.4), we have $E\{M_i(t;\boldsymbol{\eta}_0)\} = 0$ for all $i = 1, \ldots, n$. The estimating function proposed by Lin et al. (1998) is

$$\mathbf{S}_n^R(\boldsymbol{\eta}) = \sum_{i=1}^{n} \int_0^{\infty} W(t;\boldsymbol{\eta}) \left[\mathbf{Z}_i - \frac{\sum_{i=1}^{n} \mathbf{Z}_i Y_i(\boldsymbol{\eta})}{\sum_{j=1}^{n} Y_i(\boldsymbol{\eta})} \right] d\tilde{N}_i(t;\boldsymbol{\eta}). \tag{2.5}$$

where $W(\cdot;\boldsymbol{\eta})$ is a weight function. By solving $\mathbf{S}_n^R(\boldsymbol{\eta}) = 0$, we obtain a solution $\hat{\boldsymbol{\eta}}$. In this case, proving asymptotic results is trickier than that in the single event case because $M_i(t;\boldsymbol{\eta}_0)$ is not a martingale unless $N_i^*(t)$ are non-homogeneous Poisson processes (Anderson and Gill, 1982). By employing the modern empirical process theory, Lin et al. (1998) showed that $\hat{\boldsymbol{\eta}}$ is strongly consistent and asymptotically normal. As the single event case, the estimator is the one that minimizes $||\mathbf{S}_n^R(\boldsymbol{\eta})||$.

2.3 Linear Transformation Model

In this section, we use the same notation as in the AFT model for single event. Let T be the time to failure, C be time to independent censoring and $\mathbf{Z}$ be $p \times 1$ vector of covariates. The observable quantities are independent and identically distributed replicates $\{\tilde{T}_i, \Delta_i, \mathbf{Z}_i\}_{i=1}^n$ where $\tilde{T}_i = T_i \wedge C_i$ and $\Delta_i = I(T_i \leq C_i)$, $(i = 1, ..., n)$. It is assumed that T_i is independent of C_i and the survival function of C_i, say G, does not depend on covariates $\mathbf{Z}_i$. Cheng et al. (1995) proposed a generalized form of Cox model, which is called a linear transformation model. The form of the linear transformation model is

$$g\{S_{\mathbf{Z}}(t)\} = h(t) + \mathbf{Z}^T\boldsymbol{\beta},$$

where $S_{\mathbf{Z}}(\cdot)$ is survival function of T given $\mathbf{Z}$, $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression coefficients, $h(\cdot)$ is a unknown increasing function and $g(\cdot)$ is a known decreasing function.The form of this transformation model is equivalent to

$$h(T) = -\mathbf{Z}^T\boldsymbol{\beta} + \epsilon, \tag{2.6}$$

where ϵ is error term whose distribution is completely specified as $F = 1 - g^{-1}$. If F follows the extreme value distribution, then (2.6) is the Cox proportional hazard model. If F has the standard logistic distribution, then (2.6) is a proportional odds model.

Since $h(\cdot)$ is an increasing function, there is no difference between rank of $h(T_i)$ and rank of T_i. Thus it is sensible to use the rank of $h(T_i)$ for statistical inference of regression coefficients. To perform statistical inference for $\boldsymbol{\beta}$, they considered the dichotomous variables $\{I(T_i \geq T_j), i \neq j = 1, ..., n\}$. Then

$$
\begin{aligned}
E\{I(T_i \geq T_j)|\mathbf{Z}_i, \mathbf{Z}_j\} &= P\{h(T_i) \geq h(T_j)|\mathbf{Z}_i, \mathbf{Z}_j\} \\
&= P\{\epsilon_i - \mathbf{Z}_i^T\boldsymbol{\beta} \geq \epsilon_j - \mathbf{Z}_j^T\boldsymbol{\beta}\} = P\{\epsilon_i - \epsilon_j \geq (\mathbf{Z}_i - \mathbf{Z}_j)^T\boldsymbol{\beta}\}.
\end{aligned}
$$

Let $\xi(\mathbf{Z}_{ij}^T\boldsymbol{\beta}) = P\{\epsilon_i - \epsilon_j \geq (\mathbf{Z}_i - \mathbf{Z}_j)^T\boldsymbol{\beta}\}$, where $\mathbf{Z}_{ij} = \mathbf{Z}_i - \mathbf{Z}_j$. It is desirable to derive an unbiased estimating function for $\xi(\mathbf{Z}_{ij}^T\boldsymbol{\beta})$. Then

$$
\begin{aligned}
E\left\{\frac{\Delta_j I(\tilde{T}_i \geq \tilde{T}_j)}{G^2(\tilde{T}_j)}\middle|\mathbf{Z}_i, \mathbf{Z}_j\right\} &= E\left(E\left[\frac{I(T_i \geq T_j)I\{C_i \wedge C_j \geq T_j\}}{G^2(T_j)}\middle|T_j, \mathbf{Z}_i, \mathbf{Z}_j\right]\right) \\
&= E\left(E\{I(T_i \geq T_j)|T_j, \mathbf{Z}_i, \mathbf{Z}_j\}E\left\{\frac{I(C_i \geq T_j)}{G(T_j)}\middle|T_j, \mathbf{Z}_i, \mathbf{Z}_j\right\}\right. \\
&\quad \left. \times E\left\{\frac{I(C_j \geq T_j)}{G(T_j)}\middle|T_j, \mathbf{Z}_i, \mathbf{Z}_j\right\}\right) \\
&= E\{I(T_i \geq T_j)|\mathbf{Z}_i, \mathbf{Z}_j\} = P\{h(T_i) \geq h(T_j)|\mathbf{Z}_i, \mathbf{Z}_j\} \\
&= \xi(\mathbf{Z}_{ij}^T\boldsymbol{\beta}).
\end{aligned}
$$

Then by substituting $G^2(\tilde{T}_j)$ by $\hat{G}^2(\tilde{T}_j)$, which is the Kaplan-Meier estimator of survival function for censoring variable C, an estimating function can be established by

$$
\mathbf{U}(\boldsymbol{\beta}) = \sum_{i=1}^{n}\sum_{j=1}^{n} w(\mathbf{Z}_{ij}^T\boldsymbol{\beta})\mathbf{Z}_{ij}\left\{\frac{\Delta_j I(\tilde{T}_i \geq \tilde{T}_j)}{\hat{G}^2(\tilde{T}_j)} - \xi(\mathbf{Z}_{ij}^T\boldsymbol{\beta})\right\}, \tag{2.7}
$$

where $w(\cdot)$ is a weight function. $\mathbf{U}(\boldsymbol{\beta}) = 0$ has a unique solution asymptotically if $w(\cdot)$ is positive. Let denote this solution be $\hat{\boldsymbol{\beta}}$. Since $\xi(\cdot)$ is smooth function, unlike AFT model, we can use a Newton-Raphson algorithm to obtain $\hat{\boldsymbol{\beta}}$. Let $\boldsymbol{\beta}_0$ be true value of $\boldsymbol{\beta}$. By using a Taylor expansion of $\mathbf{U}(\hat{\boldsymbol{\beta}})$ around $\boldsymbol{\beta}_0$, the authors showed that estimator $\hat{\boldsymbol{\beta}}$ is asymptotically normal. If the survival function of C depends on $\mathbf{Z}$, and $\mathbf{Z}$ can be discretized into K distinct values, then the estimating function can be changed to

$$\tilde{\mathbf{U}}(\boldsymbol{\beta}) = \sum_{i=1}^{n}\sum_{j=1}^{n} w(\mathbf{Z}_{ij}^T\boldsymbol{\beta})\mathbf{Z}_{ij}\left\{\frac{\Delta_j I(\tilde{T}_i \geq \tilde{T}_j)}{\hat{G}_{\mathbf{Z}_i}(\tilde{T}_j)\hat{G}_{\mathbf{Z}_j}(\tilde{T}_j)} - \xi(\mathbf{Z}_{ij}^T\boldsymbol{\beta})\right\}, \tag{2.8}$$

where $\hat{G}_{\mathbf{Z}}(\cdot)$ is the Kaplan-Meier estimator for the survival function of the censoring variable C from $\{\tilde{T}_l, \Delta_l\}_{l=1}^n$ where $\mathbf{Z}_l = \mathbf{Z}(l = 1, ..., n)$. Then the solution of $\tilde{\mathbf{U}}(\boldsymbol{\beta}) = 0$, say $\tilde{\boldsymbol{\beta}}$, is also asymptotically normal.

Chapter 3 Review of Dependent Censoring

3.1 Introduction

The semicompeting risks data structure has been studied for a while. This data structure provides a nice way to address the dependent censoring problem. The key feature of this framework is to deal with two distinct events simultaneously. Methodology from this framework can be easily extended to recurrent events under the dependent censoring. In this chapter, some literature related to semicompeting risks and recurrent events under the dependent censoring will be reviewed.

3.2 Semicompeting Risks Data

3.2.1 Basic Concepts and Model Without Covariates

Semicompeting risks data is first proposed by Fine et al. (2001). According to them, in the semicompeting risks data structure, there are two types of events: a nonterminal event and a terminal event. In the medical study, a nonterminal event can be any failure of the researcher's interest except death while a terminal event can be death. In this structure, a terminal event may censor the nonterminal event, but the nonterminal event does not stop the occurrence of the terminal event. In fact, semicomepting risks data have a different characteristic compared to competing risks data. In competing risks data, several distinct and terminal events exist, and we only observe minimum of several terminal events. In semicompeting risks data, a nonterminal event and a terminal event exist. It is possible to observe a nonterminal event without the terminal event, the terminal event without the nonterminal event or we can observe the both events (Ghosh, 2006). Datasets of examples in the Chapter 1 clearly fall into this semicompeting risks structure. In Example 1 from Chapter 1, death clearly censors relapse, but the relapse does not stop the observance of death. In Example 2 from Chapter 1, the withdrawal stops virologic failure, but the virologic failure does not prevent the observance of the withdrawal.

Now we will explain the structure of semicompeting risks data. Let X be time to a nonterminal event and and D be time to a terminal event. Let C be an independent censoring time, which is censoring time because of random loss to follow up and the end of study (Lin et al. 1996). It is assumed that C is independent of (X, D). In this setup, define $\tilde{V} = X \wedge D,\ \kappa = I(X \leq D),\ \Delta = I(D \leq C),\ \tilde{D} = D \wedge C,\ \tilde{X} = \tilde{V} \wedge C = X \wedge D \wedge C,\ \zeta = I(V \leq C) = I(X \wedge D \leq C)$. The observed data, which is called semi-competing risks data, consist of n independent and identically distributed collections $\{\tilde{X}_i, \tilde{D}_i, \zeta_i, \Delta_i, \kappa_i\zeta_i\}_{i=1}^n$.

This data structure contains many advantages. First, if failure is not statistically independent with censoring, the joint distribution of failure time and censoring is nonidentifiable nonparametrically (Tsiatis, 1975). However, in this semicompeting risks setting, the joint distribution of the nonterminal event and the terminal event, say X and D, is nonparametrically identifiable in the upper wedge when $X < D$. Note that there is no assumption of specific dependence structure between the nonterminal event and the terminal event. It is also noticeable that semicompeting data structure can be applied not only to the medical study, but also any studies containing a nonterminal event and a terminal event.

In the medical study, a biological marker takes an important role when making prognosis for a patient. Day et al. (1997) considered the recurrence of a disease and biomarker positivity, where the biomarker positivity means that a patient is flagged as high risk. The goal of Day et al. (1997) was to find whether the appearance of marker can change the patient's subsequent risk which is sufficient to trigger to use treatment for a disease. The useful quantity to evaluate predictive power of marker is the predictive hazard ratio, which is defined by

$$R(s,t) = \frac{\lambda(t|\{s\})}{\lambda(t|(s,\infty])},$$

where for $A \subseteq (0,\infty)$, the definition of $\lambda(t|A)$ is

$$\lambda(t|A) = \lim_{\epsilon\to 0} \frac{d}{d\epsilon} P(T_R < t+\epsilon | T_R \geq t, T_F \in A),$$

and T_R is the recurrence time of the disease and T_F is the first time marked as high risk. This predictive hazard ratio is the ratio of the recurrence hazard rate at time t of a subject marked as high risk at s to that of a subject marked as high risk after time s. It borrows from the gamma frailty model from Clayton (1978) in that T_R and T_F are conditionally independent given an unobserved frailty when this predictive hazard ratio is constant that is greater than 1. Due to the definition of the predictive hazard ratio and the purpose of prediction, it is reasonable to assume a constant value of the predictive hazard ratio in the area $0 \leq s \leq t < \infty$. In the lower wedge, it is most likely that the conditional independence of T_R and T_F does not hold so that it is not sensible to assume a constant value of the predictive ratio. They shows that this assumption is sufficient to guarantee the conditional independence of T_R and T_F in the gamma frailty model in the upper wedge.

For modeling semicompeting risks data, they extended the idea of Day et al. (1997) of using bivariate frailty models for predicting the future risk of a patient when biological marker exists. They defined a model in the area $X < D$ without covariates. For $\theta \geq 1$ and $0 \leq s \leq t \leq \infty$,

$$F(s,t) = (F_x^{1-\theta}(s) + F_d^{1-\theta}(t) - 1)^{1/(1-\theta)}, \tag{3.1}$$

where F_x and F_d satisfy the definition of survival functions. Note that by the identifiability issue, this model is identifiable in the area $X < D$ only. Based on the model, they proposed a new estimator by extending arguments of Oakes' (1982, 1986). Define $\delta_{ij} = I\{(X_i - X_j)(D_i - D_j) > 0\}$, $\tilde{X}_{ij} = X_i \wedge X_j$, $\tilde{D}_{ij} = D_i \wedge D_j$, $\tilde{C}_{ij} = C_i \wedge C_j$, $\Delta_{ij} = I(\tilde{X}_{ij} \leq \tilde{D}_{ij} \leq \tilde{C}_{ij})$, $\tilde{X}^*_{ij} = \tilde{X}_{ij} \wedge \tilde{D}_{ij} \wedge \tilde{C}_{ij}$, and $\tilde{D}^*_{ij} = \tilde{D}_{ij} \wedge \tilde{C}_{ij}$. Due to the assumption of semicompeting risks data, δ_{ij} is only calculable when $\Delta_{ij} = 1$. Under model (3.1), $E\{E(\delta_{ij}|\tilde{X}^*_{ij}, \tilde{D}^*_{ij})|\Delta_{ij} = 1\} = \theta_0/(1+\theta_0)$, where θ_0 is the true value of θ because the predictive hazard ratio takes constant value θ_0 in the upper wedge. The estimating function for θ_0 is

$$U(\theta) = \sum_{i<j} W(\tilde{X}^*_{ij}, \tilde{D}^*_{ij})\Delta_{ij}\left\{\delta_{ij} - \frac{\theta}{1+\theta}\right\}. \tag{3.2}$$

The proposed estimator is

$$\hat{\theta} = \frac{\sum_{i<j} W(\tilde{X}^*_{ij}, \tilde{D}^*_{ij})\Delta_{ij}\delta_{ij}}{\sum_{i<j} W(\tilde{X}^*_{ij}, \tilde{D}^*_{ij})(1-\delta_{ij})\Delta_{ij}}, \tag{3.3}$$

where $W(a,b)$ is a weighted random function satisfying $\sup_{a,b}|W(a,b) - \tilde{W}(a,b)|$ converges to 0, $\tilde{W}(a,b)$ is a nonrandom function and bounded for (a,b) in the support of $(\tilde{X}^*_{ij}, \tilde{D}^*_{ij})$. They also showed that this estimator is asymptotical normal.

It is also of interest to investigate an estimator of F_x. In semicomepting risks data, the usual Kaplan-Meier estimator of F_x computed from $\{\tilde{X}_i, \kappa_i\zeta_i\}_{i=1}^n$ does not converge to F_x. Fine et al. (2001) proposed the plug-in estimator of F_x and the inference procedure of F_x. They also provided asymptotic results of this estimator.

3.2.2 Model with Covariates

Although the idea of Fine et al. (2001) is pioneering and has theoretical advantage, the model does not include effects of covariates. Moreover, their model assumes the dependence structure between the nonterminal event and the terminal event. Thus its application is quite limited. Lin et al. (1996) and Peng and Fine (2006) suggested a nice solution to this issue by establishing a model incorporating effects of covariates. They proposed bivariate AFT models for the data structure which is the same as that of Fine et al. (2001).

We will discuss the data structure and notations used in Lin et al. (1996) and Peng and Fine (2006). For simplicity of notations, let X be logarithm of time to the event of interest, D be logarithm of time to the dependent censoring, C be logarithm of time to the independent censoring and $\mathbf{Z}$ be $p \times 1$ a vector of covariates. Define $\tilde{X} = X \wedge D \wedge C$, $\tilde{D} = D \wedge C$, $\Delta = I(D \leq C)$ and $\delta = I(X \leq \tilde{D})$. The observed data are independent and identical copies $(\tilde{X}_i, \tilde{D}_i, \Delta_i, \delta_i, \mathbf{Z}_i), i = 1...., n$.

The regression model is

$$\begin{pmatrix} D_i = \mathbf{Z}_i^T \boldsymbol{\eta}_0 + \epsilon_i^D \\ X_i = \mathbf{Z}_i^T \boldsymbol{\theta}_0 + \epsilon_i^X \end{pmatrix}, \quad i = 1, ..., n, \tag{3.4}$$

where $\boldsymbol{\theta}_0$ and $\boldsymbol{\eta}_0$ are $p \times 1$ vectors of regression coefficients, and $\epsilon_i \equiv (\epsilon_i^X, \epsilon_i^D)$ are independent and identically distributed error terms that are independent of $\mathbf{Z}_i$. In this case, we assume that the model is identifiable only in upper wedge $X < D$ (Fine et al. 2001; Peng and Fine, 2006). We assume that ϵ has unknown bivariate distribution F. The goal is to obtain an estimator of $\boldsymbol{\alpha}_0 = (\boldsymbol{\eta}_0^T, \boldsymbol{\theta}_0^T)^T$ without estimating the distribution of ϵ_is. We further assume that given $\mathbf{Z}$, C and (X, D) are independent, but X and D can be dependent given $\mathbf{Z}$.

Now we review the models of Lin et al. (1996) and Peng and Fine (2006). The procedure of estimating of $\boldsymbol{\eta}_0$ is straightforward because it is only related to independent censoring. The estimating function for $\boldsymbol{\eta}_0$ is

$$\mathbf{S}_n(\boldsymbol{\eta}) = n^{-1/2} \sum_{i=1}^{n} \Delta_i \left[\mathbf{Z}_i - \frac{\sum_{j=1}^{n} \mathbf{Z}_j I\{\tilde{D}_j^*(\boldsymbol{\eta}) \geq \tilde{D}_i^*(\boldsymbol{\eta})\}}{\sum_{j=1}^{n} I\{\tilde{D}_j^*(\boldsymbol{\eta}) \geq \tilde{D}_i^*(\boldsymbol{\eta})\}} \right], \tag{3.5}$$

where $\tilde{D}_i^*(\boldsymbol{\eta}) = \tilde{D}_i - \mathbf{Z}_i^T \boldsymbol{\eta}$ and the estimator for $\boldsymbol{\eta}$ is the solution of $\mathbf{S}_n(\boldsymbol{\eta}) = 0$.

For estimation of $\boldsymbol{\theta}_0$, one may want to replace $\tilde{D}_i(\boldsymbol{\eta})$ by $\tilde{X}_i - \mathbf{Z}_i^T \boldsymbol{\theta}$. However, it does not guarantee to obtain unbiased estimator of $\boldsymbol{\theta}$. We examine why this bias can occur. Let Z be a treatment indicator, i.e, $Z = 1$ if a person receives treatment and $Z = 0$ if a person does not receive treatment. Suppose that $b_Z(t)$ is a cause-specific hazard function for treatment group and control group. Then $b_Z(t)$ is

$$b_Z(t) = \lim_{\Delta t \to 0} \frac{P\{t \leq X - \theta_0 Z < t + \Delta t | X - \theta_0 Z \geq t, D - \theta_0 Z \geq t)}{\Delta t}. \tag{3.6}$$

Note that the distribution of $D - \theta_0 Z$ depends on Z unless $\theta_0 = \eta_0$, which is a violation of the assumption. This results in $b_0(t) \neq b_1(t)$, and it implies that this naive estimating function for θ_0 causes a bias.

To solve this problem, Lin et al. (1996) and Peng and Fine (2006) employed artificial censoring to remove the bias of the estimator of the model in the presence of the dependent censoring. The motivation of artificial censoring is to make observations have the common distribution which is independent of covariates so that observations can be comparable (Ding et al. 2009). In the causal inference context, Joffe (2001) showed that the estimator of the model implemented artificial censoring technique is unbiased in the simulation study. Lin et al. (1996) and Peng and Fine (2006) applied this idea on rank-based AFT model. We will see how this technique is utilized numerically when we review models of Lin et al. (1996) and Peng and Fine (2006) in detail.

Figure 3.1 and 3.2 show the mechanism of artificial censoring when the censoring variable C is absent. Suppose that we have only a treatment indicator Z and consider $\theta_0 > \eta_0$. Suppose that $\theta_0 = 2$ and $\eta_0 = 1$. In Figure 3.1 and 3.2, the horizontal axis and vertical axis are transformed times $X - \theta_0 Z$ and $D - \eta_0 Z$, respectively. In these figures, the solid line and dashed line are lines of censoring when $Z = 1$ and $Z = 0$, respectively. In Figure 3.1, when $Z = 0$, observable data points are located in the area above the dashed line. However, observable data points belong to the area above the solid line when $Z = 1$. Thus if randomly selected observations from the treatment group and the control group belong to the shaded area in Figure 3.1, they are not be comparable. The shaded area between the dashed line and the solid line should be adjusted for the comparison. In this case, for control group, originally uncensored observations may be censored because observations located above the solid line are only compatible and independent of Z. Now, let us consider the opposite situation, i.e, when $\eta_0 > \theta_0$. Suppose that $\eta_0 = 2$ and $\theta_0 = 1$. As opposite to the previous case, the area of observable data when $Z = 1$ is wider than that when $Z = 0$. Thus it is necessary to adjust the treatment group to make these observations compatible. In this case, uncensored observations in the treatment group may be censored. Figure 3.2 explains this phenomenon.

Lin et al. (1996) considered this simple case when Z is a treatment indicator and provided a solution for this dependent censoring problem. Let $g(\beta) = \theta - \eta$ if $\theta \geq \eta$ and $g(\beta) = 0$ otherwise. Then they defined $D_i - \eta Z_i - d$ and $C_i - \eta Z_i - d$. These two quantities provide new quantities, say $\tilde{X}_i^*(\alpha)$ and $\tilde{\delta}_i^*(\alpha)$, where

$$\begin{aligned}\tilde{X}_i^*(\alpha) &= (X_i - \theta Z_i) \wedge \{D_i - \eta Z_i - g(\beta)\} \wedge \{C_i - \eta Z_i - g(\beta)\}\\ \tilde{\delta}_i^*(\alpha) &= I[\{(X_i - \theta Z_i) \leq \{D_i - \eta Z_i - g(\beta)\} \wedge \{C_i - \eta Z_i - g(\beta)\}].\end{aligned}$$

Figure 3.1: Artificial censoring when $\theta_0 > \eta_0$, $\theta_0 = 2$ and $\eta_0 = 1$

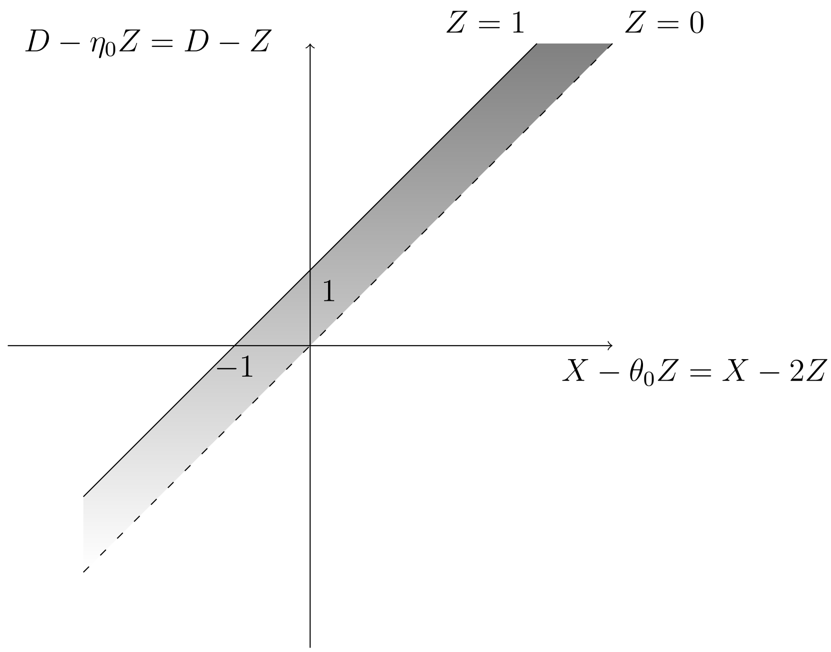
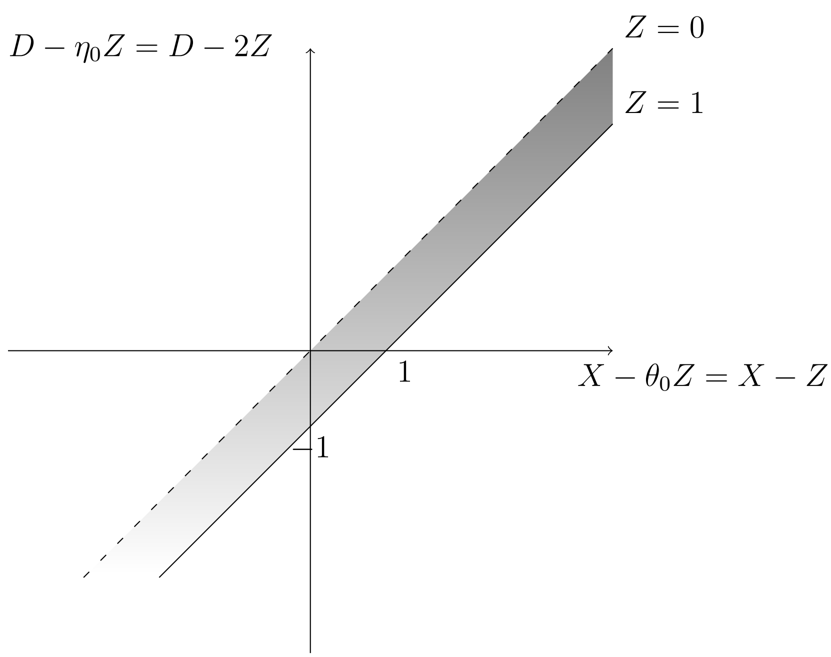


Figure 3.2: Artificial censoring when $\theta_0 < \eta_0$, $\theta_0 = 1$ and $\eta_0 = 2$



In fact, using d leads to the transformation from uncensored observations to censored ones, as can be seen in the mechanism. Moreover, this formulation of d exactly coincides the mechanism, which may censor observations in the treatment group when $\eta > \theta$ and in the control group when $\eta < \theta$ (Ghosh and Lin, 2003). As a result, $\tilde{X}_i^*(\alpha)$ and $\tilde{\delta}_i^*(\alpha)$ correct the bias caused by the dependent censoring. Then the estimating function for θ_0 according to Lin et al. (1996) is

$$U_n^L(\alpha) = n^{-1/2} \sum_{i=1}^{n} \tilde{\delta}_i^*(\alpha) \left[Z_i - \frac{\sum_{j=1}^n Z_j I\{\tilde{X}_j^*(\alpha) \geq \tilde{X}_i^*(\alpha)\}}{\sum_{j=1}^n I\{\tilde{X}_j^*(\alpha) \geq \tilde{X}_i^*(\alpha)\}} \right]. \tag{3.7}$$

The proposed estimator of θ_0 according to Lin et al. (1996) is the solution of $U_n^L\{(\hat{\eta}, \theta)^T\} = 0$, where $\hat{\eta}$ is an estimator of η_0. Let us denote this solution as $\hat{\alpha}^L = (\hat{\eta}, \hat{\theta})^T$.

For statistical inference, the key issue is calculating covariance matrix of $\hat{\alpha}$. To calculate covariance matrix of $\hat{\alpha}^L$, first it is necessary to derive covariance matrix of $U_n^L(\hat{\alpha}^L)$. In this case, the covariance matrix of $U_n^L(\hat{\alpha}^L)$ can be derived using the martingale structure of $U_n^L(\alpha)$. However, the asymptotic covariance matrix of $\hat{\alpha}^L$ is very difficult. Its asymptotic matrix has the form $\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}$, where $\mathbf{A}$ is a constant matrix and $\mathbf{B}$ is covariance matrix of $U_n^L(\alpha_0)$, where $\alpha_0 = (\eta_0, \alpha_0)^T$. Computing $\mathbf{A}$ requires the estimation of density of error terms, which is not desirable in our estimation framework (Peng and Fine, 2006).

To solve this issue, Lin et al. (1996) adapted methodology developed by Parzen et al. (1994). Parzen et al. (1994) proposed a resampling method by solving the estimating function iteratively. Let $\boldsymbol{\beta}$ be $p \times 1$ vector of parameters. The goal is to perform statistical inference of $\boldsymbol{\beta}$ by using $V(\boldsymbol{\beta})$, where $V(\boldsymbol{\beta})$ is an exactly or asymptotically pivotal estimating function, which means that we can derive the exact distribution or asymptotic distribution of $V(\boldsymbol{\beta})$ from some vector r, where $\mathbf{r}$ has a known distribution or it is possible to estimate the distribution of r consistently. Then it is possible to construct a stochastic equation $V(\boldsymbol{\beta}) = \mathbf{r}$. Let $\tilde{\boldsymbol{\beta}}$ be a solution of this equation. Under some regularity conditions, the conditional distribution of $(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}})$ given observed data is equal to the unconditional distribution of $(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$, where $\boldsymbol{\beta}_0$ is the true value of $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}$ is a consistent estimator of $\boldsymbol{\beta}_0$. To obtain the distribution of $\tilde{\boldsymbol{\beta}}$, we can generate large realizations of $\tilde{\boldsymbol{\beta}}$ by solving $V(\boldsymbol{\beta}) = \mathbf{r}$ sufficiently many times.

Adapting this approach of Parzen et al. (1994), Lin et al. (1996) suggested a resampling approach to estimate the covariance matrix of $\hat{\alpha}^L$ rather than computing $\mathbf{A}$ directly. Let $S_n(\eta)$ be an estimating function for η_0. This is same as the estimating function of (3.5) except considering a single covariate in this case. The proposed resampling approach is to solve $V_n^L(\alpha) = -u$, where

$$V_n^L(\alpha) = \begin{pmatrix} S_n(\eta) \\ U_n^L(\alpha) \end{pmatrix}, \tag{3.8}$$

and u follows normal distribution with mean 0 and covariance matrix $\hat{\mathbf{B}}$, where $\hat{\mathbf{B}}$ is covariance matrix of $U_n^L(\hat{\alpha}^L)$. Let solutions of the equation $V_n^L(\alpha) = -u$ be $\tilde{\alpha}^L$. By Parzen et al. (1994), the conditional distribution of $n^{1/2}(\tilde{\alpha}^L - \hat{\alpha}^L)$ given the data is the same as the unconditional distribution of $n^{1/2}(\hat{\alpha}^L - \alpha_0)$ asymptotically, where $\hat{\alpha}^L$ is an estimator proposed by Lin et al. (1996). Then statistical inference can be performed by these realizations, say $\tilde{\alpha}_1^L, ..., \tilde{\alpha}_B^L$. Clearly, the arguments of Lin et al. (1996) can be easily extended to the multiple covariates case. Then an estimating function of Lin et al. (1996) for multiple covariates is

$$\mathbf{U}_n^L(\boldsymbol{\alpha}) = n^{-1/2} \sum_{i=1}^{n} \tilde{\delta}_i^*(\boldsymbol{\alpha}) \left[\mathbf{Z}_i - \frac{\sum_{j=1}^{n} Z_j I\{\tilde{X}_j^*(\boldsymbol{\alpha}) \geq \tilde{X}_i^*(\boldsymbol{\alpha})\}}{\sum_{j=1}^{n} I\{\tilde{X}_j^*(\boldsymbol{\alpha}) \geq \tilde{X}_i^*(\boldsymbol{\alpha})\}} \right]. \tag{3.9}$$

where

$$\begin{aligned} g(\boldsymbol{\alpha}) &= \max_{1 \leq i \leq n} \{0, \mathbf{Z}_i^T(\boldsymbol{\theta} - \boldsymbol{\eta})\} \\ \tilde{X}_i^*(\boldsymbol{\alpha}) &= (X_i - \mathbf{Z}_i^T\boldsymbol{\theta}) \wedge \{D_i - \mathbf{Z}_i^T\boldsymbol{\eta} - g(\boldsymbol{\beta})\} \wedge \{C_i - \mathbf{Z}_i^T\boldsymbol{\eta} - g(\boldsymbol{\beta})\} \\ \tilde{\delta}_i^*(\boldsymbol{\alpha}) &= I[(X_i - \mathbf{Z}_i^T\boldsymbol{\theta}) \leq \{D_i - \mathbf{Z}_i^T\boldsymbol{\eta} - g(\boldsymbol{\beta})\} \wedge \{C_i - \mathbf{Z}_i^T\boldsymbol{\eta} - g(\boldsymbol{\beta})\}]. \end{aligned}$$

The other important contribution of Lin et al. (1996) is goodness of fit statistics. They first established score processes based on martingale residuals. To check behavior of the model, they constructed bootstrapped processes by generating standard normal random variables and using realizations from resampling. From these two processes, visual model checking can be conducted by plotting 20-30 bootstrapped processes with score processes. Empirical p-values can be calculated from Kolmogorov-Smirnov type statistics.

Peng and Fine (2006) also considered the semicompeting risks data structure from Fine et al. (2001). They considered multiple covariates in their analysis. They also applied the artificial censoring idea to the model, but used pairwise comparing methods to reduce the degree of the artificial censoring. They argued that these pairwise comparisons may lead to a more efficient estimator of $\boldsymbol{\alpha}_0$ than that in Lin et al. (1996) because the amount of artificial censoring is less than that in Lin et al. (1996) by pairwise comparisons. Estimating $\boldsymbol{\eta}_0$ is identical to that of Lin et al. (1996). For estimating $\boldsymbol{\theta}_0$, define

$$\begin{aligned} g_{ij}(\boldsymbol{\alpha}) &= \max\{0, \mathbf{Z}_i^T(\boldsymbol{\theta}-\boldsymbol{\eta}), \mathbf{Z}_j^T(\boldsymbol{\theta}-\boldsymbol{\eta})\} \\ \tilde{X}^*_{i(j)}(\boldsymbol{\alpha}) &= (X_i - \mathbf{Z}_i^T\boldsymbol{\theta}) \wedge \{D_i - \mathbf{Z}_i^T\boldsymbol{\eta} - g_{ij}(\boldsymbol{\beta})\} \wedge \{C_i - \mathbf{Z}_i^T\boldsymbol{\eta} - g_{ij}(\boldsymbol{\beta})\} \\ \tilde{\delta}^*_{i(j)}(\boldsymbol{\alpha}) &= I[(X_i - \mathbf{Z}_i^T\boldsymbol{\theta}) \le \{D_i - \mathbf{Z}_i^T\boldsymbol{\eta} - g_{ij}(\boldsymbol{\beta})\} \wedge \{C_i - \mathbf{Z}_i^T\boldsymbol{\eta} - g_{ij}(\boldsymbol{\beta})\}], \end{aligned}$$

$$\phi_{ij}(\boldsymbol{\alpha}) = \tilde{\delta}^*_{i(j)}(\boldsymbol{\alpha}) I\{\tilde{X}^*_{i(j)}(\boldsymbol{\alpha}) \le \tilde{X}^*_{j(i)}(\boldsymbol{\alpha})\} - \tilde{\delta}^*_{j(i)}(\boldsymbol{\alpha}) I\{\tilde{X}^*_{j(i)}(\boldsymbol{\alpha}) \le \tilde{X}^*_{i(j)}(\boldsymbol{\alpha})\}.$$

Then the estimating function for $\boldsymbol{\theta}_0$ according to Peng and Fine (2006) is

$$\mathbf{U}_n^P(\boldsymbol{\alpha}) = \frac{2n^{1/2}}{n(n-1)} \sum_{1 \le i < j \le n} (\mathbf{Z}_i - \mathbf{Z}_j)\phi_{ij}(\boldsymbol{\alpha}). \tag{3.10}$$

The proposed estimator of $\boldsymbol{\theta}_0$ according to Peng and Fine (2006) is a solution of $\mathbf{U}_n^P\{(\hat{\boldsymbol{\eta}}^T, \boldsymbol{\theta}^T)^T\} = 0$. They also estimated the covariance matrix for the estimator of $\boldsymbol{\alpha}$ by using Parzen et al. (1994) and proved that the estimator is strongly consistent and asymptotically normal. For estimating the covariance matrix of their estimator $\hat{\boldsymbol{\alpha}}^P$, as Lin et al. (1996), they implemented the resampling technique by Parzen et al. (1994).

As the estimating equation of AFT model for a single event under the independent censoring assumption, solutions of $\mathbf{U}_n^L(\boldsymbol{\alpha})$ and $\mathbf{U}_n^P(\boldsymbol{\alpha})$ are defined as $\hat{\boldsymbol{\theta}}^L = \arg\min_{\boldsymbol{\theta}} ||\mathbf{U}_n^L\{(\hat{\boldsymbol{\eta}}^T, \boldsymbol{\theta}^T)^T\}||$ and $\hat{\boldsymbol{\theta}}^P = \arg\min_{\boldsymbol{\theta}} ||\mathbf{U}_n^P\{(\hat{\boldsymbol{\eta}}^T, \boldsymbol{\theta}^T)^T\}||$, respectively (Peng and Fine, 2006). In computation, algorithms used to obtain estimators in the AFT model in Chapter 2 can be still applied.

The estimators proposed by Lin et al. (1996) and Peng and Fine (2006) are useful in two respects. First, their methods include effects of covariates and the resampling method is quite easy to implement. Second, these are theoretically well-justified. They proved that their estimators are strongly consistent and asymptotically normal by using martingale theory of Fleming and Harrington (2005, Chapter 5, pp. 201-228) and arguments in Ying (1993), and U-statistics theory in Honoré and Powell (1994).

It is sensible to extend this result to general function of X and D. Ding et al. (2009) proposed a generalization of the previous models. Note that the observed data are $\{\tilde{X}_i, \tilde{D}_i, \Delta_i, \delta_i, \mathbf{Z}_i\}_{i=1}^n$. The model is

$$\begin{pmatrix} h_1(X_i) = \mathbf{Z}_i^T \boldsymbol{\theta}_0 + \epsilon_i^X \\ h_2(D_i) = \mathbf{Z}_i^T \boldsymbol{\eta}_0 + \epsilon_i^D \end{pmatrix}, \quad i = 1, ..., n, \tag{3.11}$$

where $\boldsymbol{\theta}_0$ and $\boldsymbol{\eta}_0$ are $p \times 1$ vectors of regression coefficients, and h_1 and h_2 are monotone functions of X_i and D_i, respectively. h_1 is a known function but h_2 may be known or unknown. If h_2 is specified, then ϵ_i^D is unspecified. If h_2 is unspecified, to avoid the nonidentifiability issue, ϵ_i^D is specified. Let $X^H(\boldsymbol{\theta}) = h_1(X) - \mathbf{Z}^T\boldsymbol{\theta}$, $D^H(\boldsymbol{\eta}) = h_2(D) - \mathbf{Z}^T\boldsymbol{\eta}$ and $C^H(\boldsymbol{\eta}) = h_2(C) - \mathbf{Z}^T\boldsymbol{\eta}$. Define $\tilde{D}^H(\boldsymbol{\eta}) = D^H(\boldsymbol{\eta}) \wedge C^H(\boldsymbol{\eta})$. The estimator for $\boldsymbol{\eta}_0$ is obtained by solving $\mathbf{S}_n^H(\boldsymbol{\eta}) = 0$, where

$$\mathbf{S}_n^H(\boldsymbol{\eta}) = n^{-1/2} \sum_{i=1}^n \Delta_i \left[\mathbf{Z}_i - \frac{\sum_{j=1}^n \mathbf{Z}_j I\{\tilde{D}_j^H(\boldsymbol{\eta}) \geq \tilde{D}_i^H(\boldsymbol{\eta})\}}{\sum_{j=1}^n I\{\tilde{D}_j^H(\boldsymbol{\eta}) \geq \tilde{D}_i^H(\boldsymbol{\eta})\}} \right]. \tag{3.12}$$

where $\tilde{D}_i^H(\boldsymbol{\eta})$ are realizations of $\tilde{D}^H(\boldsymbol{\eta})$. Denote $\boldsymbol{\alpha} = (\boldsymbol{\eta}^T, \boldsymbol{\theta}^T)^T$. In this case, since general functions $h_1(t)$ and $h_2(t)$ are considered, in order to employ the artificial censoring mechanism of Lin et al. (1996), a new censoring time should be defined. They established the artificial censoring mechanism for general functions $h_1(t)$ and $h_2(t)$, say $R_{\boldsymbol{\alpha}}(t)$, which is defined by

$$R_{\boldsymbol{\alpha}}(t) = \inf_{z \in \Omega} h_1 \circ h_2^{-1}(t + \mathbf{z}^T\boldsymbol{\theta}) - \mathbf{z}^T\boldsymbol{\eta}.$$

Note that when $h_1(t) = h_2(t) = \log(t)$, $R_{\boldsymbol{\alpha}}(t)$ reduces the artificial censoring scheme by Lin et al. (1996). By using this general scheme, they defined several new quantities to adjust the dependent censoring, where

$$\tilde{D}_i^{H*}(\boldsymbol{\alpha}) = R_{\boldsymbol{\alpha}}\{D_i^H(\boldsymbol{\eta}) \wedge C_i^H(\boldsymbol{\eta})\} = R_{\boldsymbol{\alpha}}\{\tilde{D}_i^H(\boldsymbol{\eta})\}$$
$$\tilde{X}_i^{H*}(\boldsymbol{\alpha}) = \{h_1(X_i) - \boldsymbol{Z}_i^T\boldsymbol{\theta}\} \wedge \tilde{D}_i^{H*}(\boldsymbol{\alpha})$$
$$\tilde{\delta}_i^{H*}(\boldsymbol{\alpha}) = I[\{h_1(X_i) - \boldsymbol{Z}_i^T\boldsymbol{\theta}\} \leq \tilde{D}_i^{H*}(\boldsymbol{\alpha})]$$

Then, similar to Lin et al (1996), the estimating function for unbiased estimator of $\boldsymbol{\theta}_0$ is given by

$$\mathbf{U}_n^H(\boldsymbol{\alpha}) = n^{-1/2}\sum_{i=1}^{n}\tilde{\delta}_i^{H*}(\boldsymbol{\alpha})\Bigg[\mathbf{Z}_i - \frac{\sum_{j=1}^{n}\mathbf{Z}_j I\{\tilde{X}_j^{H*}(\boldsymbol{\alpha}) \geq \tilde{X}_i^{H*}(\boldsymbol{\alpha})\}}{\sum_{j=1}^{n} I\{\tilde{X}_j^{H*}(\boldsymbol{\alpha}) \geq \tilde{X}_i^{H*}(\boldsymbol{\alpha})\}}\Bigg]. \tag{3.13}$$

Then they implemented the resampling technique which is similar to Lin et al. (1996) for inference of their estimators and proposed a new computational method for multiple covariates by modifying linear programming technique from Jin et al. (2003).

Moreover, Ding et al. (2009) proposed an inference procedure if D has the linear transformation model. The step for this inference is as follows. Let survival function of D be $\tilde{S}_D(t) = P(\epsilon^D > t)$ and baseline survival function $S_D(t) = P(D > t|Z = 0)$. The procedure is as follows:

1. Estimate a uniformly consistent estimator of survival function $S_D(t)$, say $\hat{S}_D(t)$. Calculate $\hat{h}_2(t) = \tilde{S}_D^{-1} \circ \hat{S}_D(t)$ and $\hat{R}_{\boldsymbol{\alpha}}(t) = \inf_{z\in\Omega} h_1[\hat{S}_D^{-1}\{\tilde{S}_D(t+\mathbf{z}^T\boldsymbol{\eta})\}]-\mathbf{z}^T\boldsymbol{\theta}$.

2. Replace $h_2(t)$ with $\hat{h}_2(t)$ and $R_{\boldsymbol{\alpha}}(t)$ with $\hat{R}_{\boldsymbol{\alpha}}(t)$ in $\mathbf{S}_n^H(\boldsymbol{\eta})$ and $\mathbf{U}_n^H(\boldsymbol{\alpha})$, respectively. Then we will get new estimating functions for $\boldsymbol{\eta}$ and $\boldsymbol{\theta}$ from these quantities. Denote these new functions to be $\tilde{\mathbf{S}}_n^H(\boldsymbol{\eta})$ and $\tilde{\mathbf{U}}_n^L(\boldsymbol{\alpha})$. Then by solving $\tilde{\mathbf{S}}_n^H(\boldsymbol{\eta}) = 0$ and $\tilde{\mathbf{U}}_n^H(\boldsymbol{\alpha}) = 0$, estimators for $\boldsymbol{\eta}$ and $\boldsymbol{\alpha}$ under the new model can be obtained.

Considering general functions h_1 and h_2, many combinations of two models can be considered. The issue is to choose the best combination of the models. To choose best models, they used p-values from goodness of fit test statistics.

3.3 Model for Recurrent Events in the Presence of Dependent Censoring

In this section, we will discuss models of recurrent events under the existence of dependent censoring. These models have been proposed by extending results in the Lin et al. (1996) and Peng and Fine (2006).

Ghosh and Lin (2003) extended the approach of Lin et al. (1998) which established AFT model in recurrent events under independent censoring. Moreover, this paper is a generalization of Lin et al. (1996) to the case of recurrent events. Let $N^*(t)$ be the number of recurrent events that occur over the time interval $[0, t]$ without censoring, D be time to the terminal event, C be time to independent censoring and $\mathbf{Z}$ be $p \times 1$ a vector of covariates. It is assumed that C is independent with both $N^*(\cdot)$ and D given $\mathbf{Z}$, but $N^*(\cdot)$ and D can be dependent even given $\mathbf{Z}$. Let $\{N_i^*(\cdot), D_i, C_i, \mathbf{Z}_i\}_{i=1}^n$ be independent and identically distributed copies of $\{N^*(\cdot), D, C, \mathbf{Z}\}$. Given $\mathbf{Z}_i$, $\{D_i e^{-\mathbf{Z}_i^T \boldsymbol{\eta}_0}, N_i^*(te^{\mathbf{Z}_i^T \boldsymbol{\theta}_0})\}$ are independent and identically distributed, but their joint distribution is unknown, where $\boldsymbol{\eta}_0$ and $\boldsymbol{\theta}_0$ are $p \times 1$ vectors of regression coefficients. This joint distribution is expressed as follows:

$$\begin{pmatrix} D_i e^{-\mathbf{Z}_i^T \boldsymbol{\eta}_0} \\ N_i^*(te^{\mathbf{Z}_i^T \boldsymbol{\theta}_0}) \end{pmatrix} \overset{d}{=} \begin{pmatrix} D_0 \\ N_0^*(t) \end{pmatrix}, i = 1, ..., n, \tag{3.14}$$

where $\{D_0, N_0^*(t)\}^T$ is a bivariate distribution and $\overset{d}{=}$ means equal in distribution. The observable data are $\{N_i(\cdot), \tilde{D}_i, \Delta_i, \mathbf{Z}_i\}_{i=1}^n$, where $N_i(t) = N_i^*(t \wedge D_i \wedge C_i)$, $\tilde{D}_i = D_i \wedge C_i$ and $\Delta_i = I(D_i \leq C_i)$. As discussed before, $\boldsymbol{\eta}$ can be estimated by employing the same method as the rank-based estimation approach of AFT model in Chapter 2.

Now we will discuss the estimation procedure for $\boldsymbol{\theta}$ in Ghosh and Lin (2003) As defined in Lin et al. (1998), let T_{ik} be the time to the kth recurrent event ($k = 1, 2, \ldots$) for the ith subject ($i = 1, \ldots, n$). Then similar to Section 2.2.2 from Chapter 2, a counting process for the number of events without censoring until time t is defined by $N_i^*(t) = \sum_{i=1}^{\infty} I(T_{ik} \leq t)$ and a counting process for observed number of events is $N_i(t) = \sum_{k=1}^{\infty} I(T_{ik} \leq t \wedge \tilde{D}_i)$. Define $\tilde{T}_{ik}(\boldsymbol{\theta}) = T_{ik} e^{-\mathbf{Z}_i^T \boldsymbol{\theta}}$ and $\tilde{D}_i(\boldsymbol{\eta}) = \tilde{D}_i e^{-\mathbf{Z}_i^T \boldsymbol{\eta}}$. Let $\boldsymbol{\alpha} = (\boldsymbol{\eta}^T, \boldsymbol{\theta}^T)^T$ and $\boldsymbol{\alpha}_0 = (\boldsymbol{\eta}_0^T, \boldsymbol{\theta}_0^T)^T$. By adapting Lin et al. (1998)'s approach and the artificial censoring approach from Lin et al. (1996), quantities for the artificial censoring and new censoring time are, $d = \max_{1 \leq i \leq n} \{0, \mathbf{Z}_i^T(\boldsymbol{\theta} - \boldsymbol{\eta})\}$ and $\tilde{D}_i^*(\boldsymbol{\alpha}) = \tilde{D}_i(\boldsymbol{\eta}) e^{-\mathbf{Z}_i^T \boldsymbol{\eta} - d}$. The estimating function for $\boldsymbol{\theta}_0$ is

$$\mathbf{U}_n^{GL}(\boldsymbol{\alpha}) = n^{-1/2} \sum_{i=1}^{n} \int_0^{\infty} \left[\mathbf{Z}_i - \frac{\sum_{j=1}^n \mathbf{Z}_j I\{\tilde{D}_j^*(\boldsymbol{\eta}) \geq t\}}{\sum_{j=1}^n I\{\tilde{D}_j^*(\boldsymbol{\eta}) \geq t\}} \right] dN_{2i}(t; \boldsymbol{\alpha}), \tag{3.15}$$

where

$$N_{2i}(t;\boldsymbol{\alpha}) = \sum_{k=1}^{\infty} I\{\tilde{T}_{ik}(\boldsymbol{\theta}) \le t \wedge \tilde{D}_i^*(\boldsymbol{\alpha})\}.$$

Let the solution of equation $\mathbf{U}_n^{GL}(\boldsymbol{\alpha}) = 0$ be $\hat{\boldsymbol{\alpha}}^{GL}$. By using multivariate central limit theorem and empirical process theory, Ghosh and Lin (2003) showed that $n^{1/2}(\hat{\boldsymbol{\alpha}}^{GL} - \boldsymbol{\alpha}_0)$ is asymptotically mean zero and covariance matrix.

There are also several approaches that have made for extension of methodology of Peng and Fine (2006) to recurrent events with the terminal event. Ghosh (2010) and Hsieh et al. (2011) are these approaches.

Ghosh (2010) followed the same model and data structure as Ghosh and Lin (2003) and only considered a treatment indicator for Z_i as covariate. He demonstrated the connection between dependent censoring and truncation. From this connection, he also showed that the data structure $\{N_i(\cdot), \tilde{D}_i, \Delta_i, Z_i\}_{i=1}^n$ is equivalent to $\{K_i, (T_{ij}, j = 1, ..., K_i), \tilde{D}_i, \Delta_i, Z_i\}_{i=1}^n$ where T_{ij} is observable only $T_{ij} \le \tilde{D}_i$. Note that in this case, K_i is an observed number of recurrent events which depends on index of each individual. Let $S_n(\eta)$ be the estimating function for η_0. Note that $S_n(\eta)$ is a one-dimensional version of function (3.5). By using the artificial censoring technique that is identical to Peng and Fine (2006), Ghosh (2010) proposed the following quantities:

$$\begin{aligned} g_{ij}(\alpha) &= \max\{0, (\theta - \eta)Z_i, (\theta - \eta)Z_j\} \\ \tilde{T}_{i(j)k}(\alpha) &= T_{ik}e^{-\theta Z_i} \wedge (D_i \wedge C_i)e^{-\eta Z_i - g_{ij}(\alpha)} \\ \tilde{\delta}_{i(j)k}(\alpha) &= I\{T_{ik}e^{-\theta Z_i} \le (D_i \wedge C_i)e^{-\eta Z_i - g_{ij}(\alpha)}\} \end{aligned}$$

$$\phi_{ij}^G(\alpha) = \sum_{k=1}^{K_i}\sum_{l=1}^{K_j}[\tilde{\delta}_{i(j)k}(\alpha)I\{\tilde{T}_{i(j)k}(\alpha) \le \tilde{T}_{j(i)l}(\alpha)\} - \tilde{\delta}_{j(i)l}(\alpha)I\{\tilde{T}_{j(i)l}(\alpha) \le \tilde{T}_{i(j)k}(\alpha)\}].$$

The estimating function is

$$U_n^G(\alpha) = \frac{2n^{1/2}}{n(n-1)} \sum_{1 \le i < j \le n} (Z_i - Z_j)\phi_{ij}^G(\alpha). \tag{3.16}$$

Let $\hat{\alpha}^G = (\hat{\eta}, \hat{\theta}^G)^T$ be an estimator from solving $U_n^G(\alpha) = 0$. The next step is calculating covariance matrix for $\hat{\alpha}^G$. To calculate the covariance matrix, instead of implementing resampling approach by Parzen et al. (1994), Ghosh (2010) used a resampling approach by Zeng and Lin (2008). This algorithm is faster than that in Parzen et al. (1994) because it does not require to solve the estimating equations iteratively. The algorithm is as follows:

1. Generate two standard normal random variables $\mathbf{G} = (G_\eta, G_\theta)$.

2. Calculate $\{S_n(\hat{\eta} + n^{-1/2}G_\eta), U_n^G(\hat{\alpha}^G + n^{-1/2}\mathbf{G})\}$.

3. Repeat step 1 and step 2 M times.

4. Regress $S_n(\hat{\eta} + n^{-1/2}G_\eta)$ on G_η.

5. Regress $U_n^G(\hat{\alpha}^G + n^{-1/2}\mathbf{G})$ on G_η and G_θ.

6. Let $\hat{\boldsymbol{a}}_1$ be a regression coefficient from the regression in step 4 and $\hat{\boldsymbol{a}}_2$ be a regression coefficient from regression in step 5. Then it is possible to construct $\hat{\boldsymbol{A}}^{-1}\hat{\boldsymbol{B}}\hat{\boldsymbol{A}}^{-1}$, where $\hat{\boldsymbol{B}}$ is covariance matrix of $[\{S_n(\hat{\eta})\}^T, \{U_n^G(\hat{\alpha}^G)\}^T]^T$ and $\hat{\boldsymbol{A}}$ is a matrix such that the first row of $\hat{\boldsymbol{A}}$ is $\hat{\boldsymbol{a}}_1$ and the second row of $\hat{\boldsymbol{A}}$ is $\hat{\boldsymbol{a}}_2$.

