Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Jing Qian Date

Analysis of Outcomes Subject to Induced Dependent Censoring: Medical Cost and Successive Durations

By

Jing Qian

Doctor of Philosophy

Biostatistics

Yijian Huang, Ph.D.
Advisor

Joseph Lipscomb, Ph.D.
Committee Member

Amita K. Manatunga, Ph.D.
Committee Member

Limin Peng, Ph.D.
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

Analysis of Outcomes Subject to Induced Dependent Censoring: Medical Cost and Successive Durations

by

Jing Qian

B.S., Renmin University of China, 2002

M.S., Emory University, 2009

Advisor: Yijian Huang, Ph.D.

An abstract of

A dissertation submitted to the Faculty of the

James T. Laney School of Graduate Studies of Emory University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in Biostatistics

2009

Abstract

Analysis of Outcomes Subject to Induced Dependent Censoring: Medical Cost and Successive Durations

By Jing Qian

In medical studies for chronic diseases, survival time, the usual primary outcome of interest, may not be adequate to assess the treatment or covariate effects on the disease process. To conduct a more comprehensive evaluation, secondary outcomes capturing other features of the disease process are often assessed simultaneously. Typical examples include the lifetime medical cost and successive durations in disease process. Analysis of secondary outcomes is complicated by induced dependent censoring and identifiability issue, arising from the incomplete follow-up data in clinical trials. In this dissertation, two novel statistical methods accommodating the features of these secondary outcomes are proposed.

The first method focuses on the analysis of censored lifetime medical cost. Currently available approaches are incapable of addressing lifetime medical cost distribution for a defined group. To this end, we propose a copula-based semiparametric regression model, which parameterizes the association of the bivariate error term on time and cost scales through a normal copula function, leaving the marginal error distributions completely unspecified. We develop estimation procedure for the regression coefficients and the normal copula association parameter. The resulting estimators are shown to be consistent and asymptotically normal. Simulation studies and a lung cancer data analysis are conducted to evaluate the finite sample performance of the method.

The second approach is motivated by a colon cancer study where patients progress through cancer-free and cancer-recurrence states. Scientific interests lie in the successive durations in this bi-state progressive disease process. For the one-sample problem with incomplete follow-up data, recent investigations have focused on nonparametric inference. However, in many practical situations, the distribution of the second duration is nonparametrically nowhere identifiable. To address this issue, we suggest a semiparametric model that postulates normal copula for the association between the two durations, while leaving the marginals unspecified. Motivated by the colon cancer data, we allow our model to accommodate the situation where the second duration has a probability mass at zero. We propose an inference procedure and study the asymptotic properties of the resulting estimators. Finite sample performance of the proposed method is evaluated via the simulation studies and illustrated with colon cancer study.

Analysis of Outcomes Subject to Induced Dependent Censoring: Medical Cost and Successive Durations

by

Jing Qian

Advisor: Yijian Huang, PhD

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Biostatistics
2009

Acknowledgement

I would like to express my gratitude to all those who support me during my study at Emory.

I am deeply indebted to my advisor Professor Eugene Huang, who has guided me throughout my graduate study at Emory. It is from Eugene that I truly learned how to conduct scientific research. During these times when I felt lost on the way, Eugene was always there to point out the right direction and help me get back on track. Eugene's dedication to research and the breadth and depth of his scientific knowledge have greatly amazed me. I greatly appreciate the financial support from Eugene for my dissertation research, and the freedom Eugene gives me to make plans, and to use my time.

I am grateful to Professor Limin Peng, Professor Amita Manatunga and Professor Joseph Lipscomb for the precious time they spent serving on my dissertation committee. Their helpful comments and suggestions have improved the presentation of my dissertation.

The faculty and staff in the Department of Biostatistics and Bioinformatics supported me in my graduate study. I want to thank them for all their help and support. Especially, I am thankful to Professor John Hanfelt, who provided me with my first research experience at Emory, to Professor Amita Manatunga, who trusts me and let me help on grant revision, to Professor Limin Peng, who offered me the opportunity to work with

her on censored quantile regression, and to Professor Kirk Easley, who have guided me on statistical consulting. It has been an extremely rewarding experience to collaborate with these talented and generous people on many interesting topics. I also want to thank Professor Brent Johnson and Mr. Vernard Martin for their assistance on programming and computing. Thanks to the department administrators who have taken care of many things in support of my research and study.

Last but not least, my most heartfelt gratitude goes to my parents and my wife Xiaoxia whose patient love enabled me to complete this work.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In medical studies for chronical diseases, survival time is typically the primary outcome of interest. However, such an outcome is inadequate to assess the treatment or covariate effects on the disease process. To make more comprehensive evaluation, a few secondary outcomes capturing other features of the disease process are often assessed simultaneously. Typical examples include the lifetime medical cost, quality-adjusted survival times and sojourn times in successive disease states. Information on these secondary outcomes is important for health care evaluation and health policy administration, and thus has broad social and economic impact. However, analysis of these secondary outcomes has been a statistical challenge. The main difficulty arises from the incomplete follow-up data, which are typically inevitable in clinical trials.

The goal of this dissertation research is to develop statistical methods which are suitable for the aforementioned secondary outcomes. We focus on two types of secondary outcomes: the lifetime medical cost and the successive durations in disease progression. In this chapter, two motivating examples are presented first, and then the statistical challenges of analyzing these secondary outcomes are pointed out. After this, we give a

literature review on statistical methods for censored medical cost and successive durations in disease progression. Finally, the organization of this dissertation is outlined.

1.1 Motivating Examples

The first example, which motivates our research for the lifetime medical cost, is a lung cancer clinical trial conducted by the Southwest Oncology Group (SWOG). Our investigation on the successive durations in disease progression is motivated by the second example, a national intergroup colon cancer clinical trial.

1.1.1 SWOG Lung Cancer Clinical Trial

Lung cancer is the leading cause of cancer-related death for both men and women in the United States. Based on the estimation from the National Cancer Institute (NCI), 215 020 people will be diagnosed with and 161 840 men and women will die of cancer of the lung in 2008 (Ries et al. 2007). The estimated national direct medical cost for lung cancer is $ 4.68 billion annually (Brown, Lipscomb, and Snyder 2001). The randomized lung cancer two-arm clinical trial conducted by the Southwest Oncology Group (SWOG) was designed to compare paclitaxel plus carboplatin versus vinorelbine plus cisplatin treatments for patients with advanced non-small-cell lung cancer (Kelly et al. 2001). From April 1996 through January 1998, 444 patients from 108 sites were enrolled in the trial. Of these, 36 (8.11%) were ineligible. For the 408 eligible patients, 206 patients were randomly assigned to receive paclitaxel plus carboplatin treatment and the remaining 202 patients received vinorelbine plus cisplatin treatment.

Whereas survival time was the primary outcome, one secondary endpoint was resource utilization which consisted of supportive care medications, blood products, medical pro-

cedures, protocol and non-protocol related treatments, and medical care inpatient days or outpatient visits. Cost was assigned to each resource using national databases with adjustment to 1998 US dollars following the medical care component of the Consumer Price Index. The cost data were collected every 3 months during the first 6 months, and every 6 months thereafter, up to 24 months.

One important research question here is: How does the lifetime medical cost influenced by the baseline characteristics of patients? The lifetime medical cost here means the accumulated cost of medical care for persons with lung cancer from the time of enrollment until death. A further question that can be asked is: Given the baseline characteristic information of a group of patients, can we estimate the distribution of their lifetime medical cost? The method we develop in Chapter 2 is motivated by answering these questions.

1.1.2 A National Intergroup Colon Cancer Clinical Trial

As a cause of death due to cancer, colon cancer is second only to lung cancer in the United States. In approximately 80% of the patients with colon cancer in the United States, the diagnosis is made at a sufficiently early stage when all apparent diseased tissue can be surgically removed. Those who have regional nodal involvement that is clinically completely resected are referred to as having Duke's Stage C disease (Dukes 1932). Unfortunately, about one-half of these patients have residual cancer existing in an occult and probably microscopic stage, which leads to recurrence of disease and death within 5 years. A national intergroup trial was conducted in 1980's to evaluate the effect of the drugs levamisole and fluorouracil, as adjuvant therapy for resected colorectal carcinoma (Moertel et al. 1990).

In this trial, 929 eligible patients with Stage C disease were randomized to three study

arms. Of those, 315, 310 and 304 patients received observation, levamisole alone, and levamisole combined with fluorouracil treatments, respectively. The patient enrollment was begun in March 1984 and was completed in October 1987. The dataset available on the Mayo Clinic website contains much richer long-term information than that used in the original report by Moertel et al. (1990), with a maximum follow up of more than 8 years. By the end of the study, 177 patients in the observation arm had cancer recurrence, among whom 155 died; 172 patients in the levamisole alone arm had cancer recurrence, among whom 151 died; while in the levamisole plus fluorouracil arm 119 patients had cancer recurrence, among whom 108 died. In addition, 38 of the 929 patients in the trial died without cancer recurrence. Among those, 13, 10 and 15 patients belonged to observation, levamisole alone, and levamisole combined with fluorouracil treatments, respectively.

Moertel et al. (1990) demonstrated that therapy with levamisole plus fluorouracil produced an unequivocal advantage over observation, and delayed the time to cancer recurrence as well as time to death since randomization. On the other hand, therapy with levamisole alone produced no detectable effect. Another important issue which was not addressed by Moertel et al. (1990) is whether or not the therapy with levamisole plus fluorouracil has any benefit on survival after cancer recurrence. To answer this question, we need to estimate the distribution of the sojourn time between cancer recurrence and death. The method we develop in Chapter 3 will be able to answer this question.

1.2 Statistical Challenges

1.2.1 Induced Dependent Censoring

In clinical trials or observational studies, due to the limited time of study duration, it is inevitable that some patients are not followed until the endpoint of interest. This phenomenon is referred to as "censoring", which is well known for survival time data. For those censored individuals, their secondary outcomes are not fully observed. The censoring on secondary outcome is induced from the censoring on time.

Statistical methods for handling censoring in survival data have been well developed. As to the secondary outcomes, for example, censored medical costs, we consider cost-to-event as opposed to time-to-event. It may be presumed that standard survival analysis methods, such as the Kaplan–Meier estimator (Kaplan and Meier 1958), log-rank test(Mantel 1966) and Cox proportional hazard regression (Cox 1972) could be applied straightforward to medical cost problems. Actually, this strategy had been attempted by some medical researchers before (see Quesenberry et al. (1989), Hiatt et al. (1990), Dudley et al. (1993), and Fenn et al. (1995)). Unfortunately, the strategy is generally invalid. The main problem is the requirement in standard survival analysis techniques that the time of death and the corresponding time of censoring must be independent. On medical cost scale, the total cost at the time of event (*e.g.* death) and the total cost at the time of censoring correspond to the event time and the censoring time in standard survival analysis, respectively. However, the total cost at the time of death is not independent of the total cost at the time of censoring, even if the time of death and time of censoring are themselves independent. This induced informative censoring feature of cost data could be demonstrated by a simple example.

Example 1.1. *Let T and C be the survival time and censoring times respectively, due*

to random censorship on time scale, we have $C \perp T$, where $\perp$ represents independence. Then on the medical cost scale, the counterparts to T and C are $N(T)$ and $N(C)$, respectively. However, $N(T)$ and $N(C)$ are correlated because $N(\cdot)$ is a random process in general. In other words, $N(T) \perp N(C)$ if $N(\cdot)$ is a deterministic function, which means everyone has the same cost accumulation pattern. Unfortunately, cost accumulation patterns vary among individuals in practice.

The induced informative censoring characteristic of medical cost data precludes the straightforward application of standard survival analysis approaches on censored medical costs. Otherwise, biased estimation of mean medical cost would occur. This point has already been noticed by Gelber et al. (1989) and Glasziou et al. (1990) in the context of quality adjusted survival time (Cox et al. 1992), and by Lin et al. (1997), Hallstrom and Sullivan (1998), Lipscomb et al. (1998) and Diehr et al. (1999) in medical cost studies. As Lin et al. (1997) indicated in their paper, "A subject who has high costs per month will usually have high total costs at both the time of death and time of censoring, whereas a subject with low costs per month will tend to have low total costs at both times. Thus, the total cost at the time of death tends to be positively correlated with the total cost at the time of censoring. This likely violation of the independent censoring assumption implies that censored costs should not be analyzed by standard survival methods."

In the colon cancer clinical trial example we described above, after the surgery, a patient potentially progresses over cancer-free and cancer-relapse states before reaching death. A bi-state progressive disease process is a reasonable model for this kind of chronic diseases. The two successive durations in this bi-state progressive disease process are time to cancer relapse as well as time between cancer relapse and death. let T_1 and T_2 be the two durations in order, and C be the censoring time. Note that the censoring mechanism affects on the two successive durations serially. The censoring time on the

first duration is C, and the censoring time on the second duration, which is induced by T_1, is $(C - T_1)^+ \equiv \max\{(C - T_1), 0\}$. Because T_1 and T_2 are typically dependent, T_2 and its induced censoring time $(C - T_1)^+$ would be dependent, even if C is independent of $\{T_1, T_2\}$.

1.2.2 Identifiability Issue

Clinical studies, in general, have a limited duration, which is typically shorter than the longest survival time. Thus, the cost accumulation process beyond the end of the study is simply unobservable due to the administrative censoring. As pointed by Huang (2002), if a certain portion of the study population would incur zero cost within the study duration and survive beyond the study duration, then their lifetime medical cost is completely unknown. In fact, it is not uncommon to observe zero cost accumulation for an appreciable proportion of participants during a study. Subsequently, the distribution of lifetime medical cost could be nowhere identifiable in a one-sample nonparametric setting.

Similar issue of identifiability exists for the bi-state progressive disease process. The typical finite study duration means that the maximum support point of the censoring time, $\tau^C \equiv \sup\{t : \Pr(C \leq t) < 1\}$, is finite. Thus, if the maximum support point of the first duration T_1 is greater than the maximum support point of the censoring time C, the distribution of T_2 given $T_1 > \tau^C$ is not identifiable. Consequently, the marginal distribution of T_2 is nonparametrically nowhere identifiable in this situation.

1.3 Literature Review

1.3.1 Medical Cost

Cost assessment has become an important component in health care evaluation and gained much attention in recent years. For instance, health policy makers may be concerned about the cost of providing health care for patients with acquired immunodeficiency syndrome (AIDS) in the whole country and its economic impact, in order to draw out a sensible budget plan for AIDS therapy. As another example, to make comprehensive evaluation on alternative treatments on certain disease when one treatment has a demonstrated health benefit over another yet is more expensive, cost-effectiveness analysis is usually being adopted. Health care providers and insurance company may also be desirable to quantify the variation in medical costs between individuals due to patients' characteristics, in order to construct an accurate probabilistic decision model.

Despite the tremendous interests in and increasing demands for the evaluation of medical costs, appropriate statistical methods which capture the underlying properties of medical cost data are not straightforward to develop. The main challenges of analyzing cost data come from three aspects. Besides the induced dependent censoring and nowhere identifiable issue we discussed earlier, medical costs data also tend to have a highly right-skewed distribution.

A few modeling techniques have been successfully developed in recent years, accommodating the aforementioned features of lifetime medical cost data. Meanwhile, Some other statistical methods dealing with similar mark variables subject to induced informative censoring have also been formulated.