$\hat{\alpha}^G$ is strongly consistent and asymptotically normal (Ghosh, 2010). Clearly, it is possible to extend this model to a case of multivariate covariates. Another variation of approach of Peng and Fine (2006) for recurrent events with the terminal event is proposed by Hsieh et al. (2011). They extended model in Ding et al. (2009) by applying the pairwise comparisons approach of Peng and Fine (2006). The data structure is $\{N^*(\cdot), T_k, C, D, \mathbf{Z}\}$, where T_k is time to the kth recurrent event ($k = 1, 2, \ldots$). The assumptions are the same as those of other models for recurrent events mentioned in this section. The proposed model is

$$\begin{pmatrix} h_1(T_k) = \mathbf{Z}^T\boldsymbol{\theta}_0 + \epsilon_k \\ h_2(D) = \mathbf{Z}^T\boldsymbol{\eta}_0 + \xi^D \end{pmatrix}, \tag{3.17}$$

where $\boldsymbol{\eta}_0$ and $\boldsymbol{\theta}_0$ are $p \times 1$ vectors of regression coefficients, $h_1(\cdot)$ is a known monotone function and $h_2(\cdot)$ is another monotone function which may be specified or unspecified. If $h_2(\cdot)$ is unknown, then ξ^D is known and vice versa. If $h_1(t) = h_2(t) = \log(t)$, the model is AFT model. If $h_2(\cdot)$ is unknown and ξ follows the extreme value distribution, as discussed in Chapter 2, it is the Cox proportional hazard model. Thus model (3.17) has high flexibility. In terms of the counting process, the model (3.17) is equivalent to

$$\begin{pmatrix} N^*\{h_1^{-1}(t+\mathbf{Z}^T\boldsymbol{\theta}_0)\} \\ h_2(D)-\mathbf{Z}^T\boldsymbol{\eta}_0 \end{pmatrix} \stackrel{d}{=} \begin{pmatrix} N_\epsilon^*(t) \\ \xi^D \end{pmatrix} \tag{3.18}$$

where $N_\epsilon^*(t) = \sum_{k=1}^{\infty} I(\epsilon_k \le t)$. Let $\{N_i^*(\cdot), T_{ik}, C_i, D_i, \mathbf{Z}_i\}_{i=1}^n$ be independent and identical copies of $\{N^*(\cdot), T_k, C, D, \mathbf{Z}\}$ for $k = 1, 2, \ldots$ and $i = 1, ..., n$. The observed process is $N_i(t) = N_i^*(t \wedge D_i \wedge C_i)$. Define $\epsilon_{ik}(\boldsymbol{\theta}) = h_1(T_{ik}) - \boldsymbol{Z}_i^T\boldsymbol{\theta}$ and $\xi_i^D(\boldsymbol{\eta}) = h_2(D_i) - \mathbf{Z}_i^T\boldsymbol{\eta}$. Then clearly, $\{\epsilon_{ik}(\boldsymbol{\theta}), \xi_i^D(\boldsymbol{\eta})\}$ are also independent and identically distributed. We assume that they do not depend on $\mathbf{Z}_i$.

Estimating $\boldsymbol{\eta}_0$ is identical as what was done in Ding et al. (2009). To estimate $\boldsymbol{\theta}_0$, the censoring mechanism is more complex than that in Peng and Fine (2006) due to considering general functions h_1 and h_2. In this case, ϵ_{ik} is censored by $\epsilon_i^C(\boldsymbol{\alpha}) = h_1[h_2^{-1}\{(\xi_i^D(\boldsymbol{\eta}) \wedge \xi_i^C(\boldsymbol{\eta})) + \mathbf{Z}_i^T\boldsymbol{\eta}\}] - \mathbf{Z}_i^T\boldsymbol{\theta}$ where $\xi_i^C(\boldsymbol{\eta}) = h_2(C_i) - \mathbf{Z}_i^T\boldsymbol{\eta}$. As in Ding et al. (2009), the need for artificial censoring arises to deal with this censoring mechanism. Define $\tilde{\epsilon}_i^C(\boldsymbol{\alpha}) = R_{\boldsymbol{\alpha}}\{\xi_i^D(\boldsymbol{\eta}) \wedge \xi_i^C(\boldsymbol{\eta})\}$, where $R_{\boldsymbol{\alpha}}(t) = \inf_z h_1\{h_2^{-1}(t + \boldsymbol{z}^T\boldsymbol{\eta})\} - \mathbf{z}^T\boldsymbol{\theta}$ and $\boldsymbol{\alpha} = (\boldsymbol{\eta}^T, \boldsymbol{\theta}^T)^T$. This time is the new censoring time to adjust dependent censoring in the model. Then the first proposed estimating function is

$$\mathbf{U}_n^{HL}(\boldsymbol{\alpha}) = \int_0^\infty \left[\mathbf{Z}_i - \frac{\sum_{j=1}^n I\{\tilde{\epsilon}_j^C(\boldsymbol{\alpha}) \ge t\}\mathbf{Z}_j}{\sum_{j=1}^n I\{\tilde{\epsilon}_j^C(\boldsymbol{\alpha}) \ge t\}}\right] d\tilde{N}_{\epsilon i}(t;\boldsymbol{\alpha}), \tag{3.19}$$

where $\tilde{N}_{\epsilon i}(t;\boldsymbol{\alpha}) = \sum_{k=1}^\infty I\{\epsilon_{ik}(\boldsymbol{\theta}) \le t \wedge \tilde{\epsilon}_j^C(\boldsymbol{\alpha})\}$. If $h_1(t) = h_2(t) = \log(t)$, then this function is exactly the same as the function (3.15). For the pairwise comparisons approach, define $\tilde{\epsilon}_{i(j)}^C(\boldsymbol{\alpha}) = R_{\boldsymbol{\alpha}}^{ij}\{\xi_i^D(\boldsymbol{\eta}) \wedge \xi_i^C(\boldsymbol{\eta})\}$ where $R_{\boldsymbol{\alpha}}^{ij}(t) = \inf_{\mathbf{z}=\mathbf{z}_i,\mathbf{z}_j} h_1\{h_2^{-1}(t + \boldsymbol{z}^T\boldsymbol{\eta})\} - \mathbf{z}^T\boldsymbol{\theta}$. $R_{\boldsymbol{\alpha}}^{ij}(t)$ is an artificial censoring mechanism for pairwise comparisons and $\tilde{\epsilon}_{i(j)}^C(\boldsymbol{\alpha})$ is a new censoring variable to adjust dependent censoring in pairwise comparisons. Let $\tilde{\epsilon}_{i(j)}^k(\boldsymbol{\alpha}) = \epsilon_{ik}(\boldsymbol{\theta}) \wedge \tilde{\epsilon}_{i(j)}^C(\boldsymbol{\alpha})$ and $\tilde{\delta}_{i(j)}^k(\boldsymbol{\alpha}) = I\{\epsilon_{ik}(\boldsymbol{\theta}) \le \tilde{\epsilon}_{i(j)}^C(\boldsymbol{\alpha})\}$. Then second proposed estimating function is

$$\mathbf{U}_n^H(\boldsymbol{\alpha}) = \frac{2n^{1/2}}{n(n-1)} \sum_{1 \le i < j \le n} (\mathbf{Z}_i - \mathbf{Z}_j)\phi_{ij}^H(\boldsymbol{\alpha}), \tag{3.20}$$

where

$$\phi_{ij}^{H}(\boldsymbol{\alpha}) = \sum_{k} [\tilde{\delta}_{i(j)}^{k}(\boldsymbol{\alpha}) I\{\tilde{\epsilon}_{i(j)}^{k}(\boldsymbol{\alpha}) \leq \tilde{\epsilon}_{j(i)}^{k}(\boldsymbol{\alpha})\} - \tilde{\delta}_{j(i)}^{k}(\boldsymbol{\alpha}) I\{\tilde{\epsilon}_{j(i)}^{k}(\boldsymbol{\alpha}) \leq \tilde{\epsilon}_{i(j)}^{k}(\boldsymbol{\alpha})\}].$$

Note that in this case, the same index for recurrent events when comparing transformed times of an individual i and j while Ghosh (2010) considered different indexes for recurrent events when comparing these transformed times. If $k = 1$ and $h_1(t) = h_2(t) = \log(t)$, this is equal to the function (3.10). The other estimating function proposed by Hsieh et al. (2011) is

$$\mathbf{U}_n^{LH}(\boldsymbol{\alpha}) = \frac{2n^{1/2}}{n(n-1)} \sum_{1 \leq i < j \leq n} (\mathbf{Z}_i - \mathbf{Z}_j) \phi_{ij}^{LH}(\boldsymbol{\alpha}), \tag{3.21}$$

where

$$\phi_{ij}^{LH}(\boldsymbol{\alpha}) = \sum_{k} [\tilde{\delta}_{i(j)}^{k}(\boldsymbol{\alpha}) I\{\tilde{\epsilon}_{i(j)}^{k}(\boldsymbol{\alpha}) \leq \tilde{\epsilon}_{j(i)}^{C}(\boldsymbol{\alpha})\} - \tilde{\delta}_{j(i)}^{k}(\boldsymbol{\alpha}) I\{\tilde{\epsilon}_{j(i)}^{C}(\boldsymbol{\alpha}) \leq \tilde{\epsilon}_{i(j)}^{k}(\boldsymbol{\alpha})\}].$$

In this last estimating function, $\tilde{\epsilon}_{j(i)}^{C}(\boldsymbol{\alpha})$ is used instead of $\tilde{\epsilon}_{j(i)}^{k}(\boldsymbol{\alpha})$ in $\phi_{ij}^{LH}(\boldsymbol{\alpha})$. It implies that information in new censoring time of $\tilde{\epsilon}_{i(j)}^{k}(\boldsymbol{\alpha})$ is included in statistical inference for regression coefficients. In fact, this approach is adapting the method of creating the new indicator function to adjust the dependent censoring in Lin et al. (1996) in context of the pairwise comparisons. They used the resampling approach from Parzen et al. (1994) to perform statistical inference. As in Ding et al. (2009), they also proposed an estimating function for $\boldsymbol{\theta}$ when D has a linear transformation model.

Chapter 4
Weighted Estimation of the Accelerated Failure Time Model in the Presence of Dependent Censoring

4.1 Introduction

As can be seen in the previous chapters, semicompeting risks data have been widely studied in the past decade. The semiparametric regression approach by Lin et al. (1996) and Peng and Fine (2006) has a lot of advantages. The model in their approaches is linear model. Second, for estimation of treatment effect, bias of survival function does not occur (Varadhan et al. 2014).

However, none of these papers fully discussed optimality of the estimator. In this case, choosing an estimator that is optimal from an efficiency viewpoint is a major issue for consideration. Here, we adapt the idea of Wei et al. (1989), which proposed an optimal estimator whose form is a linear combination of estimators for multivariate failure time data. They used idea of Wei and Johnson (1985), which proposed using combinations of dependent tests in the presence of missing values. Idea of Wei and Johnson (1985) is to create a test which can maximize power based on linear combination of test statistics. Approach of Wei and Johnson (1985) is simple and flexible, so it is sensible to apply their method in our case.

In this chapter, we propose a weighted estimator by using methodology from

Wei et al. (1989). The weighted estimator combines those of Lin et al. (1996) and Peng and Fine (2006). The structure of this chapter is as follows. In Section 4.2, we describe details on our new weighted estimator. In Section 4.3, model checking procedure is briefly discussed. In Section 4.4, results of simulation studies will be given. Application of our method to a real data example is presented in Section 4.5. Some discussion concludes Section 4.6. In this chapter, for Lin et al. (1996) and Peng and Fine (2006) procedures, we use the same notations as Section 3.2.2 in Chapter 3.

4.2 Weighted estimator

Given two estimation procedures as described in Section 3.2.2 in Chapter 3, it is natural to consider their efficiencies with respect to standard error. However, in this point of view, neither estimator is superior to the other. Moreover, these estimators may not be optimal estimators with respect to the standard error. There is an argument that the estimator of Peng and Fine (2006) gains more efficiency than that of Lin et al. (1996) because pairwise comparisons lead to less artificial censoring than that in Lin et al. (1996). However, this logic only holds when we look at performance of estimators in the view of bias and variance across the estimators in simulation study. Concentrating on standard error of an estimator in a single dataset, the estimator by Peng and Fine (2006) may not provide better estimator than that of Lin et al. (1996). This will be seen in the real data analysis section.

The reason for this is due to estimation procedure of Peng and Fine (2006). As discussed Ghosh (2010), for n samples, the number of comparisons of Lin et al. (1996) for artificial censoring is of order n, while that of Peng and Fine (2006) is of order n^2. By definition of $g_{ij}(\boldsymbol{\alpha})$, different degrees of artificial censoring is applied to observations. It may lead more variation in estimation of the standard error for the estimator of the regression parameters. Hence, standard error of the estimator by Peng and Fine (2006) may be larger than that of Lin et al. (1996).

Having discussed our data structure and estimators from Lin et al. (1996) and Peng and Fine (2006), we now describe the proposed estimation in this paper. Let $\hat{\boldsymbol{\eta}} = (\hat{\eta}_1, \ldots, \hat{\eta}_k)^T$ be estimator of $\boldsymbol{\eta}_0$, $\hat{\boldsymbol{\theta}}^L = (\hat{\theta}_1^L, \ldots, \hat{\theta}_k^L)^T$ be estimator of $\boldsymbol{\theta}_0$ by Lin et al. (1996) and $\hat{\boldsymbol{\theta}}^P = (\hat{\theta}_1^P, \ldots, \hat{\theta}_k^P)^T$ be estimator of $\boldsymbol{\theta}_0$ by Peng and Fine (2006).

$\hat{\boldsymbol{\theta}}^L$ and $\hat{\boldsymbol{\theta}}^P$ are asymptotically unbiased estimators of $\boldsymbol{\theta}_0$.

We extend the scope of estimators which provide consistent estimation of $\boldsymbol{\theta}_0$. The natural extension of estimators of Lin et al. (1996) and Peng and Fine (2006) is to consider collections of estimators that are linear combination of these two estimators with sum of weights being 1. By choosing proper weights, we can expect that the variance of the new combined estimator is smaller than that of each individual estimator in $\hat{\boldsymbol{\theta}}^L$ and $\hat{\boldsymbol{\theta}}^P$.

The goal is to find weights such that the variance of the new estimator is smaller than the minimum of variance of the estimators by Lin et al. (1996) and Peng and Fine (2006), which have good theoretical properties. To obtain the estimator that yields smallest variance with these properties, we can use the idea of Wei et al. (1989), which was applied to the problem of modeling multivariate failure times.

In Wei et al. (1989), the joint distribution of estimators $\hat{\boldsymbol{\gamma}} = \{\hat{\gamma}_{mr}\}$ is considered, where $m = 1, \ldots, k$ and $r = 1 \ldots R$. In this case, m indicates index of regression parameters and r stands for index of the rth event. For obtaining an optimal estimator, they applied arguments from Wei and Johnson (1985) which derived a linear combination of test statistic to maximize power against every alternative hypothesis. Let $\hat{\mathbf{H}}$ be the covariance matrix for the estimators $\hat{\boldsymbol{\gamma}}$. Then we fix m and define $\hat{\mathbf{H}}_m$ be covariance matrix of $\hat{\boldsymbol{\gamma}}_m = (\hat{\gamma}_{m1}, \ldots, \hat{\gamma}_{mR})$. It can be obtained from the entire covariance matrix by selecting the part corresponding to $\hat{\boldsymbol{\gamma}}$ for $r = 1, \ldots, R$ under fixed m. Now we can define $\sum_{r=1}^{R} d_r \hat{\gamma}_{mr}$, where $\mathbf{d} = (d_1, d_2, \ldots, d_R)$ satisfies $\sum_{r=1}^{R} d_r = 1$ (Wei et al 1989). Then $\mathbf{d} \equiv (\mathbf{e}^T \hat{\mathbf{H}}_m^{-1} \mathbf{e})^{-1} \hat{\mathbf{H}}_m^{-1} \mathbf{e}$ is a vector of weights which leads the best estimator among linear combinations of estimators of $\hat{\boldsymbol{\gamma}}_m$ where $\mathbf{e}$ is a vector consisting of R ones (Wei and Johnson, 1985; Wei et al. 1989).

We now apply the argument in the previous paragraph to our model by considering the joint distribution of $\hat{\boldsymbol{\beta}} = \{\hat{\boldsymbol{\eta}}^T, (\hat{\boldsymbol{\theta}}^L)^T, (\hat{\boldsymbol{\theta}}^P)^T\}^T$. Let $\boldsymbol{\beta}_0 = (\boldsymbol{\eta}_0^T, \boldsymbol{\theta}_0^T, \boldsymbol{\theta}_0^T)^T$ and $\mathbf{G}_n(\boldsymbol{\beta}) = [\mathbf{S}_n^T(\boldsymbol{\eta}), \{\mathbf{U}_n^L(\boldsymbol{\alpha})\}^T, \{\mathbf{U}_n^P(\boldsymbol{\alpha})\}^T]^T$ where $\mathbf{S}_n(\boldsymbol{\eta}), \mathbf{U}_n^L(\boldsymbol{\alpha})$ and $\mathbf{U}_n^P(\boldsymbol{\alpha})$ are estimating functions for $\boldsymbol{\beta}_0$ which are introduced in Section 3.2.2 in Chapter 3. The strong consistency and asymptotic joint distribution of three estimators, described in following theorems, play a crucial role in our methodology.

To prove asymptotic results, several regularity conditions are required. As stated in Ghosh (2010) and Peng and Fine (2006), define

$$F(a,b,c,d,e) = P(\epsilon_1^X - \epsilon_1^D \leq a, \epsilon_1^X - C_1 \leq b, \epsilon_1^X - \epsilon_2^X \leq c, \epsilon_1^X - \epsilon_2^D \leq d, \epsilon_1^X - C_2 \leq e | \mathbf{Z}_1, \mathbf{Z}_2)$$

Let $\boldsymbol{\alpha}_0 = (\boldsymbol{\eta}_0^T, \boldsymbol{\theta}_0^T)^T$. Define

$$\begin{aligned}
T_1(\mathbf{Z}_1, \mathbf{Z}_2) = & \frac{\partial F}{\partial a}\{g_{12}(\boldsymbol{\alpha}_0), -\mathbf{Z}_1^T\boldsymbol{\eta}_0 - g_{12}(\boldsymbol{\alpha}_0), 0, g_{12}(\boldsymbol{\alpha}_0), -\mathbf{Z}_2^T\boldsymbol{\eta}_0 - g_{12}(\boldsymbol{\alpha}_0)\} \\
& + \frac{\partial F}{\partial b}\{g_{12}(\boldsymbol{\alpha}_0), -\mathbf{Z}_1^T\boldsymbol{\eta}_0 - g_{12}(\boldsymbol{\alpha}_0), 0, g_{12}(\boldsymbol{\alpha}_0), -\mathbf{Z}_2^T\boldsymbol{\eta}_0 - g_{12}(\boldsymbol{\alpha}_0)\} \\
& + \frac{\partial F}{\partial d}\{g_{12}(\boldsymbol{\alpha}_0), -\mathbf{Z}_1^T\boldsymbol{\eta}_0 - g_{12}(\boldsymbol{\alpha}_0), 0, g_{12}(\boldsymbol{\alpha}_0), -\mathbf{Z}_2^T\boldsymbol{\eta}_0 - g_{12}(\boldsymbol{\alpha}_0)\} \\
& + \frac{\partial F}{\partial e}\{g_{12}(\boldsymbol{\alpha}_0), -\mathbf{Z}_1^T\boldsymbol{\eta}_0 - g_{12}(\boldsymbol{\alpha}_0), 0, g_{12}(\boldsymbol{\alpha}_0), -\mathbf{Z}_2^T\boldsymbol{\eta}_0 - g_{12}(\boldsymbol{\alpha}_0)\} \\
& + 2\frac{\partial F}{\partial c}\{g_{12}(\boldsymbol{\alpha}_0), -\mathbf{Z}_1^T\boldsymbol{\eta}_0 - g_{12}(\boldsymbol{\alpha}_0), 0, g_{12}(\boldsymbol{\alpha}_0), -\mathbf{Z}_2^T\boldsymbol{\eta}_0 - g_{12}(\boldsymbol{\alpha}_0)\}
\end{aligned}$$

and

$$T_2(\mathbf{Z}_1, \mathbf{Z}_2) = T_1(\mathbf{Z}_1, \mathbf{Z}_2) - 2\frac{\partial F}{\partial c}\{g_{12}(\boldsymbol{\alpha}_0), -\mathbf{Z}_1^T\boldsymbol{\eta}_0 - g_{12}(\boldsymbol{\alpha}_0), 0, g_{12}(\boldsymbol{\alpha}_0), -\mathbf{Z}_2^T\boldsymbol{\eta}_0 - g_{12}(\boldsymbol{\alpha}_0)\}$$

From the Appendix in Peng and Fine (2006), the additional conditions are as follows:

1. The parameter space $\mathcal{W}$ is compact, and the true parameter $\boldsymbol{\alpha}_0$ is an interior point of $\mathcal{W}$.

2. $\boldsymbol{\theta}_0$ is the only solution of the estimating equation $E\{n^{-1/2}\mathbf{U}_n^P(\boldsymbol{\eta}_0, \boldsymbol{\theta})\} = 0$.

3. $E(||\mathbf{Z}||^2) < \infty$, where $|| \cdot ||$ is Euclidean norm and there exists a positive constant K such that partial derivatives of F are bounded by K and there exists positive constant K^* such that marginal probability density of F is bounded by K^* almost surely.

4. $cov[(\mathbf{Z}_1 - \mathbf{Z}_2)\{T_1(\mathbf{Z}_1, \mathbf{Z}_2)\}^{1/2}]$ and $cov[(\mathbf{Z}_1 - \mathbf{Z}_2)\{T_2(\mathbf{Z}_1, \mathbf{Z}_2)\}^{1/2}]$ are positive definite.

In many parts of proofs, we adapt arguments from Lin et al. (1996) and Peng and Fine (2006).

Theorem 4.1. By conditions of $C1 - C3$ in Appendix of Peng and Fine (2006) and conditions in Ying (1993), $\hat{\boldsymbol{\beta}}$ is (strongly) consistent.

Proof. Let $\hat{\boldsymbol{\beta}} = \{\hat{\boldsymbol{\eta}}^T, (\hat{\boldsymbol{\theta}}^L)^T, (\hat{\boldsymbol{\theta}}^P)^T\}^T$. It suffices to show that $\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\theta}}^L$ and $\hat{\boldsymbol{\theta}}^P$ are strongly consistent, respectively. Let $\boldsymbol{\alpha} = (\boldsymbol{\eta}^T, \boldsymbol{\theta}^T)^T$. Note that we have compact region, say $\mathcal{W}$ and we assume regularity conditions in Ying (1993). By Ying (1993), there exists nonrandom function m_1 such that $\sup_{\boldsymbol{\eta}\in\mathcal{N}_0} ||n^{-1/2}\mathbf{S}_n(\boldsymbol{\eta}) - \mathbf{m}_1(\boldsymbol{\eta})||$ converges to 0 with probability 1 where $\mathcal{N}_0$ is a neighborhood of $\boldsymbol{\eta}_0$. Thus $\hat{\boldsymbol{\eta}}$ is strongly consistent. Similarly, we have another nonrandom function $\mathbf{m}_2$ such that $\sup_{\boldsymbol{\alpha}\in\mathcal{N}_1} ||n^{-1/2}\mathbf{U}_n^L(\boldsymbol{\alpha}) - \mathbf{m}_2(\boldsymbol{\alpha})||$ converges to 0 with probability 1 where $\mathcal{N}_1$ is a neighborhood of $\boldsymbol{\alpha}_0$. Hence by Ying (1993), $\hat{\boldsymbol{\alpha}}^L$ is strongly consistent.

For $\hat{\boldsymbol{\theta}}^P$, by argument in Appendix of Peng and Fine (2006), note that by the U-statistics version of the law of large numbers, for all $\boldsymbol{\alpha} \in \mathcal{W}$, $||n^{-1/2}\mathbf{U}_n^P(\boldsymbol{\alpha}) - \boldsymbol{\gamma}(\boldsymbol{\alpha})||$ converges to 0 in probability where $\boldsymbol{\gamma}(\boldsymbol{\alpha}) = E\{n^{-1/2}\mathbf{U}_n^P(\boldsymbol{\alpha})\}$. We can partition our compact space as $\mathcal{W}_1, ..., \mathcal{W}_k$ so that $\mathcal{W} \in \cup_{j=1}^k \mathcal{W}_j$. Clearly, then for $\{\boldsymbol{\alpha}^j \in \mathcal{W}_j, j = 1, ..., k\}$, $\max_{1\le j\le k} ||n^{-1/2}\mathbf{U}_n^P(\boldsymbol{\alpha}^j) - \boldsymbol{\gamma}(\boldsymbol{\alpha}^j)||$ converges to 0 in probability. Then by Appendix of Peng and Fine (2006),

$$\begin{aligned}&\sup_{||\boldsymbol{\alpha}-\boldsymbol{\alpha}^{**}||\le\xi} n^{-1/2}||\mathbf{U}_n^P(\boldsymbol{\alpha}) - \mathbf{U}_n^P(\boldsymbol{\alpha}^{**})|| \\ \le& \frac{2}{n(n-1)} \sum_{1\le i<j\le n} ||\mathbf{Z}_i - \mathbf{Z}_j|| \sup_{||\boldsymbol{\alpha}-\boldsymbol{\alpha}^{**}||\le\xi} |\phi_{ij}(\boldsymbol{\alpha}) - \phi_{ij}(\boldsymbol{\alpha}^{**})|\end{aligned}$$

and for all $\epsilon > 0$, there exists $\xi > 0$ such that

$$\lim_{n\to\infty} P\left(\frac{2}{n(n-1)} \sum_{1\le i<j\le n} ||\mathbf{Z}_i - \mathbf{Z}_j|| \sup_{||\boldsymbol{\alpha}-\boldsymbol{\alpha}^{**}||\le\xi} |\phi_{ij}(\boldsymbol{\alpha}) - \phi_{ij}(\boldsymbol{\alpha}^{**})| \ge \epsilon\right) = 0$$

Hence

$$\lim_{n\to\infty} P(\sup_{||\boldsymbol{\alpha}-\boldsymbol{\alpha}^{**}||\le\xi} n^{-1/2}||\mathbf{U}_n^P(\boldsymbol{\alpha}) - \mathbf{U}_n^P(\boldsymbol{\alpha}^{**})|| \ge \epsilon) = 0$$

Thus $\hat{\boldsymbol{\theta}}^P$ is strongly consistent and clearly, $\hat{\boldsymbol{\beta}}$ is strongly consistent. □

Theorem 4.2. Assuming certain technical conditions from Ying (1993) and Peng and Fine (2006), $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ is asymptotically normal with mean zero vector and covariance matrix $\boldsymbol{\Sigma}_0$ where $\boldsymbol{\Sigma}_0 = \boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^{-1}$, where $\boldsymbol{\Gamma}_0$ is a nonsingular matrix and $\boldsymbol{\Omega}_0$ is the asymptotic covariance matrix of $\mathbf{G}_n(\boldsymbol{\beta}_0)$.

Proof. As consistency, we assume the same regularity conditions as in Ying (1993). Let $\boldsymbol{\beta}_0 = (\boldsymbol{\eta}_0^T, \boldsymbol{\theta}_0^T, \boldsymbol{\theta}_0^T)^T$ and $\mathbf{G}_n(\boldsymbol{\beta}) = [\mathbf{S}_n^T(\boldsymbol{\eta}), \{\mathbf{U}_n^L(\boldsymbol{\alpha})\}^T, \{\mathbf{U}_n^P(\boldsymbol{\alpha})\}^T]^T$. Similar to Lin et al. (1996), let $\lambda_0^{(1)}(t)$ be the cause-specific hazard function for the $\tilde{D}_i^*(\boldsymbol{\eta})$ and let $\lambda_0^{(2)}(t)$ be the cause-specific hazard function for $\tilde{X}_i^*(\boldsymbol{\alpha})$ under dependent censoring. Define

$$M_{1i}(t) = N_{1i}(t;\boldsymbol{\eta}_0) - \int_{-\infty}^{t} I\{\tilde{D}_i^*(\boldsymbol{\eta}_0) \geq u\}\lambda_0^{(1)}(u)du \tag{4.1}$$

$$M_{2i}(t) = N_{2i}(t;\boldsymbol{\alpha}_0) - \int_{-\infty}^{t} I\{\tilde{X}_i^*(\boldsymbol{\alpha}_0) \geq u\}\lambda_0^{(2)}(u)du \tag{4.2}$$

Then M_{1i} and M_{2i} are martingales (Fleming and Harrington, 2005, p.26; Lin et al. 1996). By adapting a proof in the Appendix in Lin et al. (1996), Rebdolledo's martingale central limit theorem (Fleming and Harrington, 2005, pp.227-228) gives

$$\mathbf{S}_n(\boldsymbol{\eta}_0) = n^{-1/2}\sum_{i=1}^{n}\int_{-\infty}^{\infty}\{\mathbf{Z}_i - \bar{\mathbf{Z}}^{(1)}(u)\}dM_{1i}(u) + o_p(1)$$

$$\mathbf{U}_n^L(\boldsymbol{\alpha}_0) = n^{-1/2}\sum_{i=1}^{n}\int_{-\infty}^{\infty}\{\mathbf{Z}_i - \bar{\mathbf{Z}}^{(2)}(u)\}dM_{2i}(u) + o_p(1)$$

where $\bar{\mathbf{Z}}^{(1)}(u) = \lim_{n\to\infty}[\sum_{j=1}^n I\{\tilde{D}_j^*(\boldsymbol{\eta}_0) \geq u\}\mathbf{Z}_j]/[\sum_{j=1}^n I\{\tilde{D}_j^*(\boldsymbol{\eta}_0) \geq u\}]$ and $\bar{\mathbf{Z}}^{(2)}(u) = \lim_{n\to\infty}[\sum_{j=1}^n I\{\tilde{X}_j^*(\boldsymbol{\alpha}_0) \geq u\}\mathbf{Z}_j]/[\sum_{j=1}^n \{I(\tilde{X}_j^*(\boldsymbol{\alpha}_0) \geq u)\}]$. From the Appendix of Peng and Fine (2006),

$$\mathbf{U}_n^P(\boldsymbol{\alpha}_0) = n^{-1/2}\sum_{i=1}^{n} 2\mathbf{h}_1(\mathbf{V}_i, \boldsymbol{\alpha}_0) + o_p(1)$$

where $2\mathbf{h}_1(\mathbf{v}, \boldsymbol{\alpha}_0) = 2E[\mathbf{h}(\mathbf{v}, \mathbf{V}_2, \boldsymbol{\alpha}_0)]$. For $j = 1, ..., n$, $M_{1j}(t)$ is the martingale associated with ϵ_j^D, while $M_{2j}(t)$ is the martingale associated with ϵ_j^X and $\mathbf{h}(\mathbf{V}_i, \mathbf{V}_j, \boldsymbol{\alpha}) = (\mathbf{Z}_i - \mathbf{Z}_j)\phi_{ij}(\boldsymbol{\alpha})$ (Lin et al. 1996; Peng and Fine, 2006). For $j = 1, \ldots, n$, define

$$\mathbf{a}_{0j} = \int_{-\infty}^{\infty}\{\mathbf{Z}_j - \bar{\mathbf{Z}}^{(1)}(u)\}dM_{1j}(u) \qquad \mathbf{a}_{1j} = \int_{-\infty}^{\infty}\{\mathbf{Z}_j - \bar{\mathbf{Z}}^{(2)}(u)\}dM_{2j}(u)$$

$$\mathbf{a}_{2j} = 2\mathbf{h}_1(\mathbf{V}_j, \boldsymbol{\alpha}_0).$$

By the Cramér-Wold theorem, $\mathbf{G}_n(\boldsymbol{\beta}_0)$ has an asymptotically normal distribution with mean zero and covariance matrix $\mathbf{\Omega}_0$, where

$$\mathbf{\Omega}_0 = E\begin{pmatrix} \mathbf{a}_{01}\mathbf{a}_{01}^T & \mathbf{a}_{01}\mathbf{a}_{11}^T & \mathbf{a}_{01}\mathbf{a}_{21}^T \\ \mathbf{a}_{11}\mathbf{a}_{01}^T & \mathbf{a}_{11}\mathbf{a}_{11}^T & \mathbf{a}_{11}\mathbf{a}_{21}^T \\ \mathbf{a}_{21}\mathbf{a}_{01}^T & \mathbf{a}_{21}\mathbf{a}_{11}^T & \mathbf{a}_{21}\mathbf{a}_{21}^T \end{pmatrix}$$

Note that $E\{n^{-1/2}\mathbf{U}_n^P(\boldsymbol{\alpha})\} = \boldsymbol{\gamma}(\boldsymbol{\alpha})$. As stated in the Appendix of Peng and Fine (2006), under conditions of $N1 - N3$ from Honoré and Powell (1994), there exists an open neighborhood of $\boldsymbol{\alpha}_0$, say K_0, such that

$$\sup_{\boldsymbol{\alpha}\in K_0} \frac{||\mathbf{U}_n^P(\boldsymbol{\alpha}) - \mathbf{U}_n^P(\boldsymbol{\alpha}_0) - n^{1/2}\boldsymbol{\gamma}(\boldsymbol{\alpha})||}{1 + n^{1/2}||\boldsymbol{\gamma}(\boldsymbol{\alpha})||} = o_p(1) \tag{4.3}$$

Using a Taylor series expansion of $\gamma(\boldsymbol{\alpha})$ around $\boldsymbol{\alpha}_0$,

$$\gamma(\boldsymbol{\alpha}) = \boldsymbol{\gamma}(\boldsymbol{\alpha}_0) + \left.\frac{\partial\boldsymbol{\gamma}(\boldsymbol{\alpha})}{\partial\boldsymbol{\eta}}\right|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}(\boldsymbol{\eta}-\boldsymbol{\eta}_0) + \left.\frac{\partial\boldsymbol{\gamma}(\boldsymbol{\alpha})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}(\boldsymbol{\theta}-\boldsymbol{\theta}_0) + o(||\boldsymbol{\alpha}-\boldsymbol{\alpha}_0||) \tag{4.4}$$

With these two results (4.3) and (4.4), by Appendix of Peng and Fine (2006),

$$\mathbf{U}_n^P(\boldsymbol{\alpha}) = \mathbf{U}_n^P(\boldsymbol{\alpha}_0) + n^{1/2}\left.\frac{\partial\gamma(\boldsymbol{\alpha})}{\partial\boldsymbol{\eta}}\right|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}(\boldsymbol{\eta}-\boldsymbol{\eta}_0) + n^{1/2}\left.\frac{\partial\gamma(\boldsymbol{\alpha})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}(\boldsymbol{\theta}-\boldsymbol{\theta}_0) + o_p(1+n^{1/2}||\boldsymbol{\alpha}-\boldsymbol{\alpha}_0||) \tag{4.5}$$

From Ying (1993), we have that

$$\mathbf{S}_n(\boldsymbol{\eta}) = \mathbf{S}_n(\boldsymbol{\eta}_0) + n^{1/2}\mathbf{P}_0(\boldsymbol{\eta}-\boldsymbol{\eta}_0) + o_p(1) \tag{4.6}$$

for any $\boldsymbol{\eta}$ in the small neighborhood of $\boldsymbol{\eta}_0$, where $\mathbf{P}_0$ is $k \times k$ nonsingular matrix. From the Appendix in Lin et al. (1996), for $\mathbf{J}_{1n}(\boldsymbol{\alpha}) = [\mathbf{S}_n^T(\boldsymbol{\eta}), \{\mathbf{U}_n^L(\boldsymbol{\alpha})\}^T]^T$,

$$\mathbf{J}_{1n}(\boldsymbol{\alpha}) = \mathbf{J}_{1n}(\boldsymbol{\alpha}_0) + n^{1/2}\mathbf{L}_{10}(\boldsymbol{\alpha}-\boldsymbol{\alpha}_0) + o_p(1) \tag{4.7}$$

for any $\boldsymbol{\alpha}$ in the small neighborhood of $\boldsymbol{\alpha}_0$, where $\mathbf{L}_{10}$ is defined as

$$\mathbf{L}_{10} = \begin{pmatrix} \mathbf{P}_0 & 0 \\ \mathbf{M}_0 & \mathbf{H}_0 \end{pmatrix},$$

a $2k \times 2k$ nonsingular matrix, and $\mathbf{M}_0$ and $\mathbf{H}_0$ are $k \times k$ constant matrices. Define

$\mathbf{J}_{2n}(\boldsymbol{\alpha}) = [\mathbf{S}_n^T(\boldsymbol{\eta}), \{\mathbf{U}_n^P(\boldsymbol{\alpha})\}^T]^T$. Using the expansion from Peng and Fine (2006), for any $\boldsymbol{\alpha}$ in the small neighborhood of $\boldsymbol{\alpha}_0$,

$$\mathbf{J}_{2n}(\boldsymbol{\alpha}) = \mathbf{J}_{2n}(\boldsymbol{\alpha}_0) + n^{1/2}\mathbf{L}_{20}(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0) + o_p(1) \tag{4.8}$$

$$\mathbf{L}_{20} = \begin{pmatrix} \mathbf{P}_0 & 0 \\ \mathbf{R}_0 & \mathbf{V}_0 \end{pmatrix}$$

where $\mathbf{R}_0 = \frac{\partial\gamma(\boldsymbol{\alpha})}{\partial\boldsymbol{\eta}}|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}$ and $\mathbf{V}_0 = \frac{\partial\gamma(\boldsymbol{\alpha})}{\partial\boldsymbol{\theta}}|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}$. Combining expansions of (4.6), (4.7) and (4.8), we have

$$\mathbf{G}_n(\boldsymbol{\beta}) = \mathbf{G}_n(\boldsymbol{\beta}_0) + n^{1/2}\boldsymbol{\Gamma}_0(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + o_p(1)$$

for any $\boldsymbol{\beta}$ in the small neighborhood of $\boldsymbol{\beta}_0$, where $\boldsymbol{\Gamma}_0$ is defined as

$$\boldsymbol{\Gamma}_0 = \begin{pmatrix} \mathbf{P}_0 & 0 & 0 \\ \mathbf{M}_0 & \mathbf{H}_0 & 0 \\ \mathbf{R}_0 & 0 & \mathbf{V}_0 \end{pmatrix}$$

. The results from Honoré and Powell (1994) and Ying (1993), along with the consistency of $\hat{\boldsymbol{\beta}}$, imply that

$$n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = -\boldsymbol{\Gamma}_0^{-1}\mathbf{G}_n(\boldsymbol{\beta}_0) + o_p(1).$$

By combining the above results with Slutsky's theorem, $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ has an asymptotically normal distribution with mean zero and covariance matrix $\boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^{-1}$. $\square$

Theorem 4.2 implies the asymptotic normality of $\hat{\boldsymbol{\beta}}$ with the form of $\boldsymbol{\Sigma}_0$ being

$$\boldsymbol{\Sigma}_0 = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \boldsymbol{\Sigma}_{13} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} & \boldsymbol{\Sigma}_{23} \\ \boldsymbol{\Sigma}_{31} & \boldsymbol{\Sigma}_{32} & \boldsymbol{\Sigma}_{33} \end{pmatrix}.$$

Let $\hat{\boldsymbol{\Sigma}}$ be the estimated covariance matrix of $\boldsymbol{\Sigma}_0$. In this covariance matrix, $\hat{\boldsymbol{\Sigma}}_{11}$ is a $k \times k$ covariance matrix for $\hat{\boldsymbol{\eta}}$, $\hat{\boldsymbol{\Sigma}}_{22}$ is a $k \times k$ covariance matrix for $\hat{\boldsymbol{\theta}}^L$ and $\hat{\boldsymbol{\Sigma}}_{33}$ is a $k \times k$ covariance matrix for $\hat{\boldsymbol{\theta}}^P$. Moreover, $\hat{\boldsymbol{\Sigma}}_{12}$ and $\hat{\boldsymbol{\Sigma}}_{13}$ represent covariance terms between $\hat{\boldsymbol{\eta}}$ and $\hat{\boldsymbol{\theta}}^L$ and between $\hat{\boldsymbol{\eta}}$ and $\hat{\boldsymbol{\theta}}^P$, respectively. Define $\hat{\boldsymbol{\Sigma}}_{23}$ as the covariance matrix between $\hat{\boldsymbol{\theta}}^L$ and $\hat{\boldsymbol{\theta}}^P$. Clearly, $\hat{\boldsymbol{\Sigma}}_{21} = \hat{\boldsymbol{\Sigma}}_{12}^T$, $\hat{\boldsymbol{\Sigma}}_{31} = \hat{\boldsymbol{\Sigma}}_{13}^T$ and $\hat{\boldsymbol{\Sigma}}_{32} = \hat{\boldsymbol{\Sigma}}_{23}^T$.

The issue remains of how to obtain the matrix corresponding to $\hat{\mathbf{H}}_m^{-1}$ in our context. Note that $\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\theta}}^L$ and $\hat{\boldsymbol{\theta}}^P$ are correlated with each other. The estimating function structure implies that $\hat{\boldsymbol{\theta}}^L$ and $\hat{\boldsymbol{\theta}}^P$ cannot be estimated separately from $\hat{\boldsymbol{\eta}}$. Thus our matrix corresponding to $\hat{\mathbf{H}}_m^{-1}$ should include the effect of $\hat{\boldsymbol{\eta}}$. To obtain the matrix, we need to invert whole matrix and extract the submatrix corresponding to $\hat{\boldsymbol{\theta}}^L$ and $\hat{\boldsymbol{\theta}}^P$. There are two approaches to obtain the submatrix.

The first approach is to invert $\hat{\boldsymbol{\Sigma}}$ and obtain the submatrix of $\hat{\boldsymbol{\Sigma}}^{-1}$ corresponding to $\hat{\theta}_m^L$ and $\hat{\theta}_m^P$. Let us denote this matrix as $\hat{\boldsymbol{\Sigma}}_m^*$. Clearly, this matrix is 2×2 and also positive definite. Then we can calculate $\hat{\mathbf{c}}_m = (\hat{c}_{m1}, \hat{c}_{m2})^T = (\mathbf{h}^T \hat{\boldsymbol{\Sigma}}_m^* \mathbf{h})^{-1} \hat{\boldsymbol{\Sigma}}_m^* \mathbf{h}$, where $\mathbf{h} = (1,1)^T$. By using the form of the optimal estimator in Wei et al. (1989), we obtain new weighted estimator for mth covariate, say $\hat{\theta}_m^{MWE}$, where

$$\hat{\theta}_m^{MWE} = \hat{c}_{m1}\hat{\theta}_m^L + \hat{c}_{m2}\hat{\theta}_m^P.$$

We can repeat this step for the other regression coefficients. Then we obtain $\hat{\boldsymbol{\theta}}^{MWE} = (\hat{\theta}_1^{MWE}, \ldots, \hat{\theta}_k^{MWE})^T$. In this first approach, weights are generated through using k number of 2×2 matrices. We can refer this first approach as 'marginal approach'.

Sometimes it is desirable to consider entire covariates all at once when obtaining weights. The second approach is to obtain the corresponding submatrix of $\hat{\boldsymbol{\Sigma}}^{-1}$ for $\{(\hat{\boldsymbol{\theta}}^L)^T, (\hat{\boldsymbol{\theta}}^P)^T\}^T$. We denote this matrix as $\hat{\boldsymbol{\Sigma}}^{**}$. This approach is different from first one in that $\hat{\boldsymbol{\Gamma}}_m$ consists of elements of the covariance matrix from $\hat{\theta}_m^L$ and $\hat{\theta}_m^P$ but now $\hat{\boldsymbol{\Sigma}}^{**}$ has elements of covariance matrix from corresponding entire $\{(\hat{\boldsymbol{\theta}}^L)^T, (\hat{\boldsymbol{\theta}}^P)^T\}^T$. This approach reflects the effect of $\{(\hat{\boldsymbol{\theta}}^L)^T, (\hat{\boldsymbol{\theta}}^P)^T\}^T$ jointly on our new estimator. Let $\mathbf{E}$ be a $2k \times k$ matrix such that

$$\mathbf{E} = \begin{pmatrix} 1, 0, \ldots, 0 \\ 0, 1, \ldots, 0 \\ \vdots \\ 0, 0, \ldots, 1 \\ 1, 0, \ldots, 0 \\ 0, 1, \ldots, 0 \\ \vdots \\ 0, 0, \ldots, 1 \end{pmatrix}.$$

$\mathbf{E}$ is a multivariate extension of $\mathbf{h}$. Note that $\mathbf{E}$ is concatenation of two $k \times k$ identity matrices by row. Entries that are 1 in these two $k \times k$ identity matrices are source of weights for $\hat{\boldsymbol{\theta}}^L$ and $\hat{\boldsymbol{\theta}}^P$. The next step is to construct $\hat{\mathbf{B}}$, which is

$$\hat{\mathbf{B}} = \{(\mathbf{E}^T \hat{\boldsymbol{\Sigma}}^{**} \mathbf{E})^{-1} \mathbf{E}^T \hat{\boldsymbol{\Sigma}}^{**}\}^T.$$

Then $\hat{\mathbf{B}}$ has the form

$$\begin{pmatrix} \hat{c}^*_{1,1} & \ldots & \hat{c}^*_{1,k} \\ \hat{c}^*_{2,1} & \ldots & \hat{c}^*_{2,k} \\ \vdots & \vdots & \vdots \\ \hat{c}^*_{(k+1),1} & \ldots & \hat{c}^*_{(k+1),k} \\ \vdots & \vdots & \vdots \\ \hat{c}^*_{2k,1} & \ldots & \hat{c}^*_{2k,k} \end{pmatrix}.$$

This matrix is a multivariate extension of $\hat{\mathbf{c}}_m$ from the first approach. This matrix is a contrast matrix in the sense that $\hat{c}^*_{m,m} + \hat{c}^*_{(k+m),m} = 1$ for the mth regression coefficient of $\hat{\boldsymbol{\theta}}^L$ and $\hat{\boldsymbol{\theta}}^P$. Moreover, $\hat{c}^*_{p,p} + \hat{c}^*_{(k+p),p} = 0$ for $p \neq m = 1, \ldots, k$. Using a vector form, from this approach our new estimator, say $\hat{\boldsymbol{\theta}}^{JWE}$,

$$\hat{\boldsymbol{\theta}}^{JWE} = (\hat{\theta}^{JWE}_1, \ldots, \hat{\theta}^{JWE}_k)^T = (\hat{c}^*_{1,1}\hat{\theta}^L_1 + \hat{c}^*_{(k+1),1}\hat{\theta}^P_1, \ldots, \hat{c}^*_{k,k}\hat{\theta}^L_k + \hat{c}^*_{(2k),k}\hat{\theta}^P_k)^T.$$

We can also refer this approach as the 'joint approach'.

Now the key step is to obtain $\hat{\boldsymbol{\Sigma}}$. We use the resampling approach of Parzen et al. (1994), which was also used in Lin et al. (1996) and Peng and Fine (2006). Let $\hat{\boldsymbol{\alpha}}^L = \{\hat{\boldsymbol{\eta}}^T, (\hat{\boldsymbol{\theta}}^L)^T\}^T$ and $\hat{\boldsymbol{\alpha}}^P = \{\hat{\boldsymbol{\eta}}^T, (\hat{\boldsymbol{\theta}}^P)^T\}^T$. From Lin et al. (1996) and Peng and

Fine (2006), we have

$$\mathbf{W}_i^{(1)} = \Delta_i \left[\mathbf{Z}_i - \frac{\sum_{j=1}^n I\{\tilde{D}_j^*(\hat{\boldsymbol{\eta}}) \geq \tilde{D}_i^*(\hat{\boldsymbol{\eta}})\}\mathbf{Z}_j}{\sum_{j=1}^n I\{\tilde{D}_j^*(\hat{\boldsymbol{\eta}}) \geq \tilde{D}_i^*(\hat{\boldsymbol{\eta}})\}} \right] - \sum_{l=1}^n \frac{\Delta_l I\{\tilde{D}_i^*(\hat{\boldsymbol{\eta}}) \geq \tilde{D}_l^*(\hat{\boldsymbol{\eta}})\}}{\sum_{j=1}^n I\{\tilde{D}_j^*(\hat{\boldsymbol{\eta}}) \geq \tilde{D}_l^*(\hat{\boldsymbol{\eta}})\}}$$
$$\times \left[\mathbf{Z}_i - \frac{\sum_{j=1}^n I\{\tilde{D}_j^*(\hat{\boldsymbol{\eta}}) \geq \tilde{D}_l^*(\hat{\boldsymbol{\eta}})\}\mathbf{Z}_j}{\sum_{j=1}^n I\{\tilde{D}_j^*(\hat{\boldsymbol{\eta}}) \geq \tilde{D}_l^*(\hat{\boldsymbol{\eta}})\}} \right],$$

$$\mathbf{W}_i^{(2)} = \tilde{\delta}_i^*(\hat{\boldsymbol{\alpha}}^L) \left[\mathbf{Z}_i - \frac{\sum_{j=1}^n I\{\tilde{X}_j^*(\hat{\boldsymbol{\alpha}}^L) \geq \tilde{X}_i^*(\hat{\boldsymbol{\alpha}}^L)\}\mathbf{Z}_j}{\sum_{j=1}^n I\{\tilde{X}_j^*(\hat{\boldsymbol{\alpha}}^L) \geq \tilde{X}_i^*(\hat{\boldsymbol{\alpha}}^L)\}} \right] - \sum_{l=1}^n \frac{\tilde{\delta}_l^*(\hat{\boldsymbol{\alpha}}^L) I\{\tilde{X}_i^*(\hat{\boldsymbol{\alpha}}^L) \geq \tilde{X}_l^*(\hat{\boldsymbol{\alpha}}^L)\}}{\sum_{j=1}^n I\{\tilde{X}_j^*(\hat{\boldsymbol{\alpha}}^L) \geq \tilde{X}_l^*(\hat{\boldsymbol{\alpha}}^L)\}}$$
$$\times \left[\mathbf{Z}_i - \frac{\sum_{j=1}^n I\{\tilde{X}_j^*(\hat{\boldsymbol{\alpha}}^L) \geq \tilde{X}_l^*(\hat{\boldsymbol{\alpha}}^L)\}\mathbf{Z}_j}{\sum_{j=1}^n I\{\tilde{X}_j^*(\hat{\boldsymbol{\alpha}}^L) \geq \tilde{X}_l^*(\hat{\boldsymbol{\alpha}}^L)\}} \right],$$

and

$$\mathbf{W}_i^{(3)} = \frac{2}{n-1} \sum_{j=1}^n (\mathbf{Z}_i - \mathbf{Z}_j)\phi_{ij}(\hat{\boldsymbol{\alpha}}^P).$$

Define

$$\mathbf{W}_i = \begin{pmatrix} \mathbf{W}_i^{(1)} \\ \mathbf{W}_i^{(2)} \\ \mathbf{W}_i^{(3)} \end{pmatrix}.$$

A consistent estimator of $\boldsymbol{\Omega}_0$ is

$$\hat{\boldsymbol{\Omega}} = \frac{1}{n} \sum_{i=1}^n \mathbf{W}_i \mathbf{W}_i^T.$$

We then solve the estimating equation

$$\mathbf{G}_n(\boldsymbol{\beta}) = -n^{-1/2} \sum_{i=1}^n \mathbf{W}_i Q_i, \tag{4.9}$$

where Q_i $(i = 1, \ldots, n)$ represent standard normal random variables. Note that $\mathbf{G}_n(\boldsymbol{\beta}) = [\mathbf{S}_n^T(\boldsymbol{\eta}), \{\mathbf{U}_n^L(\boldsymbol{\alpha})\}^T, \{\mathbf{U}_n^P(\boldsymbol{\alpha})\}^T]^T$ is the joint estimating function for $(\boldsymbol{\eta}_0^T, \boldsymbol{\theta}_0^T, \boldsymbol{\theta}_0^T)^T$. By solving equation (4.9), we obtain many realizations of $\hat{\boldsymbol{\beta}}$s, say $\hat{\boldsymbol{\beta}}^R = \{(\hat{\boldsymbol{\eta}}^*)^T, (\hat{\boldsymbol{\theta}}^{L*})^T, (\hat{\boldsymbol{\theta}}^{P*})^T\}^T$ where $\{(\hat{\boldsymbol{\eta}}^*)^T, (\hat{\boldsymbol{\theta}}^{L*})^T, (\hat{\boldsymbol{\theta}}^{P*})^T\}^T$ are solutions from (4.9). The next theorem, combined with Theorem 4.2, justifies the resampling approach for calculating $\hat{\boldsymbol{\Sigma}}$.

Theorem 4.3. Based on the technical conditions in Parzen et al. (1994), the unconditional distribution of $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ is same asymptotically as the conditional distribution of $n^{1/2}(\hat{\boldsymbol{\beta}}^R - \hat{\boldsymbol{\beta}})$ where $\hat{\boldsymbol{\beta}}^R$ are realizations of $\hat{\boldsymbol{\beta}}$ from resampling.

Proof. Recall that for any $\boldsymbol{\beta}$ in the small neighborhood of $\boldsymbol{\beta}_0$, we have

$$\mathbf{G}_n(\boldsymbol{\beta}) = \mathbf{G}_n(\boldsymbol{\beta}_0) + n^{1/2}\boldsymbol{\Gamma}_0(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + o_p(1) \tag{4.10}$$

Note that $\hat{\boldsymbol{\beta}}^R$ are solutions of equation (4.9). By conditioning on observed data and using expansion (4.10) as well as by adapting arguments in Lin et al. (1996) and Parzen et al. (1994),

$$\mathbf{G}_n(\hat{\boldsymbol{\beta}}^R) = \mathbf{G}_n(\hat{\boldsymbol{\beta}}) + n^{1/2}\boldsymbol{\Gamma}_0(\hat{\boldsymbol{\beta}}^R - \hat{\boldsymbol{\beta}}) + o_p(1)$$

and hence,

$$n^{1/2}(\hat{\boldsymbol{\beta}}^R - \hat{\boldsymbol{\beta}}) = -\boldsymbol{\Gamma}_0^{-1} n^{-1/2} \sum_{i=1}^{n} \mathbf{W}_i Q_i + o_p(1)$$

Note that $n^{-1/2}\sum_{i=1}^{n} \mathbf{W}_i Q_i$ is asymptotically normal with covariance matrix $\boldsymbol{\Sigma}_0$. Then given observed data, the distribution of $n^{1/2}(\hat{\boldsymbol{\beta}}^R - \hat{\boldsymbol{\beta}})$ is asymptotically normal with covariance matrix $\boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Sigma}_0\boldsymbol{\Gamma}_0^{-1}$. Hence conditional distribution of $n^{1/2}(\hat{\boldsymbol{\beta}}^R - \hat{\boldsymbol{\beta}})$ on observed data is asymptotically same as unconditional distribution of $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$. □

For $m = 1, \ldots k$ and $j = 1, \ldots, M$, let $(\hat{\eta}_m^*)^{(j)}$, $(\hat{\theta}_m^{L*})^{(j)}$ and $(\hat{\theta}_m^{P*})^{(j)}$ be jth realizations of an element $\hat{\eta}_m, \hat{\theta}_m^L$ and $\hat{\theta}_m^P$ corresponding to mth covariate, respectively. The algorithm for the first approach is as follows.

1. By resampling, calculate the covariance matrix $\hat{\boldsymbol{\Sigma}}$ using realizations $(\hat{\eta}_m^*)^{(j)}$, $(\hat{\theta}_m^{L*})^{(j)}$ and $(\hat{\theta}_m^{P*})^{(j)}$, ($m = 1, \ldots, k$ and $j = 1, \ldots, M$).

2. From $\hat{\boldsymbol{\Sigma}}^{-1}$, obtain the covariance matrix corresponding to $\hat{\theta}_m^L$ and $\hat{\theta}_m^P$, say $\hat{\boldsymbol{\Sigma}}_m^*$.

3. Calculate $\hat{\mathbf{c}}_m = (\hat{c}_{m1}, \hat{c}_{m2})^T = (\mathbf{h}^T\hat{\boldsymbol{\Sigma}}_m^*\mathbf{h})^{-1}\hat{\boldsymbol{\Sigma}}_m^*\mathbf{h}$ where $\mathbf{h} = (1, 1)^T$ and obtain the new estimate $\hat{\theta}_m^{MWE} = \hat{c}_{m1}\hat{\theta}_m^L + \hat{c}_{m2}\hat{\theta}_m^P$.

4. Repeat step 3 for all covariates.

The algorithm for the second approach is as follows.

1. By resampling, calculate the covariance matrix $\hat{\mathbf{\Sigma}}$ using realizations $(\hat{\eta}_m^*)^{(j)}$, $(\hat{\theta}_m^{L*})^{(j)}$ and $(\hat{\theta}_m^{P*})^{(j)}$ $(m = 1, \ldots, k$ and $j = 1, \ldots, M)$.

2. Obtain $\hat{\mathbf{\Sigma}}^{**}$ from $\hat{\mathbf{\Sigma}}$.

3. From $\hat{\mathbf{\Sigma}}^{**}$ and $\mathbf{E}$, obtain $\hat{\mathbf{B}}$.

4. Calculate the new estimate $\hat{\theta}_m^{JWE} = \hat{c}_{m,m}^* \hat{\theta}_m^L + \hat{c}_{k+m,m}^* \hat{\theta}_m^P$, where $\hat{c}_{j,l}^*$ be the element of jth row and lth column of $\hat{\mathbf{B}}$.

By Theorem 4.1 and Theorem 4.2, our new estimators are consistent and asymptotically normal.

4.3 Model checking

For assessing the adequacy of the model, since our weight estimator is based on estimators from Lin et al. (1996) and Peng and Fine (2006), it is reasonable to consider entire processes from Lin et al. (1996) and Peng and Fine (2006). In this case, we extend model checking technique from Lin et al. (1996). As defined in Lin et al. (1996), Let $N_{1i}(t;\boldsymbol{\eta}) = \Delta_i I\{\tilde{D}_i^*(\eta) \leq t\}$ and $N_{2i}(t;\boldsymbol{\alpha}) = \tilde{\delta}_i^*(\boldsymbol{\alpha}) I\{\tilde{X}_i^*(\boldsymbol{\alpha}) \leq t\}$, where $i = 1, ..., n$. Then Nelson-Aalen estimators for the event of interest and dependent censoring are

$$\hat{\Lambda}_0^{(1)}(u;\boldsymbol{\eta}) = \int_{-\infty}^{t} \frac{\sum_{i=1}^n dN_{1i}(u;\boldsymbol{\eta})}{\sum_{j=1}^n I\{\tilde{D}_j^*(\boldsymbol{\eta}) \geq u\}} \qquad \hat{\Lambda}_0^{(2)}(u;\boldsymbol{\alpha}) = \int_{-\infty}^{t} \frac{\sum_{i=1}^n dN_{2i}(u;\boldsymbol{\alpha})}{\sum_{j=1}^n I\{\tilde{X}_j^*(\boldsymbol{\alpha}) \geq u\}}.$$

Note that by (4.1) and (4.2), martingale residuals are defined as

$$\hat{M}_{1i}(t;\hat{\boldsymbol{\eta}}) = N_{1i}(t;\hat{\boldsymbol{\eta}}) - \int_{-\infty}^{t} I\{\tilde{D}_i^*(\hat{\boldsymbol{\eta}}) \geq u\} d\hat{\Lambda}_0^{(1)}(u,\hat{\boldsymbol{\eta}})$$

$$\hat{M}_{2i}(t;\hat{\boldsymbol{\alpha}}) = N_{2i}(t;\hat{\boldsymbol{\alpha}}) - \int_{-\infty}^{t} I\{\tilde{X}_i^*(\hat{\boldsymbol{\alpha}}) \geq u\} d\hat{\Lambda}_0^{(2)}(u,\hat{\boldsymbol{\alpha}}),$$

where $\hat{\boldsymbol{\alpha}}$ can be either $\hat{\boldsymbol{\alpha}}^L = \{\hat{\boldsymbol{\eta}}^T, (\hat{\boldsymbol{\theta}}^L)^T\}^T$ or $\hat{\boldsymbol{\alpha}}^P = \{\hat{\boldsymbol{\eta}}^T, (\hat{\boldsymbol{\theta}}^P)^T\}^T$. Then as defined in Lin et al. (1996),

$$\mathbf{S}_n(s;\boldsymbol{\eta}) = n^{-1/2} \sum_{i=1}^n \mathbf{Z}_i \hat{M}_{1i}(s;\boldsymbol{\eta}) \quad \mathbf{U}_n(t;\boldsymbol{\alpha}) = n^{-1/2} \sum_{i=1}^n \mathbf{Z}_i \hat{M}_{2i}(t;\boldsymbol{\alpha}).$$

Then similar to Lin et al. (1996) and Peng and Fine (2006), we can substitute $\hat{\boldsymbol{\eta}}$ on $\mathbf{S}_n(s;\boldsymbol{\eta})$, $\hat{\boldsymbol{\alpha}}^L$ and $\hat{\boldsymbol{\alpha}}^P$ on $\mathbf{U}_n(t;\boldsymbol{\alpha})$. $[\mathbf{S}_n^T(s;\hat{\boldsymbol{\eta}}), \{\mathbf{U}_n(t;\hat{\boldsymbol{\alpha}}^L)\}^T, \{\mathbf{U}_n(t;\hat{\boldsymbol{\alpha}}^P)\}^T]^T$ are called observed score processes with respect to dependent censoring and the event of interest, respectively (Ghosh, 2010; Lin et al. 1996; Peng and Fine, 2006). We can construct $[\hat{\mathbf{S}}_n^T(s;\hat{\boldsymbol{\eta}}^*), \{\hat{\mathbf{U}}_n^L(t;\hat{\boldsymbol{\alpha}}^{L*})\}^T, \{\hat{\mathbf{U}}_n^P(v;\hat{\boldsymbol{\alpha}}^{P*})\}^T]^T$ (Lin et al. 1996; Peng and Fine, 2006), where

$$\begin{aligned}\hat{\mathbf{S}}_n(s;\hat{\boldsymbol{\eta}}^*) = n^{-1/2}\sum_{i=1}^n \int_{-\infty}^s \left[\mathbf{Z}_i - \frac{\sum_{j=1}^n I\{\tilde{D}_j^*(\hat{\boldsymbol{\eta}}) \geq w\}\mathbf{Z}_j}{\sum_{j=1}^n I\{\tilde{D}_j^*(\hat{\boldsymbol{\eta}}) \geq w\}}\right] d\hat{M}_{1i}(w;\hat{\boldsymbol{\eta}})Q_i \\ + \mathbf{S}_n(s;\hat{\boldsymbol{\eta}}^*) - \mathbf{S}_n(s;\hat{\boldsymbol{\eta}})\end{aligned}$$

$$\begin{aligned}\hat{\mathbf{U}}_n^L(t;\hat{\boldsymbol{\alpha}}^{L*}) &= n^{-1/2}\sum_{i=1}^n \int_{-\infty}^t \left[\mathbf{Z}_i - \frac{\sum_{j=1}^n I\{\tilde{X}_j^*(\hat{\boldsymbol{\alpha}}^L) \geq w\}\mathbf{Z}_j}{\sum_{j=1}^n I\{\tilde{X}_j^*(\hat{\boldsymbol{\alpha}}^L) \geq w\}}\right] d\hat{M}_{2i}(w;\hat{\boldsymbol{\alpha}}^L)Q_i \\ &\quad + \mathbf{U}_n(t;\hat{\boldsymbol{\alpha}}^{L*}) - \mathbf{U}_n(t;\hat{\boldsymbol{\alpha}}^L) \\ \hat{\mathbf{U}}_n^P(v;\hat{\boldsymbol{\alpha}}^{P*}) &= n^{-1/2}\sum_{i=1}^n \int_{-\infty}^v \left[\mathbf{Z}_i - \frac{\sum_{j=1}^n I\{\tilde{X}_j^*(\hat{\boldsymbol{\alpha}}^P) \geq w\}\mathbf{Z}_j}{\sum_{j=1}^n I\{\tilde{X}_j^*(\hat{\boldsymbol{\alpha}}^P) \geq w\}}\right] d\hat{M}_{2i}(w;\hat{\boldsymbol{\alpha}}^P)Q_i \\ &\quad + \mathbf{U}_n(v;\hat{\boldsymbol{\alpha}}^{P*}) - \mathbf{U}_n(v;\hat{\boldsymbol{\alpha}}^P),\end{aligned}$$

where $\hat{\boldsymbol{\alpha}}^{L*} = \{(\hat{\boldsymbol{\eta}}^*)^T, (\hat{\boldsymbol{\theta}}^{L*})^T\}^T$ and $\hat{\boldsymbol{\alpha}}^{P*} = \{(\hat{\boldsymbol{\eta}}^*)^T, (\hat{\boldsymbol{\theta}}^{P*})^T\}^T$. These three processes are called bootstrapped processes (Ghosh, 2010; Lin et al. 1996; Peng and Fine, 2006). We can plot the observed process with bootstrapped processes by randomly selecting 20 or 30 observations. Standard tests for goodness of fit can be performed by calculating Kolmogorov-Smirnov type test statistics. Test statistics are then defined by $\sup_s ||\mathbf{S}_n(s;\hat{\boldsymbol{\eta}})||, \sup_t ||\mathbf{U}_n(t;\hat{\boldsymbol{\alpha}}^L)||$, and $\sup_v ||\mathbf{U}_n(v;\hat{\boldsymbol{\alpha}}^P)||$. To calculate the null distribution of the test statistics, first we obtain jth realizations of bootstrap samples $(\hat{\boldsymbol{\eta}}^*)^{(j)}, (\hat{\boldsymbol{\theta}}^{L*})^{(j)}$ and $(\hat{\boldsymbol{\theta}}^{P*})^{(j)}$. Then we compute $BS_j = \sup_s ||\hat{\mathbf{S}}_n(s;(\hat{\boldsymbol{\eta}}^*)^{(j)})||, BS_j^L = \sup_t ||\hat{\mathbf{U}}_n^L(t;(\hat{\boldsymbol{\alpha}}^{L*})^{(j)})||$ and $BS_j^P = \sup_v ||\hat{\mathbf{U}}_n^P(v;(\hat{\boldsymbol{\alpha}}^{P*})^{(j)})||$, respectively for $j = 1, \ldots, M$, where $(\hat{\boldsymbol{\alpha}}^{L*})^{(j)}$ and $(\hat{\boldsymbol{\alpha}}^{P*})^{(j)}$ are jth realizations of bootstrap samples of $\hat{\boldsymbol{\alpha}}^{L*}$ and $\hat{\boldsymbol{\alpha}}^{P*}$. The p-values can be defined by

$$p_1 = \frac{1}{M}\sum_{j=1}^M I\{BS_j \geq \sup_s ||\mathbf{S}_n(s;\hat{\boldsymbol{\eta}})||\}$$

$$p_2 = \frac{1}{M}\sum_{j=1}^{M} I\{BS_j^L \geq \sup_t ||\mathbf{U}_n(t;\hat{\boldsymbol{\alpha}}^L)||\}$$

$$p_3 = \frac{1}{M}\sum_{j=1}^{M} I\{BS_j^P \geq \sup_v ||\mathbf{U}_n(v;\hat{\boldsymbol{\alpha}}^P)||\}.$$

(Hsieh et al. 2011). If a p-value is smaller than predetermined level, we reject the null hypothesis, which means that data does not have appropriate fit on our bivariate model. Note that a multiple testing problem arises for testing the models for $\boldsymbol{\theta}$. We address this by adjusting p-values based on a Bonferroni correction with two tests.

4.4 Simulation Studies

We consider two simulation settings. In first simulation setting, the errors follow a bivariate normal distribution with mean (0,1.2) with variance 1 and correlation ρ=0,0.25. The independent censoring time C is generated from $\log(U^*)$, where U^* has uniform distribution with minimum value 0 and maximum value 20. Covariate is $Z \sim Bernoulli(0.5)$, where $Bernoulli(0.5)$ is Bernoulli distribution with success probability 0.5. We run 500 simulation runs. Within each simulation run, 500 resampling runs are tried for covariance matrix calculation. Sample sizes are $N = 150$ and $N = 300$. If there is only one covariate in the model, the first and the second method of the weighted estimation are equivalent. Let this common weighed estimator be $\hat{\theta}^{WE}$. We calculate bias (Bias), mean squared error (MSE), mean of standard error (SEE), 95% coverage rate (Coverage). The coverage is based on the normal approximation. Moreover, to evaluate robustness of estimators, we also compute median of difference of the estimator from true value (Dmedian), median of squared error of estimates (Mediansq), and median of standard errors (Sdmedian). Results are summarized on Table 4.1 and Table 4.2.

In second simulation setting, we generate Gamma random variable ν with mean μ=1 and variance σ^2=0 or 1, then create $W = \exp(\epsilon^X)$, which is an exponential random variable with rate $4\nu^{-1}$ and $\exp(\epsilon^D)$ with an exponential random variable with rate ν^{-1}. Then we generate time to the event of interest by $\exp(X) = \exp(\boldsymbol{\theta}_0^T\mathbf{Z})\exp(\epsilon^X)$ and time to the dependent censoring by $\exp(D) = \exp(\boldsymbol{\eta}_0^T\mathbf{Z})\exp(\epsilon^D)$ (By notation in our paper, X, D and C are already

Table 4.1: Simulation result when $N = 150$ and $N = 300$, $\rho = 0$ with covariate $Bernoulli(0.5)$.

	$N = 150$			
	Bias (Dmedian[1])	MSE (Mediansq[2])	SEE (Sdmedian[3])	Coverage
$\hat{\theta}^L$	0.018 (0.018)	0.04 (0.017)	0.204 (0.2)	0.95
$\hat{\theta}^P$	0.021 (0.014)	0.036 (0.017)	0.193 (0.19)	0.96
$\hat{\theta}^{WE}$	0.016 (0.006)	0.036 (0.015)	0.188 (0.185)	0.95
	$N = 300$			
	Bias (Dmedian[1])	MSE (Mediansq[2])	SEE (Sdmedian[3])	Coverage
$\hat{\theta}^L$	-0.002 (-0.003)	0.017 (0.006)	0.140 (0.140)	0.95
$\hat{\theta}^P$	-0.001 (0.002)	0.016 (0.007)	0.133 (0.132)	0.95
$\hat{\theta}^{WE}$	-0.004 (-0.002)	0.016 (0.007)	0.130 (0.129)	0.94

Estimators : $\hat{\theta}^L$: the estimator by Lin et al. (1996); $\hat{\theta}^P$: the estimator by Peng and Fine (2006); $\hat{\theta}^{WE}$: the weighted estimator by the proposed approach (Note that the marginal approach and the joint approach are equal in one variable case)

[1] median of difference of the estimator from true value

[2] median of squared error

[3] median of standard error

Table 4.2: Simulation result when $N = 150$ and $N = 300$, $\rho = 0.25$ with covariate $Bernoulli(0.5)$.

	$N = 150$			
	Bias (Dmedian[1])	MSE (Mediansq[2])	SEE (Sdmedian[3])	Coverage
$\hat{\theta}^L$	0.005 (0.01)	0.036 (0.017)	0.198 (0.197)	0.95
$\hat{\theta}^P$	0.006 (0.007)	0.032 (0.015)	0.189 (0.188)	0.95
$\hat{\theta}^{WE}$	-0.001 (-0.006)	0.033 (0.016)	0.184 (0.183)	0.94
	$N = 300$			
	Bias (Dmedian[1])	MSE (Mediansq[2])	SEE (Sdmedian[3])	Coverage
$\hat{\theta}^L$	-0.003 (0.005)	0.018 (0.008)	0.138 (0.137)	0.95
$\hat{\theta}^P$	0.001 (0.007)	0.017 (0.007)	0.131 (0.131)	0.95
$\hat{\theta}^{WE}$	-0.003 (0.002)	0.017 (0.007)	0.129 (0.128)	0.95

Estimators: $\hat{\theta}^L$: the estimator by Lin et al. (1996); $\hat{\theta}^P$: the estimator by Peng and Fine (2006); $\hat{\theta}^{WE}$: the weighted estimator by the proposed approach (Note that the marginal approach and the joint approach are equal in one variable case)

[1] median of difference of the estimator from true value

[2] median of squared error

[3] median of standard error

log-transformed times. Thus in this context, $\exp(X), \exp(D)$ and $\exp(C)$ are times in the original scale). The independent censoring time $\exp(C)$ has uniform distribution with minimum value 0 and maximum value 20. True parameter values are $\boldsymbol{\theta}_0 = (0.5, 1)^T$ and $\boldsymbol{\eta}_0 = (1, 0.5)^T$ and covariates are $Z_1 \sim U(0, 1)$, where $U(0, 1)$ is uniform distribution with minimum value 0 and maximum value 1 and $Z_2 \sim Bernoulli(0.5)$. We run 500 simulation runs. Within each simulation run, 500 resampling runs are tried for covariance matrix calculation. Let $\hat{\boldsymbol{\theta}}^{MWE}$ be weighted estimators from calculating weights marginally (the first proposed method) and let $\hat{\boldsymbol{\theta}}^{JWE}$ be weighted estimators from calculating weights jointly (the second proposed method). We compute the same quantities as we did in the first stage of the simulation study. Results are summarized on Table 4.3 and Table 4.4.

In these simulation results, we can see that our weighted estimators have good results. In both cases, bias and mean squared error of our new estimator has similar performance compared to the estimators by Lin et al. (1996) and Peng and Fine (2006). Mean of standard errors and median of standard errors are smaller than the estimators by Lin et al. (1996) and Peng and Fine (2006). Moreover, computation results for the median of difference of the estimators from true value and the median of squared error imply that our proposed estimator is comparable with the estimators from the original methods.

In the first simulation setting, the difference of standard error between our proposed estimator and $\hat{\theta}^L$ is bigger than the one between $\hat{\theta}^P$ and the proposed estimator. In the second simulation setting, the phenomenon is the opposite. Furthermore, in the first simulation setting, $\hat{\theta}^P$ has lower standard error on average than one of $\hat{\theta}^L$ while $\hat{\boldsymbol{\theta}}^L$ have better efficiency (with respect to standard error) than ones by $\hat{\boldsymbol{\theta}}^P$ in the second simulation setting. This simulation result verifies our claim, which means that neither estimator is better than another. Our proposed estimator takes advantage of smaller standard error with achieving small bias and correct coverage except $N = 150$ with $\sigma^2 = 1$ in the second simulation setting. In this scenario, empirical coverage of proposed estimators is lower than nominal 95% coverage. This is due to low coverage of $\hat{\boldsymbol{\theta}}^L$. Since we combine $\hat{\boldsymbol{\theta}}^L$ and $\hat{\boldsymbol{\theta}}^P$, if one of them has low coverage, it is highly likely that the coverage of the weighted estimator may also be below the nominal coverage.

Table 4.3: Simulation result when $N = 150$ and $N = 300$, $\sigma^2 = 0$ with two covariates ($Z_1 : U(0,1), Z_2 : Bernoulli(0.5)$).

	$N = 150$							
	Bias (Dmedian[1])		MSE (Mediansq[2])		SEE (Sdmedian[3])		Coverage	
	Z_1	Z_2	Z_1	Z_2	Z_1	Z_2	Z_1	Z_2
$\hat{\boldsymbol{\theta}}^L$	0.0001 (0.002)	0.002 (-0.005)	0.12 (0.052)	0.042 (0.018)	0.358 (0.352)	0.226 (0.222)	0.96	0.96
$\hat{\boldsymbol{\theta}}^P$	-0.003 (0.003)	-0.003 (-0.002)	0.158 (0.074)	0.051 (0.023)	0.427 (0.418)	0.243 (0.241)	0.96	0.96
$\hat{\boldsymbol{\theta}}^{MWE}$	0.003 (-0.007)	0.003 (0.003)	0.123 (0.053)	0.043 (0.019)	0.351 (0.349)	0.219 (0.218)	0.95	0.95
$\hat{\boldsymbol{\theta}}^{JWE}$	0.003 (-0.007)	0.004 (0.001)	0.123 (0.055)	0.043 (0.018)	0.351 (0.349)	0.219 (0.217)	0.94	0.95
	$N = 300$							
	Bias (Dmedian[1])		MSE (Mediansq[2])		SEE (Sdmedian[3])		Coverage	
	Z_1	Z_2	Z_1	Z_2	Z_1	Z_2	Z_1	Z_2
$\hat{\boldsymbol{\theta}}^L$	-0.012 (-0.013)	0.004 (0.001)	0.065 (0.028)	0.02 (0.01)	0.257 (0.255)	0.148 (0.148)	0.95	0.96
$\hat{\boldsymbol{\theta}}^P$	-0.014 (-0.017)	-0.001 (-0.015)	0.081 (0.035)	0.024 (0.013)	0.283 (0.281)	0.162 (0.162)	0.95	0.96
$\hat{\boldsymbol{\theta}}^{MWE}$	-0.01 (-0.012)	0.003 (-0.001)	0.064 (0.031)	0.02 (0.01)	0.252 (0.251)	0.146 (0.146)	0.95	0.96
$\hat{\boldsymbol{\theta}}^{JWE}$	-0.01 (-0.014)	0.003 (0.002)	0.064 (0.032)	0.02 (0.01)	0.251 (0.25)	0.146 (0.146)	0.95	0.96

Estimators : $\hat{\boldsymbol{\theta}}^L$: the estimator by Lin et al. (1996); $\hat{\boldsymbol{\theta}}^P$: the estimator by Peng and Fine (2006); $\hat{\boldsymbol{\theta}}^{MWE}$: the weighted estimator by the marginal approach; $\hat{\boldsymbol{\theta}}^{JWE}$: the weighted estimator by the joint approach

[1] median of difference of the estimator from true value

[2] median of squared error

[3] median of standard error

Table 4.4: Simulation result when $N = 150$ and $N = 300$, $\sigma^2 = 1$ with two covariates ($Z_1 : U(0,1), Z_2 : Bernoulli(0.5)$).

	$N = 150$							
	Bias (Dmedian[1])		MSE (Mediansq[2])		SEE (Sdmedian[3])		Coverage	
	Z_1	Z_2	Z_1	Z_2	Z_1	Z_2	Z_1	Z_2
$\hat{\boldsymbol{\theta}}^L$	-0.010 (-0.003)	-0.033 (-0.039)	0.273 (0.127)	0.086 (0.038)	0.441 (0.434)	0.315 (0.312)	0.90	0.96
$\hat{\boldsymbol{\theta}}^P$	0.002 (-0.002)	-0.032 (-0.046)	0.295 (0.142)	0.095 (0.038)	0.559 (0.552)	0.325 (0.324)	0.95	0.96
$\hat{\boldsymbol{\theta}}^{MWE}$	-0.009 (-0.008)	-0.030 (-0.031)	0.263 (0.128)	0.085 (0.041)	0.437 (0.432)	0.303 (0.301)	0.90	0.96
$\hat{\boldsymbol{\theta}}^{JWE}$	-0.009 (-0.008)	-0.030 (-0.03)	0.262 (0.128)	0.086 (0.04)	0.437 (0.432)	0.303 (0.301)	0.90	0.96
	$N = 300$							
	Bias (Dmedian[1])		MSE (Mediansq[2])		SEE (Sdmedian[3])		Coverage	
	Z_1	Z_2	Z_1	Z_2	Z_1	Z_2	Z_1	Z_2
$\hat{\boldsymbol{\theta}}^L$	-0.024 (-0.038)	0.003 (-0.006)	0.133 (0.057)	0.04 (0.019)	0.345 (0.344)	0.211 (0.21)	0.94	0.96
$\hat{\boldsymbol{\theta}}^P$	-0.016 (-0.016)	0.012 (0.003)	0.148 (0.059)	0.045 (0.02)	0.384 (0.382)	0.222 (0.222)	0.96	0.97
$\hat{\boldsymbol{\theta}}^{MWE}$	-0.024 (-0.035)	0.007 (-0.003)	0.134 (0.058)	0.04 (0.019)	0.341 (0.341)	0.207 (0.207)	0.94	0.96
$\hat{\boldsymbol{\theta}}^{JWE}$	-0.025 (-0.035)	0.007 (-0.002)	0.135 (0.058)	0.039 (0.018)	0.341 (0.341)	0.206 (0.207)	0.94	0.96

Estimators : $\hat{\boldsymbol{\theta}}^L$: the estimator by Lin et al. (1996); $\hat{\boldsymbol{\theta}}^P$: the estimator by Peng and Fine (2006); $\hat{\boldsymbol{\theta}}^{MWE}$: the weighted estimator by the marginal approach; $\hat{\boldsymbol{\theta}}^{JWE}$: the weighted estimator by the joint approach

[1] median of difference of the estimator from true value

[2] median of squared error

[3] median of standard error

4.5 Real data analysis

We applied our method to data from the AIDS Clinical Trial Group (ACTG) Study 364 (Albrecht et al. 2001), which was used in Peng and Fine (2006). This multicenter randomized study investigated patients whose plasma RNA level is at least 500 copies per ml. Subjects were assigned to three treatments, nelfinavir (NFV), efavirenz (EFV), and combination of nelfinavir and efavirenz (NFV + EFV). Details about this study can be found in Albrecht et al. (2001).

The two failure times are time to HIV RNA level greater than 2000 copies per ml and time to withdrawal of study. Let X be the first time when HIV RNA level is greater than 2000 copies per ml and D be time to withdrawal of study. We considered four covariates and 194 observations. Z_1 takes value 1 if a patient receives EFV and 0 otherwise. Z_2 takes value 1 if a patient receives NFV + EFV and 0 otherwise. Z_3 is New3TC, which takes value 1 if lamivudine is given as a new nucleoside analogue therapy to a patient and 0 otherwise. Z_4 is logarithm of RNA level at the start of the study.

Table 4.5 and Table 4.6 show the point estimates and standard errors of $\hat{\boldsymbol{\eta}}$, $\hat{\boldsymbol{\theta}}^L$, $\hat{\boldsymbol{\theta}}^P$, $\hat{\boldsymbol{\theta}}^{MWE}$ and $\hat{\boldsymbol{\theta}}^{JWE}$. Our method works well for the models with and without New3TC on all covariates. Some variables are seen to be statistically significant based on the weighted estimator while they are not by Lin et al. (1996) or Peng and Fine (2006). For example, let us consider the effect of EFV on the time to first virologic failure. By Table 4.6, the estimated effect by using approach of Lin et al. (1996) is 0.475 and its standard error is 0.250. From the approach of Peng and Fine (2006), an estimate is 0.464 and its standard error is 0.281. Based on the fact that estimators are asymptotically normal, from Wald test using Lin et al. (1996) and Peng and Fine (2006), EFV is not a statistically significant variable at the 5% significance level. On the other hand, a weighted estimate using first approach is 0.471 and its standard error is 0.222. In this case, EFV is a statistically significant variable at the 5% significance level.

Observed score process with bootstrapped processes for withdrawal of study with respect to Z_1 is shown in Figure 4.1. Figure 4.2 and Figure 4.3 show observed score processes and bootstrapped processes of the first virologic failure using $\hat{\boldsymbol{\alpha}}^L, \hat{\boldsymbol{\alpha}}^P$ with respect to Z_1. These three plots are based on the model without New3TC. They are fluctuating around zero, so it seems that there is no graphical evidence for

Table 4.5: Point estimates with standard errors of covariates in AIDS study for model without New3TC (Standard errors are shown in parenthesis).

Covariates	$\hat{\boldsymbol{\eta}}$	$\hat{\boldsymbol{\theta}}^L$	$\hat{\boldsymbol{\theta}}^P$	$\hat{\boldsymbol{\theta}}^{MWE}$	$\hat{\boldsymbol{\theta}}^{JWE}$
EFV[1]	0.753 (0.339)	0.115 (0.219)	0.375 (0.269)	0.168 (0.206)	0.2 (0.205)
NFV[2]+ EFV	0.674 (0.255)	1.128 (0.239)	1.091 (0.309)	1.120 (0.222)	1.114 (0.222)
log(RNA)[3]	-0.544 (0.154)	-0.464 (0.215)	-0.531 (0.169)	-0.507 (0.163)	-0.511 (0.162)

[1] efavirenz
[2] nelfinavir
[3] logarithm of RNA at the start of the study

Table 4.6: Point estimates with standard errors of covariates in AIDS study for model with New3TC (Standard errors are shown in parenthesis).

Covariates	$\hat{\boldsymbol{\eta}}$	$\hat{\boldsymbol{\theta}}^L$	$\hat{\boldsymbol{\theta}}^P$	$\hat{\boldsymbol{\theta}}^{MWE}$	$\hat{\boldsymbol{\theta}}^{JWE}$
EFV[1]	0.770 (0.278)	0.475 (0.250)	0.464 (0.281)	0.471 (0.222)	0.471 (0.222)
NFV[2]+ EFV	0.650 (0.260)	1.353 (0.277)	1.246 (0.338)	1.333 (0.263)	1.317 (0.261)
New3TC[3]	0.927 (0.355)	1.449 (0.296)	1.374 (0.328)	1.431 (0.267)	1.420 (0.261)
log(RNA)[4]	-0.631 (0.183)	-0.654 (0.289)	-0.661 (0.218)	-0.659 (0.216)	-0.660 (0.215)

[1] efavirenz
[2] nelfinavir
[3] lamivudine as new nucleoside analogue therapy
[4] logarithm of RNA at the start of the study

lack of fit. The p-value for the lack of fit tests of withdrawal is 0.952 and the first virologic failure using $\hat{\boldsymbol{\alpha}}^L$ and $\hat{\boldsymbol{\alpha}}^P$ are 0.918 and 0.959 respectively. With graphical checking, the p-value indicates that there is no evidence for violation of the model assumption.

For purposes of interpretation, since D represents a standard survival time, the interpretation of $\hat{\boldsymbol{\eta}}$ is in terms of covariate effect for survival time. However, since the observed time for X depends on D, interpretation of $\hat{\boldsymbol{\theta}}$ is difficult. One way to interpret $\hat{\boldsymbol{\theta}}$ is to assume that D does not exist and interpret the effect of $\hat{\boldsymbol{\theta}}$ on X only. This approach is possible if there exists a reasonable extrapolation mechanism for X (Prentice et al. 1978). However, considering the estimation structure for $\boldsymbol{\theta}$, it is difficult to separate the effect of $\hat{\boldsymbol{\theta}}$ on X from effect of $\hat{\boldsymbol{\eta}}$ on D.

Figure 4.1: Plot of observed score process and bootstrapped processes of time to withdrawal of study with respect to Z_1. The thick line is the observed process and the dashed lines are the bootstrapped processes.

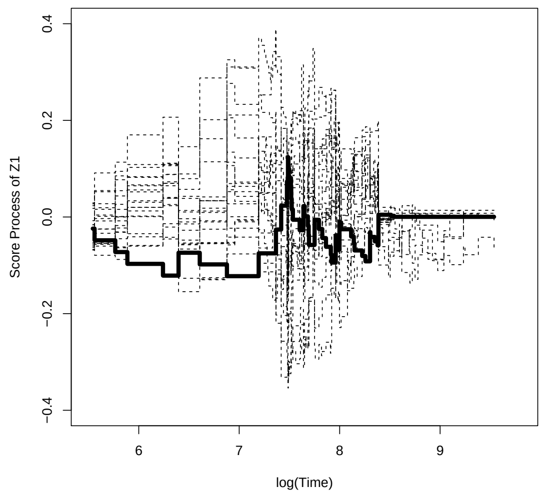
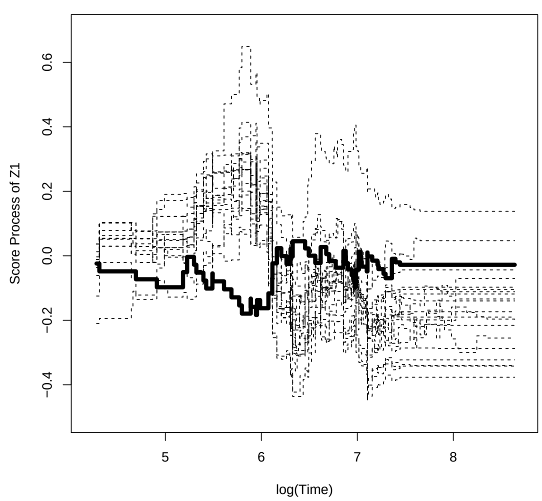


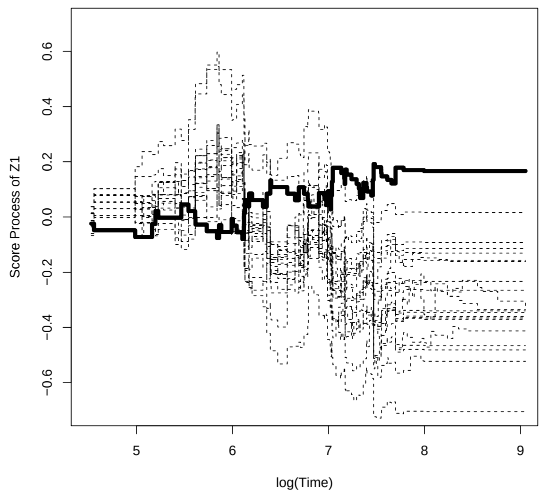
Figure 4.2: Plot of observed score process and bootstrapped processes of time to first virologic failure using $\hat{\boldsymbol{\alpha}}^L$ with respect to Z_1. The thick line is the observed process and the dashed lines are the bootstrapped processes.



4.6 Discussion

In this paper, we have proposed optimal estimators using combinations of the two estimators from Lin et al. (1996) and Peng and Fine (2006). Our methodology

Figure 4.3: Plot of observed score process and bootstrapped processes of time to first virologic failure using $\hat{\boldsymbol{\alpha}}^P$ with respect to Z_1. The thick line is the observed process and the dashed lines are the bootstrapped processes.



can be extended to a case of recurrent event with dependent censoring, which is extensively studied (Ghosh and Lin, 2003; Ghosh, 2010; Hsieh et al. 2011). We are currently working on this extension.

Optimality of the estimator has been discussed in other contexts. Recently, Lindsay et al. (2011) proposed optimal additive functions based on score functions. The main point of their method is to combine unbiased estimating functions. In our case, this would be combining estimating functions and new solution can be obtained by this estimating function. Comparing performance of this solution and our proposed estimator is of interest. This will be left open to future research.

Another way of achieving optimality is to use generalized method of moment estimator (Hansen, 1982). This estimator is a linear combination of estimating functions (Qu et al. 2000). In this case, the estimating functions have a greater dimension than the dimension of the parameter vector. The optimality is achieved by the linear combination. It is shown that the estimator from this linear combination of estimating functions is consistent and asymptotically normal (Hansen, 1982). In the literature of statistics, this idea is applied to generalized estimating equations (Qu et al. 2000). The estimating functions proposed by Qu et al. (2000) are called quadratic inference function. Recently, Xue et al. (2010) applied the quadratic inference function to Cox model.

Hansen (1982) and Qu et al. (2000) derived new estimating functions, while we combined two estimators directly. This idea of the generalized method of moments is very appealing, but the estimating functions of Lin et al. (1996) and Peng and Fine (2006) are nonsmooth. Finding derivative for the linear combination of the estimating functions, which is a key in the generalized method of moments, is challenging for our work because we cannot find the derivatives in the estimating functions proposed by Lin et al. (1996) and Peng and Fine (2006). Applying the idea of Hansen (1982) to AFT model will be interesting future research.

Our estimating functions to obtain estimators involve nonsmooth functions of $\boldsymbol{\eta}$ and $\boldsymbol{\alpha}$. Many literatures used a linear programming approach for estimating $\boldsymbol{\theta}$ (Ding et al. 2009; Jin et al. 2003). However, this linear programming method is very slow for computing estimators of $\boldsymbol{\theta}$. Thus this approach is very inefficient when implementing to solve (4.9) for estimation of $\boldsymbol{\Sigma}$. Recently, an approach called a derivative free-spectral algorithm for nonlinear equations (DF-SANE) was proposed (La Cruz et al. 2006), and there is a publication that showed that this algorithm is better than the linear programming method using an example of estimating parameters of AFT models under independent censoring (Varadhan and Gilbert, 2009). However, under dependent censoring, the artificial censoring term leads to numerical instability in estimating parameters and calculating resampled estimators. Moreover, this algorithm does not converge well under default tolerance settings using DF-SANE (Varadhan and Gilbert, 2009). Thus using this algorithm requires changing the tolerance level. Developing efficient numerical algorithms for estimating parameters is an important topic for future research.

Chapter 5
Goodness of Fit

5.1 Introduction

Goodness of fit is fundamental for assessing the appropriateness of a model. Methodology for model checking for parametric regression has been well developed (Lin et al. 2002; Klein and Moeschberger, 2003, Chapter 12, pp. 409-423). Assessing adequacy in parametric models is based on studying residuals, which capture the difference between observed and predicted part from a model (Lin et al. 2002). Residuals are an important element in model checking. They enable statisticians to perform graphical and numerical summaries for assessing model fit.

The model considered in this paper is the accelerated failure time (AFT) model, which is given by

$$T = \mathbf{Z}^T \boldsymbol{\eta}_0 + \epsilon.$$

where T is time to event of interest, $\mathbf{Z}$ is $p \times 1$ vector of covariates, $\boldsymbol{\eta}_0$ is $p \times 1$ vector of regression coefficients and ϵ is an error term. Note that for simplicity of notations, all times are log-transformed. Moreover, the distribution of ϵ is unspecified, so to estimate $\boldsymbol{\eta}_0$, nonparametric methods are used.

U-statistics, initially proposed by Hoeffding (1948), occupy an important role in the theory of statistics. For parameter vector $\boldsymbol{\theta}$ and sample $X_1, \dots X_n$, a U-statistic of order K is defined as

$$\mathbf{U}_n(\boldsymbol{\theta}) = \binom{n}{K}^{-1} \sum_{1 \le i_1 \dots i_K \le n} h(X_{i_1}, \dots X_{i_K}),$$

where $h(\cdot)$ is called the kernel. $h(\cdot)$ is usually symmetric on $(X_{i_1}, \dots X_{i_K})$. U-statistics are a critical element in semiparametric models. Denote $Y = T \wedge C$ and $\Delta = I(T \leq C)$. The observed data are n i.i.d copies of $(Y, \Delta, \mathbf{Z})$, $\quad i = 1, \dots, n$. One estimating equation for $\boldsymbol{\eta}_0$ is (Tsiatis, 1990) given by

$$\mathbf{U}_n^{indep}(\boldsymbol{\eta}) = \frac{1}{n} \sum_{i=1}^{n} \Delta_i \left[\mathbf{Z}_i - \frac{\sum_{j=1}^{n} I\{e_j(\boldsymbol{\eta}) \geq e_i(\boldsymbol{\eta})\} \mathbf{Z}_j}{\sum_{j=1}^{n} I\{e_j(\boldsymbol{\eta}) \geq e_i(\boldsymbol{\eta})\}} \right] = 0, \tag{5.1}$$

where $e_i(\boldsymbol{\eta}) = Y_i - \mathbf{Z}_i^T \boldsymbol{\eta}$. Another rank estimator, proposed by Fygenson and Ritov (1994), for the AFT model is given by the solution for the following estimating equation:

$$\mathbf{U}_n^{FR}(\boldsymbol{\eta}) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} (\mathbf{Z}_i - \mathbf{Z}_j) \Delta_i I\{e_j(\boldsymbol{\eta}) > e_i(\boldsymbol{\eta})\} = 0,$$

which can be expressed as

$$\mathbf{U}_n^{FR}(\boldsymbol{\eta}) = \frac{1}{2n(n-1)} \sum_{i \neq j} (\mathbf{Z}_i - \mathbf{Z}_j) [\Delta_i I\{e_j(\boldsymbol{\eta}) > e_i(\boldsymbol{\eta})\} - \Delta_j I\{e_i(\boldsymbol{\eta}) > e_j(\boldsymbol{\eta})\}] = 0. \tag{5.2}$$

Note that (5.1) is a U-statistic of order 1, i.e., U-statistic with $K = 1$ and (5.2) is U-statistic order of 2, i.e., U-statistic with $K = 2$. Let $\mathbf{V}_i = (\epsilon_i, C_i, \mathbf{Z}_i^T)^T$, $\quad i = 1, \dots, n$. For (5.1), $\mathbf{h}(\mathbf{V}_i, \boldsymbol{\eta}) = \mathbf{Z}_i - \dfrac{\sum_{j=1}^{n} I\{e_j(\boldsymbol{\eta}) \geq e_i(\boldsymbol{\eta})\} \mathbf{Z}_j}{\sum_{j=1}^{n} I\{e_j(\boldsymbol{\eta}) \geq e_i(\boldsymbol{\eta})\}}$ and for (5.2), $\mathbf{h}(\mathbf{V}_i, \mathbf{V}_j, \boldsymbol{\eta}) = \frac{1}{2}(\mathbf{Z}_i - \mathbf{Z}_j)[\Delta_i I\{e_j(\boldsymbol{\eta}) > e_i(\boldsymbol{\eta})\} - \Delta_j I\{e_i(\boldsymbol{\eta}) > e_j(\boldsymbol{\eta})\}]$.

Model checking techniques for censored data have been studied in many settings. Therneau et al. (1990) developed a graphical approach of checking the Cox model by using martingale residuals. Lin et al. (1993) proposed model checking based on cumulative sums of martingale residuals for the Cox proportional hazard model. Lin et al. (1996) proposed model checking procedures for the accelerated failure time (AFT) model in overall fit. Recently, León and Cai (2012) proposed checking form of covariates using 'robust residuals' based on model from León et al. (2009). They argued that when a random variable of interest and other covariates have high correlation, in the uncensored case, the approach of Lin et al. (2002) clearly fails to detect misspecification because of the high correlation.

However, the above-mentioned methodology for goodness of fit is based on

U-statistics of order 1. Many rank-based estimators arise from U-statistics of order 2. Clearly, $\mathbf{U}_n^{FR}(\boldsymbol{\eta})$ is U-statistic of order 2. In this case, performing model checking based on U-statistic of order 1 may lead to bias. In this paper, we propose methodology for goodness of fit for U-statistic order 2 principles using the AFT model. Theoretical justification is based on U-process theory from Nolan and Pollard (1987) and Nolan and Pollard (1988). In Section 5.2, method of goodness of fit for U-statistic of order 2. Proofs of theoretical results are provided in Section 5.2.2. Section 5.3 outlines the results of some simulation studies, while an application to data from an HIV clinical trial is given in Section 5.4. Some discussion concludes Section 5.5.

5.2 Checking overall fit of model

5.2.1 Independent censoring

As mentioned in the introduction, the AFT model is

$$T = \mathbf{Z}^T \boldsymbol{\eta}_0 + \epsilon. \tag{5.3}$$

In this subsection, it is assumed that failure times are independently censored. As can be seen in (5.2), estimating equation proposed by Fygenson and Ritov (1994) is a U-statistic of order 2. Let $\mathbf{V}_i = (\epsilon_i, C_i, \mathbf{Z}_i^T)^T, i = 1, \ldots, n$ and $\boldsymbol{\eta}$ be parameter of interest and $\boldsymbol{\eta}_0$ be true value. General U-statistics of order 2 with standardization to estimate $\boldsymbol{\eta}_0$ in (5.3) have the form

$$\mathbf{U}_n(\boldsymbol{\eta}) = \frac{n^{1/2}}{n(n-1)} \sum_{i \neq j} \mathbf{h}(\mathbf{V}_i, \mathbf{V}_j, \boldsymbol{\eta}),$$

where $\mathbf{h}(\cdot, \cdot, \boldsymbol{\eta})$ is a kernel function such that $E\{n^{-1/2}\mathbf{U}_n(\boldsymbol{\eta}_0)\} = 0$. Note that the kernel of the estimating equation in (5.2) is $\mathbf{h}(\mathbf{V}_i, \mathbf{V}_j, \boldsymbol{\eta}) = \dfrac{1}{2}(\mathbf{Z}_i - \mathbf{Z}_j)[\Delta_i I\{e_j(\boldsymbol{\eta}) > e_i(\boldsymbol{\eta})\} - \Delta_j I\{e_i(\boldsymbol{\eta}) > e_j(\boldsymbol{\eta})\}]$.

Under mild conditions, the estimator $\hat{\boldsymbol{\eta}}$, the solution of $\mathbf{U}_n(\boldsymbol{\eta}) = 0$, is strongly consistent and asymptotically normal (Jin et al. 2001; Honoré and Powell, 1994).

Using the assumptions from Honoré and Powell (1994),

$$\mathbf{U}_n(\boldsymbol{\eta}) = \mathbf{U}_n(\boldsymbol{\eta}_0) + n^{1/2}\mathbf{\Psi}_0(\boldsymbol{\eta} - \boldsymbol{\eta}_0) + o_p(1 + n^{1/2}||\boldsymbol{\eta} - \boldsymbol{\eta}_0||), \tag{5.4}$$

where $\mathbf{\Psi}_0$ is the derivative of $E\{n^{-1/2}\mathbf{U}_n(\boldsymbol{\eta})\}$ evaluated at $\boldsymbol{\eta} = \boldsymbol{\eta}_0$. To assess the overall fit of the model, let

$$\mathbf{U}_n(t;\boldsymbol{\eta}) = \frac{n^{1/2}}{n(n-1)}\sum_{i\neq j}\mathbf{h}(\mathbf{V}_i, \mathbf{V}_j, \boldsymbol{\eta})I\{g(\mathbf{V}_i, \mathbf{V}_j, \boldsymbol{\eta}) \leq t\},$$

where g is a function that belongs to the Euclidian class (Nolan and Pollard, 1988). One natural choice of g is maximum function. For example, in the AFT model, $g(\mathbf{V}_i, \mathbf{V}_j, \boldsymbol{\eta}) = e_i(\boldsymbol{\eta}) \vee e_j(\boldsymbol{\eta})$, where $a \vee b$ denotes maximum of a and b. Then (5.2) leads to the following expansion (Lin et al. 1996) :

$$\mathbf{U}_n(t;\boldsymbol{\eta}) = \mathbf{U}_n(t;\boldsymbol{\eta}_0) + n^{1/2}\mathbf{\Psi}_0(t)(\boldsymbol{\eta} - \boldsymbol{\eta}_0) + o_p(1 + n^{1/2}||\boldsymbol{\eta} - \boldsymbol{\eta}_0||), \tag{5.5}$$

where $\mathbf{\Psi}_0(t)$ is the slope matrix of $\mathbf{U}_n(t;\boldsymbol{\eta}_0)$ at time t. Note that when $t = \infty$, (5.5) is equal to (5.4). Since the solution of the estimating equation $\mathbf{U}_n(\boldsymbol{\eta}) = 0$ is strongly consistent, it is established that

$$\mathbf{U}_n(t;\hat{\boldsymbol{\eta}}) = \mathbf{U}_n(t;\boldsymbol{\eta}_0) + n^{1/2}\mathbf{\Psi}_0(t)(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0) + o_p(1).$$

If the model is correct, then $\mathbf{U}_n(t;\hat{\boldsymbol{\eta}})$ fluctuates around 0. $\mathbf{U}_n(t;\hat{\boldsymbol{\eta}})$ contains information on the model behavior, analogous to the martingale residuals in Lin et al. (1996) and Lin et al. (1993).

In this case, the key issue is to show that the process $\mathbf{U}_n(t;\hat{\boldsymbol{\eta}})$ converges to a Gaussian process. In this case, it is not possible to use the empirical process results from Lin et al. (1993) and Lin et al. (1996), because a sum of independent and identically distributed random variables in the estimating function does not exist (Nolan and Pollard, 1987). However, by using the U-process theory of Nolan and Pollard (1987) and Nolan and Pollard (1988), the following result can be obtained.

Theorem 5.1. Assuming that the model (5.3) is true, $\mathbf{U}_n(t;\hat{\boldsymbol{\eta}})$ converges to a zero-mean Gaussian process.

The next issue is to find the null distribution of $\mathbf{U}_n(t;\boldsymbol{\eta})$. Since the structure of process $\mathbf{h}(\mathbf{V}_i, \mathbf{V}_j, \boldsymbol{\eta})I\{g(\mathbf{V}_i, \mathbf{V}_j, \boldsymbol{\eta}) \leq t\}$ is unknown, it is very difficult to tackle the process directly. One way to solve this problem is to approximate the process by a known distribution (Lin et al. 1993). Since $\mathbf{U}_n(t; \boldsymbol{\eta})$ is nonsmooth, approximation through Taylor expansion does not work. To find an expression for the approximate distribution of $n^{1/2}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0)$, a resampling approach (Parzen et al. 1994) is used. Resampling has been used in a variety of covariance matrix estimation settings for rank regression estimators (e.g. Parzen et al. 1994; Lin et al. 1996; Peng and Fine 2006; Jin et al. 2001). In this approach,

$$\mathbf{U}_n(\boldsymbol{\eta}) = -\mathbf{u}_r, \tag{5.6}$$

with $\mathbf{u}_r$ simulated from a normal distribution whose mean is 0 and covariance matrix is $\hat{\boldsymbol{\Sigma}}$, where $\hat{\boldsymbol{\Sigma}}$ is estimated covariance matrix of $\mathbf{U}_n(\boldsymbol{\eta})$. Let the solution of (5.6) be $\boldsymbol{\eta}^*$. Under mild conditions, given observed data, $n^{1/2}(\boldsymbol{\eta}^* - \hat{\boldsymbol{\eta}})$ has the same asymptotic distribution as $n^{1/2}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0)$ (Parzen et al. 1994). Define $Q_1, \ldots, Q_n$ to be standard normal random variables.

Theorem 5.2. Assuming that the model (5.3) is true,

$$\hat{\mathbf{U}}_n(t; \boldsymbol{\eta}^*) = \frac{n^{1/2}}{n(n-1)} \sum_{i \neq j} \mathbf{h}(\mathbf{V}_i, \mathbf{V}_j, \hat{\boldsymbol{\eta}}) I\{g(\mathbf{V}_i, \mathbf{V}_j; \hat{\boldsymbol{\eta}}) \leq t\} Q_i + \mathbf{U}_n(t; \boldsymbol{\eta}^*) - \mathbf{U}_n(t; \hat{\boldsymbol{\eta}})$$

converges weakly to the same Gaussian Process limit as $\mathbf{U}_n(t; \hat{\boldsymbol{\eta}})$.