In one-sample case, Lin et al. (1997) proposed nonparametric estimators with discrete censoring time distribution, but the application of their methods is limited by the re-

quirement of discrete censoring pattern. Zhao and Tsiatis (1997) developed a consistent estimator for the distribution of quality adjusted survival time, by adopting the inverse probability of censoring weighting (IPCW) idea (Horvitz and Thompson 1952; Koul et al. 1981; Robins and Rotnitzky 1992; Robins et al. 1994). One drawback of Zhao and Tsiatis's Estimator is that the statistical properties of the estimator are established in a pointwise situation, whether the properties still hold in a uniform fashion is unclear. Therefore, the estimator may not be appropriate in developing two-sample test procedure. Bang and Tsiatis (2000) proposed simple IPCW based estimator and efficiency improved estimator to estimate the mean value of censored medical cost. Strawderman (2000) considered a general framework of stopped longitudinal process, which contains specific problems from the analysis of quality adjusted survival data to recurrent event data to lifetime medical cost data.

For two-sample comparison, by generalizing the weighted log-rank test for survival times and taking the strategy of IPCW, Zhao and Tsiatis (2001) proposed a method for comparing the survival functions of quality-adjusted lifetime from two treatments. The method could also be applied to medical cost data.

In regard to the regression analysis of lifetime medical cost distribution, Lin (2000a) proposed linear regression procedures for censored medical cost. The requirement that the covariates are all discrete with a finite number of values limits the practical application of the method. Lin (2000b) developed a proportional means regression model, which specified that the mean function for the cumulative medical cost over time, conditioning on a set of covariates, was equal to an arbitrary baseline mean function multiplied by an exponential regression function. The continuous observation of the whole cost accumulative process and the completely random censoring mechanism are required by this model. The rigorous requirements on censorship and data structure limit the application

of this method. This model is also restrictive in that it imposes common proportionate covariate effects over time, and would not be appropriate in some cases. The interpretation of the covariate effects is also not straightforward. Lin (2003) extended the prior work of Lin (2000a) and Lin (2000b) to more flexible and versatile generalized linear model framework. Bang and Tsiatis (2002) proposed the median regression method for censored cost data. It requires completely random censoring mechanism, which restricts the application of this model. A hazard regression model was developed by Jain and Strawderman (2002).

Handling appropriately both the highly positively skewed and induced informative censoring troubles with multifarious paths, the aforementioned modeling approaches take a common way to avoid the nowhere identifiability issue of the marginal distribution of lifetime medical costs. Actually, they all impose an artificial time limit to ensure the potential censoring time has positive support beyond this time limit. In this way, the target of the analysis becomes *time-restricted* medical cost instead of *lifetime* medical cost. Since the time limit is artificial and the covariates may impact survival time as well, attempt to interpret time-restricted results in terms of lifetime medical cost, as desired, is inappropriate. More seriously, it may result in the improper acceptance or rejection of certain treatments which yield benefits as well as costs over the patient's lifetime.

One way to overcome the non-identifiability issue is to consider the joint distribution of lifetime medical cost with survival time, as proposed by Huang and Louis (1998) in the one-sample nonparametric setting. Huang (2002) extended this strategy and developed the calibration regression model, which postulates linear covariate effects on possibly transformed lifetime medical cost and survival time. This model is appealing since its parameters are meaningful for lifetime medical cost. Unfortunately, it is still

impossible to quantify the covariate effect in terms of difference in dollar amount as desired unless the transformation on cost scale is linear. This is due to the fact that marginal distribution of lifetime medical cost with a set of given covariate values is still not identifiable under the model. Recently, Huang and Berry (2006) developed semiparametric estimation procedures in one-sample scenario, by constructing copula-based models that parameterize the association between survival time and cost but leave the marginals unspecified. Through parametrization of the copula function, the marginal identifiability of the lifetime medical cost is achievable.

1.3.2 Successive Durations

A bi-state progressive process is a reasonable model for the course of many chronic diseases. Examples are numerous. In a colon caner study, after surgery a participant potentially progresses over disease-free and relapse states before reaching death. The development of AIDS consists of HIV incubation period and clinical AIDS period.

In the one-sample problem with incomplete follow-up data, nonparametric estimation approaches have been developed by several recent investigations. By considering a discrete censoring mechanism, Visser (1996) derived a nonparametric maximum likelihood estimator of the bivariate survival function of two successive durations. With continuous censoring variable, Wang and Wells (1998) developed a product-limit estimator for the second duration variable and an estimator for the joint survival function, and Lin, Sun, and Ying (1999) proposed a nonparametric approach for the joint and conditional marginal distributions of the successive durations. This bi-state successive duration problem is unique in its *serial* censoring; that is, the censoring mechanism bears on the two successive durations serially. It is characteristically different from the estimation of bivariate distribution studied by Dabrowska (1988), Prentice and Cai (1992),

and van der Laan (1997) among others, where the censoring is *parallel*.

Let T_k, $k = 1, 2$, be the two durations in order, and C be the censoring time. Under serial censoring, the censoring time on the first duration is C, and the censoring time on the second duration, which is induced by T_1, is $(C - T_1)^+ \equiv \max\{(C - T_1), 0\}$. One prominent issue with serial censoring is the induced dependent censoring for the second duration. The dependence between T_2 and its induced censoring time $(C - T_1)^+$ would arise from that between T_1 and T_2, even if C is independent of $\{T_1, T_2\}$. In addition, identifiability of the distribution of T_2 is an even more thorny issue. As we discussed in Section 1.2.2, the marginal distribution of T_2 is nonparametrically nowhere identifiable in this situation. Therefore, it is not surprising that those aforementioned nonparametric approaches for successive durations have only limited successes in estimating the distribution of T_2. They either require that the maximum support point of the first duration T_1 is no greater than the maximum support point of the censoring time C, or turn to its joint distribution with T_1; the joint distribution is identifiable to a certain extent. Note that these two issues are omnipresent with the broad class of mark estimation problems as addressed in Huang and Louis (1998). A mark is a random variable associated with an event such that its observation is contingent upon the occurrence of the event. In this bi-state process problem, each of the two durations may be viewed as a mark of the terminating event of the second duration.

Furthermore, all these nonparametric approaches with the exception of Visser (1996) require the assumption that the censoring time C is completely independent of $\{T_1, T_2\}$. That means, individuals being followed at a given time t would have the same probability of being censored within $[t, t + dt)$ regardless of whether the first duration has completed and, if so, when. Lin, Sun, and Ying (1999) pointed out that this could be a practical limitation. On the other hand, the approach of Visser (1996) has its own limitation with

the requirement of a discrete time scale.

1.4 Outline

In chapter 2, we present a copula-based semiparametric regression model for censored lifetime medical cost. One of the major challenges of analyzing lifetime medical cost is that the censoring pattern on the cost scale is typically induced to be dependent. Moreover, the lifetime medical cost distribution is potentially nowhere identifiable. Currently available approaches either bypass these issues by estimating time-restricted medical cost or address the joint distribution of cost and survival time instead. Neither is capable of addressing lifetime medical cost distribution for a defined group. To this end, we propose a copula-based semiparametric regression model under which the marginal lifetime medical cost distribution for given covariates becomes identifiable. The proposed model assumes that both lifetime medical cost and survival time, on possibly transformed scales, linearly relate to the covariates. Furthermore, it parameterizes the association of the bivariate error term through a normal copula function, leaving the marginal error distributions completely unspecified. While the regression coefficients on survival time are estimated using the standard method for the accelerated failure time model, we propose a procedure to estimate the regression coefficients on lifetime medical cost and the normal copula association parameter simultaneously. The resulting estimators are shown to be consistent and asymptotically normal. Simulation studies show that the proposed method is fairly robust under copula misspecification. The proposed method is applied to a lung cancer clinical trial.

In chapter 3, we present a semiparametric inference procedure for the successive durations in this bi-state progressive disease process. The study is motivated by a colon

cancer clinical trial where patients progress through disease-free and relapse states. Scientific interests lie in the successive durations in this bi-state progressive disease process. For the one-sample problem with incomplete follow-up data, recent investigations have focused on nonparametric inference. However, in many practical situations, the distribution of the second duration is nonparametrically nowhere identifiable. Furthermore, most existing approaches require a rather restrictive censoring mechanism and have difficulty in predicting the process with given history. To address these issues, we suggest a semiparametric model that postulates normal copula for the association between the two durations, while leaving the marginals unspecified. We propose an inference procedure for estimation and establish the asymptotic properties of the proposed estimators. Finite sample performance of the proposed method is evaluated by the simulation studies and illustrated with the data from a colon cancer study.

In Chapter 4, we summarize results in this dissertation and discuss topics for future research.

Chapter 2

Copula-based Semiparametric Regression Model for Censored Lifetime Medical Cost

In this chapter, we propose a copula-based semiparametric regression model for censored lifetime medical cost, to strengthen the calibration regression model. In the one-sample problem, Huang and Berry (2006) showed that the distribution of lifetime medical cost becomes marginally identifiable upon reasonable copula parameterization between the cost and survival time. Following this line, we suggest a regression method for marginal estimation and inference of lifetime medical cost, by imposing a normal copula structure on the bivariate error in the calibration regression model. This approach targets the situation when the survival time has only a potentially small probability to exceed the study duration.

The rest of the chapter is structured as follows. In Section 2.1 we introduce the semiparametric copula regression model for lifetime medical cost subject to censoring.

In Section 2.2 we develop an inference procedure for regression coefficients as well as the association parameter in normal copula model, and we also study the asymptotic properties for the resultant estimators. In Section 2.3 we evaluate the finite-sample performance of the proposed method through simulations, followed by an application to a lung cancer clinical trial. Further discussion is given in Section 2.4. Technical details of the large-sample study on the proposed inference procedure are collected in Section 2.5.

2.1 A Copula-based Semiparametric Regression Model

Consider a sample of n subjects who are followed up to the occurrence of death or being censored, whichever comes first. For the ith individual, let T_i represent survival time, U_i be lifetime medical cost, and $\mathbf{Z}_i$ be a $p \times 1$ covariate vector. The censoring time C_i operates on T_i so that T_i, U_i and C_i are not directly observed but through the following variables:

$$X_i \equiv T_i \wedge C_i, \qquad Y_i \equiv U_i \cdot I(T_i \leq C_i), \qquad \Delta_i \equiv I(T_i \leq C_i),$$

where $\wedge$ is the minimization operator and $I(\cdot)$ is the indicator function. The observed data consist of i.i.d. observations $\{X_i, Y_i, \Delta_i, \mathbf{Z}_i\}$, $i = 1, \ldots, n$. This data structure is basic and common to various data-collection scenarios involving a mark of interest (Huang and Louis 1998). A mark refers to a random variable associated with an event such that its observation is contingent upon the occurrence of the event; the methodology developed here for medical cost may be applied to other marks of interest. Certainly there are applications where the cost accumulation process may be fully or partially

observed in addition. The method developed here neither depends on nor makes full use of such additional data; see further discussion in Section 5. For the censorship, we assume a conditional independence censoring mechanism:

$$(T_i, U_i) \text{ is independent of } C_i, \text{ given covariates } \mathbf{Z}_i. \tag{2.1}$$

The proposed copula-based semiparametric regression model postulates that $\mathbf{Z}_i$ linearly relates to $G_T(T_i)$ and $G_U(U_i)$, for known increasing (e.g., logarithmic) transformations $G_T(.)$ and $G_U(.)$,

$$\begin{pmatrix} G_T(T_i) \\ G_U(U_i) \end{pmatrix} = \begin{pmatrix} \boldsymbol{\alpha}_0^T \\ \boldsymbol{\beta}_0^T \end{pmatrix} \mathbf{Z}_i + \boldsymbol{\varepsilon}_i, \qquad i = 1, \ldots, n, \tag{2.2}$$

where $\boldsymbol{\alpha}_0$ and $\boldsymbol{\beta}_0$ are $p \times 1$ vectors of regression coefficients, and $\boldsymbol{\varepsilon}_i = (\varepsilon_i^t, \varepsilon_i^u)^T$ is the bivariate error term. Furthermore, the bivariate error term $\boldsymbol{\varepsilon}_i$ follows the normal copula model (e.g. Huang and Berry 2006),

$$\begin{pmatrix} H_T(\varepsilon_i^t) \\ H_U(\varepsilon_i^u) \end{pmatrix} \sim \text{BVN}(\rho_0),$$

for unspecified monotone increasing transformations $H_T(.)$ and $H_U(.)$, where $\text{BVN}(\rho_0)$ is the standard bivariate normal distribution with correlation coefficient ρ_0. As seen, this proposed model strengthens the calibration regression model of Huang (2002) with the normal copula structure on the bivariate error term. Apparently, this model is fairly flexible. Similar to the one-sample result of Huang and Berry (2006), the copula parameterization results in the marginal identifiability of the ε_i^u distribution. This modeling strategy is practically useful when survival time exceeds the study duration with only

a small probability, e.g., the lung cancer clinical trial in Section 2.3.3. In such a circumstance, the adopted copula structure is largely testable except for the small tail portion.

Copula modeling has been investigated before for multivariate failure time data, where there are correlated failure times and each failure time has its own censoring. In that setting, marginal estimation is generally straightforward, and adopting a copula model facilitates efficiency gain (e.g., Glidden 2000). In contrast, copula modeling in our setting leads to marginal identifiability and estimation.