These processes, which are called bootstrapped processes, are fundamental for checking the overall fit of the model. It is possible to adopt the approach of Lin et al. (1996) for graphical and numerical summaries. For a graphical summary, 20 or 30 observations from $\hat{\mathbf{U}}_n(\cdot)$ are randomly chosen and plotted with the observed process. Lack of fit can be checked by examining the behavior of the observed process and observations from the resampling processes graphically. In addition to the graphical approach, we can perform a formal test as in the case of U-statistics of order one. Similar to assessing proportional hazards (Wei, 1984; Lin et al. 1993), the test statistic for evaluating overall fit is

$$D = \sup_t ||\mathbf{U}_n(t; \hat{\boldsymbol{\eta}})||.$$

Larger values of D indicate stronger evidence for lack of fit. Let $\boldsymbol{\eta}^{i*}$ be ith value from resampling and suppose that there are M resampling values. It is possible to compute a p-value by (Hsieh et al. 2011)

$$p = \frac{1}{M} \sum_{i=1}^{M} I\{\sup_t ||\hat{\mathbf{U}}_n(t; \boldsymbol{\eta}^{i*})|| \geq D\}.$$

According to the approach above, for the estimating equation (5.2), the test statistic is $\sup_t ||\mathbf{U}_n^{FR}(t; \hat{\boldsymbol{\eta}})||$, where

$$\mathbf{U}_n^{FR}(t; \boldsymbol{\eta}) = \frac{n^{1/2}}{2n(n-1)} \sum_{i \neq j} (\mathbf{Z}_i - \mathbf{Z}_j)[\Delta_i I\{e_j(\boldsymbol{\eta}) > e_i(\boldsymbol{\eta})\} - \Delta_j I\{e_i(\boldsymbol{\eta}) > e_j(\boldsymbol{\eta})\}]$$
$$\times I(e_i(\boldsymbol{\eta}) \vee e_j(\boldsymbol{\eta}) \leq t).$$

Now it is necessary to find the null distribution of $\mathbf{U}_n^{FR}(t; \boldsymbol{\eta})$. By arguments in Ferguson and Ritov (1994), $n^{1/2}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0)$ has an asymptotically normal distribution with mean 0 and covariance matrix $\boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^{-1}$, where $\boldsymbol{\Gamma}_0$ is nonsingular and $\boldsymbol{\Omega}_0$ is an asymptotic covariance matrix of $\mathbf{U}_n^{FR}(\boldsymbol{\eta}_0)$. They proposed to use numerical derivatives for estimating $\boldsymbol{\Gamma}_0$, but these numerical derivatives involved unknown hazard functions of the event of the interest and can be numerically unstable.

We instead use the approach from Parzen et al. (1994). The empirical influence function for the asymptotic distribution of $\mathbf{U}_n^{FR}(\boldsymbol{\eta}_0)$ is given by

$$\hat{\mathbf{v}}_i = \frac{1}{n-1} \sum_{j=1}^{n} (\mathbf{Z}_i - \mathbf{Z}_j)[\Delta_i I\{e_j(\boldsymbol{\eta}) > e_i(\boldsymbol{\eta})\} - \Delta_j I\{e_i(\boldsymbol{\eta}) > e_j(\boldsymbol{\eta})\}].$$

Then we can construct

$$\mathbf{U}_n^{FR}(\boldsymbol{\eta}) = -n^{-1/2} \sum_{i=1}^{n} \hat{\mathbf{v}}_i Q_i. \tag{5.7}$$

Let the solution of the equation (5.7) be $\boldsymbol{\eta}^*$. By Parzen et al. (1994), the unconditional distribution of $n^{1/2}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0)$ has the same limiting distribution as the conditional distribution of $n^{1/2}(\boldsymbol{\eta}^* - \hat{\boldsymbol{\eta}})$. Then the bootstrapped processes are given by

$$\hat{\mathbf{U}}_n^{FR}(t;\boldsymbol{\eta}^*) = \frac{n^{1/2}}{2n(n-1)}\sum_{i\neq j}(\mathbf{Z}_i - \mathbf{Z}_j)[\Delta_i I\{e_j(\boldsymbol{\eta}) > e_i(\boldsymbol{\eta})\} - \Delta_j I\{e_i(\boldsymbol{\eta}) > e_j(\boldsymbol{\eta})\}]$$
$$\times I(e_i(\boldsymbol{\eta}) \vee e_j(\boldsymbol{\eta}) \leq t)Q_i + \mathbf{U}_n^{FR}(t;\boldsymbol{\eta}^*) - \mathbf{U}_n^{FR}(t;\hat{\boldsymbol{\eta}}).$$

These bootstrapped processes are random processes whose asymptotic distribution is identical to $\mathbf{U}_n^{FR}(t;\hat{\boldsymbol{\eta}})$. As described before, the graphical test can be performed by plotting 20 or 30 realized values of $\hat{\mathbf{U}}_n^{FR}(\cdot;\cdot)$ with the observed process $\mathbf{U}_n^{FR}(t;\hat{\boldsymbol{\eta}})$. A p-value can be computed by replications of $\boldsymbol{\eta}^*$.

Now it is important to show that the proposed test procedure is consistent. A consistent test is one whose power approaches 1 when sample size goes to infinity. Since the power is closely related to rejecting the misspecified model, the estimator under a misspecified model should converge to some constant value (Struthers and Kalbfleisch, 1986; Lin and Wei, 1989). Before proving consistency of the proposed test, it is necessary to prove the consistency of estimator under a misspecified model. Let T be the time to failure and C be independent censoring. Let $Y = T \wedge C$, $\Delta = I(T \leq C)$ and covariates be $\mathbf{W} = (\mathbf{Z}^T, \mathbf{Z}^{*T})^T$. The observed data are n i.i.d replicates of $(Y, \Delta, \mathbf{W})$. As before, all times are log-transformed. Assume that the true model is

$$T = \mathbf{W}^T\boldsymbol{\eta}_0 + \epsilon.$$

where ϵ is an i.i.d error term. Suppose that model is fitted using $\mathbf{Z}$ only, i.e., there is misspecification on model fitting. We need next theorem before proving the consistency of the test.

Theorem 5.3. Let $\hat{\boldsymbol{\eta}}^{mis}$ be the estimator from the misspecified model. Then $\hat{\boldsymbol{\eta}}^{mis}$ is a consistent estimator of $\boldsymbol{\eta}^{mis}$, which is a solution of

$$\lambda^*(\boldsymbol{\eta}) = \frac{1}{2}E\Bigg[(\mathbf{Z}_1 - \mathbf{Z}_2)\int_0^\infty \bar{G}(t + \mathbf{W}_1^T\boldsymbol{\eta}_0 - \mathbf{Z}_1^T\boldsymbol{\eta}|\mathbf{Z}_1)\bar{G}(t + \mathbf{W}_2^T\boldsymbol{\eta}_0 - \mathbf{Z}_2^T\boldsymbol{\eta}|\mathbf{Z}_2)$$
$$\times \{\bar{F}(t + \mathbf{W}_2^T\boldsymbol{\eta}_0 - \mathbf{Z}_2^T\boldsymbol{\eta}|\mathbf{Z}_2)f(t + \mathbf{W}_1^T\boldsymbol{\eta}_0 - \mathbf{Z}_1^T\boldsymbol{\eta}|\mathbf{Z}_1)$$
$$- \bar{F}(t + \boldsymbol{\beta}_0^T\mathbf{W}_1 - \mathbf{Z}_1^T\boldsymbol{\eta}|\mathbf{Z}_1)f(t + \mathbf{W}_2^T\boldsymbol{\eta}_0 - \mathbf{Z}_2^T\boldsymbol{\eta}|\mathbf{Z}_2)\}dt\Bigg].$$

where f is an error density, $\bar{F}$ is survival function of error and $\bar{G}$ is survival function of $C - \boldsymbol{W}^T \boldsymbol{\eta}_0$.

Now the following theorem shows that the proposed test is consistent.

Theorem 5.4. The test $D = \sup_t ||\mathbf{U}_n(t; \hat{\boldsymbol{\eta}})||$ is consistent against the alternative hypothesis that $\boldsymbol{\eta}$ depends on time.

5.2.2 Proofs

Before proving main results, the following assumptions are made.

1. The parameter space $\boldsymbol{\Theta}$ is compact and the true parameter $\boldsymbol{\eta}_0$ is the interior point of $\boldsymbol{\Theta}$.

2. Let $|| \cdot ||$ be Euclidean norm. The functions $\mathbf{h}(\cdot, \cdot, \boldsymbol{\eta})$ and

$$\mathbf{u}(\cdot, \cdot, \boldsymbol{\eta}, w) = \sup_{||\boldsymbol{\tau} - \boldsymbol{\eta}|| \leq w} ||\mathbf{h}(\cdot, \cdot, \boldsymbol{\tau}) - \mathbf{h}(\cdot, \cdot, \boldsymbol{\eta})||$$

   are measurable functions of $\mathbf{V}_{i_1}$ and $\mathbf{V}_{i_2}$ for $1 \leq i_1 \neq i_2 \leq n$ in some open neighborhood of $\boldsymbol{\Theta}$.

3. Let $\lambda(\boldsymbol{\eta}) = E\{n^{-1/2} \mathbf{U}_n^{FR}(\boldsymbol{\eta})\}$. Then $\lambda(\boldsymbol{\eta}_0) = 0$ and $\lambda(\boldsymbol{\eta})$ is differentiable at $\boldsymbol{\eta}_0$ with nonsingular derivative at $\boldsymbol{\eta}_0$.

4. For $1 \leq i_1 \neq i_2 \leq n$, there exist positive constant a_0, b_0 and c_0 such that $E\{\mathbf{u}(\mathbf{V}_{i_1}, \mathbf{V}_{i_2}, \boldsymbol{\eta}, r)\} \leq a_0 r$ and $E\{\mathbf{u}(\mathbf{V}_{i_1}, \mathbf{V}_{i_2}, \boldsymbol{\eta}, r)^2\} \leq b_0 r$ for all $r \leq c_0$ and all $\boldsymbol{\eta}$ in an open neighborhood of $\boldsymbol{\eta}_0$.

5. There exists $K > 0$ such that $||\mathbf{Z}|| \leq K$, i.e., $\mathbf{Z}$ is uniformly bounded by constant K.

6. The error distribution has finite Fisher information and the distribution of $\mathbf{Z}$ given $\Delta = 1$ is not concentrated on a proper hyperplane on $\mathbb{R}^p$.

7. The information bound (Bickel et al. 1993, Chapter 2, p23) for estimating $\boldsymbol{\eta}_0$ is finite and invertible.

In the proofs, results for the estimating function in Fygenson and Ritov (1994) are proved only. We first prove tightness of $\mathbf{U}_n^{FR}(t; \hat{\boldsymbol{\eta}})$.

Lemma 5.1. $\mathbf{U}_n^{FR}(t;\hat{\boldsymbol{\eta}})$ is tight.

Proof. Let N_0 be an open neighborhood of $\boldsymbol{\eta}_0$. By Lemma 2 of Honoré and Powell (1994),

$$\sup_{\boldsymbol{\eta}\in N_0} \frac{||\mathbf{U}_n^{FR}(\boldsymbol{\eta}) - \mathbf{U}_n^{FR}(\boldsymbol{\eta}_0) - n^{1/2}\lambda(\boldsymbol{\eta})||}{1+n^{1/2}||\lambda(\boldsymbol{\eta})||} = o_p(1).$$

Then by Taylor expansion and consistency of $\hat{\boldsymbol{\eta}}$,

$$\mathbf{U}_n^{FR}(t;\hat{\boldsymbol{\eta}}) = \mathbf{U}_n^{FR}(t;\boldsymbol{\eta}_0) + n^{1/2}\boldsymbol{\Gamma}_0(t)(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0) + o_p(1),$$

where $\boldsymbol{\Gamma}_0(t)$ is slope matrix of $\mathbf{U}_n^{FR}(t;\hat{\boldsymbol{\eta}})$. We will start by showing $\mathbf{U}_n^{FR}(t;\hat{\boldsymbol{\eta}})$ is tight. Clearly, $n^{1/2}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0)$ converges in distribution, so it is tight. Note that $g(\mathbf{V}_i, \mathbf{V}_j, \boldsymbol{\eta}_0) = e_i(\boldsymbol{\eta}_0) \vee e_j(\boldsymbol{\eta}_0)$ for the AFT model assuming independent censoring. For each t, a class of functions $g_t\{e_i(\boldsymbol{\eta}_0), e_j(\boldsymbol{\eta}_0)\} = e_i(\boldsymbol{\eta}_0) \vee e_j(\boldsymbol{\eta}_0) - t$ is a polynomial class by Lemma 18 of Nolan and Pollard (1987) (Note that for each t, $e_i(\boldsymbol{\eta}_0) \vee e_j(\boldsymbol{\eta}_0) - t$ is an element of a finite dimensional vector space of real functions). Then by argument of Nolan and Pollard (1987), a class of functions $g_t\{e_i(\boldsymbol{\eta}_0), e_j(\boldsymbol{\eta}_0)\}$ is Euclidean with envelope 1. Let $\mathbf{h}(\mathbf{V}_i, \mathbf{V}_j, \boldsymbol{\eta}_0, t) = \dfrac{1}{2}(\mathbf{Z}_i - \mathbf{Z}_j)[\Delta_i I\{e_j(\boldsymbol{\eta}_0) > e_i(\boldsymbol{\eta}_0)\} - \Delta_j I\{e_i(\boldsymbol{\eta}_0) > e_j(\boldsymbol{\eta}_0)\}] I\{e_i(\boldsymbol{\eta}_0) \vee e_j(\boldsymbol{\eta}_0) \le t\}$. By assumption, $||\mathbf{Z}_i|| \le K$ for all i. Since $\mathbf{Z}_i$s and $[\Delta_i I\{e_j(\boldsymbol{\eta}_0) > e_i(\boldsymbol{\eta}_0)\} - \Delta_j I\{e_i(\boldsymbol{\eta}_0) > e_j(\boldsymbol{\eta}_0)\}]$ are bounded, by Lemma 22 in Nolan and Pollard (1987), $\mathbf{h}(\mathbf{V}_i, \mathbf{V}_j, \boldsymbol{\eta}_0, t)$ is also Euclidean with some positive constant envelope M.

Let $\mathcal{G}$ be Euclidean class for envelope G and the metric $d_{Q,p,G}$ which is defined on $\mathcal{G}$ is

$$d_{Q,p,G}(f,g) = \left[\frac{T_n|f-g|^p}{Q(M^p)}\right]^{1/p} \qquad f,g \in \mathcal{G}$$

where Q is a measure on space $\mathcal{X} \otimes \mathcal{X}$ which satisfies $0 < Q(G^p) < \infty$. Define

$$J_n(s, Q, \mathcal{G}, G) = \int_0^s \log N_2(x, Q, \mathcal{G}, G)dx,$$

where $N_2(x, Q, \mathcal{G}, G)$ is the covering number $N(x, d_{Q,2,\mathcal{G}})$ (Nolan and Pollard, 1987). Then

$$\sup_n E\{J_n(s, Q, \mathcal{G}, G)\} = \sup_n E\bigg\{\int_0^s \log N_2(x, Q, \mathcal{G}, G)dx\bigg\}.$$

Let $y_1, \ldots y_{2n}$ be a sample from measure P. Define T_n to be the measure which assigns mass one at each of the $4n(n-1)$ pairs of y_v, y_w in function g_{ij} for $u \in \mathcal{G}$,

where

$$g_{ij} = u(y_{2i}, y_{2j}) - u(y_{2i}, y_{2j-1}) - u(y_{2i-1}, y_{2j}) + u(y_{2i-1}, y_{2j-1}),$$

Now we will apply argument in previous paragraph to complete the proof. Let $\mathcal{F}$ be function space for $\mathbf{h}(\mathbf{V}_i, \mathbf{V}_j, \boldsymbol{\eta}_0, t)$ and F be the envelope of $\mathcal{F}$. Then $\mathcal{F}$ is a class of function in $\mathcal{L}^2(C \times [0, a])$, where a is a positive constant. By previous argument, the envelope F is constant M. The metric for $\mathcal{F}$ is

$$d_{T_n,2,M}(f^*, g^*) = \left[\frac{T_n|f^* - g^*|^2}{T_n(M^2)}\right]^{1/2} \qquad f^*, g^* \in \mathcal{F}$$

Let $P\mathcal{F}$ be the class of functions of $E\{h(\mathbf{v}, \mathbf{V}, \boldsymbol{\eta}_0, t)\}$. Clearly, $PF = M$. Moreover, $E\{\mathbf{h}(\mathbf{v}, \mathbf{V}, \boldsymbol{\eta}_0, t)\}$ is also bounded for all $\mathbf{v}$. By Corollary 21 in Nolan and Pollard (1987), $P\mathcal{F}$ is also Euclidean with envelope M. By Nolan and Pollard (1987), $N_2(\epsilon, T_n, \mathcal{F}, M)$ is the smallest cardinality for a subclass $\mathcal{F}^*$ of $\mathcal{F}$ such that

$$\min_{\mathcal{F}^*} T_n|f - f^*|^2 \leq \epsilon^2 T_n(M^2),$$

for each function in $\mathcal{F}$. Clearly, $T_n(M^2) = M^2 n(n-1)$. Hence,

$$\min_{\mathcal{F}^*} T_n|f - f^*|^2 \leq \epsilon^2 M^2 n(n-1).$$

Note that $0 < T_n(F) = T_n(M^2) = n(n-1)M^2 < \infty$. By argument in Nolan and Pollard (1987) about Euclidean class, there exists positive constant A_1 and B_1,

$$\int_0^1 \log N_2(x, T_n, \mathcal{F}, M)dx \leq \int_0^1 (\log A_1 + B_1 \log 4 - 2B_1 \log y)dy$$
$$= \log A_1 + B_1 \log 4 - 2B_1(y \log y - y)|_0^1 = \log A_1 + B_1 \log 4 + 2B_1 < \infty$$

Clearly, $\sup_n E\{J_n(1, T_n, \mathcal{F}, M)\}^2 < \infty$. Let P_n be an empirical measure on sample $\mathbf{V}_1, \ldots, \mathbf{V}_n$. Thus $0 < P_n(M^2) < \infty$. Since $P\mathcal{F}$ is also Euclidean, by using similar arguments as the previous paragraph,

$$\sup_n E\{J(1, P_n, P\mathcal{F}, M)\}^2 < \infty.$$

Similarly,

$$J(1, P \otimes P, \mathcal{F}, M) < \infty.$$

Thus it is enough to show that for every $\zeta > 0$ and $\delta > 0$, we can find $\nu > 0$ such that

$$\limsup_{n\to\infty} E\{J(\nu, P_n, P\mathcal{F}, M) > \zeta\} < \delta. \tag{5.8}$$

Since $P\mathcal{F}$ is also Euclidean, clearly $0 < P_n M < \infty$,

$$J(\gamma, P_n, P\mathcal{F}, M) = \int_0^\nu \log N_2(x, P_n, P\mathcal{F}, M)dx \leq 1.$$

For $\zeta > 0$, by taking ν to be the solution of

$$\int_0^\nu \log N_2(x, P_n, P\mathcal{F}, M)dx = \zeta.$$

Thus (5.8) holds. Hence by Theorem 5 of Nolan and Pollard (1988), $\mathbf{U}_n^{FR}(t; \boldsymbol{\eta}_0)$ is tight. Hence $\mathbf{U}_n^{FR}(t; \hat{\boldsymbol{\eta}})$ is also tight.

□

Now we will prove Theorem 5.1.

Proof of Theorem 5.1. Let $\mathbf{h}(\mathbf{V}_i, \mathbf{V}_j, \boldsymbol{\eta}_0) = \frac{1}{2}(\mathbf{Z}_i - \mathbf{Z}_j)[\Delta_i I\{e_j(\boldsymbol{\eta}_0) > e_i(\boldsymbol{\eta}_0)\} - \Delta_j I\{e_i(\boldsymbol{\eta}_0) > e_j(\boldsymbol{\eta}_0)\}]$. Define

$$2\mathbf{h}_1(\mathbf{v}, \boldsymbol{\eta}_0, t) = 2E\{\mathbf{h}(\mathbf{v}, \mathbf{V}_2, \boldsymbol{\eta}_0, t)\}$$

where $\mathbf{h}(\mathbf{v}, \mathbf{V}_2, \boldsymbol{\eta}_0, t) = \mathbf{h}(\mathbf{v}, \mathbf{V}_2, \boldsymbol{\eta}_0)I\{g(\mathbf{v}, \mathbf{V}_2, \boldsymbol{\eta}_0) \leq t\}$ and $2\mathbf{h}_1(\mathbf{v}, \boldsymbol{\eta}_0) = 2E\{\mathbf{h}(\mathbf{v}, \mathbf{V}_2, \boldsymbol{\eta}_0)\}$. By arguments in the Appendix of Lin et al. (1996) and the Appendix of Peng and Fine (2006),

$$\begin{aligned} \mathbf{U}_n^{FR}(t; \hat{\boldsymbol{\eta}}) &= \mathbf{U}_n^{FR}(t; \boldsymbol{\eta}_0) - \boldsymbol{\Gamma}(t)\boldsymbol{\Gamma}_0^{-1}\mathbf{U}_n^{FR}(\boldsymbol{\eta}_0) + o_p(1) \\ &= n^{-1/2}\sum_{i=1}^n \mathbf{H}_i(t) - \boldsymbol{\Gamma}(t)\boldsymbol{\Gamma}_0^{-1} n^{-1/2}\sum_{i=1}^n \mathbf{H}_i + o_p(1), \end{aligned} \tag{5.9}$$

where

$$\mathbf{H}_i(t) = 2\mathbf{h}_1(\mathbf{V}_i, \boldsymbol{\eta}_0, t) \quad \mathbf{H}_i = 2\mathbf{h}_1(\mathbf{V}_i, \boldsymbol{\eta}_0).$$

Let $\mathbf{U}_n^{FR*}(t; \boldsymbol{\eta}_0) = n^{-1/2}\sum_{i=1}^n\{\mathbf{H}_i(t) - \boldsymbol{\Gamma}(t)\boldsymbol{\Gamma}_0^{-1}\mathbf{H}_i\}$. By the tightness of $\mathbf{U}_n^{FR}(t; \hat{\boldsymbol{\eta}})$,

$\mathbf{U}_n^{FR*}(t;\boldsymbol{\eta}_0)$ converges to a Gaussian process with mean zero and covariance matrix

$$E[\{\mathbf{H}_1(t)-\boldsymbol{\Gamma}(t)\boldsymbol{\Gamma}_0^{-1}\mathbf{H}_1\}\{\mathbf{H}_1(t)-\boldsymbol{\Gamma}(t)\boldsymbol{\Gamma}_0^{-1}\mathbf{H}_1\}^T]. \tag{5.10}$$

Thus Theorem 5.1 is proved. □

Proof of Theorem 5.2. We only prove the Fygenson and Ritov (1994) case. Note that

$$\begin{aligned}
\hat{\mathbf{U}}_n^{FR}(t;\boldsymbol{\eta}^*) &= \frac{n^{1/2}}{n(n-1)}\sum_{i\neq j}\mathbf{h}(\mathbf{V}_i,\mathbf{V}_j,\hat{\boldsymbol{\eta}})I\{g(\mathbf{V}_i,\mathbf{V}_j,\hat{\boldsymbol{\eta}})\le t\}Q_i \\
&- \boldsymbol{\Gamma}(t)n^{1/2}(\boldsymbol{\eta}^*-\hat{\boldsymbol{\eta}})+o_p(1) \\
&= \frac{n^{1/2}}{n(n-1)}\sum_{i\neq j}\mathbf{h}(\mathbf{V}_i,\mathbf{V}_j,\hat{\boldsymbol{\eta}})I\{g(\mathbf{V}_i,\mathbf{V}_j,\hat{\boldsymbol{\eta}})\le t\}Q_i \\
&- \boldsymbol{\Gamma}(t)\boldsymbol{\Gamma}_0^{-1}\frac{n^{1/2}}{n(n-1)}\sum_{i\neq j}\mathbf{h}(\mathbf{V}_i,\mathbf{V}_j,\hat{\boldsymbol{\eta}})Q_i+o_p(1).
\end{aligned}$$

It is clear that given the observed data, $\hat{\mathbf{U}}_n^{FR}(t;\boldsymbol{\eta}^*)$ is also a Gaussian process. It is only necessary to show that the limiting covariance matrix of $\hat{\mathbf{U}}_n^{FR}(t;\boldsymbol{\eta}^*)$ is same as that of $\mathbf{U}_n^{FR*}(t;\boldsymbol{\eta}_0)$. By the U-statistic strong law of large numbers (Serfling, 1980, Chapter 5, p190),

$$\begin{aligned}
&\frac{n^{1/2}}{n(n-1)}\sum_{i\neq j}\mathbf{h}(\mathbf{V}_i,\mathbf{V}_j,\hat{\boldsymbol{\eta}}) = n^{-1/2}\sum_{1\le i<j\le n}\frac{2}{n-1}h(\mathbf{V}_i,\mathbf{V}_j,\hat{\boldsymbol{\eta}}) \overset{a.s.}{\to} 2E\{\mathbf{h}_1(\mathbf{V}_1,\boldsymbol{\eta}_0)\} \\
&\frac{n^{1/2}}{n(n-1)}\sum_{i\neq j}\mathbf{h}(\mathbf{V}_i,\mathbf{V}_j,\hat{\boldsymbol{\eta}})I\{g(\mathbf{V}_i,\mathbf{V}_j,\hat{\boldsymbol{\eta}})\le t\} \\
&= n^{-1/2}\sum_{1\le i<j\le n}\frac{2}{n-1}\mathbf{h}(\mathbf{V}_i,\mathbf{V}_j,\hat{\boldsymbol{\eta}})I\{g(\mathbf{V}_i,\mathbf{V}_j,\hat{\boldsymbol{\eta}})\le t\} \overset{a.s.}{\to} 2E\{\mathbf{h}_1(\mathbf{V}_1,\boldsymbol{\eta}_0,t)\},
\end{aligned}$$

where $\overset{a.s.}{\to}$ denotes almost sure convergence. Hence the asymptotic covariance function of $\hat{\mathbf{U}}_n^{FR}(t;\boldsymbol{\eta}^*)$ is $E(\mathbf{L}\mathbf{L}^T)$, where

$$\mathbf{L} = 2\mathbf{h}_1(\mathbf{V}_1,\boldsymbol{\eta}_0,t)-\boldsymbol{\Gamma}(t)\boldsymbol{\Gamma}_0^{-1}2\mathbf{h}_1(\mathbf{V}_1,\boldsymbol{\eta}_0).$$

The limiting covariance matrix of $\hat{\mathbf{U}}_n^{FR}(t;\boldsymbol{\eta}^*)$ conditional on the observed data is the same as that of $\mathbf{U}_n^{FR*}(t;\boldsymbol{\eta}_0)$. This concludes the proof.

$\square$

Proof of Theorem 5.3. Let $e_i^*(\boldsymbol{\eta}) = Y_i - \mathbf{Z}_i^T\boldsymbol{\eta}$. Then the estimating equation is

$$\mathbf{U}_n^{FRmis}(\boldsymbol{\eta}) = \frac{n^{1/2}}{2n(n-1)}\sum_{i\neq j}(\mathbf{Z}_i - \mathbf{Z}_j)[\Delta_i I\{e_j^*(\boldsymbol{\eta}) > e_i^*(\boldsymbol{\eta})\} - \Delta_j I\{e_i^*(\boldsymbol{\eta}) > e_j^*(\boldsymbol{\eta})\}] = 0. \tag{5.11}$$

By Theorem 2.1(i) in Fygenson and Ritov (1994), the solution of equation (5.11) exists. Denote this solution by $\hat{\boldsymbol{\eta}}^{mis}$. By the strong law of large numbers,

$$n^{-1/2}\mathbf{U}_n^{FRmis}(\boldsymbol{\eta}) = \lambda^*(\boldsymbol{\eta}) + o(1).$$

Assume that $\lambda^*(\boldsymbol{\eta})$ has a unique solution $\boldsymbol{\eta}^{mis}$. Without loss of generality, it is assumed that $\boldsymbol{\eta}_0 = 0$. If $\boldsymbol{\eta} \to \boldsymbol{\eta}^{mis}$, by Fygenson and Ritov (1994),

$$\begin{aligned}\lambda^*(\boldsymbol{\eta}) = \frac{1}{2}E\Bigg[&(\mathbf{Z}_1 - \mathbf{Z}_2)(\mathbf{Z}_1 - \mathbf{Z}_2)^T \times \int_{-\infty}^{\infty} -\bar{G}(t - \mathbf{Z}_1^T\boldsymbol{\eta}^{mis}|\mathbf{Z}_1)\bar{G}(t - \mathbf{Z}_2^T\boldsymbol{\eta}^{mis}|\mathbf{Z}_2)\\ &\times f(t - \mathbf{Z}_1^T\boldsymbol{\eta}^{mis})f(t - \mathbf{Z}_2^T\boldsymbol{\eta}^{mis})dt + (\mathbf{Z}_1 - \mathbf{Z}_2)\int_{-\infty}^{\infty}\bar{G}(t - \mathbf{Z}_1^T\boldsymbol{\eta}^{mis}|\mathbf{Z}_1)\bar{G}(t - \mathbf{Z}_2^T\boldsymbol{\eta}^{mis}|\mathbf{Z}_2)\\ &\times\{\mathbf{Z}_2^T\bar{F}(t - \mathbf{Z}_1^T\boldsymbol{\eta}^{mis})f'(t - \mathbf{Z}_2^T\boldsymbol{\eta}^{mis}) - \mathbf{Z}_1^T\bar{F}(t - \mathbf{Z}_2^T\boldsymbol{\eta}^{mis})\times\\ &f'(t - \mathbf{Z}_1^T\boldsymbol{\eta}^{mis})\}\Bigg](\boldsymbol{\eta} - \boldsymbol{\eta}^{mis}) + o(\boldsymbol{\eta} - \boldsymbol{\eta}^{mis}).\end{aligned}$$

By Fygenson and Ritov (1994)'s argument, it is linear in a neighborhood of $\boldsymbol{\eta}^{mis}$. Moreover, since $\mathbf{U}_n^{FRmis}(\boldsymbol{\eta})$ and $\lambda^*(\boldsymbol{\eta})$ are monotone with respect to $\boldsymbol{\eta}$, $\hat{\boldsymbol{\eta}}^{mis}$ is a consistent estimator of $\boldsymbol{\eta}^{mis}$. $\square$

Now we will prove consistency of the test.

Proof of Theorem 5.4. Let T be time to the event of interest, C be time to the independent censoring and $\mathbf{Z}$ be a vector of covariates. As before, these times are log-transformed. Suppose that the observed data are n i.i.d replicates of $(Y, \Delta, \mathbf{Z})$, where $Y = T \wedge C$ and $\Delta = I(T \leq C)$ and the alternative hypothesis is that $\boldsymbol{\eta}$ in the AFT model depends on time, i.e.,

$$T = \mathbf{Z}^T\boldsymbol{\eta}(s) + \epsilon. \tag{5.12}$$

Let $\hat{\boldsymbol{\eta}}^{mt}$ be estimator of $\boldsymbol{\eta}$ assuming AFT model that has time independent parameters while in the true model parameters actually depend on time. Then by applying similar arguments for the misspecified AFT model, $\hat{\boldsymbol{\eta}}^{mt}$ converges almost surely to constant vector, say $\boldsymbol{\eta}^{mt}$. To show consistency of test, it suffices to show that $n^{-1/2}\mathbf{U}_n^{FR}(t;\hat{\boldsymbol{\eta}}^{mt})$ converges to nonzero limit (Lin et al. 1993; Arbogast and Lin, 2004) against the alternative hypothesis. Under the alternative hypothesis, by strong law of large number of U-statistics, $n^{-1/2}\mathbf{U}_n^{FR}(t;\hat{\boldsymbol{\eta}}^{mt})$ converges almost surely to

$$\begin{aligned}&\tfrac{1}{2}E[(\mathbf{Z}_1-\mathbf{Z}_2)\times\\ &E[I\{e_1(\boldsymbol{\eta}^{mt})\vee e_2(\boldsymbol{\eta}^{mt})\le t\}(\Delta_1 I\{e_2(\boldsymbol{\eta}^{mt})>e_1(\boldsymbol{\eta}^{mt})\}-\Delta_2 I\{e_1(\boldsymbol{\eta}^{mt})>e_2(\boldsymbol{\eta}^{mt})\})|\mathbf{Z}_1,\mathbf{Z}_2].\end{aligned} \tag{5.13}$$

Then given $e_1(\boldsymbol{\eta}^{mt})\vee e_2(\boldsymbol{\eta}^{mt})\le t$, the inner expectation of (5.12) is $P[\Delta_1 I\{e_2(\boldsymbol{\eta}^{mt})>e_1(\boldsymbol{\eta}^{mt})\}-\Delta_2 I\{e_1(\boldsymbol{\eta}^{mt})>e_2(\boldsymbol{\eta}^{mt})\}]$. Then,

$$\begin{aligned}&P[\Delta_1 I\{e_2(\boldsymbol{\eta}^{mt})>e_1(\boldsymbol{\eta}^{mt})\}-\Delta_2 I\{e_1(\boldsymbol{\eta}^{mt})>e_2(\boldsymbol{\eta}^{mt})\}]\\ &=P\{(T_1-\mathbf{Z}_1^T\boldsymbol{\eta}^{mt})\le(T_2-\mathbf{Z}_2^T\boldsymbol{\eta}^{mt})\wedge(C_1-\mathbf{Z}_1^T\boldsymbol{\eta}^{mt})\wedge(C_2-\mathbf{Z}_2^T\boldsymbol{\eta}^{mt})\}\\ &-P\{(T_2-\mathbf{Z}_2^T\boldsymbol{\eta}^{mt})\le(T_1-\mathbf{Z}_1^T\boldsymbol{\eta}^{mt})\wedge(C_1-\mathbf{Z}_1^T\boldsymbol{\eta}^{mt})\wedge(C_2-\mathbf{Z}_2^T\boldsymbol{\eta}^{mt})\}\\ &=P[\{\epsilon_1+\mathbf{Z}_1^T(\boldsymbol{\eta}(s)-\boldsymbol{\eta}^{mt})\}\le[\epsilon_2+\mathbf{Z}_2^T(\boldsymbol{\eta}(s)-\boldsymbol{\eta}^{mt})]\wedge(C_1-\mathbf{Z}_1^T\boldsymbol{\eta}^{mt})\wedge(C_2-\mathbf{Z}_2^T\boldsymbol{\eta}^{mt})]\\ &-P[\{\epsilon_2+\mathbf{Z}_2^T(\boldsymbol{\eta}(s)-\boldsymbol{\eta}^{mt})\}\le\{\epsilon_1+\mathbf{Z}_1^T(\boldsymbol{\eta}(s)-\boldsymbol{\eta}^{mt})\}\wedge(C_1-\mathbf{Z}_1^T\boldsymbol{\eta}^{mt})\wedge(C_2-\mathbf{Z}_2^T\boldsymbol{\eta}^{mt})].\end{aligned} \tag{5.14}$$

Since $\boldsymbol{\eta}(\cdot)$ depends on time and covariates, $\epsilon_1+\mathbf{Z}_1^T(\boldsymbol{\eta}(s)-\boldsymbol{\eta}^{mt})$ and $\epsilon_2+\mathbf{Z}_2^T(\boldsymbol{\eta}(s)-\boldsymbol{\eta}^{mt})$ do not have same distribution, thus the probability in expression (5.13) is not 0. For the expression in (5.13) to be 0, the distribution of $\epsilon_1+\mathbf{Z}_1^T(\boldsymbol{\eta}(s)-\boldsymbol{\eta}^{mt})$ and $\epsilon_2+\mathbf{Z}_2^T(\boldsymbol{\eta}(s)-\boldsymbol{\eta}^{mt})$ should be same. Thus for the function in (5.13) to be 0, $\boldsymbol{\eta}(s)=\boldsymbol{\eta}^{mt}$. □

5.2.3 Dependent Censoring

In the previous section, independent censoring was assumed. However, in semicompeting risks data structure, this independent censoring assumption is violated. We will briefly review the model and the procedure of Peng and Fine (2006) in Chapter 3. Let X be the time to event of interest, D be time to dependent censoring, C be time to independent censoring and $\mathbf{Z}$ be $p\times 1$ vector of covariates. As the previous section all times are logarithm scale. Define $\tilde{X}=X\wedge D\wedge C$, $\tilde{D}=D\wedge C$, $\xi=I(D\le C)$,

$\delta = I(X \leq D \wedge C)$. The observed data is $(\tilde{X}_i, \tilde{D}_i, \xi, \delta, \mathbf{Z}_i), i = 1, \ldots, n$. Now the model is a bivariate AFT model (Lin et al. 1996; Peng and Fine, 2006):

$$\begin{pmatrix} X_i = \mathbf{Z}_i^T \boldsymbol{\theta}_0 + \epsilon_i^X \\ D_i = \mathbf{Z}_i^T \boldsymbol{\eta}_0 + \epsilon_i^D \end{pmatrix} \quad i = 1 \ldots n,$$

where $\boldsymbol{\gamma}_0 = (\boldsymbol{\eta}_0^T, \boldsymbol{\theta}_0^T)^T$ is $2p \times 1$ vector of true value $\boldsymbol{\gamma} = (\boldsymbol{\eta}^T, \boldsymbol{\theta}^T)^T$ and $\epsilon_i = (\epsilon_i^X, \epsilon_i^D)^T$ are independent and identically distributed with unspecified survival function F. Let $\mathbf{S}_n(\boldsymbol{\eta}) = n^{1/2}\mathbf{U}_n^{indep}(\boldsymbol{\eta})$ in (5.1). Since D only depends on independent censoring, from approach by Tsiatis (1990), estimator for $\boldsymbol{\eta}_0$ is obtained by solving $\mathbf{S}_n(\boldsymbol{\eta}) = 0$. For the event of the interest, it is necessary to adjust for the effect of dependent censoring to remove bias. To adjust for it, Peng and Fine (2006) used an artificial censoring technique. Define

$$\begin{aligned} d_{ij}(\boldsymbol{\gamma}) &= \max\{0, \mathbf{Z}_i^T(\boldsymbol{\theta} - \boldsymbol{\eta}), \mathbf{Z}_j^T(\boldsymbol{\theta} - \boldsymbol{\eta})\}, \\ \tilde{X}_{i(j)}^*(\boldsymbol{\gamma}) &= (X_i - \mathbf{Z}_i^T\boldsymbol{\theta}) \wedge (D_i - \mathbf{Z}_i^T\boldsymbol{\eta} - d_{ij}(\boldsymbol{\gamma})) \wedge (C_i - \mathbf{Z}_i^T\boldsymbol{\eta} - d_{ij}(\boldsymbol{\gamma})), \\ \tilde{\delta}_{i(j)}^*(\boldsymbol{\gamma}) &= I\{(X_i - \mathbf{Z}_i^T\boldsymbol{\theta}) \leq (D_i - \mathbf{Z}_i^T\boldsymbol{\eta} - d_{ij}(\boldsymbol{\gamma})) \wedge (C_i - \mathbf{Z}_i^T\boldsymbol{\eta} - d_{ij}(\boldsymbol{\gamma}))\}, \\ \psi_{ij}(\boldsymbol{\gamma}) &= \tilde{\delta}_{i(j)}^*(\boldsymbol{\gamma}) I\{\tilde{X}_{i(j)}^*(\boldsymbol{\gamma}) \leq \tilde{X}_{j(i)}^*(\boldsymbol{\gamma})\} - \tilde{\delta}_{j(i)}^*(\boldsymbol{\gamma}) I\{\tilde{X}_{j(i)}^*(\boldsymbol{\gamma}) \leq \tilde{X}_{i(j)}^*(\boldsymbol{\gamma})\}. \end{aligned}$$

The estimating function proposed by Peng and Fine (2006) is

$$\mathbf{U}_n^P(\boldsymbol{\gamma}) = \frac{2n^{1/2}}{n(n-1)} \sum_{1 \leq i < j \leq n} (\mathbf{Z}_i - \mathbf{Z}_j)\psi_{ij}(\boldsymbol{\gamma}).$$

According to the discussion in previous chapters, Peng and Fine (2006) also used a martingale approach to check model fit although their estimating function is also U-statistic of order 2. However, the estimating function of Peng and Fine (2006) does not have a martingale structure. Moreover, the artificial censoring applied in the assessment of model fit is one by Lin et al. (1996), which differs from that in Peng and Fine (2006). Thus applying a model assessment method using the Lin et al. (1996) approach for $\mathbf{U}_n^P(\boldsymbol{\gamma})$ is problematic. By using a similar approach to Fygenson and Ritov (1994), the score process is

$$\mathbf{U}_n^P(t; \hat{\boldsymbol{\gamma}}) = \frac{n^{1/2}}{n(n-1)} \sum_{i \neq j} (\mathbf{Z}_i - \mathbf{Z}_j)\psi_{ij}(\hat{\boldsymbol{\gamma}}) I\{\tilde{X}_{i(j)}^*(\hat{\boldsymbol{\gamma}}) \vee \tilde{X}_{j(i)}^*(\hat{\boldsymbol{\gamma}}) \leq t\}.$$

To derive a p-value, as in the previous section, a resampling approach is used to derive the null distribution. Let $\mathbf{U}_n^{all}(\boldsymbol{\gamma}) = [\{\mathbf{S}_n(\boldsymbol{\eta})\}^T, \{\mathbf{U}_n^P(\boldsymbol{\gamma})\}^T]^T$. Let $\hat{\boldsymbol{\gamma}}$ be estimator of $\boldsymbol{\gamma}_0$. Then by Theorem 2 in Peng and Fine (2006), $n^{1/2}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)$ has an asymptotically normal distribution with mean 0 and covariance matrix $\boldsymbol{\Upsilon}_0^{-1}\boldsymbol{\Xi}_0\boldsymbol{\Upsilon}_0^{-1}$, where $\boldsymbol{\Upsilon}_0$ is nonsingular matrix and $\boldsymbol{\Xi}_0$ is covariance matrix of $\lim_{n\to\infty}\mathbf{U}_n^{all}(\boldsymbol{\gamma})$. By Peng and Fine (2006), the empirical influence function for the asymptotic distribution of $\mathbf{U}_n^{all}(\boldsymbol{\gamma}_0)$ is

$$\begin{aligned}
\mathbf{J}_i^{(1)} = \xi_i \left[\mathbf{Z}_i - \frac{\sum_{j=1}^n I\{\tilde{D}_j^*(\hat{\boldsymbol{\eta}}) \geq \tilde{D}_i^*(\hat{\boldsymbol{\eta}})\}\mathbf{Z}_j}{\sum_{j=1}^n I\{\tilde{D}_j^*(\hat{\boldsymbol{\eta}}) \geq \tilde{D}_i^*(\hat{\boldsymbol{\eta}})\}}\right] - \sum_{l=1}^n \frac{\xi_l I\{\tilde{D}_i^*(\hat{\boldsymbol{\eta}}) \geq \tilde{D}_l^*(\hat{\boldsymbol{\eta}})\}}{\sum_{j=1}^n I\{\tilde{D}_j^*(\hat{\boldsymbol{\eta}}) \geq \tilde{D}_l^*(\hat{\boldsymbol{\eta}})\}} \\
\times \left[\mathbf{Z}_i - \frac{\sum_{j=1}^n I\{\tilde{D}_j^*(\hat{\boldsymbol{\eta}}) \geq \tilde{D}_l^*(\hat{\boldsymbol{\eta}})\}\mathbf{Z}_j}{\sum_{j=1}^n I\{\tilde{D}_j^*(\hat{\boldsymbol{\eta}}) \geq \tilde{D}_l^*(\hat{\boldsymbol{\eta}})\}}\right], \\
\mathbf{J}_i^{(2)} = \frac{2}{n-1}\sum_{j=1}^n (\mathbf{Z}_i - \mathbf{Z}_j)\phi_{ij}(\hat{\boldsymbol{\gamma}}).
\end{aligned}$$

Let $\mathbf{J}_i = [\{\mathbf{J}_i^{(1)}\}^T, \{\mathbf{J}_i^{(2)}\}^T]^T$. To apply the resampling approach of Parzen et al. (1994), perturbed terms need to be generated. The perturbed term is generated by constructing linear combinations of $\mathbf{J}_i$s and Q_is. $\boldsymbol{\gamma}^*$ can be obtained by solving equations

$$\begin{pmatrix} \mathbf{S}_n(\boldsymbol{\eta}) = -n^{-1/2}\sum_{i=1}^n \mathbf{J}_i^{(1)} Q_i \\ \mathbf{U}_n^P(\boldsymbol{\gamma}) = -n^{-1/2}\sum_{i=1}^n \mathbf{J}_i^{(2)} Q_i \end{pmatrix}.$$

Then $n^{1/2}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)$ has the same asymptotic distribution as $n^{1/2}(\boldsymbol{\gamma}^* - \hat{\boldsymbol{\gamma}})$ (Parzen et al. 1994). By using a similar approach as in Section 5.2.1 on this chapter, we can show that joint process $[\{\mathbf{S}_n(t;\hat{\boldsymbol{\eta}})\}^T, \{\mathbf{U}_n^P(s;\hat{\boldsymbol{\gamma}})\}^T]^T$ is approximated by $[\{\hat{\mathbf{S}}_n(u;\boldsymbol{\eta}^*)\}^T, \{\hat{\mathbf{U}}_n^P(s;\boldsymbol{\gamma}^*)\}^T]^T$, where

$$\begin{aligned}
\hat{\mathbf{S}}_n(t;\boldsymbol{\eta}^*) &= n^{-1/2}\sum_{i=1}^n \int_{-\infty}^u \left[\mathbf{Z}_i - \frac{\sum_{j=1}^n I\{\tilde{D}_j^*(\hat{\boldsymbol{\eta}}) \geq v\}\mathbf{Z}_j}{\sum_{j=1}^n I\{\tilde{D}_j^*(\hat{\boldsymbol{\eta}}) \geq v\}}\right] d\hat{M}_i(v;\hat{\boldsymbol{\eta}})Q_i + \mathbf{S}_n(t;\boldsymbol{\eta}^*) - \mathbf{S}_n(t;\hat{\boldsymbol{\eta}}) \\
\hat{\mathbf{U}}_n^P(s;\boldsymbol{\beta}^*) &= \frac{n^{1/2}}{n(n-1)}\sum_{i\neq j}(\mathbf{Z}_i - \mathbf{Z}_j)\psi_{ij}(\hat{\boldsymbol{\gamma}})I\{\tilde{X}_{i(j)}^*(\hat{\boldsymbol{\gamma}}) \vee \tilde{X}_{j(i)}^*(\hat{\boldsymbol{\gamma}}) \leq s\}Q_i + \mathbf{U}_n^P(s;\boldsymbol{\gamma}^*) - \mathbf{U}_n^P(s;\hat{\boldsymbol{\gamma}})
\end{aligned}$$

Both $[\mathbf{S}_n(t;\hat{\boldsymbol{\eta}})^T, \{\mathbf{U}_n^P(s;\hat{\boldsymbol{\gamma}})\}^T]^T$ and $[\hat{\mathbf{S}}_n(u;\boldsymbol{\eta}^*)^T, \{\hat{\mathbf{U}}_n^P(s;\boldsymbol{\gamma}^*)\}^T]^T$ converge weakly to the same bivariate Gaussian process. The testing procedure based on this bivariate process is the same as for the case of independent censoring.

Remark. As can be seen in this section, unlike modeling in the independent censoring, joint modeling of failure of interest and dependent censoring is required when there exists dependence between failure of interest and censoring. This leads derivation of joint processes of failure of interest and dependent censoring for evaluation of the model fit. However, numerical summaries (test statistic and p-value) can be computed for failure of interest and dependent censoring, respectively.

5.3 Simulation Studies

We first considered simulation studies using the estimating function from Fygenson and Ritov (1994). The error term is distributed as $\epsilon \sim N(0, 1)$. For covariates, We first generated $(A_1, A_2)^T$ from a bivariate normal distribution with mean $(0, 0)^T$ and covariance matrix $\begin{pmatrix} 1 & 0.25 \\ 0.25 & 1 \end{pmatrix}$.

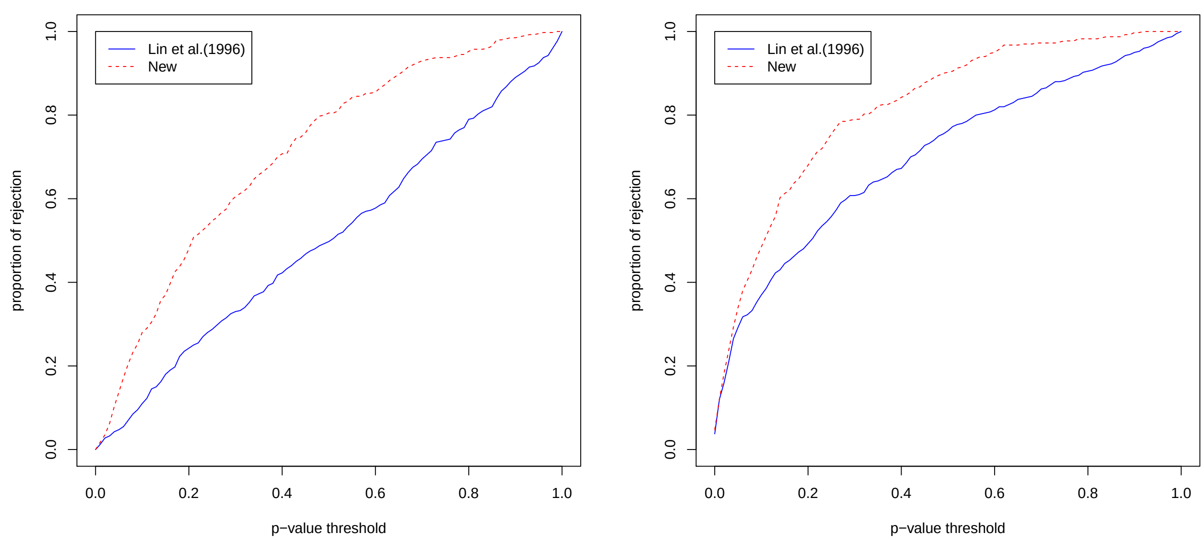
Next, we define $Z_1 = A_1$ and $Z_2 = \sum_j jA_2^2 I(b[j-1] < A_2 \le b[j]), j = 1, \ldots 21$, where $b[j]$ is $5(j-1)\%$ quantile of W_2. $b[1]$ is minimum of $b[\cdot]$ and $b[21]$ is maximum of $b[\cdot]$. Let $b[0] = -\infty$. Censoring variable is uniformly distributed with minimum value 0 and maximum value 150. True regression coefficient values are $\boldsymbol{\beta}_0 = (0.2, 1)^T$. We run 400 simulations with sample size $n = 50, 100$ and 200. In each simulation run, 200 resampling runs are performed. We fit the model by using only Z_1. For comparison, the new testing procedure is compared to that of Lin et al. (1996).

The proportion of rejections from the proposed method is higher than that from the Lin et al. (1996) method. Figures 5.1 and 5.2 show the power corresponding to threshold values of p-value. The plot shows that the proposed method performs better than the Lin et al. (1996)'s method. Table 5.1 shows power comparison between the new method and the Lin et al. (1996) method. Numerical results indicate that the proposed approach has higher power than that of Lin et al. (1996).

Table 5.1: Power comparison between the new method and that of Lin et al. (1996) for the independent censoring case

p-values / $n = 50$	Cutoff values			
	0.05	0.10	0.15	0.2
Lin et al. (1996)	0.0475	0.11	0.18	0.2425
Proposed method	0.1375	0.28	0.37	0.48
p-values / $n = 100$	Cutoff values			
	0.05	0.10	0.15	0.2
Lin et al. (1996)	0.2925	0.37	0.445	0.4925
Proposed method	0.34	0.485	0.6125	0.68
p-values / $n = 200$	Cutoff values			
	0.05	0.10	0.15	0.2
Lin et al. (1996)	0.5425	0.6475	0.71	0.7525
Proposed method	0.595	0.71	0.8025	0.845

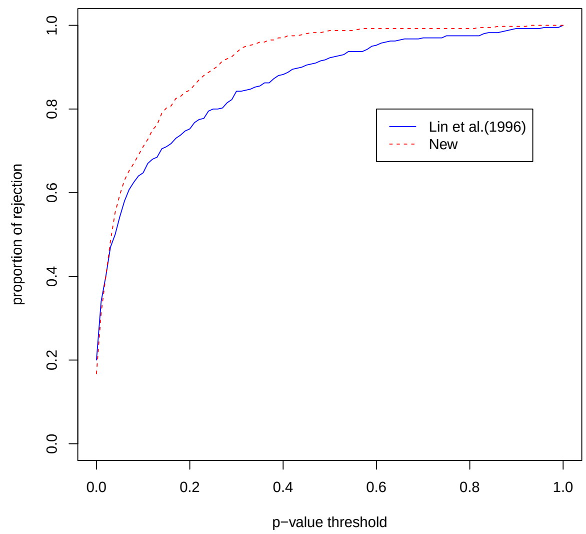
Figure 5.1: Plot of proportion of rejection according to threshold p-values when $n = 50$ (left) and $n = 100$ (right) for the independent censoring case



Next, we applied the proposed method to the dependent censoring case. Steps for data generation are shown below:

1. Generate $W = (W_1, W_2)^T \sim N\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.25 \\ 0.25 & 1 \end{pmatrix} \right\}.$

Figure 5.2: Plot of proportion of rejection according to threshold p-values when $n = 200$ for the independent censoring case



2. Set $R_1 = I(W_1 > 0)$ and $R_2 = \sum_j jW_2^2 I(b[j-1] < W_2 \leq b[j]), j = 1, \ldots 21$, where $b[j]$ is $5(j-1)\%$ quantile of W_2. $b[1]$ is minimum of $b[\cdot]$ and $b[21]$ is maximum of $b[\cdot]$. Let $b[0] = -\infty$.

3. Generate $\epsilon = (\epsilon^X, \epsilon^D) \sim N\left\{ \begin{pmatrix} 0 \\ 1.2 \end{pmatrix}, \begin{pmatrix} 1 & 0.25 \\ 0.25 & 1 \end{pmatrix} \right\}$.

4. Set $\boldsymbol{\theta}_0 = (1, 0.5)$ and $\boldsymbol{\eta}_0 = (0.5, 1)$ and generate $X = \mathbf{R}^T\boldsymbol{\theta}_0 + \epsilon^X$ and $D = \mathbf{R}^T\boldsymbol{\eta}_0 + \epsilon^D$, where $\mathbf{R} = (R_1, R_2)^T$.

Independent censoring time C is uniformly distributed with minimum value 0 and maximum value 100. We fit the misspecified model from Section 5.2.2 of this chapter, which only employs R_1, and compute the statistical power using the method of Lin et al. (1996) and the proposed method for the model of the event of interest X. In each simulation run, 200 resampling runs are tried. Table 5.2 shows the results when $n = 50$ based on 400 simulation runs and for $n = 100$ based on 200 simulation runs. Figure 5.3 shows a plot of proportion of rejection when $n = 50$ and $n = 100$. The plots in Figure 5.3 and numerical summaries from Table 5.2 lead to the same conclusion as the independent censoring case. The proposed method performs better than that of the Lin et al. (1996).

Figure 5.3: Plot of proportion of rejection according to threshold p-values when $n = 50$ (left) and $n = 100$ (right) for the model of the event of interest in the presence of dependent censoring

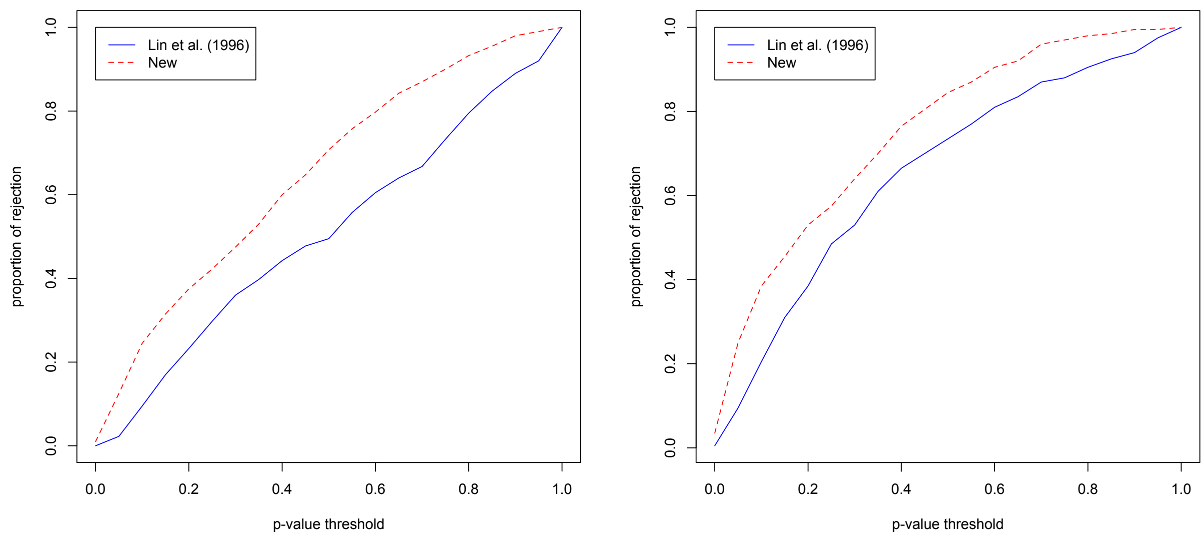


Table 5.2: Power comparison between the new method and Lin et al. (1996)'s method for the model of the event of interest in the presence of the dependent censoring

p-values / $n = 50$	Cutoff values			
	0.05	0.10	0.15	0.2
Lin et al. (1996)	0.0225	0.095	0.17	0.2325
Proposed method	0.125	0.245	0.315	0.375
p-values / $n = 100$	Cutoff values			
	0.05	0.10	0.15	0.2
Lin et al. (1996)	0.095	0.205	0.31	0.385
Proposed method	0.25	0.385	0.455	0.53

5.4 Real Data Analysis

We applied the proposed method to data from AIDS Clinical Trial Study 364 (Albrecht et al. 2001), which was previously analyzed by Peng and Fine (2006) and Chapter 4. As can be seen in Chapter 1 and Chapter 4, in this study, the plasma RNA level of every patient is at least 500 copies per ml. The event of interest is time to first viologic failure, which is defined as the first time to HIV level$\geq$ 2000. Patients will leave the study due to deterioration of heath status as time progresses (Peng and Fine, 2006). Hence dependence between failure of interest and censoring (withdrawal) exists.

In this dataset, 3 levels of treatment are considered : nelfinavir (NFV), efavirenz (EFV), and combination of nelfinavir and efavirenz (NFV + EFV). We consider three covariates. Z_1 takes value 1 if treatment assignment of a patient is EFV and 0 otherwise. Z_2 takes value 1 if treatment assignment of a patient is NFV + EFV and 0 otherwise and Z_3 is log(RNA) level. In Chapter 4, the dependent censoring and the event of interest were analyzed using the Lin et al. (1996) and Peng and Fine (2006) approaches jointly. The approach based on Lin et al. (1996) for both the Lin et al. (1996) estimator and the Peng and Fine (2006) estimator is used for the model checking.

Figure 5.4 shows a goodness of fit plot of 20 bootstrapped processes along with the observed process. The observed process is moving around zero and bootstrapped processes suggest that there is no substantial deviation of model fit.

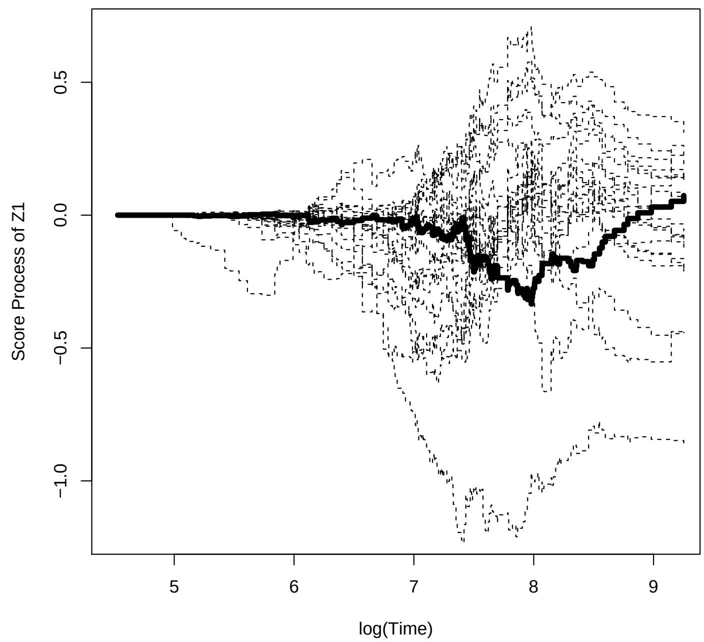
The p-value from the analysis in the Chapter 4 is 0.959. The p-value using the new approach is 0.51. Although both p-values show that there is no evidence of lack of fit for the model, substantial decrease is made on the proposed method, suggestive of higher power.

5.5 Discussion

In this chapter, we have developed a new goodness of fit approach. Using U-process theory by Nolan and Pollard (1987) and Nolan and Pollard (1988), we adapt the resampling approach from Parzen et al. (1994) and Lin et al. (1996) to derive numerical summaries and graphical tests. The new approach can be applied to estimating functions based on U-statistics of order two.

In this chapter, our attention has been on checking the overall fit of the model. Other goodness of fit techniques which can be considered are checking functional form of covariates and linearity of the model. Lin et al. (1993) proposed method for these scenarios based on the Cox model. However, direct application of Lin et al. (1993)'s approach to the semiparametric AFT model is impossible because the estimating function is nonsmooth. By mimicking the approach in this paper

Figure 5.4: Observed process (bold line) and bootstrapped processes (dashed lines) for the first virologic failure



and Lin et al. (1993), for the procedure of Fygenson and Ritov (1994), one may consider the observed process

$$\mathbf{U}_{2k}(x;\boldsymbol{\eta}) = \frac{n^{1/2}}{n(n-1)} \sum_{i \neq j} I(Z_{ki} \vee Z_{kj} \leq x)(\mathbf{Z}_i - \mathbf{Z}_j)[\Delta_i I\{e_j(\boldsymbol{\eta}) > e_i(\boldsymbol{\eta})\} - \Delta_j I\{e_i(\boldsymbol{\eta}) > e_j(\boldsymbol{\eta})\}]$$

to check the form of covariates. Developing details about checking the functional form of covariates and the linearity of the model will be communicated in separate reports.

It is also worthwhile to apply the idea of León and Cai (2012) on checking overall fit in the U-statistics of order 2 case under observational studies. For U-statistics of order 2, however, there is no concept of residuals. Thus developing a tool similar to ‘robust residuals’ can be important. This will be also communicated in separate reports.

Chapter 6
Covariate adjustment using propensity scores for dependent censoring problems

6.1 Introduction

As can be seen in Chapter 3, an artificial censoring is an important element in the marginal regression method by Lin et al. (1996) and Peng and Fine (2006). The advantage to the use of the artificial censoring is that one can estimate regression parameters based on modified estimating equations without assuming a parametric structure between the event of interest and the dependent censoring. However, excessive artificial censoring may censor many observations and thus may lead to unstable estimation when continuous covariates are included in the model. Estimators have very small variability due to excessive artificial censoring. It seems that small variability of the estimators from model using all covariates may be desirable, but it implies that the variabilities of the estimators are not properly estimated because of the excessive artificial censoring. As can be seen in the simulation studies in this chapter, excessive artificial censoring leads to incorrect coverage.

For some situations, estimation of the treatment effect might be of interest in observational studies. With semicompeting risks data, estimation of the treatment effect in observational studies is especially problematic because one must adjust for confounders. Even in randomized studies, when researchers are interested in subgroups, then it is highly likely that secondary factors that define subgroups are not randomized. Thus these secondary factors, which are covariates, can be confounders. Although it is not necessary to adjust confounders if we are interested in evaluating treatment effects on different subgroups, it may be still worthwhile to adjust the imbalances between treatment and control groups (VanderWeele and Knol, 2011). The similar logic can be applied to evaluate the treatment effect within the subgroups. The subgroup analysis of RTOG 9413 (Radiation Therapy Oncology Group) study is a good example. RTOG 9413 was a multicenter randomized phase III trial for clinically localized intermediate-risk and/or high-risk prostate cancer patients. One of the primary hypotheses was to compare combined androgen suppression (CAS) and whole pelvic radiotherapy (WPRT) followed by a boost to the prostate with CAS and prostate only radiotherapy (PORT). The protocol primary endpoint was progression-free survival, defined as time from randomization to the first occurrence of local progression, regional/nodal failure, distant failure, biochemical failure or death from any cause. While the initial reporting did not find that WPRT improves PFS, a subgroup analysis suggested WPRT may prolong PFS among intermediate risk patients (determined by the prostate specific antigen (PSA) and Gleason score (GS) at randomization)(Roach et al. 2003). Given the nature of subgroup analysis and the fact that progression was dependently censored by death, it is of great interest to obtain an unbiased treatment effect estimate for the time to the first occurrence of any disease failure (local, regional, distant, biochemical) within the semi-competing risks framework. In addition to initial PSA (ng/ml) and Gleason score, other potential prognostic variables include tumor size, T stage and age. The data being analyzed are based on the updated reporting (Roach et al. 2013).

Our approach to the problem involves propensity scores, proposed by Rosenbaum and Rubin (1983). While their use has been of interest in causal inference, they also satisfy a balancing property that corresponds to the distribution of covariates being equal for both treatment group and control group given the propensity score (Williamson et al. 2014). Thus, the propensity score can provide enough information to balance the covariates between treatment group and control group. This can be an important tool for reducing the artificial censoring needed for estimating a treatment effect.

In this chapter, we propose methodology for estimation of treatment effects adjusting for covariates using the propensity score. The chapter is organized as follows. In Section 6.2, we introduce the data structure and modeling assumption. Section 6.3 shows the methodology for estimation of treatment effect. Theoretical results and details about inference using the proposed method are demonstrated in Section 6.4 and Section 6.5. In Section 6.6, we apply the proposed methodology to the ACTG 364 study and the RTOG 9413 study. Simulation studies are shown in Section 6.7. Concluding remarks and discussion are in Section 6.8.

6.2 Preliminaries

6.2.1 Data and Model

Let X be time to the event of interest, D be time to the dependent censoring, and C be time to the independent censoring. Denote by $I(A)$ the indicator function for the event A, and let $a \wedge b$ be the minimum of a and b. Define $\boldsymbol{W} = (\boldsymbol{V}^T, Z^T)^T$ to be a vector of k variables, where $\boldsymbol{V}$ is a collection of confounder variables and Z is a binary treatment variable. Define

$$\tilde{X} = X \wedge D \wedge C, \quad \tilde{D} = D \wedge C \quad \delta = I(X \leq \tilde{D}), \quad \xi = I(D \leq C).$$

All these times are log-transformed. The data consist of n independent observations $(\tilde{X}_i, \tilde{D}_i, \boldsymbol{W}_i, \delta_i, \xi_i)$, $i = 1, \ldots, n$.The model is

$$\begin{pmatrix} X_i = \boldsymbol{\theta}_0^T \boldsymbol{W}_i + \epsilon_i^X \\ D_i = \boldsymbol{\eta}_0^T \boldsymbol{W}_i + \epsilon_i^D \end{pmatrix}, \quad i = 1, \ldots, n,$$

where $\boldsymbol{\beta}_0 = (\boldsymbol{\theta}_0^T, \boldsymbol{\eta}_0^T)^T$ is a $2k \times 1$ vector and $\epsilon \equiv (\epsilon^X, \epsilon^D)$ is error with unknown bivariate distribution F. Let $\boldsymbol{\theta}_0 = \{\theta_0^{tr}, (\boldsymbol{\theta}_0^{cfd})^T\}^T$ and $\boldsymbol{\eta}_0 = \{\eta_0^{tr}, (\boldsymbol{\eta}_0^{cfd})^T\}^T$, where θ_0^{tr} and η_0^{tr} are the subcomponents of $\boldsymbol{\theta}_0$ and $\boldsymbol{\eta}_0$ corresponding to Z. Similarly, $\boldsymbol{\theta}_0^{cfd}$ and $\boldsymbol{\eta}_0^{cfd}$ are the components of $\boldsymbol{\theta}_0$ and $\boldsymbol{\eta}_0$ corresponding to $\boldsymbol{V}$. We can rewrite the model as

$$\begin{pmatrix} X_i = \theta_0^{tr} Z_i + (\boldsymbol{\theta}_0^{cfd})^T \boldsymbol{V}_i + \epsilon_i^X \\ D_i = \eta_0^{tr} Z_i + (\boldsymbol{\eta}_0^{cfd})^T \boldsymbol{V}_i + \epsilon_i^D \end{pmatrix}, \quad i = 1, \ldots, n.$$

We assume that the model is identifiable only in the upper wedge where $X < D$ and C is independent with (X, D) given $\boldsymbol{W}$, but X and D can be dependent given

$\boldsymbol{W}$ (Fine et al. 2001; Peng and Fine, 2006).

6.3 Proposed Methodology

Using artificial censoring techniques, Lin et al. (1996) and Peng and Fine (2006) proposed two different estimating functions. Lin et al. (1996) used a single comparison of each residual time and Peng and Fine (2006) compared different pairs of residual times for the artificial censoring. In Lin et al. (1996) approach, the same degree of the artificial censoring is applied to every residual time so that the estimator from this approach may be inefficient when many covariates are included in the model. The method of Peng and Fine (2006) is clearly better than that in Lin et al. (1996) in the sense that their artificial censoring is smaller than that in Lin et al. (1996), but including continuous covariates with large variabilities may still cause the excessive artificial censoring in their method.

We now propose an estimation procedure to avoid excessive artificial censoring by using the propensity scores. In this case, our goal is to estimate the treatment effect. Although the effect of confounders is not estimated, the propensity score provides rich information for estimation of treatment effect, so obtaining unbiased treatment effect is still possible.

Assume that the model constructed by logistic regression with parameter $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_k)$ for binary treatment is the true model. Let $\boldsymbol{H}_i = (1, \boldsymbol{V}_i^T)^T, i = 1, \ldots, n$. Define the propensity score to be

$$e_i(\boldsymbol{\alpha}) = P(Z_i = 1|\boldsymbol{H}_i) = \frac{\exp(\boldsymbol{\alpha}^T \boldsymbol{H}_i)}{1 + \exp(\boldsymbol{\alpha}^T \boldsymbol{H}_i)}.$$

The weight is defined by

$$w_i(\boldsymbol{\alpha}) = \frac{Z_i}{e_i(\boldsymbol{\alpha})} + \frac{1 - Z_i}{1 - e_i(\boldsymbol{\alpha})}. \tag{6.1}$$

This weight takes value $1/e_i(\boldsymbol{\alpha})$ if $Z_i = 1$ and $1/(1 - e_i(\boldsymbol{\alpha}))$ otherwise. In the causal inference literature, the typical technique is to apply weight $w_i(\boldsymbol{\alpha})$ to estimate the average treatment effect (Lunceford and Davidian, 2004; Williamson et al. 2014; Zhu, 2013 ; Zhu et al., 2014). In our case, this weight is incorporated into estimating functions. Let η^{tr} and θ^{tr} be the treatment effect parameters with respect to D and

X, respectively. By using the weight, the proposed estimating function for η^{tr} is

$$S_n(\eta^{tr}, \boldsymbol{\alpha}) = n^{-1/2} \sum_{i=1}^{n} \xi_i w_i(\boldsymbol{\alpha}) \left[Z_i - \frac{\sum_{j=1}^{n} I\{\tilde{D}_j^*(\eta^{tr}) \geq \tilde{D}_i^*(\eta^{tr})\} w_j(\boldsymbol{\alpha}) Z_j}{\sum_{j=1}^{n} I\{\tilde{D}_j^*(\eta^{tr}) \geq \tilde{D}_i^*(\eta^{tr})\} w_j(\boldsymbol{\alpha})} \right], \quad (6.2)$$

where $\tilde{D}_i^*(\eta^{tr}) = \tilde{D}_i - \eta^{tr} Z_i$. Let $\beta^{tr} = (\eta^{tr}, \theta^{tr})^T$. The proposed estimating function for β^{tr} using Lin et al. (1996) is

$$U_n^L(\beta^{tr}, \boldsymbol{\alpha}) = n^{-1/2} \sum_{i=1}^{n} \tilde{\delta}_i^*(\beta^{tr}) w_i(\boldsymbol{\alpha}) \left[Z_i - \frac{\sum_{j=1}^{n} I\{\tilde{X}_j^*(\beta^{tr}) \geq \tilde{X}_i^*(\beta^{tr})\} w_j(\boldsymbol{\alpha}) Z_j}{\sum_{j=1}^{n} I\{\tilde{X}_j^*(\beta^{tr}) \geq \tilde{X}_i^*(\beta^{tr})\} w_j(\boldsymbol{\alpha})} \right], \quad (6.3)$$

where

$$\begin{aligned} d(\beta^{tr}) &= \max_{1 \leq i \leq n} \{0, (\theta^{tr} - \eta^{tr}) Z_i\} \\ \tilde{X}_i^*(\beta^{tr}) &= (X_i - \theta^{tr} Z_i) \wedge \{(D_i \wedge C_i) - \eta^{tr} Z_i - d(\beta^{tr})\} \\ \tilde{\delta}_i^*(\beta^{tr}) &= I[(X_i - \theta^{tr} Z_i) \leq \{(D_i \wedge C_i) - \eta^{tr} Z_i - d(\beta^{tr})]. \end{aligned}$$

Similarly, the proposed estimating function based on Peng and Fine (2006) is

$$U_n^P(\beta^{tr}, \boldsymbol{\alpha}) = \frac{2n^{1/2}}{n(n-1)} \sum_{1 \leq i < j \leq n} (Z_i - Z_j) w_i(\boldsymbol{\alpha}) w_j(\boldsymbol{\alpha}) \phi_{ij}(\beta^{tr}), \quad (6.4)$$

where

$$\begin{aligned} d_{ij}(\beta^{tr}) &= \max \{0, (\theta^{tr} - \eta^{tr}) Z_i, (\theta^{tr} - \eta^{tr}) Z_j\} \\ \tilde{X}_{i(j)}^*(\beta^{tr}) &= (X_i - \theta^{tr} Z_i) \wedge \{(D_i \wedge C_i) - \eta^{tr} Z_i - d_{ij}(\beta^{tr})\} \\ \tilde{\delta}_{i(j)}^*(\beta^{tr}) &= I[(X_i - \theta^{tr} Z_i) \leq \{(D_i \wedge C_i) - \eta^{tr} Z_i - d_{ij}(\beta^{tr})] \\ \phi_{ij}(\beta^{tr}) &= \tilde{\delta}_{i(j)}^*(\beta^{tr}) I\{\tilde{X}_{i(j)}^*(\beta^{tr}) \leq \tilde{X}_{j(i)}^*(\beta^{tr})\} - \tilde{\delta}_{j(i)}^*(\beta^{tr}) I\{\tilde{X}_{j(i)}^*(\beta^{tr}) \leq \tilde{X}_{i(j)}^*(\beta^{tr})\}. \end{aligned}$$

Let $\boldsymbol{G}_n(\boldsymbol{\alpha})$ be the score function for $\boldsymbol{\alpha}$, where

$$\boldsymbol{G}_n(\boldsymbol{\alpha}) = n^{-1/2} \sum_{i=1}^{n} \boldsymbol{H}_i \left\{ Z_i - \frac{\exp(\boldsymbol{\alpha}^T \boldsymbol{H}_i)}{1 + \exp(\boldsymbol{\alpha}^T \boldsymbol{H}_i)} \right\}$$

Let $\boldsymbol{\gamma} = (\boldsymbol{\alpha}^T, \eta^{tr}, \theta^{tr}, \theta^{tr})^T$. We solve

$$\boldsymbol{Q}_n(\boldsymbol{\gamma}) = [\boldsymbol{G}_n^T(\boldsymbol{\alpha}), S_n^T(\eta^{tr}, \boldsymbol{\alpha}), \{U_n^L(\beta^{tr}, \boldsymbol{\alpha})\}^T, \{U_n^P(\beta^{tr}, \boldsymbol{\alpha})\}^T]^T = 0, \quad (6.5)$$

to obtain estimators of the true value $\boldsymbol{\gamma}_0 = (\boldsymbol{\alpha}_0^T, \eta_0^{tr}, \theta_0^{tr}, \theta_0^{tr})^T$. Solutions for $\boldsymbol{\gamma}$ can be obtained by solving the estimating equations sequentially. First we estimate the propensity score for all samples, denoted as $\{e_i(\hat{\boldsymbol{\alpha}})\}_{i=1}^n$, where $\hat{\boldsymbol{\alpha}}$ is the root of $\boldsymbol{G}_n(\boldsymbol{\alpha}) = 0$. Next step is to solve $S_n(\eta^{tr}, \hat{\boldsymbol{\alpha}}) = 0$ by including the estimated weights. Denote the estimator of η_0^{tr} be as $\hat{\eta}^{catr}$. Incorporating $\hat{\eta}^{catr}$ and the estimated weights, the two estimators of θ_0^{tr} are obtained through solving $U_n^L(\theta^{tr}, \hat{\eta}^{catr}, \hat{\boldsymbol{\alpha}}) = 0$ and $U_n^P(\theta^{tr}, \hat{\eta}^{catr}, \hat{\boldsymbol{\alpha}}) = 0$. These solutions are denoted by $\hat{\theta}^{Lcatr}$ and $\hat{\theta}^{Pcatr}$ for the Lin et al. (1996) and Peng and Fine (2006) approaches, respectively.

This methodology works because the propensity score adjusts the distribution of confounders between treatment group and control group. Although only the treatment variable is utilized in construction of the estimating functions, the proposed weights contain information on confounders so that they correct bias of not using them. By using the balancing property of the propensity score, it is expected to obtain unbiased estimators of the treatment effect on the times to the event of interest and to the dependent censoring. The joint distribution of $(X_i - \theta_0^{tr} Z_i, D_i - \eta_0^{tr} Z_i), i = 1, \ldots, n$ have a common distribution not depend on Z_i given the propensity score since the distribution of confounder is same for both treatment group and control group given the same propensity score value. Moreover, since only the treatment variable is utilized in the estimation procedure, it is expected to have very small artificial censoring compared to original estimation procedures which employ all variables. This point is seen in the numerical study in Section 6.7.

Another advantage of the proposed method is ease of computation. Numerically, this involves solving a one-dimensional equation, which is much faster and easier to do than the multidimensional case.

6.4 Theoretical Results and Inference

It is of interest whether the estimated treatment has good theoretical justification. In this case, since the propensity score provides enough information to balance distribution of covariates between treatment group and control group, given the true propensity score, we have martingale structure for Lin et al. (1996) type estimating function. For Peng and Fine (2006) type estimating function, given the propensity score, we have the same setup as the proof of Appendix of Peng and Fine (2006).

In Appendix, we show that $E\{\boldsymbol{Q}_n(\boldsymbol{\gamma}_0)\} = 0$. In this proof, the assumption that propensity model is true enables us to construct martingale structure of estimating equations for η^{tr} and β^{tr} in Lin et al. (1996). For estimating equation based on Peng and Fine (2006), the similar technique is used in Appendix of Peng and Fine (2006).

Without the propensity score, using arguments similar to partial likelihood case, one may argue that the estimated treatment effect converges to constant value in probability (Boyd et al. 2012; Struthers and Kalbfleisch, 1986; Lin and Wei, 1989). However, in our case, due to the balancing property of the propensity score and strong consistency of $\hat{\boldsymbol{\alpha}}$, the proposed estimator converges to the *true value of parameters* almost surely.

Let $\hat{\boldsymbol{\gamma}} = (\hat{\boldsymbol{\alpha}}^T, \hat{\eta}^{catr}, \hat{\theta}^{Lcatr}, \hat{\theta}^{Pcatr})^T$ be a solution of $\boldsymbol{Q}_n(\boldsymbol{\gamma}) = 0$. Then it is important to investigate asymptotic properties of $\hat{\boldsymbol{\gamma}}$. Proofs of the following theorems are in the Appendix.

Theorem 6.1. Assuming the regularity conditions in Ying (1993), Peng and Fine (2006) and Theorem 17 of Ferguson (1996), $\hat{\boldsymbol{\gamma}}$ is strongly consistent.

Theorem 6.2. Under the regularity conditions by Ying (1993) and Peng and Fine (2006) and by Theorem 1, $n^{1/2}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)$ has an asymptotic normal distribution with mean 0 and covariance matrix $\boldsymbol{\Lambda}_0^{-1}\boldsymbol{\Omega}_0\boldsymbol{\Lambda}_0^{-1}$. where $\boldsymbol{\Lambda}_0$ is a nonsingular matrix, and $\boldsymbol{\Omega}_0$ is the limiting covariance matrix of $\boldsymbol{Q}_n(\boldsymbol{\gamma}_0)$.

Note that convergence in joint distribution of $(\hat{\boldsymbol{\alpha}}^T, \hat{\eta}^{catr}, \hat{\theta}^{Lcatr}, \hat{\theta}^{Pcatr})^T$ is proved. It is also possible to consider $\hat{\boldsymbol{\gamma}}^L = (\hat{\boldsymbol{\alpha}}^T, \hat{\eta}^{catr}, \hat{\theta}^{Lcatr})^T$ and $\hat{\boldsymbol{\gamma}}^P = (\hat{\boldsymbol{\alpha}}^T, \hat{\eta}^{catr}, \hat{\theta}^{Pcatr})^T$ separately, but this separation causes problems in statistical inference for $\hat{\boldsymbol{\alpha}}$ and $\hat{\eta}^{catr}$. This unified convergence result also implies that $\hat{\boldsymbol{\gamma}}^L$ and $\hat{\boldsymbol{\gamma}}^P$ are strongly consistent and asymptotically normal. The price for proving the joint convergence is additional assumptions. These assumptions are specified in the Appendix.

For inference, estimation of the asymptotic covariance matrix is crucial. In practice, it may be more convenient to use the data bootstrap to avoid the possibly over-complicated numerical issues.

In this data bootstrap approach, first step is to bootstrap the data, then solve estimating equations in (6.5) using the bootstrapped data. In this case, the weight based on propensity score should be also updated from the bootstrap. From large number of solutions by the bootstrap, the covariance matrix can be estimated.

Another way to estimate covariance matrix is to extend Parzen et al.'s (1994) approach. In this case, the first step is to estimate $\mathbf{\Omega}_0$. Let $\hat{\beta}^{Lcatr} = (\hat{\eta}^{catr}, \hat{\theta}^{Lcatr})$ and $\hat{\beta}^{Pcatr} = (\hat{\eta}^{catr}, \hat{\theta}^{Pcatr})$. Adapting the work of Lin et al. (1996) and Peng and Fine (2006), we now propose weighted empirical influence functions of $\boldsymbol{Q}_n(\boldsymbol{\gamma}_0)$.

$$
\begin{aligned}
\hat{\boldsymbol{v}}_{1i} &= \boldsymbol{H}_i \left\{ Z_i - \frac{\exp(\hat{\boldsymbol{\alpha}}^T \boldsymbol{H}_i)}{1 + \exp(\hat{\boldsymbol{\alpha}}^T \boldsymbol{H}_i)} \right\} \\
\hat{v}_{2i}^{(1)} &= w_i(\hat{\boldsymbol{\alpha}}) \Bigg(\xi_i \left[Z_i - \frac{\sum_{j=1}^n I\{\tilde{D}_j^*(\hat{\eta}^{catr}) \ge \tilde{D}_i^*(\hat{\eta}^{catr})\} w_j(\hat{\boldsymbol{\alpha}}) Z_j}{\sum_{j=1}^n I\{\tilde{D}_j^*(\hat{\eta}^{catr}) \ge \tilde{D}_i^*(\hat{\eta}^{catr})\} w_j(\hat{\boldsymbol{\alpha}})} \right] \\
&\quad - \sum_{l=1}^n \frac{w_l(\hat{\boldsymbol{\alpha}}) \xi_l I\{\tilde{D}_i^*(\hat{\eta}^{catr}) \ge \tilde{D}_l^*(\hat{\eta}^{catr})\}}{\sum_{j=1}^n w_j(\hat{\boldsymbol{\alpha}}) I\{\tilde{D}_j^*(\hat{\eta}^{catr}) \ge \tilde{D}_l^*(\hat{\eta}^{catr})\}} \\
&\quad \times \left[Z_i - \frac{\sum_{j=1}^n I\{\tilde{D}_j^*(\hat{\eta}^{catr}) \ge \tilde{D}_l^*(\hat{\eta}^{catr})\} w_j(\hat{\boldsymbol{\alpha}}) Z_j}{\sum_{j=1}^n I\{\tilde{D}_j^*(\hat{\eta}^{catr}) \ge \tilde{D}_l^*(\hat{\eta}^{catr})\} w_j(\hat{\boldsymbol{\alpha}})} \right] \Bigg) \\
\hat{v}_{2i}^{(2)} &= w_i(\hat{\boldsymbol{\alpha}}) \Bigg(\tilde{\delta}_i^*(\hat{\beta}^{Lcatr}) \left[Z_i - \frac{\sum_{j=1}^n I\{\tilde{X}_j^*(\hat{\beta}^{Lcatr}) \ge \tilde{X}_i^*(\hat{\beta}^{Lcatr})\} w_j(\hat{\boldsymbol{\alpha}}) Z_j}{\sum_{j=1}^n I\{\tilde{X}_j^*(\hat{\beta}^{Lcatr}) \ge \tilde{X}_i^*(\hat{\beta}^{Lcatr})\} w_j(\hat{\boldsymbol{\alpha}})} \right] \\
&\quad - \sum_{l=1}^n \frac{w_l(\hat{\boldsymbol{\alpha}}) \tilde{\delta}_l^*(\hat{\beta}^{Lcatr}) I\{\tilde{X}_i^*(\hat{\beta}^{Lcatr}) \ge \tilde{X}_l^*(\hat{\beta}^{Lcatr})\}}{\sum_{j=1}^n w_j(\hat{\boldsymbol{\alpha}}) I\{\tilde{X}_j^*(\hat{\beta}^{Lcatr}) \ge \tilde{X}_l^*(\hat{\beta}^{Lcatr})\}} \\
&\quad \times \left[Z_i - \frac{\sum_{j=1}^n I\{\tilde{X}_j^*(\hat{\beta}^{Lcatr}) \ge \tilde{X}_l^*(\hat{\beta}^{Lcatr})\} w_j(\hat{\boldsymbol{\alpha}}) Z_j}{\sum_{j=1}^n I\{\tilde{X}_j^*(\hat{\beta}^{Lcatr}) \ge \tilde{X}_l^*(\hat{\beta}^{Lcatr})\} w_j(\hat{\boldsymbol{\alpha}})} \right] \Bigg) \\
\hat{v}_{2i}^{(3)} &= \frac{2}{n-1} \sum_{j=1}^n w_i(\hat{\boldsymbol{\alpha}}) w_j(\hat{\boldsymbol{\alpha}}) (Z_i - Z_j) \phi_{ij}(\hat{\beta}^{Pcatr}).
\end{aligned}
$$

Let $\hat{\boldsymbol{v}}_{2i} = \{\hat{v}_{2i}^{(1)}, \hat{v}_{2i}^{(2)}, \hat{v}_{2i}^{(3)}\}^T$. and $\hat{\boldsymbol{v}}_i = (\hat{\boldsymbol{v}}_{1i}^T, \hat{\boldsymbol{v}}_{2i}^T)^T$. The estimator of $\mathbf{\Omega}_0$ is

$$
\hat{\mathbf{\Omega}} = \frac{1}{n} \sum_{i=1}^n \hat{\boldsymbol{v}}_i \hat{\boldsymbol{v}}_i^T. \tag{6.6}
$$

It is important to consider the variability arising from the propensity score modeling. In nonrandomized studies, the true propensity score adjusts for imbalance between confounders. However, the estimated propensity score is indeed a random variable, so the variability from propensity score modeling always exists. Since the empirical influence functions contain weights based on the estimated propensity score, the weights impact the variance of $(\hat{\eta}^{catr}, \hat{\theta}^{Lcatr}, \hat{\theta}^{Pcatr})^T$. Ignoring these weights results in inflation of variance of the estimators of interest $(\hat{\eta}^{catr}, \hat{\theta}^{Lcatr}, \hat{\theta}^{Pcatr})^T$ because if the propensity score is treated as known, it leads to a decrease in precision for $(\hat{\eta}^{catr}, \hat{\theta}^{Lcatr}, \hat{\theta}^{Pcatr})^T$.

Now let us consider estimation of covariance matrix of $\hat{\boldsymbol{\gamma}}$. In this case, the main issue is estimation of $\boldsymbol{\Lambda}_0$. As discussed in Lin et al. (1996) and Peng and Fine (2006), direct estimation of $\boldsymbol{\Lambda}_0$ involves estimation of unknown hazard function of error terms, which is numerically unstable. Although the estimating functions are continuous with respect to $\boldsymbol{\alpha}$, the derivatives of estimating functions with respect to $\boldsymbol{\alpha}$ have a very complicated form. The resampling approach from Parzen et al. (1994) is an appealing approach to estimate the covariance matrix. Their approach is to solve a stochastic equation in a large number of times and to use the solutions to estimate the covariance matrix. This approach does not require estimation of asymptotic slope matrix $\boldsymbol{\Lambda}_0$, so it is a suitable approach for nonsmooth estimating equations. We extended Parzen et al. (1994)'s approach to estimate the covariance of $\hat{\boldsymbol{\gamma}}$. As discussed before, it is important to include the effect of $\hat{\boldsymbol{\alpha}}$ in the variability of $(\hat{\eta}^{catr}, \hat{\theta}^{Lcatr}, \hat{\theta}^{Pcatr})^T$ in this approach. Let $\boldsymbol{U}_n(\boldsymbol{\gamma}) = [S_n^T(\eta^{tr}, \boldsymbol{\alpha}), \{U_n^L(\beta^{tr}, \boldsymbol{\alpha})\}^T, \{U_n^P(\beta^{tr}, \boldsymbol{\alpha})\}^T]^T$. Given the data, the following stochastic equations are solved.

$$\begin{pmatrix} \boldsymbol{G}_n(\boldsymbol{\alpha}) = -n^{1/2} \sum_{i=1}^{n} \hat{\boldsymbol{v}}_{1i} A_i \\ \boldsymbol{U}_n(\boldsymbol{\gamma}) = -n^{1/2} \sum_{i=1}^{n} \hat{\boldsymbol{v}}_{2i} A_i \end{pmatrix}, \tag{6.7}$$

where $A_i, i = 1, \ldots, n$ are standard normal random variables. Then the covariance matrix is estimated from solutions of equations in (6.7). The first step is to obtain $\boldsymbol{\alpha}^*$, the solution of the first equation in (6.7), then obtain solutions of the second equation in (6.7), say $\boldsymbol{\tau}^* = \{\eta^{catr*}, \theta^{Lcatr*}, \theta^{Pcatr*}\}^T$ by computing new weight $w_i(\boldsymbol{\alpha}^*), i = 1, \ldots, n$, where $\eta^{catr*}, \theta^{Lcatr*}, \theta^{Pcatr*}$ are individual solutions from the second equation in (6.7) corresponding to $\boldsymbol{U}_n(\boldsymbol{\gamma})$, respectively. Let $\boldsymbol{\gamma}^* = \{(\boldsymbol{\alpha}^*)^T, (\boldsymbol{\tau}^*)^T\}^T$. By repeating these two steps sufficiently large number of times, the covariance matrix can be estimated. Note that in this procedure, the only random part is A_i while the observed data is treated as fixed (Lin et al. 1996). Moreover, the variability of $\hat{\boldsymbol{\alpha}}$ is included to the estimators of interest through new weights.

The question is whether given data, the solution from the stochastic equation has the same asymptotic distribution as the one we want, which is $n^{1/2}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)$. The following theorem demonstrates theoretical justification of proposed procedure for covariance matrix estimation. The proof of the following theorem is also in Appendix.

Theorem 6.3. By the regularity conditions from Ying (1993) and Peng and Fine (2006), conditional on observed data, the asymptotic distribution of $n^{1/2}(\boldsymbol{\gamma}^* - \hat{\boldsymbol{\gamma}})$ is same as the unconditional distribution of $n^{1/2}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)$.

6.5 Goodness of fit

Checking the overall fit of a model is one crucial part in model diagnostics. In this section, we suggest a goodness of fit procedure for the estimator from the proposed procedure. Let λ_{10} and λ_{20} be the true baseline hazard function for the transformed time of dependent censoring and transformed time of the event of interest by the artificial censoring. As denoted in Appendix, martingales for the dependent censoring and the event of interest are defined by (Lin et al. 1996)

$$M_{1i}(t;\eta_0^{tr},\boldsymbol{\alpha}_0) = w_i(\boldsymbol{\alpha}_0)[\xi_i I\{\tilde{D}_i^*(\eta_0^{tr}) \le t\} - \textstyle\int_{-\infty}^t I\{\tilde{D}_i^*(\eta_0^{tr}) \ge u\}\lambda_{10}(u)du].$$
$$M_{2i}(t;\beta_0^{tr},\boldsymbol{\alpha}_0) = w_i(\boldsymbol{\alpha}_0)[\tilde{\delta}_i^*(\beta_0^{tr}) I\{\tilde{X}_i^*(\beta_0^{tr}) \le t\} - \textstyle\int_{-\infty}^t I\{\tilde{X}_i^*(\beta_0^{tr}) \ge u\}\lambda_{20}(u)du],$$

Let $N_{1i}(t;\eta^{tr}) = \xi_i I\{\tilde{D}_i^*(\eta^{tr}) \le t\}$ and $N_{2i}(t;\beta^{tr}) = \tilde{\delta}_i^*(\beta^{tr}) I\{\tilde{X}_i^*(\beta^{tr}) \le t\}$. Moreover, let $N_{1i}(t;\eta^{tr},\boldsymbol{\alpha}) = w_i(\boldsymbol{\alpha})\xi_i I\{\tilde{D}_i^*(\eta^{tr}) \le t\}$ and $N_{2i}(t;\beta^{tr},\boldsymbol{\alpha}) = w_i(\boldsymbol{\alpha})\tilde{\delta}_i^*(\beta^{tr}) I\{\tilde{X}_i^*(\beta^{tr}) \le t\}$. Then estimated martingales are defined by

$$\begin{aligned}\hat{M}_{1i}(t;\hat{\eta}^{catr},\hat{\boldsymbol{\alpha}}) &= w_i(\hat{\boldsymbol{\alpha}})[N_{1i}(t;\hat{\eta}^{catr}) - \int_{-\infty}^t I\{\tilde{D}_i^*(\hat{\eta}^{catr}) \ge u\} d\hat{\Lambda}_{10}(u;\hat{\eta}^{catr},\hat{\boldsymbol{\alpha}})du] \\ &= N_{1i}(t;\hat{\eta}^{catr},\hat{\boldsymbol{\alpha}}) - \int_{-\infty}^t w_i(\hat{\boldsymbol{\alpha}}) I\{\tilde{D}_i^*(\hat{\eta}^{catr}) \ge u\} d\hat{\Lambda}_{10}(u;\hat{\eta}^{catr},\hat{\boldsymbol{\alpha}})du\end{aligned}$$

$$\begin{aligned}\hat{M}_{2i}(t;\hat{\beta}^{catr},\hat{\boldsymbol{\alpha}}) &= w_i(\hat{\boldsymbol{\alpha}})[N_{2i}(t;\hat{\beta}^{catr}) - \int_{-\infty}^t I\{\tilde{X}_i^*(\hat{\beta}^{catr}) \ge u\} d\hat{\Lambda}_{20}(u;\hat{\beta}^{catr},\hat{\boldsymbol{\alpha}})du] \\ &= N_{2i}(t;\hat{\beta}^{catr},\hat{\boldsymbol{\alpha}}) - \int_{-\infty}^t w_i(\hat{\boldsymbol{\alpha}}) I\{\tilde{X}_i^*(\hat{\beta}^{catr}) \ge u\} d\hat{\Lambda}_{20}(u;\hat{\beta}^{catr},\hat{\boldsymbol{\alpha}})du\end{aligned}$$

where $\hat{\beta}^{catr}$ is either $\hat{\beta}^{Lcatr}$ or $\hat{\beta}^{Pcatr}$ and

$$\hat{\Lambda}_{10}(t;\eta^{tr},\boldsymbol{\alpha}) = \int_{-\infty}^{t} \frac{\sum_{l=1}^{n} w_l(\boldsymbol{\alpha}) dN_{1l}(t;\eta^{tr})}{\sum_{j=1}^{n} w_j(\boldsymbol{\alpha}) I\{\tilde{D}_j^*(\eta^{tr}) \geq u\}} = \int_{-\infty}^{t} \frac{\sum_{l=1}^{n} dN_{1l}(t;\eta^{tr},\boldsymbol{\alpha})}{\sum_{j=1}^{n} w_j(\boldsymbol{\alpha}) I\{\tilde{D}_j^*(\eta^{tr}) \geq u\}}$$
$$\hat{\Lambda}_{20}(t;\eta^{tr},\boldsymbol{\alpha}) = \int_{-\infty}^{t} \frac{\sum_{l=1}^{n} w_l(\boldsymbol{\alpha}) dN_{2l}(t;\beta^{tr})}{\sum_{j=1}^{n} w_j(\boldsymbol{\alpha}) I\{\tilde{X}_j^*(\beta^{tr}) \geq u\}} = \int_{-\infty}^{t} \frac{\sum_{l=1}^{n} dN_{2l}(t;\beta^{tr},\boldsymbol{\alpha})}{\sum_{j=1}^{n} w_j(\boldsymbol{\alpha}) I\{\tilde{X}_j^*(\beta^{tr}) \geq u\}},$$

Here, $\hat{\Lambda}_{10}(t;\hat{\eta}^{catr},\hat{\boldsymbol{\alpha}})$ and $\hat{\Lambda}_{20}(t;\hat{\beta}^{Lcatr},\hat{\boldsymbol{\alpha}})$ are weighted Nelson-Aalen estimators for the dependent censoring and the event of interest. Our interest is to evaluate the overall fit of the model across time. Define

$$\boldsymbol{\Psi}_i(\boldsymbol{\alpha}) = \boldsymbol{H}_i\left[Z_i - \frac{\exp(\boldsymbol{\alpha}^T \boldsymbol{H}_i)}{1+\exp(\boldsymbol{\alpha}^T \boldsymbol{H}_i)}\right].$$

We define

$$\boldsymbol{G}_n(\boldsymbol{\alpha}) = n^{-1/2}\textstyle\sum_{i=1}^{n} \boldsymbol{\Psi}_i(\boldsymbol{\alpha});$$
$$S_n(t;\eta^{tr},\boldsymbol{\alpha}) = n^{-1/2}\textstyle\sum_{i=1}^{n} Z_i \hat{M}_{1i}(t;\eta^{tr},\boldsymbol{\alpha}) \quad U_n(u;\beta^{tr},\boldsymbol{\alpha}) = n^{-1/2}\textstyle\sum_{i=1}^{n} Z_i \hat{M}_{2i}(u;\beta^{tr},\boldsymbol{\alpha}).$$

In this case, $S_n(t;\hat{\eta}^{catr},\hat{\boldsymbol{\alpha}})$ is the observed process for the dependent censoring and $U_n(u;\hat{\beta}^{Lcatr},\hat{\boldsymbol{\alpha}})$ and $U_n(v;\hat{\beta}^{Pcatr},\hat{\boldsymbol{\alpha}})$ are the observed processes for the event of interest (Lin et al. 1996). Note that $\boldsymbol{G}_n(\hat{\boldsymbol{\alpha}})$=0 is included in the observed processes. As discussed in Lin et al. (1996), the null distribution of $\boldsymbol{Q}_n(s,t,u;\hat{\boldsymbol{\gamma}}) = \{\boldsymbol{G}_n^T(\hat{\boldsymbol{\alpha}}), S_n(s;\hat{\eta}^{catr},\hat{\boldsymbol{\alpha}}), U_n(t;\hat{\beta}^{Lcatr},\hat{\boldsymbol{\alpha}}), U_n(u;\hat{\beta}^{Pcatr},\hat{\boldsymbol{\alpha}})\}^T$ can be approximated by

$$\hat{\boldsymbol{G}}_n = n^{-1/2}\textstyle\sum_{i=1}^{n} \boldsymbol{\Psi}_i(\hat{\boldsymbol{\alpha}})A_i + n^{-1/2}\sum_{i=1}^{n} \boldsymbol{\Psi}_i(\boldsymbol{\alpha}^*)$$
$$\hat{S}_n(s) = n^{-1/2}\sum_{i=1}^{n}\int_{-\infty}^{s}\left[Z_i - \frac{\sum_{j=1}^{n} I\{\tilde{D}_j^*(\hat{\eta}^{catr}) \geq w\} Z_j w_j(\hat{\boldsymbol{\alpha}})}{\sum_{j=1}^{n} I\{\tilde{D}_j^*(\hat{\eta}^{catr}) \geq w\} w_j(\hat{\boldsymbol{\alpha}})}\right] d\hat{M}_{1i}(w;\hat{\eta}^{catr},\hat{\boldsymbol{\alpha}})A_i$$
$$+ S_n(s;\eta^{catr*},\boldsymbol{\alpha}^*) - S_n(s;\hat{\eta}^{catr},\hat{\boldsymbol{\alpha}})$$
$$\hat{U}_n^L(t) = n^{-1/2}\sum_{i=1}^{n}\int_{-\infty}^{t}\left[Z_i - \frac{\sum_{j=1}^{n} I\{\tilde{X}_j^*(\hat{\beta}^{Latr}) \geq w\} Z_j w_j(\hat{\boldsymbol{\alpha}})}{\sum_{j=1}^{n} I\{\tilde{X}_j^*(\hat{\beta}^{Lcatr}) \geq w\} w_j(\hat{\boldsymbol{\alpha}})}\right] d\hat{M}_{2i}(w;\hat{\beta}^{Lcatr},\hat{\boldsymbol{\alpha}})A_i$$
$$+ U_n(t;\beta^{Lcatr*},\boldsymbol{\alpha}^*) - U_n(t;\hat{\beta}^{Lcatr},\hat{\boldsymbol{\alpha}})$$
$$\hat{U}_n^P(u) = n^{-1/2}\sum_{i=1}^{n}\int_{-\infty}^{u}\left[Z_i - \frac{\sum_{j=1}^{n} I\{\tilde{X}_j^*(\hat{\beta}^{Patr}) \geq w\} Z_j w_j(\hat{\boldsymbol{\alpha}})}{\sum_{j=1}^{n} I\{\tilde{X}_j^*(\hat{\beta}^{Pcatr}) \geq w\} w_j(\hat{\boldsymbol{\alpha}})}\right] d\hat{M}_{2i}(w;\hat{\beta}^{Pcatr},\hat{\boldsymbol{\alpha}})A_i$$
$$+ U_n(u;\beta^{Pcatr*},\boldsymbol{\alpha}^*) - U_n(u;\hat{\beta}^{Pcatr},\hat{\boldsymbol{\alpha}}).$$

These processes $\{\hat{\boldsymbol{G}}_n^T, \hat{S}_n(t), \hat{U}_n^L(u), \hat{U}_n^P(v)\}^T$ are called bootstrapped processes (Peng and Fine, 2006). Since our interest is to examine model behavior using the testing procedure, $\hat{S}_n(t), \hat{U}_n^L(u)$ and $\hat{U}_n^P(v)$ are crucial elements. However, as the observed processes, $\hat{\boldsymbol{G}}_n$ should be included in the bootstrapped processes because estimators of the treatment effect depends on propensity score (In fact, $\hat{\boldsymbol{G}}_n = 0$). As Lin et al. (1996) and Peng and Fine (2006), we can simulate $A_1, \ldots A_n$ and compute the test statistic $\sup_s |S_n(s; \hat{\eta}^{atr}, \hat{\boldsymbol{\alpha}})|, \sup_t |U_n(t; \hat{\beta}^{Lcatr}, \hat{\boldsymbol{\alpha}})|$ and $\sup_u |U_n(u; \hat{\beta}^{Pcatr}, \hat{\boldsymbol{\alpha}})|$. Then by using realizations from resampling, say $(\boldsymbol{\alpha}^{T*}, \eta^{catr*}, \theta^{Lcatr*}, \theta^{Pcatr*})^T$, we can compute p-value as Hsieh et al. (2011). For a graphical description, it is suggested to plot 20 or 30 bootstrapped processes with the observed processes.

6.6 Real Data Analysis

We analyzed two datasets alluded to in the Introduction: the RTOG (Radiation Therapy Oncology Group) 9413 dataset and the HIV dataset (Albrecht et al. 2001; Peng and Fine, 2006). In the RTOG study, the key interest is to compare Whole Pelvic Radiotherapy (WPRT) and Prostate Only Radiotherapy (PORT) in the intermediate risk patients. In addition to treatment assignment (RT), there are several variables which potentially affected survival and progression : age, Karnofsky performance status (KPS), pretreatment prostate specific antigen (ipsa), pretreatment Gleason Score (igs), pretreatment tumor size (itsize) and T stage status (Stage). For data analysis, categorical variables are coded numerically. Treatment was coded 1 for PORT and 0 otherwise. Stage was coded as two dummy variables. After cleaning the dataset, 677 observations are used in the analysis. 200 times of resampling and bootstrap runs are tried.

We first consider two conventional methods: fitting the model employing all covariates and fitting the model using treatment variable only using original Lin et al. (1996) and Peng and Fine (2006) approach. In the model employing all covariates, for time to death, a treatment estimate is 0.076 with standard error 0.019. For time to first occurrence of disease progression (time to first occurrence of local failure, distant failure or biochemical failure), the Lin et al. (1996) and Peng and Fine (2006) methods give estimates of -0.059 and -0.062 with standard errors 0.004 and 0.011, respectively. In 200 resampling runs, 69 resampling runs give nonconvergence, which implies that the original Lin et al. (1996) and Peng and Fine (2006) approaches are not stable.

We also only consider the treatment variable only without propensity score in the estimation procedure. Without the propensity score adjustment, the treatment effect for death is 0.076 with standard error 0.065. For progression, Lin et al. (1996) and Peng and Fine (2006) method give estimates of -0.05 (0.119) and -0.091 (0.101), respectively where numbers in parenthesis are standard errors for the corresponding estimates.

Table 6.1 shows the proposed method. For our proposed method, all 200 resampling runs are employed. We computed the standard error of estimators in three methods : the resampling method which ignores variability of the propensity score (Naive), the bootstrap method (Bootstrap) and the resampling method incorporating variability of the propensity score (Resamp). The proposed method is more stable than the original Lin et al. (1996) and Peng and Fine (2006) approach. For the coefficients in the logistic regression model, we compare standard errors by data bootstrap method and the proposed resampling method (including variability of the propensity score) with standard errors by using R command `glm`. The standard errors by R command `glm` are shown in SE (glm) of Table 6.1. Results in Table 6.1 indicate that the standard errors by our proposed method are similar to ones by R command `glm`.