2.2 Inference Procedure

2.2.1 A Compound Procedure

This proposed model is a compound one of calibration regression model and the normal copula model. Thus, one may use the calibration regression procedure established by Huang (2002) to obtain estimators for regression coefficients on time and cost scales. First, the estimation of $\boldsymbol{\alpha}_0$ is a standard problem under the accelerated failure time (AFT) model. Define counting process $N_i(t;\boldsymbol{\alpha}) \equiv I(\varepsilon_i^x(\boldsymbol{\alpha}) \leq t)\Delta_i$ and at-risk process $R_i(t;\boldsymbol{\alpha}) \equiv I(\varepsilon_i^x(\boldsymbol{\alpha}) \geq t)$, where $\varepsilon_i^x(\boldsymbol{\alpha}) \equiv G_T(X_i) - \boldsymbol{\alpha}^T\mathbf{Z}_i$. A rank-based approach is adopted here in which a weighted log-rank test $\boldsymbol{\Psi}_{1,n}(\boldsymbol{\alpha})$ (Tsiatis 1990; Wei, Ying, and Lin 1990) is constructed based on $\varepsilon_i^x(\boldsymbol{\alpha})$. The estimating function takes the form

$$\boldsymbol{\Psi}_{1,n}(\boldsymbol{\alpha}) \equiv n^{-1}\sum_{i=1}^{n}\int_{-\infty}^{+\infty}\varphi_{1,n}(s;\boldsymbol{\alpha})\left[\mathbf{Z}_i - \frac{\sum_{j=1}^{n}\mathbf{Z}_j\,R_j(s;\boldsymbol{\alpha})}{\sum_{j=1}^{n}R_j(s;\boldsymbol{\alpha})}\right]\mathrm{d}\,N_i(s;\boldsymbol{\alpha})\ , \qquad (2.3)$$

where $\varphi_{1,n}(t;\boldsymbol{\alpha})$ is a data-dependent weight function. Weights $\varphi_{1,n}(t;\boldsymbol{\alpha}) = 1$ and $\varphi_{1,n}(t;\boldsymbol{\alpha}) = n^{-1}\cdot\sum_{i=1}^{n} I(\varepsilon_i^x(\boldsymbol{\alpha}) \geq t)$ correspond to the log-rank and Gehan (1965)

estimating functions, respectively. Then, for given $\boldsymbol{\alpha}_0$, the estimation of $\boldsymbol{\beta}_0$ can be obtained by generalizing the weighted log-rank estimating function to the marked point process framework, i.e.,

$$\boldsymbol{\Upsilon}_{2,n}(\boldsymbol{\alpha},\boldsymbol{\beta}) \equiv n^{-1}\sum_{i=1}^{n}\int_{-\infty}^{+\infty}\varphi_{2,n}(s;\boldsymbol{\alpha},\boldsymbol{\beta})\left[\mathbf{Z}_i-\frac{\sum_{j=1}^{n}\mathbf{Z}_j\,R_j(s;\boldsymbol{\alpha})}{\sum_{j=1}^{n}R_j(s;\boldsymbol{\alpha})}\right]\mathrm{d}\,N_{2i}(s;\boldsymbol{\alpha},\boldsymbol{\beta})\ ,$$

where $\varphi_{2,n}(t;\boldsymbol{\alpha},\boldsymbol{\beta})$ is a weight function similar to $\varphi_{1,n}(t;\boldsymbol{\alpha})$; $N_{2i}(t;\boldsymbol{\alpha},\boldsymbol{\beta}) \equiv N_i(t;\boldsymbol{\alpha})\cdot\xi(Y_i-\boldsymbol{\beta}^T\mathbf{Z}_i)$ is the *mark process*, which has a random jump size rather than constant 1; and $\xi(\cdot)$ is a known continuous and strictly monotone function. After obtaining the estimator $\widehat{\boldsymbol{\alpha}}_0$ for $\boldsymbol{\alpha}_0$ as a root of $\boldsymbol{\Psi}_{1,n}(\boldsymbol{\alpha})$, we can then solve $\boldsymbol{\Upsilon}_{2,n}(\widehat{\boldsymbol{\alpha}}_0,\boldsymbol{\beta})$ to find the estimator $\widehat{\boldsymbol{\beta}}_0$ for $\boldsymbol{\beta}_0$.

The inference procedure of Huang and Berry (2006) for normal copula model could then be applied to the residuals of survival time and lifetime medical cost, for the estimation of the normal copula association parameter and further the marginal baseline distribution of lifetime medical cost. Here, we briefly describe the inference procedure of Huang and Berry (2006) for normal copula model (3.1) in the one-sample case (i.e., $\boldsymbol{\alpha}_0$ and $\boldsymbol{\beta}_0$ are known). The idea is to estimate the association parameter ρ_0 in the bivariate normal copula first, and then the marginal distribution of the baseline cost distribution, $F_{\varepsilon^u}(\cdot)$, can be achieved. Let $\varepsilon_i^y(\boldsymbol{\beta}) \equiv G_U(Y_i)-\boldsymbol{\beta}^T\mathbf{Z}_i$. By normal distribution theory, model (3.1) is equivalent to

$$H_U\{\varepsilon_i^y(\boldsymbol{\beta}_0)\}\ \mid\ \{X_i,\mathbf{Z}_i,\Delta_i=1\}\ \sim\ \text{Normal}\,(\rho_0\,H_T\{\varepsilon_i^x(\boldsymbol{\alpha}_0)\},\,1-\rho_0^2),$$

which implies the following linear transformation model,

$$(1-\rho_0^2)^{-1/2}H_U\{\varepsilon_i^y(\boldsymbol{\beta}_0)\}\,\big|\,\{X_i,\mathbf{Z}_i,\Delta_i=1\}\ \sim\ \theta_0 H_T\{\varepsilon_i^x(\boldsymbol{\alpha}_0)\}+\epsilon, \tag{2.4}$$

where $\theta = \rho(1-\rho^2)^{-1/2}$, and ϵ is a standard normal error. Let $W_i(\boldsymbol{\alpha}) \equiv H_T\{\varepsilon_i^x(\boldsymbol{\alpha})\}$. By invoking pairwise comparisons, we obtain the following identity that is free of transformation $H_U(\cdot)$:

$$\Pr\left(\varepsilon_i^y(\boldsymbol{\beta}_0) \geq \varepsilon_j^y(\boldsymbol{\beta}_0) \mid X_i, X_j, \mathbf{Z}_i, \mathbf{Z}_j, \Delta_i = 1, \Delta_j = 1\right) = \Phi_2(\theta_0 W_{ij}(\boldsymbol{\alpha}_0)) \qquad \text{for } i \neq j, \tag{2.5}$$

where $W_{ij}(\boldsymbol{\alpha}) \equiv W_i(\boldsymbol{\alpha}) - W_j(\boldsymbol{\alpha})$, and $\Phi_2(\cdot)$ is the cumulative distribution function of a normal distribution with mean 0 and variance 2. Given $\boldsymbol{\alpha}_0$ and $\boldsymbol{\beta}_0$, an estimating function for θ_0 based on identity (2.5) is:

$$\boldsymbol{\Upsilon}_{3,n}(\theta) \equiv n^{-2} \sum_{i,j=1}^{n} \Delta_i \Delta_j \, \widehat{W}_{ij}(\boldsymbol{\alpha}_0) \left[I\{\varepsilon_i^y(\boldsymbol{\beta}_0) \geq \varepsilon_j^y(\boldsymbol{\beta}_0)\} - \Phi_2(\theta \, \widehat{W}_{ij}(\boldsymbol{\alpha}_0)) \right], \tag{2.6}$$

where $\widehat{W}_i(\boldsymbol{\alpha}_0) \equiv \widehat{H}_T(\varepsilon_i^x(\boldsymbol{\alpha}_0)) = \Phi^{-1}\{\widehat{F}_{\varepsilon^t}(\varepsilon_i^x(\boldsymbol{\alpha}_0))\}$, and $\widehat{F}_{\varepsilon^t}(t)$ is the Kaplan-Meier estimator of $F_{\varepsilon^t}(t)$. Write the estimator $\widehat{\theta}_0$ for θ_0 as the root of $\boldsymbol{\Upsilon}_{3,n}(\theta)$, an estimator of ρ_0 is obtained as $\widehat{\rho}_0 = \widehat{\theta}_0(1+\widehat{\theta}_0^2)^{-\frac{1}{2}}$. Now, we can estimate the marginal distribution $F_{\varepsilon^u}(\cdot)$ by taking advantage of $\widehat{\rho}_0$ and $\widehat{H}_T(\varepsilon_i^x(\boldsymbol{\alpha}_0))$. From identity (2.4), we can estimate the probability $\Pr\{H_U(\varepsilon_i^y(\boldsymbol{\beta}_0)) \leq s \big| X_i, \mathbf{Z}_i, \Delta_i = 1\}$. By some simple algebra, the estimator $\widehat{H}_U(\cdot)$ for the transformation function $H_U(\cdot)$ can be obtained. Noting the one-to-one mapping relationship between $H_U(\cdot)$ and $F_{\varepsilon^u}(\cdot)$, i.e., $H_U(\cdot) = \Phi^{-1}\{F_{\varepsilon^u}(\cdot)\}$, we can then obtain the estimated marginal distribution $\widehat{F}_{\varepsilon^u}(\cdot)$ by reassigning the probability mass of $\widehat{H}_U(\cdot)$ through that one-to-one mapping. The detailed procedures refer to Section 3.2 in Huang and Berry (2006).

This compound procedure, as referred to, is very straightforward and easy to be conducted. However, it does not take advantage of the copula model in the estimation of the cost regression coefficients. In the following, we will develop a new estimation procedure which potentially achieves better efficiency.

2.2.2 Proposed New Estimation Procedure

We still start with the estimation of $\boldsymbol{\alpha}_0$, which is a standard problem under the AFT model on the time scale and can be solved by using the weighted log-rank estimation function (2.3) described above.

Applying the Kaplan-Meier estimation procedure to $\{\varepsilon_i^x(\boldsymbol{\alpha}), \Delta_i\}$, $i = 1, \ldots, n$, gives the estimator $1 - \widehat{F}_T(t\,; \boldsymbol{\alpha})$. If $\boldsymbol{\alpha}_0$ is known, $\widehat{F}_T(t\,; \boldsymbol{\alpha}_0)$ is the Kaplan-Meier estimator of the distribution function of $\varepsilon_i^t(\boldsymbol{\alpha}_0)$, i.e., $Pr\{\varepsilon_i^t(\boldsymbol{\alpha}_0) \leq t\}$. Obviously $\boldsymbol{\alpha}_0$ is unknown. Nonetheless, one may use $\widehat{F}_T(t\,; \widetilde{\boldsymbol{\alpha}})$ instead, where $\widetilde{\boldsymbol{\alpha}}$ is an estimator of $\boldsymbol{\alpha}_0$. This estimator will be needed in later estimation.

For the estimation of $\boldsymbol{\beta}_0$ and ρ_0, we extend the result in Huang and Berry (2006). Following the notations introduced in Section 2.2.2, we still use the identities (2.4) and (2.5) to motivate our new estimation procedure.

Given $\widehat{F}_T(t\,; \widetilde{\boldsymbol{\alpha}})$, $H_T(t)$ can be estimated by $\widehat{H}_T(t; \widetilde{\boldsymbol{\alpha}}) \equiv \Phi^{-1}\{\widehat{F}_T(t\,; \widetilde{\boldsymbol{\alpha}})\}$. To avoid technical difficulties arising from the unboundedness of $\Phi^{-1}(\cdot)$ at 0 and 1 and from the tail instability of $\widehat{F}_T(t\,; \widetilde{\boldsymbol{\alpha}})$, we only consider individuals with $\varepsilon_i^x(\widetilde{\boldsymbol{\alpha}})$ bounded away from its lower and upper limits. For technical purposes of asymptotic study, we adopt a smooth function of $\varepsilon_i^x(\boldsymbol{\alpha})$ with bounded derivative, denoted by $\iota(\varepsilon_i^x(\boldsymbol{\alpha}))$, to approximate this selection. Write $\Delta_i^\circ(\boldsymbol{\alpha}) \equiv \Delta_i \cdot \iota(\varepsilon_i^x(\boldsymbol{\alpha}))$, $\widehat{W}_i(\boldsymbol{\alpha}) \equiv \widehat{H}_T\{\varepsilon_i^x(\boldsymbol{\alpha})\}$ and $\widehat{W}_{ij}(\boldsymbol{\alpha}) \equiv \widehat{W}_i(\boldsymbol{\alpha}) - \widehat{W}_j(\boldsymbol{\alpha})$. We propose to estimate $\boldsymbol{\beta}_0$ and θ_0 simultaneously by adopting the following estimating function based on pairwise comparisons:

$$\boldsymbol{\Psi}_{2,n}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta) \equiv n^{-2} \sum_{i,j=1}^{n} \Delta_i^\circ(\boldsymbol{\alpha}) \Delta_j^\circ(\boldsymbol{\alpha}) \begin{pmatrix} \widehat{W}_{ij}(\boldsymbol{\alpha}) \\ \mathbf{Z}_{ij} \end{pmatrix} \Big[I\{\varepsilon_i^y(\boldsymbol{\beta}) \geq \varepsilon_j^y(\boldsymbol{\beta})\} - \Phi_2(\theta\, \widehat{W}_{ij}(\boldsymbol{\alpha})) \Big]. \tag{2.7}$$

As shown later, $\boldsymbol{\Psi}_{2,n}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \theta_0)$ is centered around 0 asymptotically.

Combining $\boldsymbol{\Psi}_{2,n}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta)$ with the aforementioned log-rank estimating function $\boldsymbol{\Psi}_{1,n}(\boldsymbol{\alpha})$, we obtain following estimating function for $(\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T, \theta)^T$:

$$\boldsymbol{\Psi}_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta) \equiv \left\{ \boldsymbol{\Psi}_{1,n}(\boldsymbol{\alpha})^T,\ \boldsymbol{\Psi}_{2,n}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta)^T \right\}^T \tag{2.8}$$

After obtaining the estimator $\widehat{\boldsymbol{\alpha}}_0$ from $\boldsymbol{\Psi}_{1,n}(\boldsymbol{\alpha}) = 0$, we can then find the estimator $(\widehat{\boldsymbol{\beta}}_0^T, \widehat{\theta}_0)^T$ for $(\boldsymbol{\beta}_0^T, \theta_0)^T$ as a zero-crossing of $\boldsymbol{\Psi}_{2,n}(\widehat{\boldsymbol{\alpha}}_0, \boldsymbol{\beta}, \theta)$. Since $\rho = \theta(1+\theta^2)^{1/2}$, an estimator of ρ_0 is obtained as $\widehat{\rho}_0 = \widehat{\theta}_0(1+\widehat{\theta}_0^2)^{-1/2}$. Note that $\boldsymbol{\Psi}_{2,n}(\widehat{\boldsymbol{\alpha}}_0, \boldsymbol{\beta}, \theta)$ is not a monotone random field (cf. Fygenson and Ritov 1994) in the domain of $(\boldsymbol{\beta}^T, \theta)^T$, i.e., $\mathbb{R}^{p+1}$. Multiple zero-crossings may exist. We resolve this issue by defining the estimators as the zero-crossing closest to the estimator from the compound procedure. The latter is a consistent estimator.

Given the estimators for regression coefficients $\boldsymbol{\alpha}_0$, $\boldsymbol{\beta}_0$ and copula association parameter ρ_0, we can obtain the estimation of the marginal error distribution on cost scale $\widehat{F}_{\varepsilon^U}(.)$ by following the estimation procedure developed in Huang and Berry (2006, Section 3.2). Furthermore, given specific covariate values $\mathbf{Z} = \mathbf{z}$, we may estimate the marginal distribution of lifetime medical cost for the group defined by $\mathbf{z}$, say, $\widehat{F}_U(.|\mathbf{z})$ and the corresponding mean and median of lifetime medical cost.

2.2.3 Asymptotic Properties

The consistency and asymptotic normality of the estimators $(\widehat{\boldsymbol{\alpha}}_0^T, \widehat{\boldsymbol{\beta}}_0^T, \widehat{\theta}_0)^T$ are established in Theorems 2.1 and 2.2, with detailed proof collected in Appendices 2.5.1 and 2.5.2. In the following theorems, we assume that the censoring mechanism (2.1) and regression model (2.2) hold, and that observations $\{X_i, Y_i, \Delta_i, \mathbf{Z}_i\}$, $i = 1, \ldots, n$, are i.i.d.. For the convenience of expression, we write $\boldsymbol{\eta} \equiv (\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T, \theta)^T$, $\boldsymbol{\eta}_0 \equiv (\boldsymbol{\alpha_0}^T, \boldsymbol{\beta_0}^T, \theta_0)^T$ and

$\widehat{\boldsymbol{\eta}}_0 \equiv (\widehat{\boldsymbol{\alpha}}_0^T, \widehat{\boldsymbol{\beta}}_0^T, \widehat{\theta}_0)^T$.

Theorem 2.1. *Suppose that regularity conditions C1-C6 in Appendix 2.5.1 hold. There exists a zero-crossing of $\boldsymbol{\Psi}_n(\boldsymbol{\eta})$, say $\widehat{\boldsymbol{\eta}}_0$, that converges to $\boldsymbol{\eta}_0$ in probability.*

Theorem 2.2. *Under the regularity conditions C1-C6 in Appendix 2.5.1, $n^{1/2}(\widehat{\boldsymbol{\eta}}_0 - \boldsymbol{\eta}_0)$ is asymptotically normal with mean 0.*

In terms of the interval estimation for $\boldsymbol{\eta}_0$, we have two approaches available in general. One is to derive the influence curves of its maps from the distribution of $\{X_i, Y_i, \Delta_i, \mathbf{Z}_i\}$. However, the derivation is algebraically complex. An alternative is through resampling. Here, we adopt nonparametric bootstrap method, which may be justified similarly to Huang and Berry (2006) upon establishing the Hadamard-differentiability of the mapping from the distribution of $\{X, Y, \mathbf{Z}, \Delta\}$ to $\widehat{\boldsymbol{\eta}}_0$ (Gill 1989, Theorem 5; van der Vaart and Wellner 1996, Section 3.9.3).

2.3 Numerical Studies

To evaluate the proposed inference procedures with small and moderate samples, we conducted extensive simulation studies. The performance of our proposed estimation procedure and the compound procedure described at the beginning of Section 3 were compared under the normal copula model. We then investigated the robustness of proposed procedure under misspecification of the error copula. Finally, we applied this procedure to the estimation of lifetime medical cost with data from a lung cancer clinical trial. In our numerical studies, we set $\Delta_i^\circ(\boldsymbol{\alpha})$ in estimating function (2.7) equal to Δ_i.

One challenge to implement the proposed estimation procedure is to find a consistent zero-crossing of nonsmooth estimating function $\boldsymbol{\Psi}_{1,n}(\boldsymbol{\alpha})$. As suggested by Lin, Wei,

and Ying (1998) and Jin, Lin, Wei, and Ying (2003), this root-finding problem can be transformed to a minimization problem, which can be solved via the linear programming technique.

Similarly, $\boldsymbol{\Psi}_{2,n}(\widehat{\boldsymbol{\alpha}}_0, \boldsymbol{\beta}, \theta)$ is also a nonsmooth estimating function in $\boldsymbol{\beta}$. But the linear programming techniques may no longer be applicable. One way to overcome this difficulty is to approximate the discontinuous estimating function with a smooth function. Specifically, we propose to use the sigmoid function $s(x) \equiv 1/(1 + \exp(-x))$ to approximate the indicator function in (2.7). For large $|x|$, $s(x)$ is a good approximation to $I(x \geq 0)$. However, for x around zero, this approximation is not accurate enough. As indicated by Ma and Huang (2005) recently, an effective way to improve the accuracy is by introducing a sequence of positive numbers σ_k converging to 0, and using $s_k(x) = s(x/\sigma_k)$ to approximate $I(x \geq 0)$. We use the following iteration algorithm to find the estimator for $\boldsymbol{\beta}_0$ and θ_0. We first choose the initial value $\sigma_0 = 1/\sqrt{n}$, where n is the sample size, and find an initial estimators $\widehat{\boldsymbol{\beta}}_0^{(0)}$ and $\widehat{\theta}_0^{(0)}$ from the resulting continuous estimating function by using the Newton-Raphson type algorithm. By updating $\sigma_k = 0.5 \times \sigma_{k-1} (k = 1, 2 \ldots,)$ in the kth iteration, we may obtain the corresponding estimators $\widehat{\boldsymbol{\beta}}_0^{(k)}$ and $\widehat{\theta}_0^{(k)}$. The iteration stops when $\widehat{\boldsymbol{\beta}}_0^{(k)}$ and $\widehat{\theta}_0^{(k)}$ converge. In our simulations, this iteration algorithm performs well.

2.3.1 Monte-Carlo Simulations Under Normal Copula Model

A number of settings were investigated under moderate censoring, with single- and double-covariate. Both the transformation $G_T(.)$ on survival time and the transformation $G_U(.)$ on lifetime medical cost in (2.2) were specified as logarithm. The marginal baseline distributions of survival time and lifetime medical cost were set as either the standard lognormal distribution or the standard exponential distribution. Given the baseline

marginal distributions $F_{\varepsilon^t}(.)$ and $F_{\varepsilon^u}(.)$, the corresponding random samples were generated by marginally transforming random samples from the standard bivariate normal distribution; that is, $\varepsilon_i^t = F_{\varepsilon^t}^{-1}\{\Phi(S_i)\}$ and $\varepsilon_i^u = F_{\varepsilon^u}^{-1}\{\Phi(V_i)\}$, where $(S_i, V_i) \sim \texttt{BN}(\rho_0)$. In each case, various association levels with $\rho_0 = 0.8, 0.4, 0, -0.4, -0.8$ were considered, corresponding to strong positive, moderate positive, independent, moderate negative and strong negative associations between the two outcomes. Note that different associations as considered would only impact the estimation of $\boldsymbol{\beta}_0$ and ρ_0, but not that of $\boldsymbol{\alpha}_0$.

For comparison, we investigated the performance of both the proposed copula-based semiparametric estimation procedure and the compound procedure. Both the Gehan weight function and log-rank weight function were considered for the estimating function (2.3) for $\boldsymbol{\alpha}_0$, the regression coefficients on survival time. In the compound procedure, we adopted the same weight functions in the two estimating functions with the calibration regression model, following the suggestion of Huang (2002). Two sets of simulations are reported.

With a single covariate, the first set involved four different combinations of the marginal baseline distributions on time scale and cost scale, either the standard exponential or the standard lognormal. The covariate Z was uniformly distributed in $[-1, 1]$. We set $\alpha_0 = 1$ and $\beta_0 = 1$; other values for (α_0, β_0) were also considered, resulting in similar results. The follow-up time T was subject to right censoring. The censoring time C, which was independent of T and Z, followed an exponential distribution with rate 0.19 but curtailed at 2.2 or with rate 0.12 but curtailed at 3.3, according to exponential or lognormal baseline distribution on time scale. Thus, the top 15% of T was censored and the overall censoring rate was approximately 25%. The sample size was set to 100.

Table 2.1 presents the simulation results for the first set, with $\rho_0 = 0$, 0.4 and 0.8. Results for $\rho_0 = -0.4$ and -0.8, which were not shown here, exhibited features similar

to the results for $\rho_0 = 0.4$ and 0.8, respectively. Each scenario was simulated with 1000 replications. The nonparametric bootstrap with size 100 was used to obtain estimated standard deviation and 95% Wald-type confidence interval. As shown, the proposed estimator for β_0 from copula-based semiparametric estimation procedure is essentially unbiased and the estimated standard deviation tracks the empirical standard deviation well. The coverage probability of 95% Wald-type confidence interval is fairly accurate. By comparing the empirical standard deviation of the calibration regression estimator for β_0 with that of the copula-based semiparametric estimator for β_0, we could find that the latter one is more efficient in most cases. This might be expected since we imposed a parametric normal copula structure for the two outcomes in the copula-based semiparametric regression model. With stronger association, we observed more efficiency gain. At the same time, the magnitude of efficiency gain does depend on the marginal distributions. For example, with exponential baseline distribution on time scale and lognormal baseline distribution on cost scale, the efficiency gain is bigger as compared to the case when the baseline distributions on both time scale and cost scale are lognormal. With Gehan weighted and log-rank weighted estimator $\widehat{\alpha}_0$, the performances of the resultant copula-based semiparametric estimators for β_0 are quite similar. For calibration regression, the results based on log-rank weight show better efficiency than their counterparts based on Gehan weight. In the following, we will only present the results based on log-rank weight.

Table 2.2 reports the simulation results for the estimation of $\rho_0 = 0$, 0.4 and 0.8 in the first set. Four different scenarios with exponential or lognormal baseline distribution on time scale or cost scale are demonstrated. The corresponding mean and median of marginal lifetime medical cost distribution for the specific group with the covariate value 0.5 are also presented. As shown, the proposed copula-based semiparametric

Table 2.1: *Simulation Summary Statistics for Regression Coefficients With Single Covariate, and Marginal Exponential or Lognormal Baseline Distribution on Time Scale and on Cost Scale*

			Lognormal Baseline on Cost Scale						*Exponential Baseline on Cost Scale*					
	Weight	$\widehat{\alpha}_0$	$\rho_0 = 0$		$\rho_0 = 0.4$		$\rho_0 = 0.8$		$\rho_0 = 0$		$\rho_0 = 0.4$		$\rho_0 = 0.8$	
			$\widehat{\beta}_0$		$\widehat{\beta}_0$		$\widehat{\beta}_0$		$\widehat{\beta}_0$		$\widehat{\beta}_0$		$\widehat{\beta}_0$	
			Exponential Baseline on Time Scale											
B	Gehan	0	13	11	7	9	-2	4	19	12	13	10	0	4
	log-rank	3	13	12	7	10	-2	7	20	12	13	12	6	2
D	Gehan	237	240	226	242	217	252	198	311	258	357	260	405	249
	log-rank	219	220	227	222	217	227	191	283	259	304	259	322	235
$\widehat{D}$	Gehan	243	234	224	238	219	251	203	298	257	342	264	388	257
	log-rank	220	216	225	219	217	228	194	275	258	297	262	316	240
C	Gehan	95.5	93.8	94.3	93.4	94.7	94.3	94.6	94.2	94.5	93.2	95.3	94.2	95.0
	log-rank	94.7	94.9	94.4	94.1	94.7	94.3	94.5	94.3	94.5	94.2	95.0	94.1	95.0
			Lognormal Baseline on Time Scale											
B	Gehan	-4	13	10	7	8	-1	4	18	11	13	10	2	7
	log-rank	-2	13	10	7	10	-2	6	19	11	13	11	1	10
D	Gehan	191	236	222	236	216	238	202	306	253	347	258	385	252
	log-rank	202	216	223	216	219	215	209	277	254	296	261	310	261
$\widehat{D}$	Gehan	194	230	222	233	219	237	206	294	255	334	264	375	260
	log-rank	201	213	224	214	221	216	211	272	257	290	267	309	266
C	Gehan	95.4	93.9	94.6	93.1	94.7	94.2	94.8	94.4	95.2	93.4	95.4	94.0	95.0
	log-rank	93.6	94.9	94.9	94.2	94.7	95.1	94.1	94.4	95.3	94.6	95.0	94.8	94.2

NOTE: Two columns under $\widehat{\beta}_0$: calibration regression (left) and the proposed procedure. B: Empirical bias ($\times 10^3$); D: Empirical standard deviation ($\times 10^3$); $\widehat{D}$: Average of estimated standard deviation ($\times 10^3$); C: Empirical coverage (%) of the 95% Wald-type confidence interval.

estimator for ρ_0 is essentially unbiased. The proposed mean estimator is essentially unbiased when the marginal baseline distribution on cost scale is exponential. With a long-tailed marginal distribution on cost scale, such as lognormal, especially in the case of a strong positive association between lifetime medical cost and survival time, the bias of the proposed mean estimator is relatively large. This might be expected since the mean is sensitive to a long tail, which is largely censored under a strong positive association. The performance of the proposed median estimator is satisfactory under both types of baseline distributions on cost scale. The estimated standard deviation tracks the empirical standard deviation well, and 95% Wald-type confidence intervals achieve reasonably accurate coverage probability except for those of $\widehat{\mu}$ in the case of lognormal baseline distribution on cost scale and strong positive association. The lower coverage phenomena here are mainly caused by the skewness the cost distribution. The efficiency gain of copula-based semiparametric estimators over the estimators based on compound procedure is similar to our findings from regression coefficients estimation in Table 2.1 above.

The second set mimicked a two-arm clinical trial, with two independent covariates. The first covariate took values of 1 and -1 with equal probabilities, and the second one followed a uniform distribution in $[-1, 1]$. We set $(\boldsymbol{\alpha}_0^T, \boldsymbol{\beta}_0^T)^T = (1, 1, 1, 1)$. The standard exponential baseline distribution on time scale and the standard lognormal baseline distribution on cost scale were specified. The censoring time, independent of the covariates, had an exponentially distribution with rate 0.18 but curtailed at 3.5. Thus, the top 15% of survival time was censored and the overall censoring rate was 25%. Sample size 100 was studied.

Table 2.3 presents the simulation summaries for the second set with $\rho_0 = 0$, 0.4 and 0.8, where each scenario was simulated with 1,000 replications. Simulation results of

Table 2.2: *Simulation Results for Normal Copula Association Parameter ρ_0, Mean μ and Median ν of Lifetime Medical Cost Distribution of Defined Group With the Covariate Value 0.5*

	$\rho_0 = 0$						$\rho_0 = 0.4$						$\rho_0 = 0.8$					
	$\widehat{\rho}_0$		$\widehat{\mu}$		$\widehat{\nu}$		$\widehat{\rho}_0$		$\widehat{\mu}$		$\widehat{\nu}$		$\widehat{\rho}_0$		$\widehat{\mu}$		$\widehat{\nu}$	
Exponential Baseline on Time Scale and Lognormal Baseline on Cost Scale																		
B	3	4	27	31	43	43	-9	-4	-5	7	20	31	-14	-8	-109	-101	15	28
D	143	140	603	617	356	364	118	115	648	653	342	344	49	48	644	575	318	301
$\widehat{D}$	139	140	571	593	369	380	117	118	615	622	365	369	54	53	610	532	347	326
C	92.5	93.5	91.7	92.2	95.0	95.1	93.2	94.0	90.2	90.3	94.2	94.9	96.9	95.3	86.0	86.0	94.6	95.1
Exponential Baseline on Time Scale and Exponential Baseline on Cost Scale																		
B	6	4	41	25	32	29	-8	-4	40	24	17	21	-15	-8	18	-5	14	24
D	144	140	377	352	291	284	119	115	433	382	296	276	50	48	493	364	291	250
$\widehat{D}$	141	140	382	354	304	292	117	118	457	394	316	293	55	53	546	379	321	270
C	92.0	93.4	94.5	94.1	94.7	95.0	93.5	93.6	92.9	93.5	94.2	94.6	96.9	95.2	91.9	92.3	94.6	95.0
Lognormal Baseline on Time Scale and Lognormal Baseline on Cost Scale																		
B	4	5	25	28	40	41	-9	-4	-9	6	18	27	-15	-8	-138	-121	15	16
D	147	143	590	603	351	361	120	117	635	649	336	339	49	48	617	572	310	311
$\widehat{D}$	137	138	562	588	363	375	118	119	597	623	356	371	54	53	565	547	333	338
C	92.4	93.1	92.0	91.8	94.8	95.0	93.1	93.6	90.0	90.1	93.9	95.0	96.6	95.3	85.1	86.2	95.0	94.5
Lognormal Baseline on Time Scale and Exponential Baseline on Cost Scale																		
B	7	5	38	23	29	27	-7	-4	35	23	17	18	-16	-8	2	-5	15	29
D	147	143	366	342	287	282	120	117	416	377	291	271	50	49	468	380	279	257
$\widehat{D}$	142	142	373	350	299	289	118	119	435	396	305	294	55	53	494	405	303	281
C	92.0	93.2	94.5	93.8	94.5	94.6	93.4	93.7	92.6	92.2	94.2	94.4	96.9	95.0	91.7	91.7	94.8	94.4

NOTE: Two columns under each estimator correspond to compound procedure (left) and the proposed procedure (right), respectively. The symbols B, D, $\widehat{D}$ and C have the same meanings as those in Table 2.1.