To examine how many observations are artificially censored, We computed artificial censoring rate. Let CR_D be the censoring rate subject to independent censoring, which is defined as $1 - \sum_{i=1}^{n} \xi_i/n$ and CR_X be the censoring rate subject to the dependent censoring, which is defined as $1 - \sum_{i=1}^{n} \delta_i/n$. Let ACR_{FL} be the artificial censoring rate from Lin et al. (1996) approach and ACR_{FP} be the one from Peng and Fine (2006) approach considering all covariates. Let $\hat{\boldsymbol{\beta}}^{FL}$ and $\hat{\boldsymbol{\beta}}^{FP}$ are estimators by Lin et al. (1996) and Peng and Fine (2006) approach from including all covariates in the model. Using arguments in Hsieh et al. (2011), mathematical definitions of these quantities are

$$ACR_{FL} = 1 - \frac{\sum_{i=1}^{n} \tilde{\delta}_i^{full*}(\hat{\boldsymbol{\beta}}^{LF})}{\sum_{i=1}^{n} \delta_i}$$

$$ACR_{FP} = 1 - \frac{\sum_{i=1}^{n} \sum_{j\neq i} \tilde{\delta}_{i(j)}^{full*}(\hat{\boldsymbol{\beta}}^{PF})}{(n-1)\sum_{i=1}^{n} \delta_i},$$

Table 6.1: Point estimates and standard error (SE) in RTOG data analysis using proposed method

	Point estimates	SE (Naive)	SE (Bootstrap)	SE (Resamp)
$\hat{\eta}^{catr}$	0.064	0.062	0.055	0.059
$\hat{\theta}^{Lcatr}$	-0.077	0.119	0.11	0.115
$\hat{\theta}^{Pcatr}$	-0.112	0.101	0.096	0.097
Propensity score model				
	Point estimates	SE (glm)	SE (Bootstrap)	SE (Resamp)
Intercept	0.812	1.677	1.754	1.644
age	-0.003	0.012	0.012	0.012
KPS[1]	0.0008	0.012	0.014	0.013
ipsa[2]	-0.008	0.006	0.006	0.006
igs[3]	-0.071	0.094	0.099	0.093
itsize[4]	-0.002	0.008	0.008	0.008
Stage_{c1}[5]	-0.079	0.289	0.302	0.304
Stage_{c2}[6]	-0.112	0.258	0.275	0.258

Estimators - $\hat{\eta}^{catr}$: the proposed estimator of the dependent censoring ; $\hat{\theta}^{Lcatr}$: the proposed estimator by Lin et al. (1996); $\hat{\theta}^{Pcatr}$: the proposed estimator by Peng and Fine (2006)

[1] Karnofsky performance status
[2] pretreatment prostate specific antigen
[3] pretreatment Gleason Score
[4] pretreatment tumor size
[5] If T stage of a patient is (T1c,T2a), then Stage_{c1}=1 otherwise 0
[6] If T stage of a patient is (T2c-T4), Stage_{c2}=1 otherwise 0

where $\tilde{\delta}_i^{full*}$ and $\tilde{\delta}_{i(j)}^{full*}$ are artificial censoring indicators using all covariates from Lin et al. (1996) and Peng and Fine (2006) approach, respectively. Let ACR_{AL} be the artificial censoring rate from proposed method from Lin et al. (1996) approach and ACR_{AP} be the one from the proposed method of Peng and Fine (2006). Then

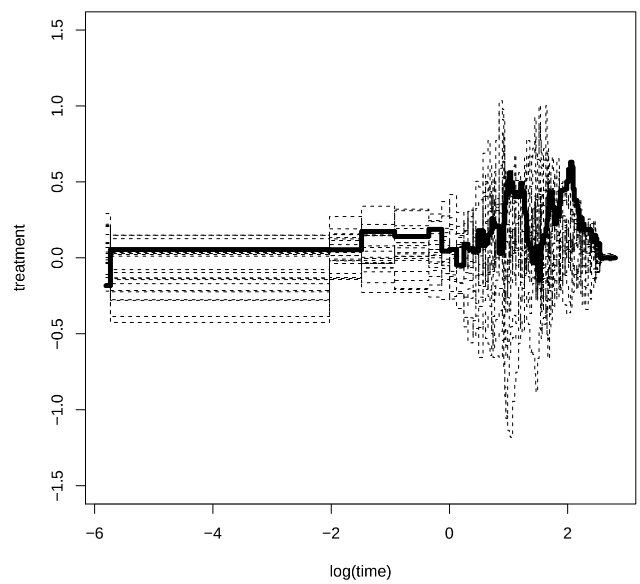
$$ACR_{AL} = 1 - \frac{\sum_{i=1}^n \tilde{\delta}_i^*(\hat{\beta}^{Lcatr})}{\sum_{i=1}^n \delta_i}$$
$$ACR_{AP} = 1 - \frac{\sum_{i=1}^n \sum_{j\neq i} \tilde{\delta}_{i(j)}^*(\hat{\beta}^{Pcatr})}{(n-1)\sum_{i=1}^n \delta_i}.$$

Based on quantities mentioned in previous paragraphs, artificial censoring rates of full model by Lin et al. (1996) and Peng and Fine (2006) are 0.806 and 0.196, respectively. Artificial censoring rates from the proposed approaches in the Section 6.3 are 0.034 and 0.048, which uses more uncensored observations than the full model approach.

Next, we apply the goodness of fit procedure described in Section 6.4. Figure 6.1, Figure 6.2 and Figure 6.3 show the observed process with 20 simulated bootstrap processes. P-values for model fit of $(\hat{\boldsymbol{\alpha}}, \hat{\eta}^{catr})$, model fit of $(\hat{\boldsymbol{\alpha}}, \hat{\eta}^{catr}, \hat{\theta}^{Lcatr})$ and model fit of $(\hat{\boldsymbol{\alpha}}, \hat{\eta}^{catr}, \hat{\theta}^{Pcatr})$ is 0.7, 0.61 and 0.97, respectively. Both the graphical and numerical results show that proposed model is adequate for the data.

Although the HIV dataset is from a randomized study, including covariates

Figure 6.1: Observed and bootstrapped processes of time to progression using $(\hat{\boldsymbol{\alpha}}, \hat{\eta}^{catr}, \hat{\theta}^{Lcatr})$ for RT



increases precision of treatment effect. Thus including covariates may result in better efficiency than model with the treatment only. In the HIV study, there are two continuous covariates, log(RNA) and CD4 count. Variation of CD4 count is very large. Thus if one analyzes the dataset including the CD4 count, excessive artificial censoring occurs. For estimating treatment effect including this covariate, our proposed method is useful.

The treatment group in the dataset has three levels: NFV only, EFV only, and NFV+EFV. Baseline treatment is NFV. In this case, we merged EFV only and

Figure 6.2: Observed and bootstrapped processes of time to progression using $(\hat{\boldsymbol{\alpha}}, \hat{\eta}^{catr}, \hat{\theta}^{Pcatr})$ (right) for RT

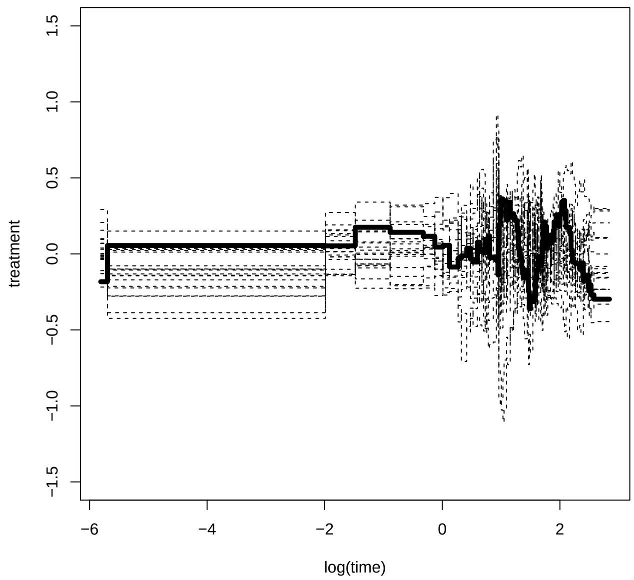
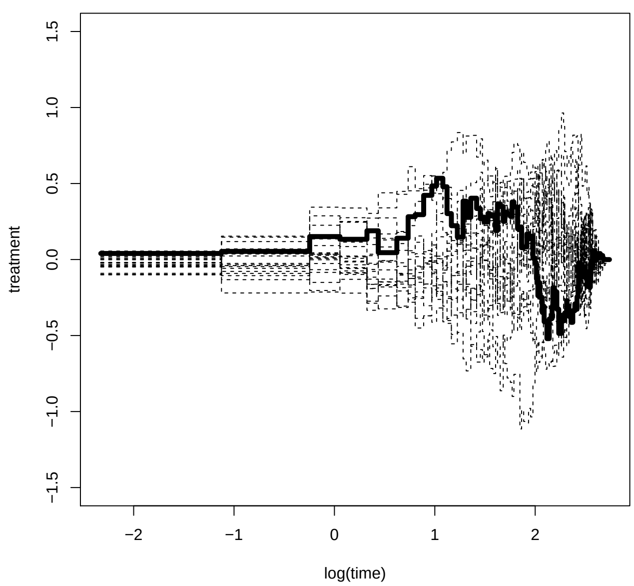


Figure 6.3: Observed and bootstrapped processes of time to death for RT



NFV+EFV into one group to apply the proposed methodology. One observed value is removed because it has CD4 count 0.

Table 6.2 shows the analysis by using proposed approach. Ignoring variability

Table 6.2: Point estimates and standard error (SE) in RTOG data analysis using proposed method

	Point estimate	SE (Naive)	SE (Bootstrap)	SE (Resamp)
$\hat{\eta}^{catr}$	0.745	0.15	0.149	0.148
$\hat{\theta}^{Lcatr}$	0.936	0.241	0.221	0.228
$\hat{\theta}^{Pcatr}$	0.766	0.239	0.217	0.225
Propensity score				
	Point estimate	SE(glm)	SE(Bootstrap)	SE (Resamp)
Intercept	1.145	0.97	1.066	0.975
log(RNA)	-0.133	0.208	0.227	0.207
CD4 count	0.00008	0.0008	0.0009	0.0008

Estimators : $\hat{\eta}^{catr}$: the proposed estimator of the dependent censoring ; $\hat{\theta}^{Lcatr}$: the proposed estimator by Lin et al. (1996); $\hat{\theta}^{Pcatr}$: the proposed estimator by Peng and Fine (2006)

of the propensity score results greater standard error than that in data bootstrap method. Moreover, incorporating variability of the propensity score provides lower standard error than that in ignoring it. As with the RTOG data analysis, the standard errors for the logistic regression coefficients are similar to ones by the proposed method.

Next, we compared the standard errors when employing treatment variable only. When we analyze data using treatment variable only, the treatment effect of time to withdrawal is 0.749 with the standard error 0.151. The treatment effect of time to the first virologic failure by the original Lin et al. (1996) is 1.021 with standard error 0.24. The treatment effect by the original Peng and Fine (2006) method provides 0.818 with standard error 0.237. We can see that the standard errors by ignoring variability of the propensity score are almost same as ones when employing the treatment variable only. Moreover, when comparing standard errors from the proposed method in Table 6.2 and fitting model using the treatment variable only, it can be seen that the use of the propensity score model increases the precision of treatment effect.

By considering all covariates, the estimator of $\boldsymbol{\eta}_0$ corresponding to Z is 0.782 with standard error 0.016. In this case, estimators of $\boldsymbol{\theta}_0$ corresponding to Z by

Lin et al. (1996) and Peng and Fine (2006) are 1.101 and 1.174. However, the standard errors of these estimators are 0 due to excessive artificial censoring. The artificial censoring rates by using all covariates are 1 for both the original Lin et al. (1996) and Peng and Fine (2006) method. On the other hand, the artificial censoring rates by using the proposed method is 0.069 for the approach based on Lin et al. (1996) and 0.013 for Peng and Fine (2006) based approach.

6.7 Simulation Studies

Next, we performed some simulations to explore the finite-sample properties of the proposed methodology. We generated a confounder $V \sim N(0,4)$ and simulated treatment variable Z as Bernoulli random variable with probability $\exp(\boldsymbol{\alpha}_0^T \boldsymbol{H})[1+\exp(\boldsymbol{\alpha}_0^T \boldsymbol{H})]^{-1}$, where $\boldsymbol{H} = (1, \boldsymbol{V}^T)^T$ and $\boldsymbol{\alpha}_0 = (\alpha_1, \alpha_2) = (0, 0.5)^T$. Then error variable $\boldsymbol{\epsilon}$ is bivariate normal with mean $(0, 1.2)^T$ and covariance matrix $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, where $\rho = 0, 0.25, 0.5$. Independent censoring times are simulated as $C \sim log(U)$, where U has uniform distribution on $(0, 200)$. True values of parameters are $\boldsymbol{\theta}_0 = (1, 0.5)^T$ and $\boldsymbol{\eta}_0 = (0.5, 1)^T$. 500 datasets are simulated and with each simulated dataset, 500 bootstrap runs and 500 resampling runs mentioned in the previous section are tried.

We calculated bias, empirical standard deviation (EMPSD), mean of standard error (SEE) and 95% coverage probability (CP) of estimators from considering all covariates and proposed estimator. In this case, the coverage probability is based on the empirical distribution based on resampling runs or bootstrap runs. As the real data analysis case, in the proposed method, we calculated the standard error of the estimators in three ways : the resampling method which ignores variability of the propensity score (Naive), the bootstrap method (Bootstrap) and the resampling method incorporating variability of the propensity score (Resamp).

Tables 6.3 and 6.4 show the numerical results considering entire covariates and the proposed methodology. Henceforth the procedure considering entire covariates is denoted as “the full covariates procedure". The estimators from utilizing all covariates for the dependent censoring and the event of interest by Lin et al. (1996) and Peng and Fine (2006) are denoted as $\hat{\boldsymbol{\eta}}^F$, $\hat{\boldsymbol{\theta}}^{LF}$ and $\hat{\boldsymbol{\theta}}^{PF}$, respectively. Since the

Table 6.3: Bias, empirical standard deviation (EMPSD), mean of standard error (SEE) and 95% coverage (CP) for estimators including all covariates when $N = 250$ and $N = 500$

$N = 250$									
		Bias		EMPSD		SEE		CP	
		V	Z	V	Z	V	Z	V	Z
$\rho = 0$	$\hat{\boldsymbol{\eta}}^F$	0.0001	0.001	0.042	0.152	0.039	0.154	0.938	0.946
	$\hat{\boldsymbol{\theta}}^{LF}$	0.019	0.007	0.078	0.187	0.089	0.132	0.899	0.833
	$\hat{\boldsymbol{\theta}}^{PF}$	0.0004	-0.01	0.04	0.149	0.042	0.154	0.944	0.957
$\rho = 0.25$	$\hat{\boldsymbol{\eta}}^F$	0.002	0.004	0.041	0.159	0.04	0.155	0.95	0.961
	$\hat{\boldsymbol{\theta}}^{LF}$	0.044	-0.01	0.113	0.199	0.094	0.122	0.817	0.753
	$\hat{\boldsymbol{\theta}}^{PF}$	-0.001	0.001	0.04	0.153	0.041	0.151	0.943	0.938
$\rho = 0.5$	$\hat{\boldsymbol{\eta}}^F$	0.002	-0.001	0.043	0.167	0.04	0.156	0.93	0.962
	$\hat{\boldsymbol{\theta}}^{LF}$	0.055	-0.021	0.105	0.192	0.097	0.109	0.727	0.702
	$\hat{\boldsymbol{\theta}}^{PF}$	-0.003	0.004	0.04	0.149	0.041	0.15	0.944	0.952
$N = 500$									
		Bias		EMPSD		SEE		CP	
		V	Z	V	Z	V	Z	V	Z
$\rho = 0$	$\hat{\boldsymbol{\eta}}^F$	0.002	-0.008	0.028	0.131	0.028	0.109	0.941	0.952
	$\hat{\boldsymbol{\theta}}^{LF}$	0.013	0.012	0.057	0.135	0.079	0.125	0.924	0.891
	$\hat{\boldsymbol{\theta}}^{PF}$	0.0003	0.002	0.028	0.104	0.029	0.109	0.963	0.961
$\rho = 0.25$	$\hat{\boldsymbol{\eta}}^F$	0.003	-0.006	0.029	0.126	0.028	0.11	0.943	0.967
	$\hat{\boldsymbol{\theta}}^{LF}$	0.037	-0.008	0.093	0.145	0.087	0.118	0.872	0.875
	$\hat{\boldsymbol{\theta}}^{PF}$	0.00009	0.002	0.027	0.101	0.029	0.107	0.955	0.965
$\rho = 0.5$	$\hat{\boldsymbol{\eta}}^F$	0.002	-0.011	0.028	0.123	0.028	0.11	0.928	0.958
	$\hat{\boldsymbol{\theta}}^{LF}$	0.063	-0.035	0.121	0.153	0.091	0.108	0.744	0.834
	$\hat{\boldsymbol{\theta}}^{PF}$	-0.0001	0.0008	0.028	0.105	0.028	0.105	0.952	0.952

Estimators - $\hat{\boldsymbol{\eta}}^F$: the estimator of dependent censoring in the full model ; $\hat{\boldsymbol{\theta}}^{LF}$: the estimator by Lin et al. (1996) in the full model; $\hat{\boldsymbol{\theta}}^{PF}$: the estimator by Peng and Fine (2006) in the full model

simulation results are based on the joint distributions of estimators for dependent censoring, and the event of interest by Lin et al. (1996) and Peng and Fine (2006) approach, simulation runs are removed if the standard errors of any one of these estimators in the simulation runs are 0. When $N = 250$, for $\rho = 0, 0.25$ and 0.5, 34, 62 and 127 results from simulation runs are removed because the standard errors of the estimators by Lin et al. (1996) for either confounder or treatment are 0. Similarly, when $N = 500$, 40, 77 and 168 results are omitted. This noncovergence problem is serious especially $\rho = 0.5$ and $N = 500$, 33.6% of entire simulation runs are abandoned, which is relatively large. However, for the proposed method, all 500 runs for the estimation of the treatment effect converge.

If simulation runs corresponding to only $\hat{\boldsymbol{\eta}}^F$ and $\hat{\boldsymbol{\theta}}^{PF}$ are considered, there are no noncovergence runs. it can be seen that the method of Lin et al. (1996) is problematic in this case. This is due to the excessive artificial censoring, as we illustrate in computing artificial censoring part.

The numerical results show that the proposed method works well. The mean of the standard errors when considering estimated propensity score as true is high so that coverage probability is conservative. However, the bootstrap and resampling approaches provide desired coverage. The proposed approach has better coverage relative to the full covariates procedure. Moreover, we compare the estimated standard error using the bootstrap and resampling approaches with the standard errors from R command `glm`. When $N = 250$, the mean of standard errors for $\hat{\alpha}_1$ and $\hat{\alpha}_2$ from `glm` are 0.14 and 0.084, respectively. For $N = 500$, the mean of standard errors for $\hat{\alpha}_1$ and $\hat{\alpha}_2$ from `glm` are 0.099 and 0.059, respectively. Results of $\hat{\alpha}_1$ and $\hat{\alpha}_2$ show that the bootstrap and resampling approaches for $\hat{\alpha}_1$ and $\hat{\alpha}_2$ provide the correct estimate of the standard error.

Table 6.5 shows the artificial censoring proportion. To examine the effect of the artificial censoring, entire simulation runs are included in this calculation. When considering all covariates, the artificial censoring rate by the original Lin et al. (1996) approach is high. However, the artificial censoring rate of the proposed method based on Lin et al. (1996) is small, which implies that loss of observations for the proposed methodology is much smaller than that in the full covariates procedure.

In practice, it is difficult to check for the adequacy of the propensity score model. Thus another simulation study is performed to explore the robustness of the proposed procedure. In this simulation study, it is not assumed that the propensity model relating confounder and treatment variable is true. We generate $\boldsymbol{J} = (J_1, J_2)^T$ from a bivariate normal distribution with mean $(0, 0)^T$ and covariance matrix $\begin{pmatrix} 4 & a \\ a & 1 \end{pmatrix}$, where $a = 0, 1$. In other words, J_1 has normal distribution with mean 0 and variance 4 marginally and J_2 has normal distribution with mean 0 and variance 1 marginally. When $a = 1$, J_1 and J_2 have correlation 0.5. Then we set

Table 6.4: Bias, empirical standard deviation (EMPSD), mean of standard error (SEE) and 95% coverage (CP) for the proposed estimator when $N = 250$ and $N = 500$

$N = 250$									
		Bias	EMPSD	SEE			CP		
				Naive	Bootstrap	Resamp	Naive	Bootstrap	Resamp
$\rho = 0$	$\hat{\eta}^{catr}$	0.006	0.196	0.249	0.191	0.192	0.98	0.924	0.934
	$\hat{\theta}^{Lcatr}$	-0.011	0.242	0.406	0.241	0.25	1	0.946	0.94
	$\hat{\theta}^{Pcatr}$	-0.015	0.192	0.382	0.197	0.219	1	0.954	0.94
	$\hat{\alpha}_1$	0.01	0.144		0.142	0.142		0.936	0.934
	$\hat{\alpha}_2$	0.004	0.079		0.086	0.087		0.954	0.954
$\rho = 0.25$	$\hat{\eta}^{catr}$	0.014	0.187	0.25	0.191	0.192	0.98	0.944	0.946
	$\hat{\theta}^{Lcatr}$	0.009	0.241	0.405	0.238	0.247	1	0.936	0.938
	$\hat{\theta}^{Pcatr}$	-0.0004	0.19	0.381	0.194	0.216	0.998	0.956	0.956
	$\hat{\alpha}_1$	0.01	0.144		0.142	0.142		0.94	0.932
	$\hat{\alpha}_2$	0.004	0.079		0.086	0.087		0.954	0.958
$\rho = 0.5$	$\hat{\eta}^{catr}$	0.015	0.187	0.25	0.191	0.193	0.98	0.944	0.948
	$\hat{\theta}^{Lcatr}$	0.014	0.237	0.402	0.235	0.245	1	0.94	0.944
	$\hat{\theta}^{Pcatr}$	0.003	0.182	0.377	0.188	0.211	0.998	0.95	0.952
	$\hat{\alpha}_1$	0.01	0.144		0.142	0.142		0.938	0.936
	$\hat{\alpha}_2$	0.004	0.079		0.086	0.087		0.956	0.948
$N = 500$									
		Bias	EMPSD	SSE			CP		
				Naive	Bootstrap	Resamp	Naive	Bootstrap	Resamp
$\rho = 0$	$\hat{\eta}^{catr}$	0.013	0.137	0.176	0.134	0.134	0.986	0.95	0.952
	$\hat{\theta}^{Lcatr}$	0.008	0.17	0.284	0.166	0.168	0.998	0.934	0.942
	$\hat{\theta}^{Pcatr}$	-0.002	0.126	0.27	0.134	0.14	1	0.966	0.956
	$\hat{\alpha}_1$	-0.001	0.096		0.099	0.099		0.958	0.968
	$\hat{\alpha}_2$	0.001	0.058		0.06	0.06		0.938	0.928
$\rho = 0.25$	$\hat{\eta}^{catr}$	0.02	0.135	0.175	0.133	0.133	0.988	0.934	0.934
	$\hat{\theta}^{Lcatr}$	0.017	0.166	0.282	0.163	0.164	0.996	0.94	0.928
	$\hat{\theta}^{Pcatr}$	0.001	0.131	0.268	0.131	0.136	1	0.956	0.948
	$\hat{\alpha}_1$	-0.001	0.096		0.099	0.1		0.964	0.962
	$\hat{\alpha}_2$	0.001	0.058		0.06	0.06		0.938	0.938
$\rho = 0.5$	$\hat{\eta}^{catr}$	0.018	0.134	0.175	0.133	0.133	0.988	0.944	0.946
	$\hat{\theta}^{Lcatr}$	0.016	0.165	0.281	0.161	0.163	0.994	0.936	0.95
	$\hat{\theta}^{Pcatr}$	0.002	0.127	0.266	0.127	0.133	1	0.942	0.94
	$\hat{\alpha}_1$	-0.001	0.096		0.1	0.099		0.96	0.954
	$\hat{\alpha}_2$	0.001	0.058		0.06	0.06		0.932	0.928

Estimators : $\hat{\eta}^{catr}$: the proposed estimator of the dependent censoring ; $\hat{\theta}^{Lcatr}$: the proposed estimator using Lin et al. (1996) approach; $\hat{\theta}^{Pcatr}$: the proposed estimator using Peng and Fine (2006) approach; $\hat{\alpha}_1$: the estimator of α_1; $\hat{\alpha}_2$: the estimator of α_2

Table 6.5: Artificial censoring proportions assuming true propensity score model

$N = 250$						
	CR_D[1]	CR_X[2]	ACR_{FL}[3]	ACR_{FP}[4]	ACR_{AL}[5]	ACR_{AP}[6]
$\rho = 0$	0.09	0.214	0.779	0.148	0.066	0.065
$\rho = 0.25$	0.09	0.195	0.843	0.144	0.064	0.064
$\rho = 0.5$	0.09	0.167	0.906	0.133	0.065	0.066
$N = 500$						
	CR_D[1]	CR_X[2]	ACR_{FL}[3]	ACR_{FP}[4]	ACR_{AL}[5]	ACR_{AP}[6]
$\rho = 0$	0.088	0.215	0.823	0.148	0.064	0.065
$\rho = 0.25$	0.089	0.193	0.883	0.142	0.066	0.068
$\rho = 0.5$	0.088	0.166	0.941	0.133	0.066	0.068

[1] the censoring rate subject to the independent censoring
[2] the censoring rate subject to the dependent censoring
[3] the artificial censoring rate from Lin et al. (1996) approach considering all covariates
[4] the artificial censoring rate from Peng and Fine (2006) approach considering all covariates
[5] the artificial censoring rate from proposed method of Lin et al. (1996) approach
[6] the artificial censoring rate from the proposed method of Peng and Fine (2006) approach

$V = J_1$ and $Z = I(J_2 > 0)$. Other parameter settings are same as before except $\rho = 0, 0.25$.
Three scenarios are considered :

(1) Dependent censoring with confounder ($a = 1$ and $\rho = 0.25$)

(2) Independent censoring with confounder ($a = 1$ and $\rho = 0$)

(3) Dependent censoring with randomized study ($a = 0$ and $\rho = 0.25$)

Table 6.6 and Table 6.7 show results from using the entire covariates in the model and from the proposed model. 54 runs, 43 runs, and 96 runs are removed for case 1, case 2 and case 3, respectively when $N = 250$. For $N = 500$, 83 runs, 32 runs and 107 runs are removed for case 1, case 2 and case 3, respectively.

Numerical results indicate that the proposed method works well. In this case, as for the true propensity model case, 3 ways of standard error calculation are used.

Table 6.6: Bias, empirical standard deviation (EMPSD), mean of standard error (SEE) and 95% coverage (CP) for using all covariates when $N = 250$ and $N = 500$

		Bias		EMPSD		SEE		CP	
$N = 250$									
		V	Z	V	Z	V	Z	V	Z
Case 1	$\hat{\boldsymbol{\eta}}^F$	0.001	-0.001	0.041	0.166	0.039	0.153	0.944	0.951
	$\hat{\boldsymbol{\theta}}^{LF}$	0.04	0.003	0.108	0.203	0.095	0.123	0.827	0.78
	$\hat{\boldsymbol{\theta}}^{PF}$	-0.005	0.011	0.041	0.151	0.041	0.151	0.944	0.953
Case 2	$\hat{\boldsymbol{\eta}}^F$	0.0002	-0.002	0.043	0.187	0.039	0.153	0.941	0.95
	$\hat{\boldsymbol{\theta}}^{LF}$	0.018	0.025	0.083	0.189	0.093	0.137	0.897	0.84
	$\hat{\boldsymbol{\theta}}^{PF}$	-0.003	0.011	0.041	0.155	0.042	0.154	0.939	0.937
Case 3	$\hat{\boldsymbol{\eta}}^F$	0.001	-0.037	0.036	0.269	0.036	0.143	0.946	0.941
	$\hat{\boldsymbol{\theta}}^{LF}$	0.041	-0.005	0.095	0.204	0.1	0.125	0.837	0.787
	$\hat{\boldsymbol{\theta}}^{PF}$	-0.005	-0.002	0.038	0.144	0.038	0.14	0.953	0.931
$N = 500$									
		Bias		EMPSD		SEE		CP	
		V	Z	V	Z	V	Z	V	Z
Case 1	$\hat{\boldsymbol{\eta}}^F$	0.002	-0.009	0.029	0.124	0.028	0.109	0.933	0.945
	$\hat{\boldsymbol{\theta}}^{LF}$	0.031	0.002	0.087	0.148	0.087	0.118	0.851	0.894
	$\hat{\boldsymbol{\theta}}^{PF}$	-0.001	-0.003	0.028	0.098	0.029	0.106	0.952	0.966
Case 2	$\hat{\boldsymbol{\eta}}^F$	0.001	-0.012	0.031	0.162	0.028	0.11	0.947	0.949
	$\hat{\boldsymbol{\theta}}^{LF}$	0.015	0.013	0.065	0.145	0.079	0.123	0.908	0.885
	$\hat{\boldsymbol{\theta}}^{PF}$	-0.0004	0.003	0.029	0.102	0.029	0.108	0.955	0.968
Case 3	$\hat{\boldsymbol{\eta}}^F$	0.001	-0.026	0.026	0.194	0.026	0.101	0.926	0.959
	$\hat{\boldsymbol{\theta}}^{LF}$	0.037	0.001	0.088	0.145	0.089	0.116	0.804	0.883
	$\hat{\boldsymbol{\theta}}^{PF}$	-0.001	-0.007	0.026	0.096	0.027	0.098	0.947	0.952

Estimators - $\hat{\boldsymbol{\eta}}^F$: the estimator of the dependent censoring in the full model; $\hat{\boldsymbol{\theta}}^{LF}$: the estimator by Lin et al. (1996) in the full model; $\hat{\boldsymbol{\theta}}^{PF}$: the estimator by Peng and Fine (2006) in the full model

The coverage probability is based on empirical distribution based on resampling runs or bootstrap runs. $\hat{\boldsymbol{\eta}}^F$, $\hat{\boldsymbol{\theta}}^{LF}$ and $\hat{\boldsymbol{\theta}}^{PF}$ are estimators from full model. As in the simulation scenario in the main paper, treating estimated propensity score as true results in large standard error of estimators $(\hat{\eta}^{catr}, \hat{\theta}^{Lcatr}, \hat{\theta}^{Pcatr})^T$. The data bootstrap and new resampling approach reflected variation of estimated propensity score into $(\hat{\eta}^{catr}, \hat{\theta}^{Lcatr}, \hat{\theta}^{Pcatr})^T$. Our methodology even works well for the randomized study. One interesting is that more simulation runs are lost in randomized study than those in observational study. Even though our primary

Table 6.7: Bias, empirical standard deviation (EMPSD), mean of standard error (SEE) and 95% coverage (CP) for proposed estimator when $N = 250$ and $N = 500$

$N = 250$									
		Bias	EMPSD	SSE			CP		
				naive	Bootstrap	Resamp	Naive	Bootstrap	Resamp
Case 1	$\hat{\eta}^{catr}$	0.027	0.183	0.242	0.186	0.186	0.986	0.958	0.958
	$\hat{\theta}^{Lcatr}$	0.033	0.232	0.388	0.227	0.234	0.992	0.94	0.948
	$\hat{\theta}^{Pcatr}$	0.025	0.19	0.373	0.191	0.21	1	0.968	0.974
Case 2	$\hat{\eta}^{catr}$	0.025	0.191	0.242	0.186	0.187	0.98	0.936	0.94
	$\hat{\theta}^{Lcatr}$	0.027	0.243	0.391	0.229	0.238	0.998	0.93	0.94
	$\hat{\theta}^{Pcatr}$	0.017	0.204	0.374	0.194	0.213	1	0.944	0.95
Case 3	$\hat{\eta}^{catr}$	0.009	0.152	0.201	0.154	0.154	0.984	0.952	0.958
	$\hat{\theta}^{Lcatr}$	0.004	0.189	0.313	0.18	0.181	0.998	0.94	0.934
	$\hat{\theta}^{Pcatr}$	0.012	0.161	0.302	0.157	0.157	1	0.946	0.94
$N = 500$									
		Bias	EMPSD	SSE			CP		
				Naive	Bootstrap	Resamp	Naive	Bootstrap	Resamp
Case 1	$\hat{\eta}^{catr}$	0.007	0.133	0.173	0.132	0.131	0.978	0.934	0.93
	$\hat{\theta}^{Lcatr}$	0.02	0.159	0.277	0.16	0.161	1	0.948	0.948
	$\hat{\theta}^{Pcatr}$	0.008	0.118	0.261	0.126	0.131	1	0.952	0.97
Case 2	$\hat{\eta}^{catr}$	0.012	0.134	0.173	0.132	0.132	0.988	0.936	0.94
	$\hat{\theta}^{Lcatr}$	0.022	0.158	0.277	0.162	0.162	1	0.956	0.962
	$\hat{\theta}^{Pcatr}$	0.009	0.122	0.262	0.13	0.134	1	0.966	0.966
Case 3	$\hat{\eta}^{catr}$	-0.0001	0.1	0.141	0.109	0.109	0.994	0.972	0.97
	$\hat{\theta}^{Lcatr}$	-0.002	0.121	0.221	0.128	0.127	0.998	0.96	0.958
	$\hat{\theta}^{Pcatr}$	-0.002	0.108	0.211	0.11	0.11	1	0.958	0.956

Estimators - $\hat{\eta}^{catr}$: the proposed estimator of the dependent censoring ; $\hat{\theta}^{Lcatr}$: the proposed estimator using Lin et al. (1996) approach; $\hat{\theta}^{Pcatr}$: the proposed estimator using Peng and Fine (2006) approach

goal is to estimate treatment effect in observational study including continuous covariates which have large variations, it also works well for randomized study including continuous covariates with large variance.

As for the simulation study assuming true logistic regression model, we compute the artificial censoring rate using entire 500 simulation runs. As can be seen the Table 6.8, the artificial censoring proportion from the proposed method is significantly lower than that of the full model case.

Table 6.8: Artificial censoring proportions not assuming propensity score model

$N = 250$						
	CR_D [1]	CR_X [2]	ACR_{FL} [3]	ACR_{FP} [4]	ACR_{AL} [5]	ACR_{AP} [6]
Case 1	0.088	0.194	0.839	0.143	0.062	0.064
Case 2	0.088	0.215	0.783	0.149	0.062	0.064
Case 3	0.078	0.2	0.887	0.158	0.054	0.053
$N = 500$						
	CR_D [1]	CR_X [2]	ACR_{FL} [3]	ACR_{FP} [4]	ACR_{AL} [5]	ACR_{AP} [6]
Case 1	0.088	0.193	0.877	0.143	0.062	0.064
Case 2	0.087	0.215	0.818	0.147	0.062	0.063
Case 3	0.077	0.2	0.917	0.157	0.053	0.053

[1] the censoring rate subject to the independent censoring
[2] the censoring rate subject to the dependent censoring
[3] the artificial censoring rate from Lin et al. (1996) approach considering all covariates
[4] the artificial censoring rate from Peng and Fine (2006) approach considering all covariates
[5] the artificial censoring rate from proposed method of Lin et al. (1996) approach
[6] the artificial censoring rate from the proposed method of Peng and Fine (2006) approach

The proposed method works well although the propensity score model is misspecified. Note that even for the randomized study, the proposed method has the advantage compared to the full covariates approach in terms of numerical stability. The artificial censoring rates not assuming the true propensity model are also shown in Table 6.8. As in the previous case, the artificial censoring in the proposed approach is smaller than that in the full covariates procedure.

In these simulation studies, the original Peng and Fine (2006) method is better than the original Lin et al. (1996) method. However, if the variability of covariates is large, even the original Peng and Fine (2006) method does not provide correct coverage. We perform a simulation study using the HIV dataset. First, by using estimates from real data analysis, the time to event of interest and time to dependent censoring are generated

$$\begin{pmatrix} X = 1.10083Z - 0.739872V_1 - 0.000008V_2 + \epsilon^X \\ D = 0.781576Z - 0.537557V_1 - 0.000849V_2 + \epsilon^D \end{pmatrix},$$

where Z indicates the bivariate treatment defined in the real data analysis, V_1 is logarithm of RNA value, and V_2 is CD4 count. Let $\mathbf{L}^* = (1, V_1^T, V_2^T)^T$. In this case, Z is generated by Bernoulli(p), where $p = \frac{\exp(\boldsymbol{\alpha}_0^T \mathbf{L}^*)}{1+\exp(\boldsymbol{\alpha}_0^T \mathbf{L}^*)}$ and $\boldsymbol{\alpha}_0$ are logistic regression coefficients $(1.145059, -0.13313, 0.000083)^T$ from real data study. Based on residual values, we generate error values from bivariate normal distribution with mean $(7.88, 8.2)^T$ and covariance matrix $\begin{pmatrix} 1.02 & 0.42 \\ 0.42 & 0.42 \end{pmatrix}$. Independent censoring time C has uniform distribution with minimum value 3 and maximum value 10. In each simulation run, 150 observations (without replacement) are selected in observations. 500 times of resampling and bootstrap runs are tried.

As in the real data study, including all covariates results in excessive artificial censoring. Only 22 runs of Lin et al. (1996) estimators give nonzero standard errors and any simulation run of Peng and Fine (2006) estimators does not provide nonzero standard errors for all covariates. Table 6.9 shows the result of the simulation study using proposed method. Standard errors are computed in three ways as other simulation studies. The proposed methodology works well except in terms of coverage for α_3, the logistic regression coefficient corresponding to V_2. As can be seen in Table 6.10, the artificial censoring rate of the full model for Lin et al. (1996) and Peng and Fine (2006) approach is 0.988 and 1, respectively. Compared to the artificial censoring rates in the full model, the artificial censoring rates by the proposed method based on Lin et al. (1996) and Peng and Fine (2006) are 0.071 and 0.048, respectively. In this study, the original Peng and Fine (2006) method does not provide the correct coverage for the estimate. The result of the simulation study shows the effectiveness of our approach when the variability of confounders is extremely large.

6.8 Discussion

In this chapter, we have proposed methodology for estimating treatment effects under a semicompeting risks data structure in the context of an observational study. In semicompeting risks data, only one nonterminal event is the event of interest. In medical study, it is common that the event of interest occurs several times. This type of events is called recurrent events. Recurrent events in the presence of dependent censoring have been extensively studied by Ghosh and Lin (2003),

Table 6.9: Bias, empirical standard deviation (EMPSD), mean of standard error (SEE) and 95% coverage (CP) for simulation study by using HIV dataset

	Bias	EMPSD	SEE (Naive)	SEE (Bootstrap)	SEE (Resamp)	CP (Naive)	CP (Bootstrap)	CP (Resamp)
$\hat{\eta}^{atr}$	-0.001	0.166	0.182	0.169	0.174	0.964	0.954	0.946
$\hat{\theta}^{Latr}$	-0.006	0.23	0.261	0.236	0.239	0.96	0.948	0.946
$\hat{\theta}^{Patr}$	-0.007	0.216	0.251	0.224	0.224	0.976	0.944	0.934
Intercept	0.065	1.232		1.235	1.208		0.942	0.934
V_1	-0.015	0.254		0.259	0.255		0.938	0.936
V_2	0.00003	0.001		0.001	0.001		0.918	0.892

Estimators - $\hat{\eta}^{catr}$: the proposed estimator of the dependent censoring ; $\hat{\theta}^{Lcatr}$: the proposed estimator using Lin et al. (1996) approach; $\hat{\theta}^{Pcatr}$: the proposed estimator using Peng and Fine (2006) approach

Table 6.10: Artificial censoring proportions for simulation study using HIV dataset

CR_D[1]	CR_X[2]	ACR_{FL}[3]	ACR_{FP}[4]	ACR_{AL}[5]	ACR_{AP}[6]
0.466	0.497	0.988	1	0.071	0.048

[1] the censoring rate subject to the independent censoring
[2] the censoring rate subject to the dependent censoring
[3] the artificial censoring rate from Lin et al. (1996) approach considering all covariates
[4] the artificial censoring rate from Peng and Fine (2006) approach considering all covariates
[5] the artificial censoring rate from proposed method of Lin et al. (1996) approach
[6] the artificial censoring rate from the proposed method of Peng and Fine (2006) approach

Ghosh (2010), and Hsieh et al. (2011). Our methodology could be applied to this situation. Currently, this extension is under investigation.

In this chapter, the propensity score is modeled by using logistic regression model. However, there are other ways to construct propensity score by nonparametric method, such as boosting (Breiman et al. 1984). Recently, Zhu (2013) and Zhu et al. (2014) proposed combining propensity scores from logistic regression and nonparametric learning method for causal inference. It can be also interesting topic to compare performance between parametric modeling and nonparametric method, and methodology of Zhu (2013) and Zhu et al. (2014).

In this paper, ideas from causal inference framework were used to motivate the estimation procedure. One may wish to give a causal interpretation as usual sense from casual inference literature. However, it is important to note that it is impossible to make any causal interpretation. This is nicely explained in Ghosh (2012).

Chapter 7
Future Work

7.1 Introduction

In this thesis, we addressed restrictions of the marginal regression by Lin et al. (1996) and Peng and Fine (2006) and provide solutions for the restrictions. Lin et al. (1996) and Peng and Fine (2006) proposed the estimation procedure for the time to event of interest in the presence of the dependent censoring. However, they did not study the efficiency with standard errors in detail and the key technique which Lin et al. (1996) and Peng and Fine (2006) employed, the artificial censoring technique, may cause an unstable estimation of covariate effects for the time to the event of the interest when the covariates have a large variability. Moreover, the goodness of fit method which Peng and Fine (2006) used does not match their U-statistic of order 2 estimating function. In Chapter 4, we proposed the weighted estimator based on resampling. In Chapter 5, the goodness of fit for U-statistics of order 2 is proposed. This new goodness of fit method provides higher power than the case of incorrect use of U-statistic of order 1 method for U-statistics of order 2 estimating functions. In Chapter 6, the covariate adjustment by the propensity score is developed. This adjustment provides for stable estimation of the treatment effect and significantly reduces the degree of artificial censoring.

With semicompeting risks data, a linear transformation model has high flexibility in terms of modeling, thus this model in the semicompeting risks data setting also can generate a rich set of models. In this chapter, we will briefly present ideas of the linear transformation model and variable selection in the semicompeting risks data as future work.

7.2 Semicompeting Risks Data in the Linear Transformation Model

The linear transformation models are a very interesting class of models in survival analysis. It includes broad a range of models, including the proportional hazard model and the proportional odds model. Its high flexibility enables statisticians to construct an appropriate model. Fine et al. (1998) modified this model to guarantee the consistency of estimators when censoring proportion of subjects is high. Cai et al. (2000) suggested this model for clustered failure time data and Cai et al. (2002) proposed a semiparametric mixture effect transformation model. The aforementioned models assume that the censoring variable is independent of covariates. Chen et al. (2002) suggested the linear transformation model which relaxes this assumption.

However, all these models assume that subjects are censored by the independent censoring mechanism. It is reasonable to consider the linear transformation model in the context of semicompeting risks data. Let X be time to the event of interest, D be the time to dependent censoring, C be the time to independent censoring and $\mathbf{Z}$ be a $p \times 1$ vector of covariates. Note that the semicompeting risks data structure is independent and identical replicates $\{\tilde{X}_i, \tilde{D}_i, \delta_i, \xi_i, \mathbf{Z}_i\}_{i=1}^n$ of $(\tilde{X}, \tilde{D}, \delta, \xi, \mathbf{Z})$, where $\tilde{X} = X \wedge D \wedge C,\ \tilde{D} = D \wedge C,\ \delta = I(X \le \tilde{D})$ and $\xi = I(D \le C)$. It is assumed that C is independent with $(X, D, \mathbf{Z})$. It is also assumed that the censoring proportion subject to C is not extreme. Then it is sensible to build the transformation model for semicompeting risks data as

$$\begin{pmatrix} h_1(D_i) = -\mathbf{Z}_i^T \boldsymbol{\eta}_0 + \epsilon_i^D \\ h_2(X_i) = -\mathbf{Z}_i^T \boldsymbol{\theta}_0 + \epsilon_i^X \end{pmatrix} \quad i = 1, \ldots, n,$$

where $h_1(\cdot)$ and $h_2(\cdot)$ are unknown but nondecreasing functions and $\boldsymbol{\eta}_0$ and $\boldsymbol{\theta}_0$ are $p \times 1$ vectors of regression coefficients. Moreover, $\epsilon_i \equiv (\epsilon_i^D, \epsilon_i^X), i = 1, ..., n$ are independent and identically distributed and unlike the assumption of the AFT model in the previous chapters, they are completely specified. Recall that the estimating function under independent censoring in the transformation model is

$$\mathbf{S}_n(\boldsymbol{\eta}) = \sum_{i=1}^{n}\sum_{j=1}^{n} w(\mathbf{Z}_{ij}^T\boldsymbol{\eta})\mathbf{Z}_{ij}\left\{\frac{\xi_j I(\tilde{D}_i \geq \tilde{D}_j)}{\hat{G}^2(\tilde{D}_j)} - \zeta(\mathbf{Z}_{ij}^T\boldsymbol{\eta})\right\}, \tag{7.1}$$

where $\hat{G}$ is the Kaplan-Meier estimate for variable C, $\mathbf{Z}_{ij} = \mathbf{Z}_i - \mathbf{Z}_j$ and $\zeta(\mathbf{Z}_{ij}^T\boldsymbol{\eta}) = P\{\epsilon_i^D - \epsilon_j^D \geq (\mathbf{Z}_i - \mathbf{Z}_j)^T\boldsymbol{\eta}\}$. As the AFT model case, the simple replacement of $\tilde{D}$ to $\tilde{X}$ does not provide an unbiased estimation of $\boldsymbol{\theta}_0$ because X is censored by $\tilde{D}$.

As can be seen on the proof of obtaining the unbiased estimator of $\zeta(\mathbf{Z}_{ij}^T\boldsymbol{\eta})$ in Cheng et al. (1995), the most important fact that is used in the proof is independence of D and C. To apply the logic similar to Cheng et al. (1995) in semicompeting risks data, independence of X and D is required. However, imposing this assumption can be a very strong assumption. More details should be studied to derive the unbiased estimating equation for $\boldsymbol{\theta}_0$.

7.3 Variable selection

Variable selection has been well studied in past decades. The basic idea is to select variables which contribute substantially to the response. Stepwise regression and subset selection are the conventions to use to select variables for applied statistics, but they are not systematic and contains errors by the selection procedure (Johnson et al. 2008). To cure this problem, several methods have been developed : the least absolute shrinkage operator (LASSO)(Tibshirani 1996), ridge regression, smoothly clipped absolute deviation (SCAD)(Fan and Li, 2001) and so on.

Originally, these methods were developed based on linear regression. Many researchers have extended these methods to various models, such as the generalized linear model and estimating functions. Johnson et al. (2008) proposed variable selection using general estimating functions. They applied the variable selection method on censored data and missing data. For censored data, their model is a semiparametric linear regression model by the Buckley-James estimator, which is the extension of least square estimator to censored regression model (Lai and Ying, 1991).

For semicompeting risks data, the variable selection in the marginal regression method can be useful because of the artificial censoring. As can be seen in Chapter

6, the excessive artificial censoring causes unstable estimation of covariate effects. To reduce the excessive artificial censoring, selecting variables which can truly affect to the time to the event of interest is important.

The methods by Johnson (2008) and Johnson et al. (2008) are sensible to apply the bivariate AFT model. However, one may be careful on applying Johnson et al. (2008) because of the existence of the dependent censoring. Moreover, Theorem 1 in Johnson et al. (2008) shows root-n consistency which is the convergence in the probability. However, based on Ying (1993)'s argument, the solution of the penalized estimating equations would converge in almost surely to the nonzero coefficients. This is left as the future work.

Appendix
Appendix for Chapter 4 and Chapter 6

1 Mathematical Proofs (Chapter 6)

(a) Proof of $E\{S_n(\eta_0^{tr}, \boldsymbol{\alpha}_0)\} = 0$, $E\{U_n^L(\beta_0^{tr}, \boldsymbol{\alpha}_0)\} = 0$ and $E\{U_n^P(\beta_0^{tr}, \boldsymbol{\alpha}_0)\} = 0$

In this part, we will prove $E\{S_n(\eta_0^{tr}, \boldsymbol{\alpha}_0)\} = 0$, $E\{U_n^L(\beta_0^{tr}, \boldsymbol{\alpha}_0)\} = 0$ and $E\{U_n^P(\beta_0^{tr}, \boldsymbol{\alpha}_0)\} = 0$. Let $\mathbf{H}_i = (1, \mathbf{V}_i), i = 1, \ldots, n$. Denote

$$e_i(\boldsymbol{\alpha}) = \frac{\exp(\boldsymbol{\alpha}^T \mathbf{H}_i)}{1 + \exp(\boldsymbol{\alpha}^T \mathbf{H}_i)}$$
$$w_i(\boldsymbol{\alpha}) = \frac{Z_i}{e_i(\boldsymbol{\alpha})} + \frac{1 - Z_i}{1 - e_i(\boldsymbol{\alpha})}$$

$$S_n(\eta^{tr}, \boldsymbol{\alpha}) = n^{-1/2} \sum_{i=1}^n \xi_i w_i(\boldsymbol{\alpha}) \left[Z_i - \frac{\sum_{j=1}^n I\{\tilde{D}_j^*(\eta^{tr}) \geq \tilde{D}_i^*(\eta^{tr})\} w_j(\boldsymbol{\alpha}) Z_j}{\sum_{j=1}^n I\{\tilde{D}_j^*(\eta^{tr}) \geq \tilde{D}_i^*(\eta^{tr})\} w_j(\boldsymbol{\alpha})} \right]$$

where $\tilde{D}_i^*(\eta^{tr}) = D \wedge C - \eta^{tr} Z_i$. The new estimating function based on Lin et al. (1996) is

$$U_n^L(\beta^{tr}, \boldsymbol{\alpha}) = n^{-1/2} \sum_{i=1}^n \tilde{\delta}_i^*(\beta^{tr}) w_i(\boldsymbol{\alpha}) \left[Z_i - \frac{\sum_{j=1}^n I\{\tilde{X}_j^*(\beta^{tr}) \geq \tilde{X}_i^*(\beta^{tr})\} w_j(\boldsymbol{\alpha}) Z_j}{\sum_{j=1}^n I\{\tilde{X}_j^*(\beta^{tr}) \geq \tilde{X}_i^*(\beta^{tr})\} w_j(\boldsymbol{\alpha})} \right],$$

where

$$\beta^{tr} = (\eta^{tr}, \theta^{tr})$$

$$
\begin{aligned}
d(\beta^{tr}) &= \max_i \{0, (\theta^{tr} - \eta^{tr})Z_i\} \\
\tilde{X}_i^*(\beta^{tr}) &= (X_i - \theta^{tr} Z_i) \wedge \{(D_i \wedge C_i) - \eta^{tr} Z_i - d(\beta^{tr})\} \\
\tilde{\delta}_i^*(\beta^{tr}) &= I[(X_i - \theta^{tr} Z_i) \leq \{(D_i \wedge C_i) - \eta^{tr} Z_i - d(\beta^{tr})].
\end{aligned}
$$

The new estimating function based on Peng and Fine (2006) is

$$
U_n^P(\beta^{tr}, \boldsymbol{\alpha}) = \frac{2n^{1/2}}{n(n-1)} \sum_{1 \leq i < j \leq n} (Z_i - Z_j) w_i(\boldsymbol{\alpha}) w_j(\boldsymbol{\alpha}) \phi_{ij}(\beta^{tr}),
$$

where

$$
\begin{aligned}
d_{ij}(\beta^{tr}) &= \max \{0, (\theta^{tr} - \eta^{tr})Z_i, (\theta^{tr} - \eta^{tr})Z_j\} \\
\tilde{X}_{i(j)}^*(\beta^{tr}) &= (X_i - \theta^{tr} Z_i) \wedge \{(D_i \wedge C_i) - \eta^{tr} Z_i - d_{ij}(\beta^{tr})\} \\
\tilde{\delta}_{i(j)}^*(\beta^{tr}) &= I[(X_i - \theta^{tr} Z_i) \leq \{(D_i \wedge C_i) - \eta^{tr} Z_i - d_{ij}(\beta^{tr})] \\
\phi_{ij}(\beta^{tr}) &= \tilde{\delta}_{i(j)}^*(\beta^{tr}) I\{\tilde{X}_{i(j)}^*(\beta^{tr}) \leq \tilde{X}_{j(i)}^*(\beta^{tr})\} - \tilde{\delta}_{j(i)}^*(\beta^{tr}) I\{\tilde{X}_{i(j)}^*(\beta^{tr}) \geq \tilde{X}_{j(i)}^*(\beta^{tr})\}.
\end{aligned}
$$

Let $\boldsymbol{\alpha}_0$ be true regression coefficient of logistic regression. We assume the following:

1. Given covariates, common density of ϵ_is is uniformly bounded. Moreover, let $f(\cdot)$ be common density of D_is. Assume that (Ghosh, 2000, Chapter 6)

$$
\int_{-\infty}^{\infty} \left(\frac{\dot{f}(t)}{f(t)} \right)^2 f(t) dt < \infty
$$

   where $\dot{f}(t) = \dfrac{df}{dt}$.
2. The solutions of $\mathbf{G}_n(\boldsymbol{\alpha}) = 0$, $S_n(\eta^{tr}, \boldsymbol{\alpha}_0) = 0$, $U_n^L(\theta^{tr}, \eta_0^{tr}, \boldsymbol{\alpha}_0)$ and $U_n^P(\theta^{tr}, \eta_0^{tr}, \boldsymbol{\alpha}_0) = 0$ are unique.
3. For $i = 1, \ldots, n$, there exists $s > 0$ and $r > 0$ such that $0 < s \leq w_i(\boldsymbol{\alpha}) \leq r < \infty$ for all $\boldsymbol{\alpha}$.
4. Denote filtration as $\mathcal{F}_t = \{N_{1i}(u; \eta^{tr}, \boldsymbol{\alpha}_0), N_{2i}(u; \eta^{tr}, \boldsymbol{\alpha}_0), Y_{1i}(u; \eta^{tr}, \boldsymbol{\alpha}_0), Y_{2i}(u; \beta^{tr}, \boldsymbol{\alpha}_0), Z_i;\ i = 1, \ldots, n; 0 \leq u < t\}$, where

$$
N_{1i}(t; \eta^{tr}, \boldsymbol{\alpha}) = w_i(\boldsymbol{\alpha}) I\{\tilde{D}_i^*(\eta^{tr}) \leq t, \xi_i = 1\} = w_i(\boldsymbol{\alpha}) \xi_i I\{\tilde{D}_i^*(\eta^{tr}) \leq t\}
$$

$$N_{2i}(t;\beta^{tr},\boldsymbol{\alpha}) = w_i(\boldsymbol{\alpha})I\{\tilde{X}_i^*(\beta^{tr}) \le t, \tilde{\delta}_i^*(\beta^{tr}) = 1\} = w_i(\boldsymbol{\alpha})\tilde{\delta}_i^*(\beta^{tr})I\{\tilde{X}_i^*(\beta^{tr}) \le t\}$$
$$Y_{1i}(t;\eta^{tr},\boldsymbol{\alpha}) = w_i(\boldsymbol{\alpha})I\{\tilde{D}_i^*(\eta^{tr}) \ge t\}$$
$$Y_{2i}(u;\beta^{tr},\boldsymbol{\alpha}) = w_i(\boldsymbol{\alpha})I\{\tilde{X}_i^*(\beta^{tr}) \ge t\}.$$

For $i = 1, \ldots, n$, define

$$L_i^{(1)}(u) = Z_i - \frac{\sum_{j=1}^n I\{\tilde{D}_j^*(\eta_0^{tr}) \ge u\}w_j(\boldsymbol{\alpha}_0)Z_j}{\sum_{j=1}^n I(\tilde{D}_j^*(\eta_0^{tr}) \ge u)w_j(\boldsymbol{\alpha}_0)}$$
$$L_i^{(2)}(u) = Z_i - \frac{\sum_{j=1}^n I\{\tilde{X}_j^*(\beta_0^{tr}) \ge u\}w_j(\boldsymbol{\alpha}_0)Z_j}{\sum_{j=1}^n I\{\tilde{X}_j^*(\beta_0^{tr}) \ge u\}w_j(\boldsymbol{\alpha}_0)}.$$

Then $L_i^{(1)}(\cdot)$ and $L_i^{(2)}(\cdot)$ are $\mathcal{F}^-$ predictable.

5. Existence of limiting quantities : For every $u > 0$, there exist $\bar{z}^{(1)}(\cdot) > 0$ and $\bar{z}^{(2)}(\cdot) > 0$ such that

$$\bar{z}^{(1)}(u) = \lim_{n\to\infty} \frac{\sum_{j=1}^n I\{\tilde{D}_j^*(\eta_0^{tr}) \ge u\}w_j(\boldsymbol{\alpha}_0)Z_j}{\sum_{j=1}^n I\{\tilde{D}_j^*(\eta_0^{tr}) \ge u\}w_j(\boldsymbol{\alpha}_0)}$$
$$\bar{z}^{(2)}(u) = \lim_{n\to\infty} \frac{\sum_{j=1}^n I\{\tilde{X}_j^*(\beta_0^{tr}) \ge u\}w_j(\boldsymbol{\alpha}_0)Z_j}{\sum_{j=1}^n I\{\tilde{X}_j^*(\beta_0^{tr}) \ge u\}w_j(\boldsymbol{\alpha}_0)}.$$

6. Let $\tilde{\lambda}(\beta^{tr},\boldsymbol{\alpha}) = E\{n^{-1/2}U_n^P(\beta^{tr},\boldsymbol{\alpha})\}$. Assume that $\tilde{\lambda}(\beta^{tr},\boldsymbol{\alpha})$ is differentiable at β_0^{tr}, and both $\left.\frac{\partial\tilde{\lambda}(\beta^{tr},\boldsymbol{\alpha})}{\partial\eta^{tr}}\right|_{\beta^{tr}=\beta_0^{tr},\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}$ and $\left.\frac{\partial\tilde{\lambda}(\beta^{tr},\boldsymbol{\alpha})}{\partial\theta^{tr}}\right|_{\beta^{tr}=\beta_0^{tr},\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}$ are nonsingular.

Moreover, we need assumptions dealing with the logistic regression model for propensity scores. These assumptions are from Ferguson (1996, Chapter 17, p114), Zhu (2013) and Zhu et al. (2014). Let

$$\mathbf{H} = (1, \mathbf{V}^T)^T, \quad e(\boldsymbol{\alpha}) = \frac{\exp(\boldsymbol{\alpha}^T\mathbf{H})}{1 + \exp(\boldsymbol{\alpha}^T\mathbf{H})}$$

(a) The parameter $\boldsymbol{\alpha}$ belongs to a compact subset of Γ. The likelihood

$$f(z,\boldsymbol{\alpha}) = e(\boldsymbol{\alpha})^z(1 - e(\boldsymbol{\alpha}))^{1-z},$$

is measurable in z for every $\boldsymbol{\alpha}$ in Γ. Moreover, f is continuous in $\boldsymbol{\alpha}$ for every z.

(b) For all z and $\boldsymbol{\alpha}$,

$$\log\left(\frac{f(z,\boldsymbol{\alpha}|\mathbf{H})}{f(z,\boldsymbol{\alpha}_0|\mathbf{H})}\right) \le h(z),$$

where $h(z)$ is a function satisfying $E_{\boldsymbol{\alpha}_0}|h(z)| < \infty$.

Under the logistic regression model assumption for treatment,

$$\begin{aligned}
S_n(\eta^{tr},\boldsymbol{\alpha}_0) &= n^{-1/2}\sum_{i=1}^{n}\xi_i w_i(\boldsymbol{\alpha_0})\Bigg[Z_i - \frac{\sum_{j=1}^{n} I\{\tilde{D}_j^*(\eta^{tr}) \ge \tilde{D}_i^*(\eta^{tr})\} w_j(\boldsymbol{\alpha_0}) Z_j}{\sum_{j=1}^{n} I(\tilde{D}_j^*(\eta^{tr}) \ge \tilde{D}_i^*(\eta^{tr})) w_j(\boldsymbol{\alpha_0})}\Bigg] \\
U_n^L(\beta^{tr},\boldsymbol{\alpha}_0) &= n^{-1/2}\sum_{i=1}^{n}\tilde{\delta}_i^*(\beta^{tr}) w_i(\boldsymbol{\alpha}_0)\Bigg[Z_i - \frac{\sum_{j=1}^{n} I\{\tilde{X}_j^*(\beta^{tr}) \ge \tilde{X}_i^*(\beta^{tr})\} w_j(\boldsymbol{\alpha}_0) Z_j}{\sum_{j=1}^{n} I\{\tilde{X}_j^*(\beta^{tr}) \ge \tilde{X}_i^*(\beta^{tr})\} w_j(\boldsymbol{\alpha}_0)}\Bigg].
\end{aligned} \tag{A.1}$$

Note that (A.1) is equivalent to

$$\begin{aligned}
S_n(\eta^{tr},\boldsymbol{\alpha}_0) &= n^{-1/2}\sum_{i=1}^{n}\int_{-\infty}^{\infty}\Bigg[Z_i - \frac{\sum_{j=1}^{n} I\{\tilde{D}_j^*(\eta^{tr}) \ge u\} w_j(\boldsymbol{\alpha}_0) Z_j}{\sum_{j=1}^{n} I\{\tilde{D}_j^*(\eta^{tr}) \ge u\} w_j(\boldsymbol{\alpha}_0)}\Bigg] dN_{1i}(u;\eta^{tr},\boldsymbol{\alpha}_0) \\
U_n^L(\beta^{tr},\boldsymbol{\alpha}_0) &= n^{-1/2}\sum_{i=1}^{n}\int_{-\infty}^{\infty}\Bigg[Z_i - \frac{\sum_{j=1}^{n} I\{\tilde{X}_j^*(\beta^{tr}) \ge u\} w_j(\boldsymbol{\alpha}_0) Z_j}{\sum_{j=1}^{n} I\{\tilde{X}_j^*(\beta^{tr}) \ge u\} w_j(\boldsymbol{\alpha}_0)}\Bigg] dN_{2i}(u;\beta^{tr},\boldsymbol{\alpha}_0).
\end{aligned}$$

Then by algebra,

$$\begin{aligned}
&\sum_{i=1}^{n}\int_{-\infty}^{\infty}\Bigg[Z_i - \frac{\sum_{j=1}^{n} I\{\tilde{D}_j^*(\eta_0^{tr}) \ge u\} w_j(\boldsymbol{\alpha}_0) Z_j}{\sum_{j=1}^{n} I\{\tilde{D}_j^*(\eta_0^{tr}) \ge u\} w_j(\boldsymbol{\alpha}_0)}\Bigg] dN_{1i}(u;\eta_0^{tr},\boldsymbol{\alpha}_0) \\
&= \sum_{i=1}^{n}\int_{-\infty}^{\infty}\Bigg[Z_i - \frac{\sum_{j=1}^{n} I\{\tilde{D}_j^*(\eta_0^{tr}) \ge u\} w_j(\boldsymbol{\alpha}_0) Z_j}{\sum_{j=1}^{n} I\{\tilde{D}_j^*(\eta_0^{tr}) \ge u\} w_j(\boldsymbol{\alpha}_0)}\Bigg] dM_{1i}(u;\eta_0^{tr},\boldsymbol{\alpha}_0) \\
&+ \sum_{i=1}^{n}\int_{-\infty}^{\infty} w_i(\boldsymbol{\alpha}_0)\Bigg[Z_i - \frac{\sum_{j=1}^{n} I\{\tilde{D}_j^*(\eta_0^{tr}) \ge u\} w_j(\boldsymbol{\alpha}_0) Z_j}{\sum_{j=1}^{n} I\{\tilde{D}_j^*(\eta_0^{tr}) \ge u\} w_j(\boldsymbol{\alpha}_0)}\Bigg] I\{\tilde{D}_i^*(\eta_0^{tr}) \ge u\}\lambda_{10}(u)du,
\end{aligned}$$

where

$$M_{1i}(u;\eta^{tr},\boldsymbol{\alpha}_0) = w_i(\boldsymbol{\alpha}_0)[\xi_i I\{\tilde{D}_i^*(\eta^{tr}) \le u\} - \int_{-\infty}^{u} I\{\tilde{D}_i^*(\eta^{tr}) \ge t\}\lambda_{10}(t)dt].$$

and $\lambda_{10}(\cdot)$ is the baseline hazard function for time to death. Then

$$\sum_{i=1}^{n}\int_{-\infty}^{\infty} w_i(\boldsymbol{\alpha}_0)Z_i I\{\tilde{D}_i^*(\eta_0^{tr}) \geq u\}\lambda_{10}(u)du - \int_{-\infty}^{\infty}\sum_{i=1}^{n} w_i(\boldsymbol{\alpha}_0) I\{\tilde{D}_i^*(\eta_0^{tr}) \geq u\}\lambda_{10}(u)$$
$$\times \frac{\sum_{j=1}^{n} I\{\tilde{D}_j^*(\eta_0^{tr}) \geq u\}w_j(\boldsymbol{\alpha}_0)Z_j}{\sum_{j=1}^{n} I\{\tilde{D}_j^*(\eta_0^{tr}) \geq u\}w_j(\boldsymbol{\alpha}_0)}du$$
$$= \sum_{i=1}^{n}\int_{-\infty}^{\infty} w_i(\boldsymbol{\alpha}_0)Z_i I\{\tilde{D}_i^*(\eta_0^{tr}) \geq u\}\lambda_{10}(u)du - \sum_{i=1}^{n}\int_{-\infty}^{\infty} w_i(\boldsymbol{\alpha}_0)Z_i I\{\tilde{D}_i^*(\eta_0^{tr}) \geq u\}\lambda_{10}(u)du$$
$$= 0,$$

Hence

$$S_n(\eta_0^{tr}, \boldsymbol{\alpha}_0) = n^{-1/2}\sum_{i=1}^{n}\int_{-\infty}^{\infty} w_i(\boldsymbol{\alpha}_0)\Bigg[Z_i - \frac{\sum_{j=1}^{n} I\{\tilde{D}_j^*(\eta_0^{tr}) \geq u\}w_j(\boldsymbol{\alpha}_0)Z_j}{\sum_{j=1}^{n} I\{\tilde{D}_j^*(\eta_0^{tr}) \geq u\}w_j(\boldsymbol{\alpha}_0)}\Bigg]dM_{1i}(u;\eta_0^{tr}, \boldsymbol{\alpha}_0).$$

Similarly,

$$U_n^L(\beta_0^{tr}, \boldsymbol{\alpha}_0) = n^{-1/2}\sum_{i=1}^{n}\int_{-\infty}^{\infty}\Bigg[Z_i - \frac{\sum_{j=1}^{n} I\{\tilde{X}_j^*(\beta_0^{tr}) \geq u\}w_j(\boldsymbol{\alpha}_0)Z_j}{\sum_{j=1}^{n} I\{\tilde{X}_j^*(\beta_0^{tr}) \geq u\}w_j(\boldsymbol{\alpha}_0)}\Bigg]dM_{2i}(u;\beta_0^{tr}, \boldsymbol{\alpha}_0)$$
$$M_{2i}(u;\beta^{tr}, \boldsymbol{\alpha}_0) = w_i(\boldsymbol{\alpha}_0)[\tilde{\delta}_i^*(\beta^{tr})I\{\tilde{X}_i^*(\beta^{tr}) \leq u\} - \int_{-\infty}^{u} I\{\tilde{X}_i^*(\beta^{tr}) \geq t\}\lambda_{20}(t)dt]$$

where $\lambda_{20}(\cdot)$ is the baseline hazard function for time to the event of interest.
By taking expectations,

$$E\{S_n(\eta_0^{tr}, \boldsymbol{\alpha}_0)\}$$
$$= n^{-1/2}E\Bigg\{\sum_{i=1}^{n}\int_{-\infty}^{\infty} E\Bigg\{\Bigg[Z_i - \frac{\sum_{j=1}^{n} I\{\tilde{D}_j^*(\eta_0^{tr}) \geq u\}w_j(\boldsymbol{\alpha}_0)Z_j}{\sum_{j=1}^{n} I\{\tilde{D}_j^*(\eta_0^{tr}) \geq u\}w_j(\boldsymbol{\alpha}_0)}\Bigg]dM_{1i}(u;\eta_0^{tr}, \boldsymbol{\alpha}_0)\Bigg| e_i(\boldsymbol{\alpha}_0)\Bigg\}\Bigg\}.$$

Given $e_i(\boldsymbol{\alpha}_0)$, $L_i^{(1)} = Z_i - \dfrac{\sum_{j=1}^{n} I\{\tilde{D}_j^*(\eta_0^{tr}) \geq u\}w_j(\boldsymbol{\alpha}_0)Z_j}{\sum_{j=1}^{n} I\{\tilde{D}_j^*(\eta_0^{tr}) \geq u\}w_i(\boldsymbol{\alpha}_0)}$ is bounded and $\mathcal{F}_t$-predictable. Hence given $e_i(\boldsymbol{\alpha}_0)$,

$$\sum_{i=1}^{n}\int_{-\infty}^{\infty}\Bigg[Z_i - \frac{\sum_{j=1}^{n} I\{\tilde{D}_j^*(\eta_0^{tr}) \geq u\}w_j(\boldsymbol{\alpha}_0)Z_j}{\sum_{j=1}^{n} I\{\tilde{D}_j^*(\eta_0^{tr}) \geq u\}w_i(\boldsymbol{\alpha}_0)}\Bigg]dM_{1i}(u;\eta_0^{tr}, \boldsymbol{\alpha}_0),$$

is a martingale and $E\{S_n(\eta_0^{tr},\boldsymbol{\alpha}_0)\}=0$ (Boyd et al. 2012; Theorem 2.5.4 in Fleming and Harrington, 2005, p.77). Similarly, $U_n^L(\beta_0^{tr},\boldsymbol{\alpha}_0)$ is also a martingale and $E\{U_n^L(\beta_0^{tr},\boldsymbol{\alpha}_0)\}=0$.

Let $\boldsymbol{\theta}_0=(\theta_0^{tr},(\boldsymbol{\theta}_0^{cfd})^T)^T$ and $\boldsymbol{\eta}_0=(\eta_0^{tr},(\boldsymbol{\eta}_0^{cfd})^T)^T$. Note that $\boldsymbol{\theta}_0^{cfd}$ is the true value of $\boldsymbol{\theta}_0$ corresponding to $\mathbf{V}$ and $\boldsymbol{\eta}_0^{cfd}$ is the true value of $\boldsymbol{\eta}_0$ corresponding $\mathbf{V}$. Under the true value of β_0^{tr} and $\boldsymbol{\alpha}_0$,

$$
\begin{aligned}
&E\{U_n^P(\beta_0^{tr},\boldsymbol{\alpha}_0)\}=E[(Z_1-Z_2)w_1(\boldsymbol{\alpha}_0)w_2(\boldsymbol{\alpha}_0)\times\{P(\epsilon_1^X+(\boldsymbol{\theta}_0^{cfd})^T\mathbf{V}_1\\
&\le\{\epsilon_1^D+(\boldsymbol{\eta}_0^{cfd})^T\mathbf{V}_1-d_{12}(\beta_0^{tr})\}\wedge\{\epsilon_2^X+(\boldsymbol{\theta}_0^{cfd})^T\mathbf{V}_2\}\wedge\{\epsilon_2^D+(\boldsymbol{\eta}_0^{cfd})^T\mathbf{V}_2-d_{12}(\beta_0^{tr})\}\\
&\wedge\{C_1-\eta_0^{tr}Z_1-d_{12}(\beta_0^{tr})\}\wedge\{C_2-\eta_0^{tr}Z_2-d_{12}(\beta_0^{tr})\}|e_1(\boldsymbol{\alpha}_0),e_2(\boldsymbol{\alpha}_0),Z_1,Z_2)\\
&-P(\epsilon_2^X+(\boldsymbol{\theta}_0^{cfd})^T\mathbf{V}_2\le\{\epsilon_2^D+(\boldsymbol{\eta}_0^{cfd})^T\mathbf{V}_2-d_{12}(\beta_0^{tr})\}\wedge\{\epsilon_1^X+(\boldsymbol{\theta}_0^{cfd})^T\mathbf{V}_1\}\\
&\wedge\{\epsilon_1^D+(\boldsymbol{\eta}_0^{cfd})^T\mathbf{V}_1-d_{12}(\beta_0^{tr})\}\wedge\{C_1-\eta_0^{tr}Z_1-d_{12}(\beta_0^{tr})\}\\
&\wedge\{C_2-\eta_0^{tr}Z_2-d_{12}(\beta_0^{tr})\}|e_1(\boldsymbol{\alpha}_0),e_2(\boldsymbol{\alpha}_0),Z_1,Z_2)\}].
\end{aligned}
$$

Since $C\perp(X,D)|\mathbf{W}$, it is clear that $C\perp(\epsilon^X,\epsilon^D)|\mathbf{W}$. By using the fact that $\{\tilde{X}_i,\tilde{D}_i,\delta_i,\xi_i,\mathbf{W}_i\}(i=1,\ldots,n)$ are i.i.d copies and the propensity score is a balancing score,

$$
\begin{aligned}
&P(\epsilon_1^X+(\boldsymbol{\theta}_0^{cfd})^T\mathbf{V}_1\le\{\epsilon_1^D+(\boldsymbol{\eta}_0^{cfd})^T\mathbf{V}_1-d_{12}(\beta_0^{tr})\}\wedge\{\epsilon_2^X+(\boldsymbol{\theta}_0^{cfd})^T\mathbf{V}_2\}\\
&\wedge\{\epsilon_2^D+(\boldsymbol{\eta}_0^{cfd})^T\mathbf{V}_2-d_{12}(\beta_0^{tr})\}\wedge\{C_1-\eta_0^{tr}Z_1-d_{12}(\beta_0^{tr})\}\\
&\wedge\{C_2-\eta_0^{tr}Z_2-d_{12}(\beta_0^{tr})\}|e_1(\boldsymbol{\alpha}_0),e_2(\boldsymbol{\alpha}_0),Z_1,Z_2)\\
&=P(\epsilon_2^X+(\boldsymbol{\theta}_0^{cfd})^T\mathbf{V}_2\le\{\epsilon_2^D+(\boldsymbol{\eta}_0^{cfd})^T\mathbf{V}_2-d_{12}(\beta_0^{tr})\}\wedge\{\epsilon_1^X+(\boldsymbol{\theta}_0^{cfd})^T\mathbf{V}_1\}\\
&\wedge\{\epsilon_1^D+(\boldsymbol{\eta}_0^{cfd})^T\mathbf{V}_1-d_{12}(\beta_0^{tr})\}\wedge\{C_1-\eta_0^{tr}Z_1-d_{12}(\beta_0^{tr})\}\\
&\wedge\{C_2-\eta_0^{tr}Z_2-d_{12}(\beta_0^{tr})\}|e_1(\boldsymbol{\alpha}_0),e_2(\boldsymbol{\alpha}_0),Z_1,Z_2).
\end{aligned}
$$

Thus $E\{U_n^P(\beta_0^{tr},\boldsymbol{\alpha}_0)\}=0$.

(b) Proof of Theorem 6.1

The next step is to prove consistency of the estimator. Since the propensity model is true, by Ferguson (1996, Chapter 17, p114), $\hat{\boldsymbol{\alpha}}$ converges to $\boldsymbol{\alpha}_0$ almost surely. Assume the regularity conditions in Ying (1993) and the

Appendix of Peng and Fine (2006). Define

$$N_1(t;\eta^{tr},\boldsymbol{\alpha}) = \sum_{i=1}^{n} N_{1i}(t;\eta^{tr},\boldsymbol{\alpha}) \quad N_1^z(t;\eta^{tr},\boldsymbol{\alpha}) = \sum_{i=1}^{n} Z_i N_{1i}(t;\eta^{tr},\boldsymbol{\alpha})$$

$$Q_{1n}^z(t;\eta^{tr},\boldsymbol{\alpha}) = \sum_{j=1}^{n} I\{\tilde{D}_j^*(\eta^{tr}) \geq t\} w_j(\boldsymbol{\alpha}) Z_j \quad Q_{1n}(t;\eta^{tr},\boldsymbol{\alpha}) = \sum_{j=1}^{n} I\{\tilde{D}_j^*(\eta^{tr}) \geq t\} w_j(\boldsymbol{\alpha})$$

$$Q_1^z(t;\eta^{tr},\boldsymbol{\alpha}) = E[w_1(\boldsymbol{\alpha}) Z_1 I\{\tilde{D}_1^*(\eta^{tr}) \geq t\}] \quad Q_1(t;\eta^{tr},\boldsymbol{\alpha}) = E[w_1(\boldsymbol{\alpha}) I\{\tilde{D}_1^*(\eta^{tr}) \geq t\}]$$

$$\tilde{N}_{1i}(t;\eta^{tr},\boldsymbol{\alpha}) = w_i(\boldsymbol{\alpha}) I\{\tilde{D}_i^*(\eta^{tr}) \geq t, \xi_i = 1\} \quad \tilde{N}_1(t;\eta^{tr},\boldsymbol{\alpha}) = \sum_{i=1}^{n} \tilde{N}_{1i}(t;\eta^{tr},\boldsymbol{\alpha})$$

$$\tilde{N}_1^z(t;\eta^{tr},\boldsymbol{\alpha}) = \sum_{i=1}^{n} \tilde{N}_{1i}(t;\eta^{tr},\boldsymbol{\alpha}) Z_i.$$

Then let

$$S_n(t;\eta^{tr},\boldsymbol{\alpha}) = n^{-1/2} \sum_{i=1}^{n} \int_{-\infty}^{t} \left[Z_i - \frac{Q_{1n}^z(u;\eta^{tr},\boldsymbol{\alpha})}{Q_{1n}(u;\eta^{tr},\boldsymbol{\alpha})} \right] dN_{1i}(u;\eta^{tr},\boldsymbol{\alpha})$$

$$m_1(t;\eta^{tr},\boldsymbol{\alpha}) = E\left\{ \int_{-\infty}^{t} \left[Z_1 - \frac{E\{Q_1^z(u;\eta^{tr},\boldsymbol{\alpha})\}}{E\{Q_1(u;\eta^{tr},\boldsymbol{\alpha})\}} \right] dN_{11}(u;\eta^{tr},\boldsymbol{\alpha}) \right\}.$$

Note that the proposed estimating function for η^{tr} is $S_n(\infty;\eta^{tr},\boldsymbol{\alpha})$. Ying (1993) argues that the expansion holds if $\boldsymbol{W}$ is treated as random. In his case, the full covariates are used for residual terms. In our case, only treatment variable is subtracted from original time to death or original time to the event of interest. Since treatment and confounder are independent given propensity score (Rosenbaum and Rubin, 1983) and the expectation of estimating equations of Lin et al. (1996) under the true value of parameters is equal to 0, arguments from Ying (1993) can still be applied to our case. Let $L(t;\eta^{tr},\boldsymbol{\alpha})$ be any one of the empirical processes $\tilde{N}, \tilde{N}^z, nQ_{1n}^z, nQ_{1n}, N_1$ and N_1^z. Take $0 \leq \zeta < 1, C_1^* > 0, K > 0, \omega > 0$. Since we assume that the propensity model is true, by Ferguson (1996, Chapter 17, p114),

$$n^{-1} \sum_{j=1}^{n} [I\{\tilde{D}_j^*(\eta^{tr}) \geq t\} w_j(\hat{\boldsymbol{\alpha}}) Z_j] \stackrel{a.s.}{\to} E[I\{\tilde{D}_1^*(\eta^{tr}) \geq t\} w_1(\boldsymbol{\alpha}_0) Z_1]$$

$$n^{-1} \sum_{j=1}^{n} [I\{\tilde{D}_j^*(\eta^{tr}) \geq t\} w_j(\hat{\boldsymbol{\alpha}})] \stackrel{a.s.}{\to} E[I\{\tilde{D}_1^*(\eta^{tr}) \geq t\} w_1(\boldsymbol{\alpha}_0)].$$

Moreover, by the strong law of large numbers,

$$n^{-1}\sum_{j=1}^{n}[I\{\tilde{D}_j^*(\eta^{tr}) \geq t\}w_j(\boldsymbol{\alpha})Z_j] \overset{a.s.}{\rightarrow} E[I\{\tilde{D}_1^*(\eta^{tr}) \geq t\}w_1(\boldsymbol{\alpha})Z_1] \qquad (A.2)$$

$$n^{-1}\sum_{j=1}^{n}[I\{\tilde{D}_j^*(\eta^{tr}) \geq t\}w_j(\boldsymbol{\alpha})] \overset{a.s.}{\rightarrow} E[I\{\tilde{D}_1^*(\eta^{tr}) \geq t\}w_1(\boldsymbol{\alpha})]. \qquad (A.3)$$

Then using Theorem 3 from Ying (1993),

$$\sup_{|\eta^{tr}|\leq C_1^*, EL(t;\eta^{tr},\boldsymbol{\alpha_0})\leq Kn^{1-\zeta}} |L(t;\eta^{tr},\hat{\boldsymbol{\alpha}}) - E\{L(t;\eta^{tr},\boldsymbol{\alpha}_0)\}| = o(n^{(1-\zeta)/2+\omega}).$$

By Ying (1993), under certain regularity conditions, for $\zeta^* > 0$

$$\sup_{|\eta^{tr}|\leq C_1^*} |n^{-1/2}S_n(\eta^{tr},\hat{\boldsymbol{\alpha}}) - m_1(\eta^{tr},\boldsymbol{\alpha}_0)| = o(n^{-1/2+\zeta^*}),$$

where $m_1(\eta^{tr},\boldsymbol{\alpha}_0) = m_1(\infty;\eta^{tr},\boldsymbol{\alpha}_0)$. Similarly, we define

$$Q_2^z(t;\beta^{tr},\boldsymbol{\alpha}) = E[I\{\tilde{X}_1^*(\beta^{tr}) \geq t\}w_1(\boldsymbol{\alpha})Z_1]$$
$$Q_2(t;\beta^{tr},\boldsymbol{\alpha}) = E[I\{\tilde{X}_1^*(\beta^{tr}) \geq t\}w_1(\boldsymbol{\alpha})].$$

Similarly to the independent censoring case, there exist nonrandom functions $m_2(\cdot;\beta^{tr},\boldsymbol{\alpha})$, where

$$m_2(t;\beta^{tr},\boldsymbol{\alpha}) = E\left\{\int_{-\infty}^{t}\left[Z_1 - \frac{E\{Q_2^z(u;\beta^{tr},\boldsymbol{\alpha})\}}{E\{Q_2(u;\beta^{tr},\boldsymbol{\alpha})\}}\right]dN_{21}(u;\beta^{tr},\boldsymbol{\alpha})\right\}.$$

Let $m_2(\beta^{tr},\boldsymbol{\alpha}) = m_2(\infty;\beta^{tr},\boldsymbol{\alpha})$. By Ferguson (1996, Chapter 17, p114),

$$n^{-1}\sum_{j=1}^{n}[I\{\tilde{X}_j^*(\beta^{tr}) \geq t\}w_j(\hat{\boldsymbol{\alpha}})Z_j] \overset{a.s.}{\rightarrow} E[I\{\tilde{X}_1^*(\beta^{tr}) \geq t\}w_1(\boldsymbol{\alpha}_0)Z_1]$$

$$n^{-1}\sum_{j=1}^{n}[I\{\tilde{X}_j^*(\beta^{tr}) \geq t\}w_j(\hat{\boldsymbol{\alpha}})] \overset{a.s.}{\rightarrow} E[I\{\tilde{X}_1^*(\beta^{tr}) \geq t\}w_1(\boldsymbol{\alpha}_0)].$$

Moreover, by the strong law of large numbers,

$$n^{-1}\sum_{j=1}^{n}[I\{\tilde{X}_j^*(\beta^{tr}) \geq t\}w_j(\boldsymbol{\alpha})Z_j] \overset{a.s.}{\to} E[I\{\tilde{X}_1^*(\beta^{tr}) \geq t\}w_1(\boldsymbol{\alpha})Z_1] \quad \text{(A.4)}$$

$$n^{-1}\sum_{j=1}^{n}[I\{\tilde{X}_j^*(\beta^{tr}) \geq t\}w_j(\boldsymbol{\alpha})] \overset{a.s.}{\to} E[I\{\tilde{X}_1^*(\beta^{tr}) \geq t\}w_1(\boldsymbol{\alpha})]. \quad \text{(A.5)}$$

For some positive constant C_2^* and for $\zeta^{**} > 0$,

$$\sup_{||\beta^{tr}|| \leq C_2^*} |n^{-1/2}U_n^L(\beta^{tr}, \hat{\boldsymbol{\alpha}}) - m_2(\beta^{tr}, \boldsymbol{\alpha}_0)| = o(n^{-1/2+\zeta^{**}}).$$

Note that due to the strong consistency of $\hat{\boldsymbol{\alpha}}$ and uniqueness of $\boldsymbol{\alpha}_0$, for any neighborhood of η^{tr} and β^{tr}, say $\mathcal{N}_0$ and $\mathcal{N}_1$, respectively, by Ying (1993),

$$\sup_{\eta^{tr} \in \mathcal{N}_0} |n^{-1/2}S_n(\eta^{tr}, \hat{\boldsymbol{\alpha}}) - m_1(\eta^{tr}, \boldsymbol{\alpha}_0)| \overset{p}{\to} 0$$

$$\sup_{\beta^{tr} \in \mathcal{N}_1} |n^{-1/2}U_n^L(\beta^{tr}, \hat{\boldsymbol{\alpha}}) - m_2(\beta^{tr}, \boldsymbol{\alpha}_0)| \overset{p}{\to} 0.$$

Moreover, by (A.2) - (A.5), it is also true that if $\boldsymbol{\alpha}$ belongs on any neighborhood of $\boldsymbol{\alpha}_0$, say $\mathcal{B}$, then any fixed $\eta^{tr} \in \mathcal{N}_0$, $n^{-1/2}S_n(\eta^{tr}, \boldsymbol{\alpha})$ can be approximated by $m_1(\eta^{tr}, \boldsymbol{\alpha})$. Hence,

$$\sup_{\boldsymbol{\alpha} \in \mathcal{B}, \eta^{tr} \in \mathcal{N}_0} |n^{-1/2}S_n(\eta^{tr}, \boldsymbol{\alpha}) - m_1(\eta^{tr}, \boldsymbol{\alpha})| \overset{p}{\to} 0$$

Similarly, for any $\beta^{tr} \in \mathcal{N}_1$ and $\boldsymbol{\alpha} \in \mathcal{B}$,

$$\sup_{\boldsymbol{\alpha} \in \mathcal{B}, \beta^{tr} \in \mathcal{N}_1} |n^{-1/2}U_n^L(\beta^{tr}, \boldsymbol{\alpha}) - m_2(\beta^{tr}, \boldsymbol{\alpha})| \overset{p}{\to} 0.$$

Thus $\hat{\eta}^{catr}$ and $\hat{\theta}^{Lcatr}$ are strongly consistent. Now we want to show strong consistency of $\hat{\theta}^{Pcatr}$. Let $\mathcal{W}$ be a compact set of parameter β^{tr} and $h(Z_i, Z_j, \mathbf{V}_i, \mathbf{V}_j, \beta^{tr}, \boldsymbol{\alpha}_0) = w_i(\boldsymbol{\alpha}_0)w_j(\boldsymbol{\alpha}_0)(Z_i - Z_j)\phi_{ij}(\beta^{tr})$. From the U-statistics law of large numbers,

$$|n^{-1/2}U_n^P(\beta^{tr}, \boldsymbol{\alpha}_0) - \lambda(\beta, \boldsymbol{\alpha}_0)| \overset{p}{\to} 0.$$

for all $\beta^{tr} \in \mathcal{W}$. By decomposing the compact set $\mathcal{W}$ into several finite subsets

$\mathcal{W}_1, \ldots, \mathcal{W}_m$ such that $\mathcal{W} \in \cup_{i=1}^m \mathcal{W}_i$, for $(\beta^{tr})^i \in \mathcal{W}_i$,

$$\max_{1\le i\le m} |n^{-1/2} U_n^P((\beta^{tr})^i, \boldsymbol{\alpha}_0) - \lambda((\beta^{tr})^i, \boldsymbol{\alpha}_0)| \xrightarrow{p} 0.$$

Since $w_i(\boldsymbol{\alpha}_0), i = 1 \ldots n$ are bounded, by Appendix of Peng and Fine (2006),

$$\sup_{||\beta^{tr}-\tilde{\beta}^{tr}||\le\nu} n^{-1/2}|U_n^P(\beta^{tr}, \boldsymbol{\alpha}_0) - U_n^P(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0)| \le \frac{2}{n(n-1)}[\sum_{1\le i<j\le n} |Z_i - Z_j| K_{ij}(\tilde{\beta}^{tr}, \nu)],$$

where

$$L_{ij}^{(1)}(\tilde{\beta}^{tr}, \nu) = w_i(\boldsymbol{\alpha}_0) w_j(\boldsymbol{\alpha}_0) I[\{\beta^{tr} : |\beta^{tr} - \tilde{\beta}^{tr}| \le \nu, \tilde{X}_{i(j)}^*(\beta^{tr}) = \tilde{X}_{j(i)}^*(\beta^{tr})\} \ne \emptyset]$$
$$L_{ij}^{(2)}(\tilde{\beta}^{tr}, \nu) = w_i(\boldsymbol{\alpha}_0) w_j(\boldsymbol{\alpha}_0) I[\{\beta : |\beta - \tilde{\beta}^{tr}| \le \nu, \tilde{\delta}_{i(j)}^*(\beta) \ne \tilde{\delta}_{i(j)}^*(\tilde{\beta}^{tr})\} \ne \emptyset]$$
$$K_{ij}(\tilde{\beta}^{tr}, \nu) = \{L_{ij}^{(1)}(\tilde{\beta}^{tr}, \nu) + L_{ij}^{(2)}(\tilde{\beta}^{tr}, \nu) + L_{ji}^{(2)}(\tilde{\beta}^{tr}, \nu)\}.$$

Let $H_{ij}(\tilde{\beta}^{tr}, \nu) = |Z_i - Z_j| K_{ij}(\tilde{\beta}^{tr}, \nu)$ and $H(\tilde{\beta}^{tr}, \nu) = \sum_{1\le i<j\le n} |Z_i - Z_j| K_{ij}(\tilde{\beta}^{tr}, \nu)$. By Hoeffding decomposition,

$$H(\tilde{\beta}^{tr}, \nu) - E\{H(\tilde{\beta}^{tr}, \nu)\} = \sum_{i=1}^n B_i(\tilde{\beta}^{tr}, \nu) + \sum_{i<j} B_{ij}(\tilde{\beta}^{tr}, \nu),$$

where

$$B_i(\tilde{\beta}^{tr}, \nu) = \sum_{j\ne i} [E\{H_{ij}(\tilde{\beta}^{tr}, \nu) | Z_i, e_i(\boldsymbol{\alpha}_0)\} - E\{H_{ij}(\tilde{\beta}^{tr}, \nu)\}]$$
$$B_{ij}(\tilde{\beta}^{tr}, \nu) = H_{ij}(\tilde{\beta}^{tr}, \nu) - E\{H_{ij}(\tilde{\beta}^{tr}, \nu) | Z_i, e_i(\boldsymbol{\alpha}_0)\} - E\{H_{ij}(\tilde{\beta}^{tr}, \nu) | Z_j, e_j(\boldsymbol{\alpha}_0)\}$$
$$+ E\{H_{ij}(\tilde{\beta}^{tr}, \nu)\}.$$

To complete the proof of consistency, we need the following lemma.

Lemma A1. There exists constants b_0 and c_0 such that $E\{H_{ij}(\tilde{\beta}^{tr}, \nu)\} \le b_0 \nu$ and $E\{H(\tilde{\beta}^{tr}, \nu)\} \le c_0 \nu n^2$

Proof. We can use arguments as in Peng and Fine (2006) and the Appendix of Hsieh et al. (2011). Note that the set

$$\{||\beta^{tr} - \tilde{\beta}^{tr}|| \le \nu, \tilde{\delta}_{i(j)}^*(\beta^{tr}) \ne \tilde{\delta}_{i(j)}^*(\tilde{\beta}^{tr})\} \in D_1(\tilde{\beta}^{tr}, \nu) \cup D_2(\tilde{\beta}^{tr}, \nu),$$

where

$$D_1(\tilde{\beta}^{tr}, \nu) = \{||\beta^{tr} - \tilde{\beta}^{tr}|| < \nu, X_i - \theta^{tr} Z_i = [D_i + \eta^{tr}(Z_j - Z_i)] - \theta^{tr} Z_j\}$$
$$D_2(\tilde{\beta}^{tr}, \nu) = \{||\beta^{tr} - \tilde{\beta}^{tr}|| < \nu, X_i - \theta^{tr} Z_i = [C_i + \eta^{tr}(Z_j - Z_i)] - \theta^{tr} Z_j\}.$$

Then

$$\begin{aligned}
&D_1(\tilde{\beta}^{tr}, \nu) \\
&= \{||\beta^{tr} - \tilde{\beta}^{tr}|| < \nu, X_i + \theta^{tr}(Z_j - Z_i) = D_i + \eta^{tr}(Z_j - Z_i)\} \\
&= \{||\beta^{tr} - \tilde{\beta}^{tr}|| < \nu, \epsilon_i^X + \boldsymbol{\theta}_0^T \mathbf{W}_i + \theta^{tr}(Z_j - Z_i) = \epsilon_i^D + \boldsymbol{\eta}_0^T \mathbf{W}_i + \eta^{tr}(Z_j - Z_i)\} \\
&= \{||\beta^{tr} - \tilde{\beta}^{tr}|| < \nu, \epsilon_i^X + \boldsymbol{\theta}_0^T \mathbf{W}_i + [\tilde{\theta}^{tr} - (\tilde{\theta}^{tr} - \theta^{tr})](Z_j - Z_i) \\
&= \epsilon_i^D + \boldsymbol{\eta}_0^T \mathbf{W}_i + [\tilde{\eta}^{tr} - (\tilde{\eta}^{tr} - \eta^{tr})](Z_j - Z_i)\} \\
&\subseteq \{||\epsilon_i^X - \epsilon_i^D + (\boldsymbol{\theta}_0 - \boldsymbol{\eta}_0)^T \mathbf{W}_i + (\tilde{\theta}^{tr} - \tilde{\eta}^{tr})(Z_j - Z_i)|| < 2\nu |Z_j - Z_i|\}.
\end{aligned}$$

Thus there exists $d_0 > 0$ such that $w_i(\boldsymbol{\alpha}_0) w_j(\boldsymbol{\alpha}_0) P\{D_1(\tilde{\beta}^{tr}, \nu) | e_i(\boldsymbol{\alpha}_0), e_j(\boldsymbol{\alpha}_0), Z_i, Z_j\} \leq 2 w_i(\boldsymbol{\alpha}_0) w_j(\boldsymbol{\alpha}_0) d_0 |Z_j - Z_i| \nu$ by assumption. Similarly,

$$w_i(\boldsymbol{\alpha}_0) w_j(\boldsymbol{\alpha}_0) P\{D_2(\tilde{\beta}^{tr}, \nu) | e_i(\boldsymbol{\alpha}_0), e_j(\boldsymbol{\alpha}_0), Z_i, Z_j\} \leq 2 w_i(\boldsymbol{\alpha}_0) w_j(\boldsymbol{\alpha}_0) |Z_j - Z_i| \nu.$$

Hence, $E\{L_{ij}^{(2)}(\tilde{\beta}^{tr}, \nu)\} \leq 2 d_0 E\{w_i(\boldsymbol{\alpha}_0) w_j(\boldsymbol{\alpha}_0) |Z_j - Z_i|\} \nu$. Similarly, there exists $d_0^* > 0$ such that $E\{L_{ji}^{(2)}(\tilde{\beta}^{tr}, \nu)\} \leq 2 d_0^* E\{w_i(\boldsymbol{\alpha}_0) w_j(\boldsymbol{\alpha}_0) |Z_j - Z_i|\} \nu$ and there exists $f_0^* > 0$ and $P\{L_{ij}^{(1)}(\tilde{\beta}^{tr}, \nu)\} \leq 2 f_0^* E\{w_i(\boldsymbol{\alpha}_0) w_j(\boldsymbol{\alpha}_0) |Z_j - Z_i|\} \nu$. Since $w_i(\boldsymbol{\alpha}_0)$ are bounded, there exists $K_0 > 0$ and $K_1 > 0$ such that $E\{K_{ij}(\tilde{\beta}^{tr}, \nu)\} \leq K_0 \nu$ and $E\{K_{ij}^2(\tilde{\beta}^{tr}, \nu)\} \leq K_1 \nu$. By the Cauchy-Schwarz inequality,

$$E\{H_{ij}(\tilde{\beta}^{tr}, \nu)\} \leq \sqrt{E\{K_{ij}^2(\tilde{\beta}^{tr}, \nu)\} E|Z_i - Z_j|^2} \leq K_1 \nu \sqrt{E|Z_i - Z_j|^2}.$$

Hence there exists $b_0 > 0$ such that $E\{H_{ij}(\tilde{\beta}^{tr}, \nu)\} \leq b_0 \nu$. Finally, $E\{H(\tilde{\beta}^{tr}, \nu)\} \leq K_1 b_0 n^2$. Thus there exists $c_0 > 0$ such that $E\{H(\tilde{\beta}^{tr}, \nu)\} \leq c_0 \nu n^2$. □

Then by Lemma A1 above, $E\{H_{ij}(\tilde{\beta}^{tr}, \nu)\} \leq b_0 \nu$ and $E\{H(\tilde{\beta}^{tr}, \nu)\} \leq c_0 \nu n^2$. Note that B_i and B_{ij} are uncorrelated. Hence there exist $v_{10} > 0$ and $v_{20} > 0$

such that

$$Var\{H(\tilde{\beta}^{tr},\nu)\} = \sum_{i=1}^{n} Var\{B_i(\tilde{\beta}^{tr},\nu)\} + \sum_{i<j} Var\{B_{ij}(\tilde{\beta}^{tr},\nu)\} \leq v_{10}n^3 + v_{20}n^2 = O(n^3).$$

Take $\epsilon > 0$ and let $0 < \nu < \epsilon/(3b_0)$. Then by Markov inequality,

$$\begin{aligned} P\{[n(n-1)]^{-1}H(\tilde{\beta}^{tr},\nu) \geq \epsilon\} &\leq P\{[n(n-1)]^{-1}[H(\tilde{\beta}^{tr},\nu) - E\{H(\tilde{\beta}^{tr},\nu)]\} \geq \epsilon/3\} \\ &\leq \frac{9Var\{H(\tilde{\beta}^{tr},\nu)\}}{[n(n-1)]^2\epsilon^2} \to 0. \end{aligned}$$

Hence

$$\sup_{||\beta^{tr}-\tilde{\beta}^{tr}||\leq\nu} n^{-1/2}|U_n^P(\beta^{tr},\boldsymbol{\alpha}_0) - U_n^P(\tilde{\beta}^{tr},\boldsymbol{\alpha}_0)| \xrightarrow{p} 0. \tag{A.6}$$

Note that for any $\nu^* > 0$,

$$\begin{aligned} &\sup_{||\boldsymbol{\alpha}-\boldsymbol{\alpha}_0||\leq\nu^*,||\beta^{tr}-\tilde{\beta}^{tr}||\leq\nu} n^{-1/2}|U_n^P(\beta^{tr},\boldsymbol{\alpha}) - U_n^P(\tilde{\beta}^{tr},\boldsymbol{\alpha}_0)| \\ &\leq \sup_{||\beta^{tr}-\tilde{\beta}^{tr}||\leq\nu,||\boldsymbol{\alpha}-\boldsymbol{\alpha}_0||\leq\nu^*} n^{-1/2}|U_n^P(\beta^{tr},\boldsymbol{\alpha}) - U_n^P(\beta^{tr},\hat{\boldsymbol{\alpha}})| \\ &+ \sup_{||\beta^{tr}-\tilde{\beta}^{tr}||\leq\nu,||\boldsymbol{\alpha}-\boldsymbol{\alpha}_0||\leq\nu^*} n^{-1/2}|U_n^P(\beta^{tr},\hat{\boldsymbol{\alpha}}) - U_n^P(\beta^{tr},\boldsymbol{\alpha}_0)| \\ &+ \sup_{||\beta^{tr}-\tilde{\beta}^{tr}||\leq\nu,||\boldsymbol{\alpha}-\boldsymbol{\alpha}_0||\leq\nu^*} n^{-1/2}|U_n^P(\beta^{tr},\boldsymbol{\alpha}_0) - U_n^P(\tilde{\beta}^{tr},\boldsymbol{\alpha}_0)|. \end{aligned}$$

For fixed β^{tr}, by Taylor expansion,

$$n^{-1/2}|U_n^P(\beta^{tr},\hat{\boldsymbol{\alpha}}) - U_n^P(\beta^{tr},\boldsymbol{\alpha}_0)| = n^{-1/2}\left|(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0)\frac{\partial}{\partial\boldsymbol{\alpha}}U_n^P(\beta^{tr},\boldsymbol{\alpha})\Big|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}\right| + o_p(1).$$

Then for fixed β^{tr}, $n^{-1/2}\frac{\partial}{\partial\boldsymbol{\alpha}}U_n^P(\beta^{tr},\boldsymbol{\alpha})\Big|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0} = O(1)$ and $\hat{\boldsymbol{\alpha}} \xrightarrow{a.s.} \boldsymbol{\alpha}_0$, we have

$$\sup_{||\beta^{tr}-\tilde{\beta}^{tr}||\leq\nu,||\boldsymbol{\alpha}-\boldsymbol{\alpha}_0||\leq\nu^*} n^{-1/2}|U_n^P(\beta^{tr},\hat{\boldsymbol{\alpha}}) - U_n^P(\beta^{tr},\boldsymbol{\alpha}_0)| \xrightarrow{p} 0. \tag{A.7}$$

Then for $\boldsymbol{\alpha} \in \mathcal{B}$, due to uniqueness of $\boldsymbol{\alpha}_0$,

$$\sup_{\boldsymbol{\alpha}\in\mathcal{B},\beta^{tr}\in\mathcal{W}} |n^{-1/2}U_n^P(\beta^{tr},\boldsymbol{\alpha}) - \lambda(\beta^{tr},\boldsymbol{\alpha}_0)| \xrightarrow{p} 0.$$

By strong consistency of $\hat{\boldsymbol{\alpha}}$, we have

$$\sup_{\beta^{tr} \in \mathcal{W}} |n^{-1/2} U_n^P(\beta^{tr}, \hat{\boldsymbol{\alpha}}) - \lambda(\beta^{tr}, \boldsymbol{\alpha}_0)| \xrightarrow{p} 0.$$

Hence

$$\sup_{||\beta^{tr} - \tilde{\beta}^{tr}|| \le \nu, ||\boldsymbol{\alpha} - \boldsymbol{\alpha}_0|| \le \nu^*} |n^{-1/2} U_n^P(\beta^{tr}, \boldsymbol{\alpha}) - n^{-1/2} U_n^P(\beta^{tr}, \hat{\boldsymbol{\alpha}})| \xrightarrow{p} 0. \tag{A.8}$$

Combining (A.6), (A.7) and (A.8) yields,

$$\sup_{|\boldsymbol{\alpha} - \boldsymbol{\alpha}_0|| \le \nu^*, ||\beta^{tr} - \tilde{\beta}^{tr}|| \le \nu} n^{-1/2} |U_n^P(\beta^{tr}, \boldsymbol{\alpha}) - U_n^P(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0)| \xrightarrow{p} 0.$$

Moreover,

$$\sup_{||\boldsymbol{\alpha} - \tilde{\boldsymbol{\alpha}}|| \le \nu^*, ||\beta^{tr} - \tilde{\beta}^{tr}|| \le \nu} n^{-1/2} |U_n^P(\beta^{tr}, \boldsymbol{\alpha}) - U_n^P(\tilde{\beta}^{tr}, \tilde{\boldsymbol{\alpha}})| \xrightarrow{p} 0.$$

Thus the consistency of $\hat{\theta}^{Pcatr}$ is proved.

Remark. In this case, the key important element is strong consistency of $\hat{\boldsymbol{\alpha}}$. Given strong consistency of $\hat{\boldsymbol{\alpha}}$, the random function $n^{-1/2} S_n(\eta^{tr}, \hat{\boldsymbol{\alpha}})$ and $n^{-1/2} U_n^L(\eta^{tr}, \hat{\boldsymbol{\alpha}})$ converge to deterministic function. Moreover, to prove strong consistency of $\hat{\theta}^{Pcatr}$, we apply arguments from Peng and Fine (2006) given $\boldsymbol{\alpha}_0$ and use strong consistency of $\hat{\boldsymbol{\alpha}}$. Then we can approximate $n^{-1/2} S_n(\eta^{tr}, \boldsymbol{\alpha})$ and $n^{-1/2} U_n^L(\eta^{tr}, \boldsymbol{\alpha})$ to nonrandom function with respect to $\boldsymbol{\alpha}$.

(c) Proof of Theorem 6.2

Let $\mathbf{G}_n(\boldsymbol{\alpha})$ be given by

$$\mathbf{G}_n(\boldsymbol{\alpha}) = n^{-1/2} \sum_{i=1}^{n} \mathbf{H}_i [Z_i - \frac{\exp(\boldsymbol{\alpha}^T \mathbf{H}_i)}{1 + \exp(\boldsymbol{\alpha}^T \mathbf{H}_i)}].$$

Let $\Psi_i(\boldsymbol{\alpha}) = \mathbf{H}_i [Z_i - \dfrac{\exp(\boldsymbol{\alpha}^T \mathbf{H}_i)}{1 + \exp(\boldsymbol{\alpha}^T \mathbf{H}_i)}]$. Then $\mathbf{G}_n(\boldsymbol{\alpha}) = n^{-1/2} \sum_{i=1}^{n} \Psi_i(\boldsymbol{\alpha})$. By martingale central limit theorem (Theorem 5.3.5 in Fleming and Harrington, 2005, pp.227-228) and U-statistic theory,

$$\begin{aligned}
S_n(\eta_0^{tr}, \boldsymbol{\alpha}_0) &= n^{-1/2} \sum_{i=1}^{n} \int_{-\infty}^{\infty} \{Z_i - \bar{z}^{(1)}(t)\} dM_{1i}(u; \eta_0^{tr}, \boldsymbol{\alpha}_0) + o_p(1) \\
U_n^L(\beta_0^{tr}, \boldsymbol{\alpha}_0) &= n^{-1/2} \sum_{i=1}^{n} \int_{-\infty}^{\infty} \{Z_i - \bar{z}^{(2)}(t)\} dM_{2i}(u; \beta_0^{tr}, \boldsymbol{\alpha}_0) + o_p(1) \\
U_n^P(\beta_0^{tr}, \boldsymbol{\alpha}_0) &= n^{-1/2} \sum_{i=1}^{n} 2h_1(Z_i, \beta_0^{tr}, \boldsymbol{\alpha}_0) + o_p(1) \\
\mathbf{G}_n(\boldsymbol{\alpha}_0) &= n^{-1/2} \sum_{i=1}^{n} \Psi_i(\boldsymbol{\alpha}_0),
\end{aligned}$$

where $h_1(z, \beta, \boldsymbol{\alpha}_0) = E\{h(z, Z_2, \mathbf{V}_1, \mathbf{V}_2, \beta, \boldsymbol{\alpha}_0)\}$. Let $\boldsymbol{\tau} = (\eta^{tr}, \theta^{tr}, \theta^{tr})^T$ and $\mathbf{U}_n(\boldsymbol{\tau}, \boldsymbol{\alpha}_0) = [S_n^T(\eta^{tr}, \boldsymbol{\alpha}_0), \{U_n^L(\beta^{tr}, \boldsymbol{\alpha}_0)\}^T, \ \{U_n^P(\beta^{tr}, \boldsymbol{\alpha}_0)\}^T]^T$. By standard asymptotic theory of maximum likelihood estimator and by Cramér-Wold theorem,

$$\begin{pmatrix} \mathbf{G}_n(\boldsymbol{\alpha}_0) \\ \mathbf{U}_n(\boldsymbol{\tau_0}, \boldsymbol{\alpha}_0) \end{pmatrix} \xrightarrow{d} N\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, E \begin{bmatrix} \mathbf{v}_1\mathbf{v}_1^T & \mathbf{v}_1\mathbf{v}_2^T \\ \mathbf{v}_2\mathbf{v}_1^T & \mathbf{v}_2\mathbf{v}_2^T \end{bmatrix} \right),$$

where

$$\mathbf{v}_1 = \Psi_1(\boldsymbol{\alpha}_0) \quad \mathbf{v}_2 = (v_{21}, v_{22}, v_{23})^T$$

$$v_{21} = \int_{-\infty}^{\infty} \{Z_i - \bar{z}^{(1)}(u)\} dM_{1i}(u; \eta_0^{tr}, \boldsymbol{\alpha}_0) \quad v_{22} = \int_{-\infty}^{\infty} \{Z_i - \bar{z}^{(2)}(u)\} dM_{2i}(u; \beta_0^{tr}, \boldsymbol{\alpha}_0)$$

$$v_{23} = 2h_1(Z_1, \beta_0^{tr}, \boldsymbol{\alpha}_0).$$

Then we have following lemma.