$\rho_0 = -0.4$ and -0.8 are similar to those of $\rho_0 = 0.4$ and 0.8 respectively, and thus are omitted. As shown, the proposed estimators are virtually unbiased, the nonparametric bootstrap based standard deviation estimation procedure behaves well and the coverage probability of their Wald-type confidence interval is fairly accurate. By comparing the empirical standard deviation of the calibration regression estimator for $\boldsymbol{\beta}_0$ with those of the copula-based semiparametric estimator for $\boldsymbol{\beta}_0$, we could find that the latter has modest efficiency gain over the former when the two outcomes are weakly associated; with stronger association, efficiency gain is more significant.

2.3.2 Monte-Carlo Simulations Under Nonnormal Copulas

We considered Clayton's and Frank's families (Shih and Louis 1995). Copulas in these two families generally differ from normal copulas except when survival time and lifetime medical cost are independent. We adopted the same simulation scenarios as described in the second set above except for the copula. These two nonnormal copula families as well as normal copula are all governed by a single parameter, which has a one-to-one mapping with the association measure Kendall's tau within each family. For comparison, corresponding to each nonzero ρ value considered in the three sets above, we chose a nonnormal copula with the same Kendall's tau. Only positive association was considered for Claytons copula, given its limitation in accommodating negative association. The results are presented in Table 2.4, including the mean and median of marginal lifetime medical cost distribution for the group with covariates $(Z_1, Z_2) = (1, 0.5)$. The results with Frank's copula corresponding to $\rho_0 = -0.4$ and -0.8 are omitted, since they are similar to those from Frank's copula corresponding to $\rho_0 = 0.4$ and 0.8 respectively. As shown, the proposed copula-based semiparametric estimator for $\boldsymbol{\beta}_0$ is fairly robust against misspecified copula structure. Nonetheless, with Clayton's family, the bias for the

Table 2.3: *Simulation Summary Statistics With Double Covariates, Marginal Exponential Baseline Distribution on Time Scale and Marginal Lognormal Baseline Distribution on Cost Scale*

			Calibration Regression		*Copula-based Semiparametric*	
	$\widehat{\alpha}_{01}$	$\widehat{\alpha}_{02}$	$\widehat{\beta}_{01}$	$\widehat{\beta}_{02}$	$\widehat{\beta}_{01}$	$\widehat{\beta}_{02}$
				$\rho_0 = 0.8$		
B	3	15	-1	11	-5	13
D	126	208	136	234	113	189
$\widehat{D}$	129	218	135	240	125	202
C	94.5	95.2	94.1	95.3	95.1	96.1
				$\rho_0 = 0.4$		
B			-4	9	-7	8
D			130	217	127	206
$\widehat{D}$			129	222	129	217
C			93.7	94.3	94.4	95.5
				$\rho_0 = 0$		
B			-5	6	-6	6
D			127	211	131	214
$\widehat{D}$			128	215	133	221
C			94.0	94.6	94.4	95.0

NOTE: $(\widehat{\alpha}_{01}, \widehat{\alpha}_{02})$ and $(\widehat{\beta}_{01}, \widehat{\beta}_{02})$ are estimators for regression coefficients on time scale and cost scale, respectively. ρ_0 is the association parameter of the baseline distribution under normal copula. The symbols B, D, $\widehat{D}$ and C have the same meanings as those in Table 2.1.

mean estimator is large, while the performance of the median estimator is reasonable. As to Frank's family, the performances of both mean and median estimators are satisfactory.

2.3.3 Application to a Lung Cancer Trial

Lung cancer is the most common cancer in terms of mortality for both men and women in the United States. It is estimated to cost society $ 4.7 billion annually in direct medical cost (Brown, Lipscomb, and Snyder 2001). Treating lung cancer with the minimum economic burden possible is important. Our study was motivated by the analysis of lifetime medical cost in a trial conducted by the Southwest Oncology Group (SWOG). This randomized SWOG trial was designed to compare paclitaxel plus carboplatin versus vinorelbine plus cisplatin treatments for patients with advanced non-small-cell lung cancer (Kelly et al. 2001). Whereas survival time was the primary outcome, one secondary endpoint was resource utilization which consisted of supportive care medications, blood products, medical procedures, protocol and non-protocol related treatments, and medical care inpatient days or outpatient visits. Cost was assigned to each resource using national databases with adjustment to 1998 US dollars following the medical care component of the Consumer Price Index.

The cost data were collected every 3 months during the first half year, and every half year after that, until the 2 year study duration was reached. Participants with insufficient documentation or no follow-up for cost data collection were excluded, and 183 (49.7%) of the remaining 368 participants in the current analysis were randomized to receive paclitaxel plus carboplatin treatment. The median follow-up time was 6.7 months and 31.0% participants were censored. The survival rate at 24 months among all participants was estimated to be 20.5%. With the natural logarithmic transformation on both survival time and lifetime medical cost, we fitted the proposed copula-based

Table 2.4: *Simulation Summary Statistics With Proposed Methods, Double Covariates, Marginal Exponential Baseline Distribution on Time Scale and Marginal Lognormal Baseline Distribution on Cost Scale*

			Compound Procedure				*Proposed Procedure*			
	$\widehat{\alpha}_{01}$	$\widehat{\alpha}_{02}$	$\widehat{\beta}_{01}$	$\widehat{\beta}_{02}$	$\widehat{\mu}$	$\widehat{\nu}$	$\widehat{\beta}_{01}$	$\widehat{\beta}_{02}$	$\widehat{\mu}$	$\widehat{\nu}$
					Clayton's copula corresponding to $\rho_0 = 0.8$					
B	1	10	-3	2	-322	13	-19	-10	-630	-95
D	134	210	127	223	2182	1136	111	194	1762	1055
$\widehat{D}$	130	219	125	228	2244	1229	113	203	1833	1112
C	93.2	95.3	93.9	94.3	87.1	92.7	95.3	96.0	84.9	93.2
					Clayton's copula corresponding to $\rho_0 = 0.4$					
B			-1	0	-315	62	-13	-5	-455	-13
D			119	206	2045	1178	118	209	1996	1189
$\widehat{D}$			120	209	2052	1241	123	211	1966	1228
C			94.9	94.9	88.5	94.7	95.3	94.8	87.3	92.6
					Frank's copula corresponding to $\rho_0 = 0.8$					
B			-2	6	301	93	-7	-7	126	24
D			140	235	2573	1284	114	190	2157	1111
$\widehat{D}$			137	242	2647	1409	116	205	2151	1213
C			93.6	94.4	92.0	92.8	95.3	96.7	92.1	94.3
					Frank's copula corresponding to $\rho_0 = 0.4$					
B			0	1	182	106	-4	-8	161	85
D			129	215	2338	1282	124	214	2365	1286
$\widehat{D}$			129	221	2317	1377	128	216	2259	1345
C			94.7	94.4	92.2	94.7	95.4	94.7	91.2	94.0

NOTE: $(\widehat{\alpha}_{01}, \widehat{\alpha}_{02})$ and $(\widehat{\beta}_{01}, \widehat{\beta}_{02})$ are estimators for regression coefficients on time scale and cost scale, respectively. ρ_0 is the normal copula association parameter. $\widehat{\mu}$ and $\widehat{\nu}$ are estimators for the mean and median of marginal lifetime medical cost distribution for the defined group with covariates $(Z_1, Z_2) = (1, 0.5)$. The symbols B, D, $\widehat{D}$ and C have the same meanings as those in Table 2.1.

semiparametric regression model (2.2) to the dataset. Three covariates of interest here were the treatment indicator, lactate dehydrogenase (LDH) normality indicator, and age at enrollment. If a patient's LDH is less than or equal to the upper limit of the normal range for serum LDH, then it has a normal LDH; otherwise, it has an abnormal LDH. The number of patients with normal LDH is 223, corresponding to 60.6% of the 368 participants. The median age of patients is at 62.0 years with the range from 32.3 to 83.8. Huang (2002) analyzed the same dataset using the calibration regression model. We repeated the analysis for comparison purposes. In regard to estimation of standard error, we used the nonparametric bootstrap method with a bootstrap size of 500. The corresponding 95% Wald-type confidence intervals were also constructed.

Table 2.5 shows the analysis results based on the proposed method and the calibration regression. The results based on calibration regression are very similar to those in Huang (2002). There was little difference in survival time between the two treatments. However, paclitaxel plus carboplatin led to significantly higher lifetime medical cost. Based on the copula-based semiparametric estimator, lifetime medical cost of paclitaxel plus carboplatin was $\exp(0.285) = 1.330$ times as much as vinorelbine plus cisplatin after adjustment for LDH and age. The result is consistent with that of the calibration regression estimator. In contrast, patients with normal LDH level tended to have better survival time but similar lifetime medical cost, compared with those with abnormal LDH level. Finally, age showed little effect on both outcomes. By comparing the estimation of standard error, we could find that those of the semiparamatric copula estimators were smaller than their counterparts of calibration regression estimators. This finding was similar to that from the simulation studies, where the semiparamatric copula estimators were shown to be more efficient than their calibration regression counterparts in most cases.

Table 2.5: *Analysis Results of the SWOG Lung Cancer Data*

	Estimation results for regression coefficients		
	tx	LDH	Age
		Estimator for $\boldsymbol{\alpha}_0$	
Estimate	.0222	.6330	-.0059
SE	.1378	.1454	.0070
95%CI	(-.2479, .2923)	(.3480, .9179)	(-.0195, .0078)
		Calibration regression estimator for $\boldsymbol{\beta}_0$	
Estimate	.3379	.1399	-.0054
SE	.1161	.1237	.0060
95%CI	(.1104, .5654)	(-.1025, .3823)	(-.0172, .0064)
		Copula-based semiparametric estimator for $\boldsymbol{\beta}_0$	
Estimate	.2850	.1794	-.0015
SE	.1052	.1076	.0050
95%CI	(.0789, .4911)	(-.0313, .3903)	(-.0114, .0084)

NOTE: tx: treatment indicator, with Vinorelbine plus Cisplatin as the reference; LDH: lactate dehydrogenase normality indicator; *SE*: the estimated standard deviation; *95%CI*: the 95% Wald-type confidence interval.

The sempiparametric copula estimator for ρ_0 was 0.775, with 95% confidence interval $(0.714, 0.836)$, showing a strong association between the baseline distribution of lifetime medical cost and survival time. Based on the proposed model, the marginal lifetime medical cost distribution for any specific group could be estimated. As an example, we considered a group of patients of age 62 (the median age of all participants), who had normal LDH and received paclitaxel plus carboplatin treatment. Figure 2.1 displays the estimated marginal survival function of lifetime medical cost for this specific group, using copula-based semiparametric estimators. The estimated mean lifetime medical cost of this group was \$61 435, with 95% confidence interval (50 522, 72 349), whereas the estimated median cost is \$52 493, with 95% confidence interval (44 895, 60 092). The substantial difference between the mean and median indicated highly positive skewness of this distribution.

2.4 Discussion

We have proposed a copula-based semiparametric regression model for lifetime medical cost with incomplete follow-up data. This conceptually simple regression model is semiparametric in the sense that the marginal error distribution of both lifetime medical cost and survival time are completely unspecified. As opposed to time-restricted quantities with artificial limits which are often considered in the literature, our proposed model is targeting at underlying distribution of interest. Also, the inference procedure accommodates the conditional independence censoring mechanism, a relatively weak censorship assumption.

The proposed copula-based semiparametric regression model strikes a balance between model identifiability and robustness. The calibration regression model in Huang

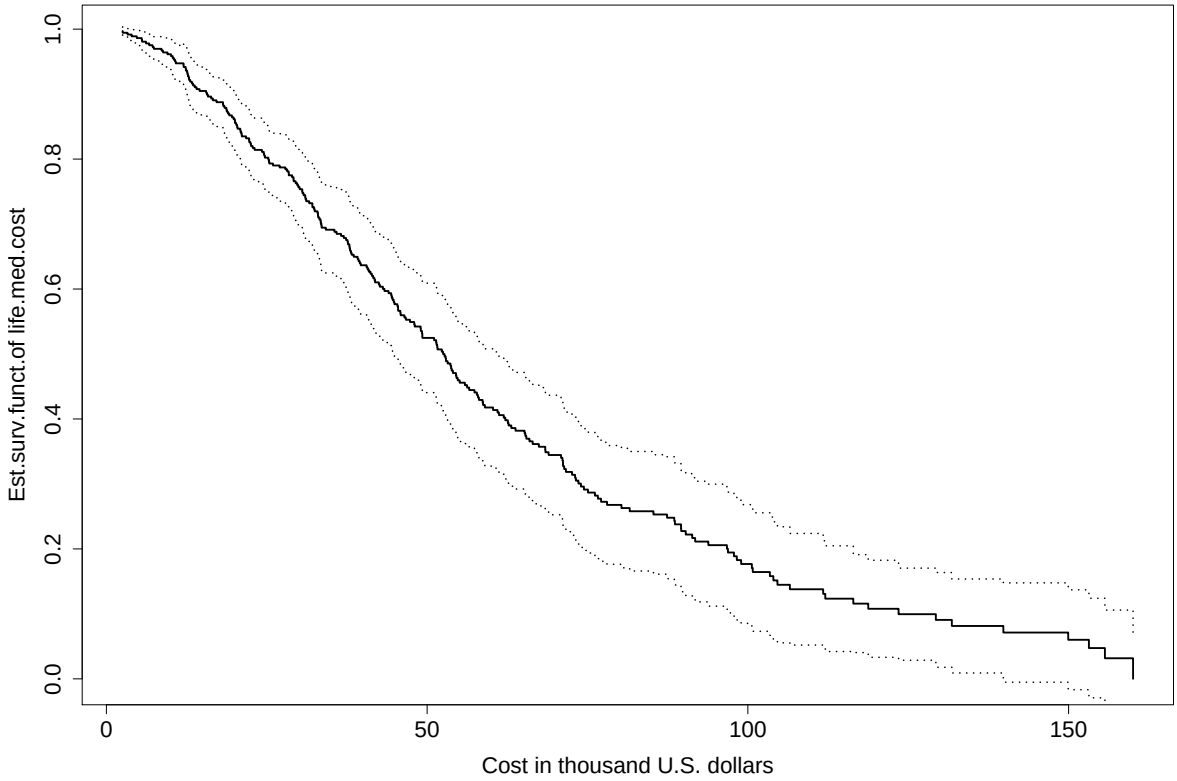

Figure 2.1: Southwest Oncology Group study. Estimated survival function of lifetime medical cost for a group of age 62, who had normal LDH and received the paclitaxel plus carboplatin therapy, along with pointwise 95% Wald-type confidence intervals.

(2002) is more robust, but we do not have the identifiability of lifetime medical cost distribution. On the other hand, it is generally undesirable to use a method which requires many assumptions. Through parameterizing the association between two marginal error distributions by normal copula function but leaving the marginal error distributions completely unspecified, we are able to identify and estimate the marginal distribution of lifetime medical cost. This copula-based semiparametric regression model is practically most useful in medical studies where only a small portion of patients survive beyond the study duration. In such a situation, the selected normal copula structure is largely testable except for the small tail portion (Huang and Berry 2006) and can serve the purpose of appropriately characterizing the association structure as well as leading to the identifiability of the marginal distribution of lifetime medical cost.

The copula-based semiparametric inference procedure offered is computationally easy and is shown to yield consistent and asymptotically normal estimators. Our simulation study shows that the proposed method performs well in samples of moderate size. Compared to the compound procedure, the copula-based semiparametric approach is more efficient in most cases, when the two outcomes are associated. Nevertheless, as we have illustrated in simulation study, the efficiency gain of the copula-based semiparametric approach is modest under certain baseline distribution, even when the two outcomes are strongly associated. As shown in our simulation studies, the estimation of regression coefficients from the copula-based semiparametric regression model is not quite sensitive to the misspecification of the error copula structure. The inference procedure appears to be quite robust with Frank's family but less robust with Clayton's family. These results suggest the importance of model checking in practice. Goodness-of-fit test under the proposed copula-based semiparametric regression model for censored lifetime medical cost is under development. The basic idea is along the same line as the goodness-of-fit

test proposed by Huang and Berry (2006) in the one-sample problem.

Our proposed model only requires and exploits uncensored lifetime medical cost in addition to the standard survival data, and thus can be applied to a wide range of cost data collection schemes. In practice, additional data might be available. For example, the cost accumulation process was observed at scheduled visits in the study we discussed in Section 2.3.3, or one may observe accumulated cost at censoring time for each censored individual. In this case, potentially, there is room for efficiency improvement by taking advantage of the additional data. Unfortunately, for this purpose, it seems necessary to model the stochastic cost accumulation process. This task would be difficult if not impossible, given that it is already challenging to model the highly skewed cost distribution.

2.5 Proofs

2.5.1 Consistency

Suppose the semiparametric copula model (2.2) and the censoring mechanism (2.1) hold. The following regularity conditions are adopted:

C1. The parameter space of $(\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T, \theta)^T$, say, $\mathcal{A} \times \mathcal{B} \times \Theta$ is compact and the true parameter $(\boldsymbol{\alpha}_0^T, \boldsymbol{\beta}_0^T, \theta_0)^T$ is an interior point of $\mathcal{A} \times \mathcal{B} \times \Theta$.

C2. The covariates $\mathbf{Z}$ are bounded.

C3. Censoring time C has a bounded density function.

C4. $\mathrm{E}\,[\,|X|^r\,] < \infty$ for some $r > 0$.

C5. The error density f_{ε^t}, f_{ε^u} and their derivatives f'_{ε^t}, f'_{ε^u} are bounded, respectively, and $\int (f'_{\varepsilon^t}(s)/f_{\varepsilon^t}(s))^2 f_{\varepsilon^t}(s)\,\mathrm{d}s < \infty$, $\int (f'_{\varepsilon^u}(s)/f_{\varepsilon^u}(s))^2 f_{\varepsilon^u}(s)\,\mathrm{d}s < \infty$, respectively.

C6. The partial derivative matrix $\partial \boldsymbol{\Psi}_2 (\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta) / \partial (\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T, \theta)^T$ evaluated at $(\boldsymbol{\alpha}_0^T, \boldsymbol{\beta}_0^T, \theta_0)^T$ is not singular, where $\boldsymbol{\Psi}_2 (\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta)$ is a deterministic function defined in Lemma 2.2.

To show the consistency of the estimator, we may first prove the uniform consistency of the estimating function $\boldsymbol{\Psi}_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta)$ in (2.8). To this end, we need the following Lemmas 2.1 and 2.2. We will use $\| \cdot \|$ to denote the Euclidean norm in $\mathbb{R}^k$.

Lemma 2.1. *Let δ_n be any positive sequence converging to 0, and $\widehat{F}_T(t\,; \boldsymbol{\alpha})$ be the Kaplan-Meier type estimator of the residual $\varepsilon^x(\boldsymbol{\alpha})$. Suppose for some $\tau_b < \infty$, $\Pr\{\varepsilon^x(\boldsymbol{\alpha}_0) > \tau_b\} > 0$. Under conditions* C1-C5*, $\widehat{F}_T(t\,; \boldsymbol{\alpha})$ converges to $F_T(t\,; \boldsymbol{\alpha}_0)$ uniformly in $t \leq \tau_b$ and $\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_0\| \leq \delta_n$, almost surely.*

Proof. First, we claim that $\widehat{F}_T(t\,; \boldsymbol{\alpha}_0)$ converges uniformly to $F_T(t\,; \boldsymbol{\alpha}_0)$ on $t \leq \tau_b$ almost surely. By considering the Kaplan-Meier estimator $\widehat{F}_T(t\,; \boldsymbol{\alpha}_0)$ of the residual $\varepsilon^x(\boldsymbol{\alpha}_0)$ as a functional of the empirical distribution of the data $\{\varepsilon_i^x(\boldsymbol{\alpha}_0), \Delta_i\}$, $i = 1, \ldots, n$, we may study the uniform consistency of $\widehat{F}_T(t\,; \boldsymbol{\alpha})$ by using empirical process theory.

Notice that the Nelson-Aalen estimator $\widehat{\Lambda}(t; \boldsymbol{\alpha}_0)$ based on the data $\{\varepsilon_i^x(\boldsymbol{\alpha}_0), \Delta_i\}$, $i = 1, \ldots, n$ is defined as

$$\widehat{\Lambda}(t; \boldsymbol{\alpha}_0) \equiv \int_{-\infty}^{t} \frac{\mathrm{d} \sum_{i=1}^n N_i(s; \boldsymbol{\alpha}_0)}{\sum_{i=1}^n R_i(s; \boldsymbol{\alpha}_0)} = \int_{-\infty}^{t} \frac{\mathrm{d}\, \widehat{\mathrm{E}}_n \left[I\{\varepsilon^x(\boldsymbol{\alpha}_0) \leq s\}\, \Delta \right]}{\widehat{\mathrm{E}}_n \left[I\{\varepsilon^x(\boldsymbol{\alpha}_0) \geq s\} \right]}, \tag{2.9}$$

where $N_i(t; \boldsymbol{\alpha})$ and $R_i(t; \boldsymbol{\alpha})$ are defined before the equation (2.3) and $\widehat{\mathrm{E}}_n$ represents sample empirical mean based on n independent and identically distributed samples. The Kaplan-Mier estimator $\widehat{F}_T(t\,; \boldsymbol{\alpha}_0)$ is related to the Nelson-Aalen estimator $\widehat{\Lambda}(t; \boldsymbol{\alpha}_0)$ by the formula

$$1 - \widehat{F}_T(t\,; \boldsymbol{\alpha}_0) = \prod\nolimits_{-\infty}^{t} [1 - \mathrm{d}\widehat{\Lambda}(s; \boldsymbol{\alpha}_0)], \tag{2.10}$$

where the symbol $\prod$ represents the product integration (Gill and Johansen 1990). Now

we follow the idea of Gill (1994, Section 6) to prove the uniform consistency of $\widehat{F}_T(t\,;\boldsymbol{\alpha})$.

From equation (2.9), we may find that the Nelson-Aalen estimator $\widehat{\Lambda}(t;\boldsymbol{\alpha}_0)$ is a functional of two empirical processes, $\mathbb{F}_1^{(n)} \equiv \widehat{\mathrm{E}}_n\left[I\{\varepsilon^x(\boldsymbol{\alpha}_0) \leq s\}\,\Delta\right]$ and $\mathbb{F}_2^{(n)} \equiv \widehat{\mathrm{E}}_n\left[I\{\varepsilon^x(\boldsymbol{\alpha}_0) \geq s\}\right]$. Define the corresponding theoretical parts to be $F_1 \equiv \mathrm{E}\left[I\{\varepsilon^x(\boldsymbol{\alpha}_0) \leq s\}\,\Delta\right]$ and $F_2 \equiv \mathrm{E}\left[I\{\varepsilon^x(\boldsymbol{\alpha}_0) \geq s\right]$, which are the expectation of $\mathbb{F}_1^{(n)}$ and $\mathbb{F}_2^{(n)}$. The *Glivenko-Cantelli* Theorem (e.g. van der Vaart (1998), P. 266) ensures the uniform convergence for both $\mathbb{F}_1^{(n)}$ and $\mathbb{F}_2^{(n)}$.

Based on equation (2.9) and (2.10), we can write

$$1 - \widehat{F}_T(t\,;\boldsymbol{\alpha}) = \prod\nolimits_{-\infty}^{t} \left(1 - \frac{\mathrm{d}\,\mathbb{F}_1^{(n)}}{\mathbb{F}_2^{(n)}}\right), \tag{2.11}$$

which can be considered as the composition of three mappings

$$(\mathbb{F}_1^{(n)}, \mathbb{F}_2^{(n)}) \;\mapsto\; \left(\mathbb{F}_1^{(n)}, \frac{1}{\mathbb{F}_2^{(n)}}\right) \mapsto \int \left(\frac{1}{\mathbb{F}_2^{(n)}}\right) \mathrm{d}\,\mathbb{F}_1^{(n)} \mapsto \prod \left(1 - \mathrm{d}\left(\int \left(\frac{1}{\mathbb{F}_2^{(n)}}\right) \mathrm{d}\,\mathbb{F}_1^{(n)}\right)\right).$$

If we consider the mappings as applying to functions on the interval $(-\infty,\, \tau_b]$, we can see that the first mapping is supremum norm continuous at pairs of functions since $\mathbb{F}_2^{(n)}$ is uniformly bounded away from zero on $(-\infty,\, \tau_b]$. The second mapping, which is an ordinary integration, is supremum norm continuous at functions of uniformly bounded variation. The same result can be established for the third mapping, the product integration, due to Theorem 7 in Gill and Johansen (1990). Thus, the mapping from $(\mathbb{F}_1^{(n)},\, \mathbb{F}_2^{(n)})$ to $1 - \widehat{F}_T(t\,;\boldsymbol{\alpha})$ is supremum norm continuous. Applied to the expectation of $(\mathbb{F}_1^{(n)}, \mathbb{F}_2^{(n)})$, i.e., (F_1, F_2) on the interval $(-\infty,\, \tau_b]$, the mappings yields the limit function of $1 - \widehat{F}_T(t\,;\boldsymbol{\alpha}_0)$, say, $1 - F_T(t\,;\boldsymbol{\alpha}_0)$. The Glivenko-Cantelli properties of $(\mathbb{F}_1^{(n)}, \mathbb{F}_2^{(n)})$ and the supremun norm continuity of the mappings therefore give us the uniform convergence of the Kaplan-Meier estimator $\widehat{F}_T(t\,;\boldsymbol{\alpha}_0)$ to its limit function $F_T(t\,;\boldsymbol{\alpha}_0)$.

Let $F_T(t, \boldsymbol{\alpha})$ be the limit function of $\widehat{F}_T(t, \boldsymbol{\alpha})$. Then, under Conditions *C1-C5* and from theorem 2 of Yang (1997), we have

$$\sup_{t \le \tau_b,\, \|\boldsymbol{\alpha}-\boldsymbol{\alpha}_0\| \le K n^{-r}} \left| \widehat{F}_T(t, \boldsymbol{\alpha}) \,-\, F_T(t, \boldsymbol{\alpha}) \,-\, \widehat{F}_T(t, \boldsymbol{\alpha}_0) \,+\, F_T(t, \boldsymbol{\alpha}_0) \right| \,=\, o(n^{-1/2-r/2+\epsilon}) \tag{2.12}$$

almost surely, for every $r \in [0, 1)$, $K > 0$ and $\epsilon > 0$, given that $\boldsymbol{\alpha} \in \mathcal{A}$. By the continuity of $F_T(t, \boldsymbol{\alpha})$ in $\boldsymbol{\alpha}$ around $\boldsymbol{\alpha}_0$ and the triangle inequality, we have

$$\sup_{t \le \tau_b,\, \|\boldsymbol{\alpha}-\boldsymbol{\alpha}_0\| \le \delta_n} |\widehat{F}_T(t\,; \boldsymbol{\alpha}) - F_T(t\,; \boldsymbol{\alpha}_0)| \to 0, \quad \text{almost surely.}$$

□

Besides τ_b above, suppose for some $\tau_a > -\infty$, $\Pr\left(\varepsilon^x(\boldsymbol{\alpha}_0) < \tau_a\right) > 0$. Since $\widehat{H}_T(t; \boldsymbol{\alpha}) \equiv \Phi^{-1}\{\widehat{F}_T(t\,; \boldsymbol{\alpha})\}$ and $H_T(t; \boldsymbol{\alpha}) \equiv \Phi^{-1}\{F_T(t\,; \boldsymbol{\alpha})\}$, and the mapping $F_T(t\,; \boldsymbol{\alpha}_0) \mapsto H_T(t\,; \boldsymbol{\alpha}_0)$ on the interval $t \in [\tau_a,\, \tau_b]$ is bounded, continuous and monotone, the following is an immediate consequence of Lemma 2.1.

Corollary 2.1. *Let δ_n be any positive sequence converging to 0. Under conditions* C1-C5*, $\widehat{H}_T(t; \boldsymbol{\alpha})$ converges to $H_T(t; \boldsymbol{\alpha}_0)$ uniformly in $t \in [\tau_a,\, \tau_b]$ and $\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_0\| \le \delta_n$, almost surely.*

Let $\boldsymbol{\Psi}^*_{2,n}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta)$ be the same as $\boldsymbol{\Psi}_{2,n}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta)$ as defined in (2.7) except that $\widehat{W}_{ij}(\boldsymbol{\alpha})$ is replaced by $W_{ij}(\boldsymbol{\alpha})$. Lemma 2.2 below states the uniform consistency of $\boldsymbol{\Psi}^*_{2,n}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta)$.

Lemma 2.2. *Under conditions* C1-C4*, there exists deterministic function $\boldsymbol{\Psi}_2(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta)$ such that $\boldsymbol{\Psi}^*_{2,n}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta)$ uniformly converges to $\boldsymbol{\Psi}_2(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta)$ in $\mathcal{A} \times \mathcal{B} \times \Theta$, in probability.*

Proof. Note that

$$\begin{aligned}\boldsymbol{\Psi}^{*}_{2,n}(\boldsymbol{\alpha},\boldsymbol{\beta},\theta) &\equiv n^{-2}\sum_{i,j=1}^{n}\Delta^{\circ}_{i}(\boldsymbol{\alpha})\Delta^{\circ}_{j}(\boldsymbol{\alpha})\begin{pmatrix} W_{ij}(\boldsymbol{\alpha}) \\ \\ \mathbf{Z}_{ij}\end{pmatrix}\left[I\{\varepsilon^{y}_{i}(\boldsymbol{\beta})\geq\varepsilon^{y}_{j}(\boldsymbol{\beta})\}-\Phi_{2}(\theta\, W_{ij}(\boldsymbol{\alpha}))\right]\\ &= \mathbf{S}^{\mathbf{I}}_{n}(\boldsymbol{\alpha},\boldsymbol{\beta})\,-\,\mathbf{S}^{\mathbf{II}}_{n}(\boldsymbol{\alpha},\theta)\,,\end{aligned}$$

where $\mathbf{S}^{\mathbf{I}}_{n}(\boldsymbol{\alpha},\boldsymbol{\beta})\equiv n^{-2}\sum_{i,j=1}^{n}\Delta^{\circ}_{i}(\boldsymbol{\alpha})\Delta^{\circ}_{j}(\boldsymbol{\alpha})\begin{pmatrix} W_{ij}(\boldsymbol{\alpha}) \\ \\ \mathbf{Z}_{ij}\end{pmatrix} I\{\varepsilon^{y}_{i}(\boldsymbol{\beta})\geq\varepsilon^{y}_{j}(\boldsymbol{\beta})\}$ and $\mathbf{S}^{\mathbf{II}}_{n}(\boldsymbol{\alpha},\theta)\equiv$ $n^{-2}\sum_{i,j=1}^{n}\Delta^{\circ}_{i}(\boldsymbol{\alpha})\Delta^{\circ}_{j}(\boldsymbol{\alpha})\begin{pmatrix} W_{ij}(\boldsymbol{\alpha}) \\ \\ \mathbf{Z}_{ij}\end{pmatrix}\Phi_{2}(\theta\, W_{ij}(\boldsymbol{\alpha}))$. In the following, we show that both $\mathbf{S}^{\mathbf{I}}_{n}(\boldsymbol{\alpha},\boldsymbol{\beta})$ and $\mathbf{S}^{\mathbf{II}}_{n}(\boldsymbol{\alpha},\theta)$ converge uniformly to their limit functions.