Lemma A2. If k_n converges to 0 in probability,

$$\sup_{||\boldsymbol{\gamma}-\boldsymbol{\gamma}_0|| \le k_n} \frac{||\mathbf{Q}_n(\boldsymbol{\gamma}) - \mathbf{Q}_n(\boldsymbol{\gamma}_0) - \boldsymbol{\Lambda}_0 n^{1/2}(\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)||}{1 + n^{1/2}||\boldsymbol{\gamma} - \boldsymbol{\gamma}_0||} = o_p(1)$$

where

$$\boldsymbol{\Lambda}_0 = \begin{pmatrix} \mathbf{L}_1 & 0 & 0 & 0 \\ \mathbf{L}_2 & E_1 & 0 & 0 \\ \mathbf{L}_3 & E_2 & E_3 & 0 \\ \mathbf{L}_4 & E_4 & 0 & E_5 \end{pmatrix}$$

$$\mathbf{L}_1 = E\left[\frac{\partial \Psi_1(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}\bigg|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}\right]$$

$$\mathbf{L}_2 = \int_{-\infty}^{\infty} E\left[\frac{\partial}{\partial \boldsymbol{\alpha}}\{Z_1 - \bar{z}^{(1*)}(t;\eta^{tr},\boldsymbol{\alpha})\}dN_{11}(t;\eta^{tr},\boldsymbol{\alpha})\right]_{\eta^{tr}=\eta_0^{tr},\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}$$

$$\mathbf{L}_3 = \int_{-\infty}^{\infty} E\left[\frac{\partial}{\partial \boldsymbol{\alpha}}\{Z_1 - \bar{z}^{(2*)}(t;\beta^{tr},\boldsymbol{\alpha})\}dN_{21}(t;\beta^{tr},\boldsymbol{\alpha})\right]_{\beta^{tr}=\beta_0^{tr},\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}$$

$$\mathbf{L}_4 = \frac{\partial \lambda(\beta^{tr},\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}\bigg|_{\beta^{tr}=\beta_0^{tr},\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}$$

$$E_1 = \int_{-\infty}^{\infty} E\left[w_1(\boldsymbol{\alpha}_0)I\{\tilde{D}_1^*(\eta_0)\geq t\}\{Z_1 - \bar{z}^{(1)}(t)\}\frac{\lambda_{10}'(t)}{\lambda_{10}(t)}f(t)dt\right]$$

$$E_2 = \int_{-\infty}^{\infty} E\left[\frac{\partial}{\partial \eta^{tr}}\{Z_1 - \bar{z}^{(2*)}(t;\beta^{tr},\boldsymbol{\alpha})\}dN_{21}(t;\beta^{tr},\boldsymbol{\alpha})\right]_{\beta^{tr}=\beta_0^{tr},\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}$$

$$E_3 = \int_{-\infty}^{\infty} E\left[\frac{\partial}{\partial \theta^{tr}}\{Z_1 - \bar{z}^{(2*)}(t;\beta^{tr},\boldsymbol{\alpha})\}dN_{21}(t;\beta^{tr},\boldsymbol{\alpha})\right]_{\beta^{tr}=\beta_0^{tr},\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}$$

$$E_4 = \frac{\partial \lambda(\beta^{tr},\boldsymbol{\alpha})}{\partial \eta^{tr}}\bigg|_{\beta^{tr}=\beta_0^{tr},\boldsymbol{\alpha}=\boldsymbol{\alpha}_0} \qquad E_5 = \frac{\partial \lambda(\beta^{tr},\boldsymbol{\alpha})}{\partial \theta^{tr}}\bigg|_{\beta^{tr}=\beta_0^{tr},\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}$$

and

$$\bar{z}^{(1*)}(t;\beta^{tr},\boldsymbol{\alpha}) = \lim_{n\to\infty}\frac{\sum_{j=1}^n I\{\tilde{D}_j^*(\eta^{tr})\geq t\}w_j(\boldsymbol{\alpha})Z_j}{\sum_{j=1}^n I\{\tilde{D}_j^*(\eta^{tr})\geq t\}w_j(\boldsymbol{\alpha})}$$

$$\bar{z}^{(2*)}(t;\beta^{tr},\boldsymbol{\alpha}) = \lim_{n\to\infty}\frac{\sum_{j=1}^n I\{\tilde{X}_j^*(\beta^{tr})\geq t\}w_j(\boldsymbol{\alpha})Z_j}{\sum_{j=1}^n I\{\tilde{X}_j^*(\beta^{tr})\geq t\}w_j(\boldsymbol{\alpha})}$$

If k_n converges to 0 almost surely,

$$\sup_{||\boldsymbol{\gamma}-\boldsymbol{\gamma}_0||\leq k_n}\frac{||\mathbf{Q}_n(\boldsymbol{\gamma}) - \mathbf{Q}_n(\boldsymbol{\gamma}_0) - \boldsymbol{\Lambda}_0 n^{1/2}(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)||}{1+n^{1/2}||\boldsymbol{\gamma}-\boldsymbol{\gamma}_0||} = o(1)$$

Proof. We will follow the approach of Ghosh (2000, Chapter 6). Let

$$\begin{pmatrix}\mathbf{G}_n(\boldsymbol{\alpha})\\ S_n(\eta^{tr},\boldsymbol{\alpha})\\ U_n^L(\beta^{tr},\boldsymbol{\alpha})\end{pmatrix} = \begin{pmatrix}\mathbf{G}_n(\boldsymbol{\alpha}_0)\\ S_n(\eta_0^{tr},\boldsymbol{\alpha}_0)\\ U_n^L(\beta_0^{tr},\boldsymbol{\alpha}_0)\end{pmatrix} + \begin{pmatrix}\mathbf{G}_n(\boldsymbol{\alpha}) - \mathbf{G}_n(\boldsymbol{\alpha}_0)\\ S_n(\eta^{tr},\boldsymbol{\alpha}) - S_n(\eta_0^{tr},\boldsymbol{\alpha}_0)\\ U_n^L(\beta^{tr},\boldsymbol{\alpha}) - U_n^L(\beta_0^{tr},\boldsymbol{\alpha}_0)\end{pmatrix}$$

Clearly,

$$\mathbf{G}_n(\boldsymbol{\alpha}) - \mathbf{G}_n(\boldsymbol{\alpha}_0) = \dot{\mathbf{G}}_n(\boldsymbol{\alpha}_0)(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0) + o_p(n^{1/2}||\boldsymbol{\alpha} - \boldsymbol{\alpha}_0||)$$

where $\dot{\mathbf{G}}_n(\boldsymbol{\alpha}_0) = [\partial \mathbf{G}_n / \partial \boldsymbol{\alpha}]_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}$. Let

$$\begin{aligned} \boldsymbol{\gamma}^{sub_1} &= (\boldsymbol{\alpha}^T, \eta^{tr})^T \quad \boldsymbol{\gamma}^{sub_2} = (\boldsymbol{\alpha}^T, \eta^{tr}, \theta^{tr})^T \\ \boldsymbol{\gamma}_0^{sub_1} &= (\boldsymbol{\alpha}_0^T, \eta_0^{tr})^T \quad \boldsymbol{\gamma}_0^{sub_2} = (\boldsymbol{\alpha}_0^T, \eta_0^{tr}, \theta_0^{tr})^T. \end{aligned}$$

By Ying (1993),

$$\begin{aligned} S_n(\eta^{tr}, \boldsymbol{\alpha}) - S_n(\eta_0^{tr}, \boldsymbol{\alpha}_0) &= n^{1/2}\{L_2(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0) + E_1(\eta^{tr} - \eta_0^{tr})\} + o_p(1 + n^{1/2}||\boldsymbol{\gamma}^{sub_1} - \boldsymbol{\gamma}_0^{sub_1}||) \\ U_n^L(\beta^{tr}, \boldsymbol{\alpha}) - U_n^L(\beta_0^{tr}, \boldsymbol{\alpha}_0) &= n^{1/2}\{L_3(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0) + E_2(\eta^{tr} - \eta_0^{tr}) + E_3(\theta^{tr} - \theta_0^{tr})\} \\ &\quad + o_p(1 + n^{1/2}||\boldsymbol{\gamma}^{sub_2} - \boldsymbol{\gamma}_0^{sub_2}||). \end{aligned}$$

By Lemma 2 of Honoré and Powell (1994),

$$\sup_{\beta^{tr} \in \mathcal{N}_1, \boldsymbol{\alpha} \in \mathcal{B}} \frac{|U_n^P(\beta^{tr}, \boldsymbol{\alpha}) - U_n^P(\beta_0^{tr}, \boldsymbol{\alpha}_0) - n^{1/2}\lambda(\beta^{tr}, \boldsymbol{\alpha})|}{1 + n^{1/2}|\lambda(\beta^{tr}, \boldsymbol{\alpha})|} = o_p(1),$$

where $\mathcal{N}_1$ is a neighborhood of β_0^{tr}. By Taylor series expansion of $\lambda(\beta^{tr}, \boldsymbol{\alpha})$ at β_0^{tr} and $\boldsymbol{\alpha}_0$,

$$\begin{aligned} \lambda(\beta^{tr}, \boldsymbol{\alpha}) = \lambda(\beta_0^{tr}, \boldsymbol{\alpha}_0) + &\begin{pmatrix} \left.\dfrac{\partial \lambda(\beta_0^{tr}, \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}\right|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0} \\ \left.\dfrac{\partial \lambda(\beta^{tr}, \boldsymbol{\alpha}_0)}{\partial \eta^{tr}}\right|_{\beta^{tr}=\beta_0^{tr}} \\ \left.\dfrac{\partial \lambda(\beta^{tr}, \boldsymbol{\alpha}_0)}{\partial \theta^{tr}}\right|_{\beta^{tr}=\beta_0^{tr}} \end{pmatrix} \begin{pmatrix} \boldsymbol{\alpha} - \boldsymbol{\alpha}_0 & \eta^{tr} - \eta_0^{tr} & \theta^{tr} - \theta_0^{tr} \end{pmatrix} \\ &+ o(||\boldsymbol{\gamma}^{sub_2} - \boldsymbol{\gamma}_0^{sub_2}||). \end{aligned}$$

Extending arguments in the Appendix of Peng and Fine (2006),

$$\begin{aligned} U_n^P(\beta^{tr}, \boldsymbol{\alpha}^{tr}) = U_n^P(\beta_0^{tr}, \boldsymbol{\alpha}_0^{tr}) &+ n^{1/2}\{\mathbf{L}_4(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0) + E_4(\eta^{tr} - \eta_0^{tr}) + E_5(\theta^{tr} - \theta_0^{tr})\} \\ &+ o_p(1 + n^{1/2}||\boldsymbol{\gamma}^{sub_2} - \boldsymbol{\gamma}_0^{sub_2}||). \end{aligned}$$

Hence,

$$\sup_{||\gamma-\gamma_0||\le k_n} \frac{||\mathbf{Q}_n(\boldsymbol{\gamma}) - \mathbf{Q}_n(\boldsymbol{\gamma}_0) - \boldsymbol{\Lambda}_0 n^{1/2}(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)||}{1+n^{1/2}||\boldsymbol{\gamma}-\boldsymbol{\gamma}_0||} = o_p(1),$$

when k_n converges in probability to zero. The second result of Lemma 2 easily follows from the result that the sequence of random variable converges in probability if and only if each subsequence of the sequence of random variables contains further subsequence which converges almost surely. $\square$

By using Lemma A2 above,

$$\mathbf{Q}_n(\boldsymbol{\gamma}) = \mathbf{Q}_n(\boldsymbol{\gamma}_0) + n^{1/2}\boldsymbol{\Lambda}_0(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0) + o_p(1+n^{1/2}||\boldsymbol{\gamma}-\boldsymbol{\gamma}_0||),$$

for $\boldsymbol{\gamma} \in \mathcal{V}$, where $\mathcal{V}$ is any neighborhood of $\boldsymbol{\gamma}_0$. Then

$$\mathbf{Q}_n(\hat{\boldsymbol{\gamma}}) = \mathbf{Q}_n(\boldsymbol{\gamma}_0) + n^{1/2}\boldsymbol{\Lambda}_0(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0) + o_p(1),$$

and by consistency of $\hat{\boldsymbol{\gamma}}$ and Lemma A2, we get

$$n^{1/2}(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0) = -\boldsymbol{\Lambda}_0^{-1}\mathbf{Q}_n(\boldsymbol{\gamma}_0) + o_p(1)$$

Then by Slusky's theorem,

$$n^{1/2}(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0) \xrightarrow{d} N(0, \boldsymbol{\Lambda}_0^{-1}\boldsymbol{\Sigma}_0\boldsymbol{\Lambda}_0^{-1}),$$

where $\boldsymbol{\Sigma}_0$ is

$$\boldsymbol{\Sigma}_0 = E\begin{bmatrix} \mathbf{v}_1\mathbf{v}_1^T & \mathbf{v}_1\mathbf{v}_2^T \\ \mathbf{v}_2\mathbf{v}_1^T & \mathbf{v}_2\mathbf{v}_2^T \end{bmatrix}.$$

(d) Proof of Theorem 6.3

To justify resampling approach in Parzen et al. (1994), two conditions A(1·1) and A(1·2) in Parzen et al. (1994) should be verified. The condition A(1·1) follows from Lemma 2 directly. The condition A(1·2) implies that the root of estimating equation should be unique. This condition is also easily satisfied by the assumption. Let $\boldsymbol{\gamma}^*$ be solution of

$$\boldsymbol{Q}_n(\boldsymbol{\gamma}) = -n^{-1/2}\sum_{i=1}^{n}\mathbf{J}_i A_i,$$

where $\mathbf{J}_i$ are natural sample estimates of $\mathbf{v} = (\mathbf{v}_1^T, \mathbf{v}_2^T)^T$ and A_i are standard normal samples. From the previous section, we showed that

$$\mathbf{Q}_n(\boldsymbol{\gamma}) = \mathbf{Q}_n(\boldsymbol{\gamma}_0) + n^{1/2}\boldsymbol{\Lambda}_0(\boldsymbol{\gamma} - \boldsymbol{\gamma}_0) + o_p(1 + n^{1/2}||\boldsymbol{\gamma} - \boldsymbol{\gamma}_0||).$$

Then

$$\begin{aligned}\mathbf{Q}_n(\boldsymbol{\gamma}^*) &= \mathbf{Q}_n(\hat{\boldsymbol{\gamma}}) + n^{1/2}\boldsymbol{\Lambda}_0(\boldsymbol{\gamma}^* - \hat{\boldsymbol{\gamma}}) + o_p(1)\\ n^{1/2}(\boldsymbol{\gamma}^* - \hat{\boldsymbol{\gamma}}) &= -\boldsymbol{\Lambda}_0^{-1}n^{-1/2}\sum_{i=1}^{n}\mathbf{J}_i A_i + o_p(1).\end{aligned}$$

Since the observed data are independent and identically distributed, given observed data, the asymptotic distribution of $n^{1/2}(\boldsymbol{\gamma}^* - \hat{\boldsymbol{\gamma}})$ is normal distribution with zero mean vector and covariance matrix $\boldsymbol{\Lambda}_0^{-1}\boldsymbol{\Sigma}_0\boldsymbol{\Lambda}_0^{-1}$. Hence the conditional distribution of $n^{1/2}(\boldsymbol{\gamma}^* - \hat{\boldsymbol{\gamma}})$ is asymptotically equal to the unconditional distribution of $n^{1/2}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)$.

2 Goodness of Fit (Chapter 4)

The goodness of fit structure proposed by Lin et al. (1996) is follows. We use notations in the Chapter 4. Let $\hat{\boldsymbol{\eta}}$ be the estimator for the time to the dependent censoring, $\hat{\boldsymbol{\alpha}}^L$ be Lin et al. (1996) estimator and $\hat{\boldsymbol{\alpha}}^P$ be Peng and Fine (2006) estimator. Define $N_{1i}(t;\boldsymbol{\eta}) = \Delta_i I\{\tilde{D}_i^*(\boldsymbol{\eta}) \le t\}$ and $N_{2i}(t;\boldsymbol{\alpha}) = \tilde{\delta}_i^*(\boldsymbol{\alpha}) I\{\tilde{X}_i^*(\boldsymbol{\alpha}) \le t\}$. Let

$$\begin{aligned}\tilde{D}_i^*(\boldsymbol{\eta}) &= \tilde{D}_i - \mathbf{Z}_i^T\boldsymbol{\eta}\\ g(\boldsymbol{\beta}) &= \max_{1\le i\le n}\{0, \mathbf{Z}_i^T(\boldsymbol{\theta} - \boldsymbol{\eta})\}\\ \tilde{X}_i^*(\boldsymbol{\alpha}) &= (X_i - \mathbf{Z}_i^T\boldsymbol{\theta}) \wedge \{D_i - \mathbf{Z}_i^T\boldsymbol{\eta} - g(\boldsymbol{\beta})\} \wedge \{C_i - \mathbf{Z}_i^T\boldsymbol{\eta} - g(\boldsymbol{\beta})\}\\ \tilde{\delta}_i^*(\boldsymbol{\alpha}) &= I[(X_i - \mathbf{Z}_i^T\boldsymbol{\theta}) \le \{D_i - \mathbf{Z}_i^T\boldsymbol{\eta} - g(\boldsymbol{\beta})\} \wedge \{C_i - \mathbf{Z}_i^T\boldsymbol{\eta} - g(\boldsymbol{\beta})\}].\end{aligned}$$

Then

$$M_{1i}(t;\boldsymbol{\eta}_0) = N_{1i}(t;\boldsymbol{\eta}_0) - \int_{-\infty}^{t} I\{\tilde{D}_i^*(\boldsymbol{\eta}_0) \geq u\}\lambda_0(u)du$$
$$M_{2i}(t;\boldsymbol{\alpha}_0) = N_{2i}(t;\boldsymbol{\alpha}_0) - \int_{-\infty}^{t} I\{\tilde{X}_i^*(\boldsymbol{\alpha}_0) \geq u\}h_0(u)du$$

are martingales where $\lambda_0(u)$ and $h_0(u)$ are baseline hazard functions for the dependent censoring and the event of interest, respectively. We can define $\hat{M}_{1i}(t;\hat{\boldsymbol{\eta}})$ and $\hat{M}_{2i}(t;\hat{\boldsymbol{\alpha}})$,where

$$\hat{M}_{1i}(t;\hat{\boldsymbol{\eta}}) = N_{1i}(t;\hat{\boldsymbol{\eta}}) - \int_{-\infty}^{t} I\{\tilde{D}_i^*(\hat{\boldsymbol{\eta}}) \geq u\}d\hat{\Lambda}_0(u)$$
$$\hat{M}_{2i}(t;\hat{\boldsymbol{\eta}}) = N_{2i}(t;\hat{\boldsymbol{\alpha}}) - \int_{-\infty}^{t} I\{\tilde{X}_i^*(\hat{\boldsymbol{\alpha}}) \geq u\}d\hat{H}_0(u)$$
$$\hat{\Lambda}_0(t) = \int_{-\infty}^{t} \frac{\sum_{l=1}^{n} dN_{1l}(\hat{\boldsymbol{\eta}};u)}{\sum_{j=1}^{n} I\{\tilde{D}_j^*(\hat{\boldsymbol{\eta}}) \geq u\}} \quad \hat{H}_0(t) = \int_{-\infty}^{t} \frac{\sum_{l=1}^{n} dN_{2l}(\hat{\boldsymbol{\alpha}};u)}{\sum_{j=1}^{n} I\{\tilde{X}_j^*(\hat{\boldsymbol{\alpha}}) \geq u\}}$$

Then observed processes are defined as

$$\mathbf{S}_n(s;\boldsymbol{\eta}) = n^{-1/2}\sum_{i=1}^{n}\mathbf{Z}_i\hat{M}_{1i}(s;\boldsymbol{\eta}) \quad \mathbf{U}_n(t;\boldsymbol{\alpha}) = n^{-1/2}\sum_{i=1}^{n} Z_i\hat{M}_{2i}(t;\boldsymbol{\alpha})$$

Note that by stochastic integral,

$$\begin{aligned}\int_{-\infty}^{t} I\{\tilde{D}_i^*(\hat{\boldsymbol{\eta}}) \geq u\}d\hat{\Lambda}_0(u) &= \sum_{\tilde{D}_l^*(\hat{\eta})\leq t} \frac{I\{\tilde{D}_i^*(\hat{\boldsymbol{\eta}}) \geq \tilde{D}_l^*(\hat{\boldsymbol{\eta}})\}\Delta_l}{\sum_{j=1}^{n} I\{\tilde{D}_j^*(\hat{\boldsymbol{\eta}}) \geq \tilde{D}_l^*(\hat{\boldsymbol{\eta}})\}} \\ &= \sum_{l=1}^{n} \frac{\Delta_l I\{\tilde{D}_l^*(\hat{\boldsymbol{\eta}}) \leq t\} I\{\tilde{D}_i^*(\hat{\boldsymbol{\eta}}) \geq \tilde{D}_l^*(\hat{\boldsymbol{\eta}})\}}{\sum_{j=1}^{n} I\{\tilde{D}_j^*(\hat{\boldsymbol{\eta}}) \geq \tilde{D}_l^*(\hat{\boldsymbol{\eta}})\}}\end{aligned}$$

Hence

$$\hat{M}_{1i}(t;\hat{\boldsymbol{\eta}}) = \Delta_i I\{\tilde{D}_i^*(\hat{\boldsymbol{\eta}}) \leq t\} - \sum_{l=1}^{n} \frac{\Delta_l I\{\tilde{D}_l^*(\hat{\boldsymbol{\eta}}) \leq t\} I\{\tilde{D}_i^*(\hat{\boldsymbol{\eta}}) \geq \tilde{D}_l^*(\hat{\boldsymbol{\eta}})\}}{\sum_{j=1}^{n} I\{\tilde{D}_j^*(\hat{\boldsymbol{\eta}}) \geq \tilde{D}_l^*(\hat{\boldsymbol{\eta}})\}}$$

Similarly,

$$\int_{-\infty}^{t} I\{\tilde{X}_i^*(\hat{\boldsymbol{\alpha}}) \geq u\} d\hat{H}_0(u) = \sum_{\tilde{D}_l^*(\hat{\boldsymbol{\eta}}) \leq t} \frac{I\{\tilde{X}_i^*(\hat{\boldsymbol{\eta}}) \geq \tilde{X}_l^*(\hat{\boldsymbol{\eta}})\}\tilde{\delta}_l^*(\hat{\boldsymbol{\alpha}})}{\sum_{j=1}^n I\{\tilde{X}_j^*(\hat{\boldsymbol{\eta}}) \geq \tilde{X}_l^*(\hat{\boldsymbol{\eta}})\}}$$

$$= \sum_{l=1}^{n} \frac{\tilde{\delta}_l^*(\hat{\boldsymbol{\alpha}}) I\{\tilde{X}_l^*(\hat{\boldsymbol{\alpha}}) \leq t\} I\{\tilde{X}_i^*(\hat{\boldsymbol{\alpha}}) \geq \tilde{X}_l^*(\hat{\boldsymbol{\alpha}})\}}{\sum_{j=1}^n I\{\tilde{X}_j^*(\hat{\boldsymbol{\alpha}}) \geq \tilde{X}_l^*(\hat{\boldsymbol{\alpha}})\}}$$

Hence

$$\hat{M}_{2i}(t;\hat{\boldsymbol{\eta}}) = \tilde{\delta}_i^*(\hat{\boldsymbol{\alpha}}) I\{\tilde{X}_i^*(\hat{\boldsymbol{\alpha}}) \leq t\} - \sum_{l=1}^{n} \frac{\tilde{\delta}_l^*(\hat{\boldsymbol{\alpha}}) I\{\tilde{X}_l^*(\hat{\boldsymbol{\alpha}}) \leq t\} I\{\tilde{X}_i^*(\hat{\boldsymbol{\alpha}}) \geq \tilde{X}_l^*(\hat{\boldsymbol{\alpha}})\}}{\sum_{j=1}^n I\{\tilde{X}_j^*(\hat{\boldsymbol{\alpha}}) \geq \tilde{X}_l^*(\hat{\boldsymbol{\alpha}})\}}$$

Then similar to Lin et al. (1996) and Peng and Fine (2006), we can substitute $\hat{\boldsymbol{\eta}}$ on $S_n(s;\boldsymbol{\eta})$, $\hat{\boldsymbol{\alpha}}^L$ and $\hat{\boldsymbol{\alpha}}^P$ on $U_n(t;\boldsymbol{\alpha})$. By arguing as Lin et al. (1996) and Peng and Fine (2006), we can construct $[\hat{S}_n^T(s), \{\hat{U}_n^L(t)\}^T, \{\hat{U}_n^P(v)\}^T]^T$, where

$$\hat{\mathbf{S}}_n(s) = n^{-1/2} \sum_{i=1}^{n} \int_{-\infty}^{s} \left[\mathbf{Z}_i - \frac{\sum_{j=1}^n I\{\tilde{D}_j^*(\hat{\boldsymbol{\eta}}) \geq w\}\mathbf{Z}_j}{\sum_{j=1}^n I\{\tilde{D}_j^*(\hat{\boldsymbol{\eta}}) \geq w\}} \right] d\hat{M}_{1i}(w;\hat{\boldsymbol{\eta}}) Q_i + \mathbf{S}_n(s;\hat{\boldsymbol{\eta}}^*) - \mathbf{S}_n(s;\hat{\boldsymbol{\eta}})$$

$$\hat{\mathbf{U}}_n^L(t) = n^{-1/2} \sum_{i=1}^{n} \int_{-\infty}^{t} \left[\mathbf{Z}_i - \frac{\sum_{j=1}^n I\{\tilde{X}_j^*(\hat{\boldsymbol{\alpha}}^L) \geq w\}\mathbf{Z}_j}{\sum_{j=1}^n I\{\tilde{X}_j^*(\hat{\boldsymbol{\alpha}}^L) \geq w\}} \right] d\hat{M}_{2i}(w;\hat{\boldsymbol{\alpha}}^L) Q_i + \mathbf{U}_n(t;\hat{\boldsymbol{\alpha}}^{L*}) - \mathbf{U}_n(t;\hat{\boldsymbol{\alpha}}^L)$$

$$\hat{\mathbf{U}}_n^P(v) = n^{-1/2} \sum_{i=1}^{n} \int_{-\infty}^{v} \left[\mathbf{Z}_i - \frac{\sum_{j=1}^n I\{\tilde{X}_j^*(\hat{\boldsymbol{\alpha}}^P) \geq w\}\mathbf{Z}_j}{\sum_{j=1}^n I\{\tilde{X}_j^*(\hat{\boldsymbol{\alpha}}^P) \geq w\}} \right] d\hat{M}_{2i}(w;\hat{\boldsymbol{\alpha}}^P) Q_i + \mathbf{U}_n(v;\hat{\boldsymbol{\alpha}}^{P*}) - \mathbf{U}_n(v;\hat{\boldsymbol{\alpha}}^P)$$

For view of calculation,

$$n^{-1/2}\sum_{i=1}^{n}\mathbf{Z}_i\hat{M}_{1i}(s;\boldsymbol{\eta}) = n^{-1/2}\sum_{i=1}^{n}\mathbf{Z}_i\Bigg[N_{1i}(t;\boldsymbol{\eta}) - \sum_{l=1}^{n}\frac{\Delta_l I\{\tilde{D}_l^*(\boldsymbol{\eta}) \le t\} I\{\tilde{D}_i^*(\boldsymbol{\eta}) \ge \tilde{D}_l^*(\boldsymbol{\eta})\}}{\sum_{j=1}^{n} I\{\tilde{D}_j^*(\boldsymbol{\eta}) \ge \tilde{D}_l^*(\boldsymbol{\eta})\}}\Bigg]$$

$$n^{-1/2}\sum_{i=1}^{n}\mathbf{Z}_i\hat{M}_{2i}(s;\boldsymbol{\alpha}) = n^{-1/2}\sum_{i=1}^{n}\mathbf{Z}_i\Bigg[N_{2i}(t;\boldsymbol{\alpha}) - \sum_{l=1}^{n}\frac{\tilde{\delta}_l^*(\boldsymbol{\alpha}) I\{\tilde{X}_l^*(\boldsymbol{\alpha}) \le t\} I\{\tilde{X}_i^*(\boldsymbol{\alpha}) \ge \tilde{X}_l^*(\boldsymbol{\alpha})\}}{\sum_{j=1}^{n} I\{\tilde{X}_j^*(\boldsymbol{\alpha}) \ge \tilde{X}_l^*(\boldsymbol{\alpha})\}}\Bigg]$$

$$\hat{\mathbf{S}}_n(s) = n^{-1/2}\sum_{i=1}^{n}\int_{-\infty}^{s}\Bigg[\mathbf{Z}_i - \frac{\sum_{j=1}^{n} I\{\tilde{D}_j^*(\hat{\boldsymbol{\eta}}) \ge w\}\mathbf{Z}_j}{\sum_{j=1}^{n} I\{\tilde{D}_j^*(\hat{\boldsymbol{\eta}}) \ge w\}}\Bigg] d\hat{M}_{1i}(w;\hat{\boldsymbol{\eta}})Q_i + \mathbf{S}_n(s;\hat{\boldsymbol{\eta}}^*) - \mathbf{S}_n(s;\hat{\boldsymbol{\eta}})$$

$$= n^{-1/2}\sum_{i=1}^{n}\Bigg(\Delta_i I\{\tilde{D}_i^*(\hat{\boldsymbol{\eta}}) \le s\}\Bigg[\mathbf{Z}_i - \frac{\sum_{j=1}^{n} I\{\tilde{D}_j^*(\hat{\boldsymbol{\eta}}) \ge \tilde{D}_i^*(\hat{\boldsymbol{\eta}})\}\mathbf{Z}_j}{\sum_{j=1}^{n} I\{\tilde{D}_j^*(\hat{\boldsymbol{\eta}}) \ge \tilde{D}_i^*(\hat{\boldsymbol{\eta}})\}}\Bigg]$$

$$- \sum_{l=1}^{n}\frac{\Delta_l I\{\tilde{D}_l^*(\hat{\boldsymbol{\eta}}) \le s\} I\{\tilde{D}_i^*(\hat{\boldsymbol{\eta}}) \ge \tilde{D}_l^*(\hat{\boldsymbol{\eta}})\}}{\sum_{j=1}^{n} I\{\tilde{D}_j^*(\hat{\boldsymbol{\eta}}) \ge \tilde{D}_l^*(\hat{\boldsymbol{\eta}})\}}\Bigg[\mathbf{Z}_i - \frac{\sum_{j=1}^{n} I\{\tilde{D}_j^*(\hat{\boldsymbol{\eta}}) \ge \tilde{D}_l^*(\hat{\boldsymbol{\eta}})\}\mathbf{Z}_j}{\sum_{j=1}^{n} I\{\tilde{D}_j^*(\hat{\boldsymbol{\eta}}) \ge \tilde{D}_l^*(\hat{\boldsymbol{\eta}})\}}\Bigg]\Bigg)Q_i$$

$$+ \mathbf{S}_n(s;\hat{\boldsymbol{\eta}}^*) - \mathbf{S}_n(s;\hat{\boldsymbol{\eta}})$$

$$\hat{\mathbf{U}}_n^L(t) = n^{-1/2}\sum_{i=1}^{n}\int_{-\infty}^{t}\Bigg[\mathbf{Z}_i - \frac{\sum_{j=1}^{n} I\{\tilde{X}_i^*(\hat{\boldsymbol{\alpha}}^L) \ge w\}\mathbf{Z}_j}{\sum_{j=1}^{n} I\{\tilde{X}_i^*(\hat{\boldsymbol{\alpha}}^L) \ge w\}}\Bigg] d\hat{M}_{2i}(w;\hat{\boldsymbol{\alpha}})Q_i + \mathbf{U}_n(t;\hat{\boldsymbol{\alpha}}^{L*}) - \mathbf{U}_n(t;\hat{\boldsymbol{\alpha}}^L)$$

$$= n^{-1/2}\sum_{i=1}^{n}\Bigg(\tilde{\delta}_i^*(\hat{\boldsymbol{\alpha}}^L) I\{\tilde{X}_i^*(\hat{\boldsymbol{\alpha}}^L) \le t\}\Bigg[\mathbf{Z}_i - \frac{\sum_{j=1}^{n} I\{\tilde{X}_i^*(\hat{\boldsymbol{\alpha}}^L) \ge \tilde{X}_i^*(\hat{\boldsymbol{\alpha}}^L)\}\mathbf{Z}_j}{\sum_{j=1}^{n} I\{\tilde{X}_i^*(\hat{\boldsymbol{\alpha}}^L) \ge \tilde{X}_i^*(\hat{\boldsymbol{\alpha}}^L)\}}\Bigg]$$

$$- \sum_{l=1}^{n}\frac{\tilde{\delta}_l^*(\hat{\boldsymbol{\alpha}}^L) I\{\tilde{X}_l^*(\hat{\boldsymbol{\alpha}}^L) \le t\} I\{\tilde{X}_i^*(\hat{\boldsymbol{\alpha}}^L) \ge \tilde{X}_l^*(\hat{\boldsymbol{\alpha}}^L)\}}{\sum_{j=1}^{n} I\{\tilde{X}_j^*(\hat{\boldsymbol{\alpha}}^L) \ge \tilde{X}_l^*(\hat{\boldsymbol{\alpha}}^L)\}}\Bigg[\mathbf{Z}_i - \frac{\sum_{j=1}^{n} I\{\tilde{X}_i^*(\hat{\boldsymbol{\alpha}}^L) \ge \tilde{X}_l^*(\hat{\boldsymbol{\alpha}}^L)\}\mathbf{Z}_j}{\sum_{j=1}^{n} I\{\tilde{X}_i^*(\hat{\boldsymbol{\alpha}}^L) \ge \tilde{X}_l^*(\hat{\boldsymbol{\alpha}}^L)\}}\Bigg]\Bigg)Q_i$$

$$+ \mathbf{U}_n(t;\hat{\boldsymbol{\alpha}}^{L*}) - \mathbf{U}_n(t;\hat{\boldsymbol{\alpha}}^L)$$

$$\hat{\mathbf{U}}_n^P(t) = n^{-1/2}\sum_{i=1}^{n}\int_{-\infty}^{t}\Bigg[\mathbf{Z}_i - \frac{\sum_{j=1}^{n} I\{\tilde{X}_i^*(\hat{\boldsymbol{\alpha}}^P) \ge w\}\mathbf{Z}_j}{\sum_{j=1}^{n} I\{\tilde{X}_i^*(\hat{\boldsymbol{\alpha}}^P) \ge w\}}\Bigg] d\hat{M}_{2i}(w;\hat{\boldsymbol{\alpha}}^P)Q_i + \mathbf{U}_n(t;\hat{\boldsymbol{\alpha}}^{P*}) - \mathbf{U}_n(t;\hat{\boldsymbol{\alpha}}^P$$

$$= n^{-1/2}\sum_{i=1}^{n}\Bigg(\tilde{\delta}_i^*(\hat{\boldsymbol{\alpha}}^P) I\{\tilde{X}_i^*(\hat{\boldsymbol{\alpha}}^P) \le t\}\Bigg[\mathbf{Z}_i - \frac{\sum_{j=1}^{n} I\{\tilde{X}_i^*(\hat{\boldsymbol{\alpha}}^P) \ge \tilde{X}_i^*(\hat{\boldsymbol{\alpha}}^P)\}\mathbf{Z}_j}{\sum_{j=1}^{n} I\{\tilde{X}_i^*(\hat{\boldsymbol{\alpha}}^L) \ge \tilde{X}_i^*(\hat{\boldsymbol{\alpha}}^L)\}}\Bigg]$$

$$- \sum_{l=1}^{n}\frac{\tilde{\delta}_l^*(\hat{\boldsymbol{\alpha}}^P) I\{\tilde{X}_l^*(\hat{\boldsymbol{\alpha}}^P) \le t\} I\{\tilde{X}_i^*(\hat{\boldsymbol{\alpha}}^P) \ge \tilde{X}_l^*(\hat{\boldsymbol{\alpha}}^P)\}}{\sum_{j=1}^{n} I\{\tilde{X}_j^*(\hat{\boldsymbol{\alpha}}^P) \ge \tilde{X}_l^*(\hat{\boldsymbol{\alpha}}^P)\}}\Bigg[\mathbf{Z}_i - \frac{\sum_{j=1}^{n} I\{\tilde{X}_i^*(\hat{\boldsymbol{\alpha}}^P) \ge \tilde{X}_l^*(\hat{\boldsymbol{\alpha}}^P)\}\mathbf{Z}_j}{\sum_{j=1}^{n} I\{\tilde{X}_i^*(\hat{\boldsymbol{\alpha}}^P) \ge \tilde{X}_l^*(\hat{\boldsymbol{\alpha}}^P)\}}\Bigg]\Bigg)Q_i$$

$$+ \mathbf{U}_n(t;\hat{\boldsymbol{\alpha}}^{P*}) - \mathbf{U}_n(t;\hat{\boldsymbol{\alpha}}^P)$$

Bibliography

Albrecht, M. A., Bosch, R. J., Hammer, S. M., Liou, S.-H., Kessler, H., Para, M. F., Eron, J., Valdez, H., Dehlinger, M., Katzenstein, D. A. and The AIDS Clinical Trials Group 364 Study Team. (2001). Nelfinavir, Efavirenz, or Both after the Failure of Nucleoside Treatment of HIV Infection. *The New England Journal of Medicine*, **345**(6), 398–407.

Arbogast, P. G. and Lin, D. Y. (2004). Goodness-of-fit Methods for Matched Case-control Studies. *The Canadian Journal of Statistics*. **32**(4), 373–386.

Andersen, P. K. and Gill, R. D. (1982), Cox's Regression Model for Counting Processes: A Large Sample Study. *The Annals of Statistics*, **10**(4), 1100–1120.

Bickel, P. J., Klaassen, C. A. J., Ritov, Y. and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models.* Johns Hopkins University Press, Baltimore.

Boyd, A. P., Kittelson, J. M. and Gillen, D. L. (2012). Estimation of Treatment Effect Under Non-proportional Hazards and Conditionally Independent Censoring. *Statistics in medicine*, **31**(28), 3504–3515.

Breiman, L., Friedman, J., Stone, C. J. and Olshen, R. A. (1984). *Classification and Regression Trees*. Chapman and Hall/CRC.

Byar, D. P. (1980). The Veterans Administration Study of Chemoprophylaxis for Recurrent Stage I Bladder Tumors: Comparisons of Placebo, Pyridoxine and Topical Thiotepa, In *Bladder Tumors and Other Topics in Urological Oncology*, 363–370, Springer US.

Cai, T., Wei, L. J. and Wilcox, M. (2000). Semiparametric Regression Analysis for Clustered Failure Time Data. *Biometrika*, **87**(4), 867–878.

Cai, T., Cheng, S. C. and Wei, L. J. (2002). Semiparametric Mixed-Effects Models for Clustered Failure Time Data. *Journal of the American Statistical Association*, **97**(458), 514–522.

Chen, K., Jin, Z. and Ying, Z. (2002). Semiparametric Analysis of Transformation Models with Censored Data. *Biometrika*, **89**(3), 659–668.

Cheng, S. C., Wei, L. J. and Ying, Z. (1995). Analysis of Transformation Models with Censored Data. *Biometrika*, **82**(4), 835–845.

Clayton, D. G. (1978). A Model for Association in Bivariate Life Tables and Its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence. *Biometrika*, **65**(1), 141–151.

Copelan, E. A., Biggs, J. C., Thompson, J. M., Crilley, P., Szer, J., Klein, J. P., Kapoor, N., Avalos, B. R., Cunningham, I., Atkinson, K., Downs, K., Harmon, G. S., Daly, M. B., Brodsky, I., Bulova, S. I., and Tutschka, P. J. (1991). Treatment for Acute Myelocytic Leukemia with Allogeneic Bone Marrow Transplantation Following Preparation with Bu/Cy. *Blood* **78**(3), 838–843.

Day, R., Bryant, J. and Lefkopoulou, M. (1997), Adaptation of Bivariate Frailty Models for Prediction, With Application to Biological Markers As Prognostic Indicators. *Biometrika*, **84**(1), 45–56.

Ding, A. A., Shi, G., Wang, W. and Hsieh, J.-J. (2009). Marginal Regression Analysis for Semi-Competing Risks Data Under Dependent Censoring. *Scandinavian Journal of Statistics*, **36**(3), 481–500.

Emoto, S. E. and Matthews, P. C. (1990). A Weibull Model for Dependent Censoring. *The Annals of Statistics*, **18**(4), 1556–1577.

Fan, J. and Li, R. (2001). Variable Selection Via Nonconcave Penalized Likelihood and Its Oracle Properties. *Journal of the American statistical Association*, **96**(456), 1348–1360.

Ferguson, T. S. (1996). *A Course in Large Sample Theory.* Chapman & Hall, London.

Fine, J. P., Ying, Z. and Wei, L. J. (1998). On the Linear Transformation Model for Censored Data. *Biometrika*, **85**(4), 980–986.

Fine, J. P., Jiang, H. and Chappell, R. (2001). On Semi-Competing Risks Data. *Biometrika*, **88**(4), 907–919.

Fleming, T. R. and Harrington, D.P. (2005). *Counting Processes and Survival Analysis.* Second Edition, John Wiley and Sons, New York.

Fygenson, M. and Ritov, Y. (1994). Monotone Estimating Equations for Censored Data. *The Annals of Statistics*, **22**(2), 732–746.

Ghosh, D. (2000) *Nonparametric and Semiparametric Analysis of Recurrent Events in the Presence of Terminal Events and Dependent Censoring*. Ph.D dissertation, University of Washington.

Ghosh, D. and Lin, D. Y. (2003). Semiparametric Analysis of Recurrent Events Data in the Presence of Dependent Censoring. *Biometrics*, **59**(4), 877–885.

Ghosh, D. (2006). Semiparametric Inferences for Association With Semi-Competing Risks Data. *Statistics in Medicine*, **25**(12), 2059–2070.

Ghosh, D. (2010). Semiparametric Analysis of Recurrent events: Artificial Censoring, Truncation, Pairwise Estimation and Inference. *Lifetime Data Analysis*, **16**(4), 509–524.

Ghosh, D. (2012). A Causal Framework for Surrogate Endpoints With Semicompeting Risks Data. *Statistics and Probability Letters*. **82**(11), 1898–1902.

Hansen, L. P. (1982). Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*, **50**(4), 1029–1054.

Hoeffding, W. (1948). A Class of Statistics With Asymptotically Normal Distribution. *The Annals of Mathematical Statistics* **19**(3), 293–325.

Honoré, B. E. and Powell, J. L. (1994). Pairwise Difference Estimators of Censored and Truncated Regression Models, *Journal of Econometrics*. **64**(1-2), 241–278.

Hsieh, J.-J., Ding, A. A. and Wang, W. (2011). Regression Analysis for Recurrent Events Data Under Dependent Censoring. *Biometrics*, **67**(3), 719–729.

Jin, Z., Ying, Z. and Wei, L. J. (2001), A Simple Resampling Method by Perturbing the Minimand. *Biometrika* **88**(2), 381-390.

Jin, Z., Lin, D. Y., Wei, L. J. and Ying, Z. (2003). Rank-based Inference for the Accelerated Failure Time Model. *Biometrika*,**90**(2), 341–353.

Joffe, M. (2001). Administrative and Artificial Censoring in Censored Regression Models. *Statistics in Medicine*, **20**(15), 2287–2304.

Johnson, B. A. (2008). Variable Selection in Semiparametric Linear Regression with Censored Data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **70**(2), 351-370.

Johnson, B. A., Lin, D. Y. and Zeng, D. (2008). Penalized Estimating Functions and Variable Selection in Semiparametric Regression Models. *Journal of the American Statistical Association*, **103**(482), 672–680.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. Second Edition, John Wiley and Sons, New Jersey.

Klein, J. P. and Moeschberger, M. L. (2003) *Survival Analysis: Techniques for Censored and Truncated Data*. Second Edition, Springer-Verlag, New York.

La Cruz, W., Martínez, J. M. and Raydan, M. (2006). Spectral Residual Method without Gradient Information for Solving Large-scale Nonlinear Systems of Equations. *Mathematics of Computation*,**75**(255), 1429–1448.

Lai, T. L. and Ying, Z. (1991). Large Sample Theory of a Modified Buckley-James Estimator for Regression Analysis With Censored Data. *The Annals of Statistics*, **19**(3)1370–1402.

León, L. F., Cai, T. and Wei, L. J. (2009). Robust Inferences for Covariate Effects on Survival Time With Censored Linear Regression Models. *Statistics in Biosciences*, **1**(1), 50–64.

León, L. F. and Cai, T. (2012). Model Checking Techniques for Assessing Functional Form Specifications in Censored Linear Regression Models. *Statistica Sinica* **22**(2), 509–530.

Lin, D. Y. and Wei, L. J. (1989). The Robust Inference for the Cox Proportional Hazards Model. *Journal of the American Statistical Association*, **84**(408), 1074–1078.

Lin, D. Y., Wei, L. J. and Ying, Z. (1993). Checking the Cox model With Cumulative Sums of Martingale-based Residuals. *Biometrika* **80**(3), 557–572.

Lin, D. Y., Robins, J. M. and Wei, L. J. (1996). Comparing Two Failure Time Distributions in the Presence of Dependent Censoring. *Biometrika*, **83**(2), 381–393.

Lin, D. Y., Wei, L. J. and Ying, Z. (1998). Accelerated Failure Time Models for Counting Process. *Biometrika*, **85**(3),605–618.

Lin, D. Y., Wei, L. J. and Ying, Z. (2002). Model-checking Techniques Based on Cumulative Residuals. *Biometrics* **58**(1), 1-12.

Lindsay, B. G., Yi, G. Y. and Sun, J. (2011). Issues and Strategies in the Selection of Composite Likelihoods. *Statistica Sinica*, **21**(1), 71–105.

Louis, T. A. (1981). Nonparametric Analysis of an Accelerated Failure Time Model, *Biometrika*, **68**(2), 381–390.

Lunceford, J. K. and Davidian, M. (2004), Stratification and Weighting Via the Propensity Score in Estimation of Causal Treatment Effects: A Comparative Study. *Statistics in Medicine*, **23**(19) 2937–2960.

Nelder, J. A., and Mead, R. (1965). A Simplex Algorithm for Function Minimization. *Computer Journal*, **7**(4), 308–313.

Nolan, D. and Pollard, D. (1987). U-processes : Rate of Convergence. *The Annals of Statistics*, **15**(2), 780–799.

Nolan, D. and Pollard, D. (1988). Functional limit theorems for U-processes. *The Annals of Probability*, **16**(3), 1291–1298.

Oakes, D. (1982). A Model for Association in Bivariate Survival Data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **44**(3), 414–422.

Oakes, D. (1986). Semiparametric Inference in a Model for Association in Bivariate Survival Data. *Biometrika*, **73**(2),353–361.

Parzen, M. I., Wei, L. J. and Ying, Z. (1994). A Resampling Method Based on Pivotal Estimating Functions. *Biometrika*, **81**(2), 341–350.

Peng, L. and Fine, J. P. (2006). Rank Estimation of Accelerated Lifetime Models with Dependent Censoring. *Journal of the American Statistical Association*, **101**(475), 1085–1093.

Prentice, R. L. (1978). Linear Rank Tests with Right Censored Data. *Biometrika*, **65**(1), 167–179.

Prentice, R. L., Kalbfleisch, J. D., Peterson, A. V., Jr., Flournoy, N., Farewell, V. T. and Breslow, N. E. (1978). The Analysis of Failure Times in the Presence of Competing Risks. *Biometrics*, **34**(4), 541–554.

Qu, A., Lindsay, B. G. and Li, B. (2000). Improving Generalised Estimating Equations Using Quadratic Inference Functions. *Biometrika*, **87**(4): 823–836.

Roach, M. 3rd., DeSilvio, M., Lawton, C., Uhl, V., Machtay, M., Seider, M. J., Rotman, M., Jones, C. Asbell, S. O., Valicenti, R. K., Han, S., Thomas, C. R., Jr. and Shipley, W. S. (2003). Phase III Trial Comparing Whole-pelvic Versus Prostate-only Radiotherapy and Neoadjuvant Versus Adjuvant Combined Androgen Suppression: Radiation Therapy Oncology Group 9413. *Journal of Clinical Oncology*, **21**(10), 1904–1911.

Roach, M. 3rd., Hunt, D., Lawton, C. A., Hsu, I., Lustig, R. A., Seider, M., Christopher, J. U., Thomas, C. R., Shipley, W. U., Sandler, H. (2013). Radiation Therapy Oncology Group (RTOG) 9413: A Randomized Trial Comparing Whole Pelvic Radiation Therapy (WPRT) to Prostate Only (PORT) and Neoadjuvant Hormonal Therapy (NHT) to Adjuvant Hormonal Therapy (AHT). *International Journal of Radiation Oncology* Biology* Physics*, **87**(2), S106-S107.

Robins, J. M. and Rotnitzky, A. (1992). Recovery of Information and Adjustment for Dependent Censoring Using Surrogate Markers. In *AIDS epidemiology-methodological issues*, N. Jewell, K. Dietz and V. Farewell (eds), 297–331, Birkhauser, Boston, MA.

Rosenbaum, P .R. and Rubin, D. B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, **70**(1), 41–55.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

Stefanski, L. A. and Boos, D. D. (2002). The Calculus of M-estimation. *The American Statistician*, **56**(1), 29–38.

Struthers, C. A. and Kalbfleisch, J. D. (1986). Misspecified Proportional Hazard Models. *Biometrika*, **73**(2), 363–369.

Therneau, T. M., Grambsch, P. M. and Fleming, T. R. (1990). Martingale-based Residuals for Survival Models. *Biometrika* **77**(1). 147–160.

Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*. **58**(1), 267–288.

Tsiatis, A. A. (1975). A Nonidentifiability Aspect of the Problem of Competing Risks, *Proceedings of National Academy of Sciences*. **72**(1), 20–22.

Tsiatis, A. A. (1990). Estimating Regression Parameters Using Linear Rank Tests for Censored Data, *The Annals of Statistics*. **18**(1), 354–372.

VanderWheele, T. J. and Knol, M. J. (2011). Interpretation of Subgroup Analyses in Randomized Trials: Heterogeneity Versus Secondary Interventions. *Annals of Internal Medicine*,**154**(10), 680–683.

Varadhan, R. and Gilbert. P. D. (2009). BB: An R Package for Solving a Large System of Nonlinear Equations and for Optimizing a High-dimensional Nonlinear Objective Function. *Journal of Statistical Software*, **32**(4), 1–26.

Wang, W. (2003). Estimating the Association Parameter for Copula Models Under Dependent Censoring. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **65**(1), 257–273.

Wei, L. J. and Gail, M. H. (1983). Nonparametric Estimation for a Scale-Change with Censored Observations. *Journal of the American Statistical Association*, **78**(382), 382–388.

Wei, L. J. (1984). Testing Goodness-of-fit for Proportional Hazards Model With Censored Observations. *Journal of the American Statistical Association*, **79**(387), 649–652.

Wei, L. J. and Johnson, W. E. (1985). Combining Dependent Tests with Incomplete Repeated Measurements. *Biometrika*, **72**(2), 359–364.

Wei, L. J., Lin, D.Y. and Weissfeld, L. (1989). Regression Analysis of Multivariate Incomplete Failure Time Data by Modeling Marginal Distributions. *Journal of the American Statistical Association*, **84**(408), 1065–73.

Williamson, E. J., Forbes, A. and White. I. R. (2014). Variance Reduction in Randomised Trials by Inverse Probability Weighting Using the Propensity Score. *Statistics in Medicine*, **33**(5), 721–737.

Xue. L, Wang, L. and Qu, A. (2010). Incorporating Correlation for Multivariate Failure Time Data When Cluster Size Is Large. *Biometrics*, **66**(2), 393–404.

Ying, Z. (1993). A Large Sample Study of Rank Estimation for Censored Regression Data. *The Annals of Statistics*, **21**(1), 76–99.

Zeng, D. and Lin, D. Y. (2008). Efficient Resampling Methods for Nonsmooth Estimating Functions. *Biostatistics*, **9**(2), 355–363.

Zhu, Y. (2013). Data-adaptive Approaches to Modeling Propensity Scores in Causal Inference Problems. Ph.D. dissertation, The Pennsylvania State University.

Zhu, Y., Ghosh, D., Mitra, N. and Mukherjee, B. (2014). A Data-adaptive Strategy for Inverse Weighted Estimation of Causal Effects. *Health Services and Outcomes Research Methodology*, **14**(3), 69–91.

Vita

Youngjoo Cho

Education

Ph.D., Statistics, The Pennsylvania State University, August 2015

M.A., Applied Statistics, University of Michigan at Ann Arbor, April 2010

B.A., History, Sogang University, August 2008

B.B.A., Business Administration, Sogang University, August 2008

Research Interests

Survival Analysis, Causal Inference, Machine learning

Publications

- Cho, Y. and Ghosh, D. (2015). Weighted Estimation of the Accelerated Failure Time Model in the Presence of Dependent Censoring. *PLOS ONE*, accepted.

- Cho, Y. and Ghosh, D. (2015). A General Approach to Goodness of Fit for U-processes. *Statistica Sinica*, under revision.

- Cho, Y., Ghosh, D. and Hu, C. (2015). Covariate Adjustment Using Propensity Scores for Dependent Censoring Problems in Accelerated Failure Time Model. in preparation.

- Cho, Y. and Ghosh, D. (2015). Covariate Adjustment Using Propensity Scores for Recurrent Events in the Presence of Dependent Censoring. in preparation.

Conference Presentations

- Weighted Estimation of the Accelerated Failure Time Model in the Presence of Dependent Censoring. Poster presented at 2013 ENAR and 2013 Rao Prize Conference. Presented at 2013 Joint Statistical Meetings and 2014 ENAR.

- Adjustment Using Propensity Scores for Artificial Censoring Problems. Poster presented at Statistical Analysis of Multi-outcome Data Workshop 2014.

- Goodness of Fit of U-processes on AFT model. Poster presented at 2014 Joint Statistical Meetings.

- Covariate Adjustment Using Propensity Scores for Recurrent Events in the Presence of Dependent Censoring. Poster presented at 2015 Atlantic Causal Inference Conference.