By algebra, we can write $\mathbf{S}^{\mathbf{I}}_{n}(\boldsymbol{\alpha},\boldsymbol{\beta})$ as a functional of four empirical processes

$$\begin{aligned}&\mathbf{S}^{\mathbf{I}}_{n}(\boldsymbol{\alpha},\boldsymbol{\beta})\\ &= n^{-1}\sum_{i=1}^{n}\Delta^{\circ}_{i}(\boldsymbol{\alpha})\left[n^{-1}\begin{pmatrix} W_{i}(\boldsymbol{\alpha}) \\ \\ \mathbf{Z}_{i}\end{pmatrix}\sum_{j=1}^{n}\Delta^{\circ}_{j}(\boldsymbol{\alpha})I\{\varepsilon^{y}_{j}(\boldsymbol{\beta})\leq\varepsilon^{y}_{i}(\boldsymbol{\beta})\}\right.\\ &\quad\left.-\,n^{-1}\sum_{j=1}^{n}\Delta^{\circ}_{j}(\boldsymbol{\alpha})\begin{pmatrix} W_{j}(\boldsymbol{\alpha}) \\ \\ \mathbf{Z}_{j}\end{pmatrix}I\{\varepsilon^{y}_{j}(\boldsymbol{\beta})\leq\varepsilon^{y}_{i}(\boldsymbol{\beta})\}\right]\\ &= \int_{-\infty}^{\infty}\left[n^{-1}\sum_{i=1}^{n}\Delta^{\circ}_{i}(\boldsymbol{\alpha})\begin{pmatrix} W_{i}(\boldsymbol{\alpha}) \\ \\ \mathbf{Z}_{i}\end{pmatrix}I\{\varepsilon^{y}_{i}(\boldsymbol{\beta})\geq s\}\right]\mathrm{d}\left[n^{-1}\sum_{j=1}^{n}\Delta^{\circ}_{j}(\boldsymbol{\alpha})I\{\varepsilon^{y}_{j}(\boldsymbol{\beta})\leq s\}\right]\\ &\quad-\int_{-\infty}^{\infty}\left[n^{-1}\sum_{i=1}^{n}\Delta^{\circ}_{i}(\boldsymbol{\alpha})I\{\varepsilon^{y}_{i}(\boldsymbol{\beta})\geq s\}\right]\mathrm{d}\left[n^{-1}\sum_{j=1}^{n}\Delta^{\circ}_{j}(\boldsymbol{\alpha})\begin{pmatrix} W_{j}(\boldsymbol{\alpha}) \\ \\ \mathbf{Z}_{j}\end{pmatrix}I\{\varepsilon^{y}_{j}(\boldsymbol{\beta})\leq s\}\right]\\ &= \int_{-\infty}^{\infty}\mathbb{A}^{(n)}_{1}\,\mathrm{d}\,\mathbb{A}^{(n)}_{2}\,-\int_{-\infty}^{\infty}\mathbb{A}^{(n)}_{3}\,\mathrm{d}\,\mathbb{A}^{(n)}_{4},\end{aligned}$$

where $\mathbb{A}_1^{(n)} \equiv \widehat{\mathrm{E}}_n \left[\Delta^{\circ}(\boldsymbol{\alpha}) \begin{pmatrix} W(\boldsymbol{\alpha}) \\ \mathbf{Z} \end{pmatrix} I\{\varepsilon^y(\boldsymbol{\beta}) \geq s\} \right]$, $\mathbb{A}_2^{(n)} \equiv \widehat{\mathrm{E}}_n \left[\Delta^{\circ}(\boldsymbol{\alpha}) I\{\varepsilon^y(\boldsymbol{\beta}) \leq s\} \right]$, $\mathbb{A}_3^{(n)} \equiv \widehat{\mathrm{E}}_n \left[\Delta^{\circ}(\boldsymbol{\alpha}) I\{\varepsilon^y(\boldsymbol{\beta}) \geq s\} \right]$, $\mathbb{A}_4^{(n)} \equiv \widehat{\mathrm{E}}_n \left[\Delta^{\circ}(\boldsymbol{\alpha}) \begin{pmatrix} W(\boldsymbol{\alpha}) \\ \mathbf{Z} \end{pmatrix} I\{\varepsilon^y(\boldsymbol{\beta}) \leq s\} \right]$, and $\widehat{\mathrm{E}}_n(X)$ denotes the empirical average. Define the corresponding theoretical parts of $\mathbb{A}_1^{(n)}$ to $\mathbb{A}_4^{(n)}$ as A_1 to A_4, respectively. Consider $\mathcal{F} \equiv \{I\{Y - \mathbf{b}^T\mathbf{Z} \leq t\} : \mathbf{b} \in \mathbb{R}^p, t \in \mathbb{R}\}$, where $(Y, \mathbf{Z}^T)^T \in \mathcal{Y} \times \mathcal{Z} \equiv \mathbb{R} \times \mathbb{R}^p$ has distribution P, for arbitrary P. Since $\mathcal{F}$ is the *Vapnik-Červonenkis class* with index $p + 3$ (Kosorok 2008, Lemma 9.12), $\mathcal{F}$ is *P-Glivenko-Cantelli* for any P. Thus, $\Delta^{\circ}(\boldsymbol{\alpha})$, $I\{\varepsilon^y(\boldsymbol{\beta}) \leq s\}$ and $I\{\varepsilon^y(\boldsymbol{\beta}) \geq s\}$ are all *P-Glivenko-Cantelli*. From condition *C2*, $\mathbf{Z}$ is *P-Glivenko-Cantelli*. Under condition *C1* and *C2*, $\{\Delta^{\circ}(\boldsymbol{\alpha})W(\boldsymbol{\alpha}), \boldsymbol{\alpha} \in \mathcal{A}\}$ is a collection of measurable functions with integrable envelope function and is indexed by a compact metric space $\mathcal{A}$, thus $\Delta^{\circ}(\boldsymbol{\alpha})W(\boldsymbol{\alpha})$ is *P-Glivenko-Cantelli* (van der Vaart 1998, p. 272). Based on Glivenko-Cantelli preservation theorem for production (Theorem 9.26 and Corollary 9.27 of Kosorok (2008)), the uniform convergence of $\mathbb{A}_1^{(n)}$ to $\mathbb{A}_4^{(n)}$ holds. Furthermore, the uniform convergence of stochastic integral like $\int_{-\infty}^{\infty} \mathbb{A}_1^{(n)} \, d\mathbb{A}_2^{(n)}$ can be established in an argument similar to Lemma 3 of Gill (1989). Thus, we obtain the uniform convergence of $\mathbf{S}_n^{\mathbf{I}}(\boldsymbol{\alpha}, \boldsymbol{\beta})$.

Next, we need to show the uniform convergence of $\mathbf{S}_n^{\mathbf{II}}(\boldsymbol{\alpha}, \theta)$. Let $\mathbf{V}$ represent the random vector $(\Delta^{\circ}(\boldsymbol{\alpha}), W, \mathbf{Z}^T)^T$ and $\mathbf{V}_i, i = 1, \ldots, n$ denote n independent and identically distributed replicates of the random vector $\mathbf{V}$. Writing the kernel function as $\mathbf{h}(\mathbf{V}_i, \mathbf{V}_j, \boldsymbol{\alpha}, \theta) \equiv \Delta_i^{\circ}(\boldsymbol{\alpha})\Delta_j^{\circ}(\boldsymbol{\alpha}) \begin{pmatrix} W_{ij}(\boldsymbol{\alpha}) \\ \mathbf{Z}_{ij} \end{pmatrix} (\Phi_2(\theta W_{ij}(\boldsymbol{\alpha})) - 1/2)$ and by the fact that $\mathbf{h}(\mathbf{V}_i, \mathbf{V}_j, \boldsymbol{\alpha}, \theta)$ is symmetric with respect to the data arguments $(\mathbf{V}_i^T, \mathbf{V}_j^T)^T$, we obtain a U-statistic $\mathbf{U}_n^{\mathbf{II}}(\boldsymbol{\alpha}, \theta) \equiv n^{-2} \sum_{i,j=1}^n \mathbf{h}(\mathbf{V}_i, \mathbf{V}_j, \boldsymbol{\alpha}, \theta)$. It is easy to see that $\mathbf{U}_n^{\mathbf{II}}(\boldsymbol{\alpha}, \theta)$ is equal to $\mathbf{S}_n^{\mathbf{II}}(\boldsymbol{\alpha}, \theta)$. Let $\mathbf{U}^{\mathbf{II}}(\boldsymbol{\alpha}, \theta) \equiv \mathrm{E}\{\mathbf{U}_n^{\mathbf{II}}(\boldsymbol{\alpha}, \theta)\} = \mathrm{E}\{\mathbf{S}_n^{\mathbf{II}}(\boldsymbol{\alpha}, \theta)\}$, under assump-

tions *C1* and *C2* above, the strong law of large number for U-statistics (e.g. Serfling 1980, Theorem 5.4.A) ensures the pointwise convergence of $\mathbf{U}_n^{\mathbf{II}}(\boldsymbol{\alpha}, \theta)$ to $\mathbf{U}^{\mathbf{II}}(\boldsymbol{\alpha}, \theta)$, i.e., $\left\|\mathbf{U}_n^{\mathbf{II}}(\boldsymbol{\alpha}, \theta) - \mathbf{U}^{\mathbf{II}}(\boldsymbol{\alpha}, \theta)\right\| \xrightarrow{\mathcal{P}} 0$, $\forall\ \boldsymbol{\alpha} \in \mathcal{A}$, $\theta \in \Theta$. It is well known that pointwise convergence and equicontinuity imply uniform convergence to a continuous function on a compact set (e.g. Rudin 1976, Exercise 7.16). The generalization of this result to stochastic situations has been well studied by some scholars in Econometric area, e.g., Honoré and Powell (1994). We take the Theorem 1 in Honoré and Powell (1994) to show the uniform convergence of $\mathbf{U}_n^{\mathbf{II}}(\boldsymbol{\alpha}, \theta)$ to $\mathbf{U}^{\mathbf{II}}(\boldsymbol{\alpha}, \theta)$. Basically, we need to check whether the three assumptions of the theorem are all satisfied. Under our assumption *C1*, the parameter space $\mathcal{A} \times \Theta$ is compact. It is easy to see that the U-statistic kernel function $\mathbf{h}(\mathbf{V}_i, \mathbf{V}_j, \boldsymbol{\alpha}, \theta)$ is measurable in $(\mathbf{V}_i^T, \mathbf{V}_j^T)^T$ for each $\boldsymbol{\alpha} \in \mathcal{A}$, $\theta \in \Theta$ and this kernel function is a continuous with respect to $\boldsymbol{\alpha} \in \mathcal{A}$, $\theta \in \Theta$ on the support of $(\mathbf{V}_i^T, \mathbf{V}_j^T)^T$. One can also find that $|\mathbf{h}(\mathbf{V}_i, \mathbf{V}_j, \boldsymbol{\alpha}, \theta)| \leq \left|\begin{pmatrix} 2\,K_0 \\ \\ \mathbf{Z}_{ij} \end{pmatrix}\right|$ for all $\boldsymbol{\alpha} \in \mathcal{A}$, $\theta \in \Theta$, and $\mathrm{E}\left|\begin{pmatrix} 2\,K_0 \\ \\ \mathbf{Z}_{ij} \end{pmatrix}\right| < \infty$ due to our assumption *C1* and *C2*, where $K_0 \equiv \max\left(\left|H_T\{\tau_a\}\right|, \left|H_T\{\tau_b\}\right|\right)$. Thus, all three assumptions of the Theorem 1 in Honoré and Powell (1994) are satisfied and we have showed the uniform convergence of $\mathbf{S}_n^{\mathbf{II}}(\boldsymbol{\alpha}, \theta)$. Combining the uniform convergence results for $\mathbf{S}_n^{\mathbf{I}}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ and $\mathbf{S}_n^{\mathbf{II}}(\boldsymbol{\alpha}, \theta)$, we have proved the Lemma 2.2. □

Theorem 2.3. *Let δ_n be any positive sequence converging to 0. Under conditions* C1-C5*, there exists deterministic function $\boldsymbol{\Psi}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}, \theta)$ such that $\boldsymbol{\Psi}_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta)$ converges to $\boldsymbol{\Psi}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}, \theta)$ uniformly in $\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_0\| \leq \delta_n$ and $\mathcal{B} \times \Theta$, in probability.*

Proof. The uniform consistency of $\boldsymbol{\Psi}_{1,n}(\boldsymbol{\alpha})$ has been established by Ying (1993). For $\boldsymbol{\Psi}_{2,n}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta)$, we first note the inequality that $\|\boldsymbol{\Psi}_{2,n}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta) - \boldsymbol{\Psi}_2(\boldsymbol{\alpha}_0, \boldsymbol{\beta}, \theta)\| \leq \left\|\boldsymbol{\Psi}_{2,n}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta) - \boldsymbol{\Psi}_{2,n}^*(\boldsymbol{\alpha}_0, \boldsymbol{\beta}, \theta)\right\| + \left\|\boldsymbol{\Psi}_{2,n}^*(\boldsymbol{\alpha}_0, \boldsymbol{\beta}, \theta) - \boldsymbol{\Psi}_2(\boldsymbol{\alpha}_0, \boldsymbol{\beta}, \theta)\right\|$. Based on Lemma

2.2, the uniform convergence of $\boldsymbol{\Psi}_{2,n}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta)$ will hold given that

$$\sup_{\|\boldsymbol{\alpha}-\boldsymbol{\alpha}_0\| \leq \delta_n,\, \boldsymbol{\beta}\in\mathcal{B},\, \theta\in\Theta} \left| \boldsymbol{\Psi}_{2,n}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta) \, - \, \boldsymbol{\Psi}^*_{2,n}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}, \theta) \right| \xrightarrow{\mathcal{P}} 0, \quad \text{as } n \to \infty.$$

The latter can be shown by using corollary 2.1. Thus the uniform convergence of $\boldsymbol{\Psi}_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta)$ holds. If we write $\boldsymbol{\Psi}_1(\boldsymbol{\alpha}_0)$ as the limiting function of $\boldsymbol{\Psi}_{1,n}(\boldsymbol{\alpha})$ for $\|\boldsymbol{\alpha}-\boldsymbol{\alpha}_0\| \leq \delta_n$, then $\boldsymbol{\Psi}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}, \theta) = \left\{ \boldsymbol{\Psi}_1(\boldsymbol{\alpha}_0)^T, \, \boldsymbol{\Psi}_2(\boldsymbol{\alpha}_0, \boldsymbol{\beta}, \theta)^T \right\}^T$. □

We need an additional regularity condition *C6* to prove Theorem 2.1.

C6. The partial derivative matrix $\partial \boldsymbol{\Psi}_2(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta) / \partial(\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T, \theta)^T$ evaluated at $(\boldsymbol{\alpha}_0^T, \boldsymbol{\beta}_0^T, \theta_0)^T$ is not singular.

Proof of Theorem 2.1. Note that the limit function $\boldsymbol{\Psi}_2(\boldsymbol{\alpha}_0, \boldsymbol{\beta}, \theta)$ has a zero-crossing at $(\boldsymbol{\alpha}_0^T, \boldsymbol{\beta}_0^T, \theta_0)^T$, according to the equality (2.5). Also, it has been established that $\boldsymbol{\Psi}_1(\boldsymbol{\alpha}_0) = 0$ (Tsiatis 1990; Ying 1993). Thus, $\boldsymbol{\Psi}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \theta_0) = 0$. Given this fact and the condition *C6*, there exists a neighborhood of $(\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T, \theta)^T$ where $\boldsymbol{\Psi}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta)$ has a unique zero-crossing at $(\boldsymbol{\alpha}_0^T, \boldsymbol{\beta}_0^T, \theta_0)^T$. Thus, the uniform convergence of the estimating function $\boldsymbol{\Psi}_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta)$ implies that there exists one zero-crossing of $\boldsymbol{\Psi}_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta)$ converging to $(\boldsymbol{\alpha}_0^T, \boldsymbol{\beta}_0^T, \theta_0)^T$ in probability. □

2.5.2 Asymptotic Normality

We now show the asymptotic normality of $(\widehat{\boldsymbol{\alpha}}_0^T, \widehat{\boldsymbol{\beta}}_0^T, \widehat{\theta}_0)^T$. First, we show the asymptotic normality of $\boldsymbol{\Psi}_n(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \theta_0)$, then we show the asymptotic linearity of the estimating function $\boldsymbol{\Psi}_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta)$ around a small neighborhood of $(\boldsymbol{\alpha}_0^T, \boldsymbol{\beta}_0^T, \theta_0)^T$.

Theorem 2.4. *Under conditions* C1-C5, $n^{1/2}\boldsymbol{\Psi}_n(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \theta_0)$ *is asymptotically normal with mean 0.*

Proof. Note that $\boldsymbol{\Psi}_n(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \theta_0) = \left\{\boldsymbol{\Psi}_{1,n}(\boldsymbol{\alpha}_0)^T,\ \boldsymbol{\Psi}_{2,n}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \theta_0)^T\right\}^T$ is a statistic based on i.i.d. sample $\{X_i, Y_i, \Delta_i, \mathbf{Z}_i\}$, $i = 1, \ldots, n$. In the following, we show that $\boldsymbol{\Psi}_n(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \theta_0)$ is a plug-in estimator in the mapping from the distribution of $\{X, Y, \mathbf{Z}, \Delta\}$ to $\boldsymbol{\Psi}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \theta_0)$, then we establish the asymptotic normality of $\boldsymbol{\Psi}_n(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \theta_0)$ by using the functional delta method (van der Vaart and Wellner 1996, Section 3.9). Define $F_{XY\mathbf{Z}\Delta}(t, u, \mathbf{z}, \delta) \equiv \Pr(X \le t,\, Y \le u,\, \mathbf{Z} \le \mathbf{z},\, \Delta \le \delta)$, $F_{X\mathbf{Z}\Delta}(t, \mathbf{z}, \delta) \equiv F_{XY\mathbf{Z}\Delta}(t, \infty, \mathbf{z}, \delta)$, $F_{X\mathbf{Z},\Delta=1}(t, \mathbf{z}) \equiv \Pr(X \le t,\, \infty,\, \mathbf{Z} \le \mathbf{z},\, \Delta = 1)$, $F_{X\mathbf{Z}}(t, \mathbf{z}) \equiv F_{XY\mathbf{Z}\Delta}(t, \infty, \mathbf{z}, \infty)$ and $F_{XY\mathbf{Z},\Delta^\circ(\boldsymbol{\alpha}_0)=1}(t, u, \mathbf{z}) \equiv \Pr(X \le t,\, Y \le u,\, \mathbf{Z} \le \mathbf{z},\, \Delta^\circ(\boldsymbol{\alpha}_0) = 1)$. Write their respective empirical counterparts as $\widehat{F}_{XY\mathbf{Z}\Delta}$, $\widehat{F}_{X\mathbf{Z}\Delta}$, $\widehat{F}_{X\mathbf{Z},\Delta=1}$, $\widehat{F}_{X\mathbf{Z}}$ and $\widehat{F}_{XY,\mathbf{Z},\Delta^\circ(\boldsymbol{\alpha}_0)=1}$. In addition, let

$$
\begin{aligned}
F_{WY\mathbf{Z},\Delta^\circ(\boldsymbol{\alpha}_0)=1}(w, u, \mathbf{z}) &\equiv \Pr(W(\boldsymbol{\alpha}_0) \le w,\, Y \le u,\, \mathbf{Z} \le \mathbf{z},\, \Delta^\circ(\boldsymbol{\alpha}_0) = 1), \\
\widehat{F}_{WY\mathbf{Z},\Delta^\circ(\boldsymbol{\alpha}_0)=1}(w, u, \mathbf{z}) &\equiv n^{-1}\sum_{i=1}^{n} I(\widehat{W}_i(\boldsymbol{\alpha}_0) \le w,\, Y_i \le u,\, \mathbf{Z}_i \le \mathbf{z},\, \Delta_i^\circ(\boldsymbol{\alpha}_0) = 1)
\end{aligned}
$$

Thus,

$$
\begin{aligned}
F_{WY\mathbf{Z},\Delta^\circ(\boldsymbol{\alpha}_0)=1}(w, u, \mathbf{z}) &= F_{XY\mathbf{Z},\Delta^\circ(\boldsymbol{\alpha}_0)=1}\{G_T^{-1}(H_T^{-1}(w) + \boldsymbol{\alpha}_0^T\mathbf{z}), u, \mathbf{z}\}, \\
\widehat{F}_{WY\mathbf{Z},\Delta^\circ(\boldsymbol{\alpha}_0)=1}(w, u, \mathbf{z}) &= \widehat{F}_{XY,\mathbf{Z},\Delta^\circ(\boldsymbol{\alpha}_0)=1}\{G_T^{-1}(\widehat{H}_T^{-1}(w) + \boldsymbol{\alpha}_0^T\mathbf{z}), u, \mathbf{z}\}.
\end{aligned}
$$

We may reexpress $\boldsymbol{\Psi}_{1,n}(\boldsymbol{\alpha}_0)$ and $\boldsymbol{\Psi}_{2,n}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \theta_0)$ as the functionals of empirical functions. Notice that $\varepsilon^x(\boldsymbol{\alpha}_0) = G_T(X) - \boldsymbol{\alpha}_0^T\mathbf{Z}$ and $\varepsilon^y(\boldsymbol{\beta}_0) = G_U(Y) - \boldsymbol{\beta}_0^T\mathbf{Z}$ with known increasing transformation $G_T(\cdot)$ and $G_U(\cdot)$ respectively, we have

$$
\boldsymbol{\Psi}_{1,n}(\boldsymbol{\alpha}_0) = \int \zeta_n\left(\varepsilon_i^x(\boldsymbol{\alpha}_0); \boldsymbol{\alpha}_0\right)\left[\mathbf{z}_1 - \frac{\int \mathbf{z}_2\, I\{\varepsilon_2^x(\boldsymbol{\alpha}_0) \ge \varepsilon_1^x(\boldsymbol{\alpha}_0)\}\, d\widehat{F}_{X\mathbf{Z}}(t_2, z_2)}{\int I\{\varepsilon_2^x(\boldsymbol{\alpha}_0) \ge \varepsilon_1^x(\boldsymbol{\alpha}_0)\}\, d\widehat{F}_{X\mathbf{Z}}(t_2, \mathbf{z}_2)}\right] d\widehat{F}_{X\mathbf{Z},\Delta=1}(t_1, \mathbf{z}_1),
$$

and its theoretical counterpart is

$$\boldsymbol{\Psi}_1(\boldsymbol{\alpha}_0) \;=\; \int \zeta_n\left(\varepsilon_i^x(\boldsymbol{\alpha}_0);\boldsymbol{\alpha}_0\right)\left[\mathbf{z}_1 - \frac{\int \mathbf{z}_2\, I\{\varepsilon_2^x(\boldsymbol{\alpha}_0) \geq \varepsilon_1^x(\boldsymbol{\alpha}_0)\}\, dF_{X\mathbf{Z}}(t_2,z_2)}{\int I\{\varepsilon_2^x(\boldsymbol{\alpha}_0) \geq \varepsilon_1^x(\boldsymbol{\alpha}_0)\}\, dF_{X\mathbf{Z}}(t_2,\mathbf{z}_2)}\right] dF_{X\mathbf{Z},\Delta=1}(t_1,\mathbf{z}_1).$$

As for $\boldsymbol{\Psi}_{2,n}(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0,\theta_0)$, similarly we have

$$\begin{aligned}\boldsymbol{\Psi}_{2,n}(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0,\theta_0) \;&=\; \int\int \begin{pmatrix} w_{12} \\ \\ \mathbf{z}_{12}\end{pmatrix}\left[I\{\varepsilon_1^y(\boldsymbol{\beta}_0) \geq \varepsilon_2^y(\boldsymbol{\beta}_0)\} - \Phi_2(\theta_0\, w_{12})\right] \\ &\qquad \cdot\; d\widehat{F}_{WY\mathbf{Z},\Delta^\circ(\boldsymbol{\alpha}_0)=1}(w_1,u_1,\mathbf{z}_1)\, d\widehat{F}_{WY\mathbf{Z},\Delta^\circ(\boldsymbol{\alpha}_0)=1}(w_2,u_2,\mathbf{z}_2),\end{aligned}$$

where $w_{12} \equiv w_1 - w_2$ and $\mathbf{z}_{12} \equiv \mathbf{z}_1 - \mathbf{z}_2$, its theoretical counterpart is

$$\begin{aligned}\boldsymbol{\Psi}_2\left(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0,\theta_0\right) \;&\equiv\; \int\int \begin{pmatrix} w_{12} \\ \\ \mathbf{z}_{12}\end{pmatrix}\left[I\{\varepsilon_1^y(\boldsymbol{\beta}_0) \geq \varepsilon_2^y(\boldsymbol{\beta}_0)\} - \Phi_2(\theta_0\, w_{12})\right] \\ &\qquad \cdot\; dF_{WY\mathbf{Z},\Delta^\circ(\boldsymbol{\alpha}_0)=1}(w_1,u_1,\mathbf{z}_1)\, dF_{WY\mathbf{Z},\Delta^\circ(\boldsymbol{\alpha}_0)=1}(w_2,u_2,\mathbf{z}_2) \\ &=\; E\left\{\Delta_1^\circ(\boldsymbol{\alpha}_0)\Delta_2^\circ(\boldsymbol{\alpha}_0)\begin{pmatrix} W_{12} \\ \\ \mathbf{Z}_{12}\end{pmatrix}\left[I\{\varepsilon_1^y(\boldsymbol{\beta}_0) \geq \varepsilon_2^y(\boldsymbol{\beta}_0)\} - \Phi_2(\theta_0\, W_{12})\right]\right\}\end{aligned}$$

Thus, we have shown that $\boldsymbol{\Psi}_n(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0,\theta_0)$ is a plug-in estimator in the mapping from the distribution of $\{X,Y,\mathbf{Z},\Delta\}$ to $\boldsymbol{\Psi}(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0,\theta_0)$. The map $F_{XY\mathbf{Z}\Delta} \mapsto \boldsymbol{\Psi}(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0,\theta_0)$ can be further decomposed as

$$F_{XY\mathbf{Z}\Delta} \mapsto \left\{\begin{array}{l} \left.\begin{array}{l} F_{X\mathbf{Z}} \\ \\ F_{X\mathbf{Z},\Delta=1}\end{array}\right\} \\ \\ F_{X\mathbf{Z}\Delta} \mapsto F_{\varepsilon^t} \text{ on } (-\infty,\tau_b\,] \mapsto H_T \text{ on } [\,\tau_a,\tau_b\,] \mapsto H_T^{-1} \text{ on } [H_T(\tau_a),H_T(\tau_b)] \\ \\ F_{XY\mathbf{Z},\Delta^\circ(\boldsymbol{\alpha}_0)=1}\end{array}\right.$$

$$\left.\begin{array}{l} \mapsto \boldsymbol{\Psi}_1(\boldsymbol{\alpha}_0) \\ \mapsto G_T^{-1} \text{ on } [H_T(\tau_a) + \boldsymbol{\alpha}_0^T\mathbf{Z},\ H_T(\tau_b) + \boldsymbol{\alpha}_0^T\mathbf{Z}] \end{array}\right\} \mapsto F_{WY\mathbf{Z},\Delta^\circ(\boldsymbol{\alpha}_0)=1} \mapsto \boldsymbol{\Psi}_2\left(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \theta_0\right)$$

$$\mapsto \boldsymbol{\Psi}\left(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \theta_0\right).$$

We then show the Hadamard-differentiability of the map $F_{XY\mathbf{Z}\Delta} \mapsto \boldsymbol{\Psi}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \theta_0)$ in appropriate spaces of univariate and bivariate cadlag functions endowed with supnorm, seeing Neuhaus (1971) and van der Vaart and Wellner (1996, Section 3.9) for example. It is obvious that the maps $F_{XY\mathbf{Z}\Delta} \mapsto \{F_{X\mathbf{Z}}, F_{X\mathbf{Z},\Delta=1}\}$, $F_{XY\mathbf{Z}\Delta} \mapsto \{F_{X\mathbf{Z}\Delta}, F_{XY\mathbf{Z},\Delta^\circ(\boldsymbol{\alpha}_0)=1}\}$ and $\{\boldsymbol{\Psi}_1(\boldsymbol{\alpha}_0), \boldsymbol{\Psi}_2\left(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \theta_0\right)\} \mapsto \boldsymbol{\Psi}\left(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \theta_0\right)$ are Hadamard-differentiable. The Hadamard-differentiability of the maps $F_{X\mathbf{Z}\Delta} \mapsto F_{\varepsilon^t}$ on $(-\infty, \tau_b\,]$, F_{ε^t} on $(-\infty, \tau_b\,] \mapsto H_T$ on $[\,\tau_a, \tau_b\,]$, H_T on $[\,\tau_a, \tau_b\,] \mapsto H_T^{-1}$ on $[\,H_T(\tau_a), H_T(\tau_b)\,]$, H_T^{-1} on $[\,H_T(\tau_a), H_T(\tau_b)\,] \mapsto G_T^{-1}$ on $[\,H_T(\tau_a) + \boldsymbol{\alpha}_0^T\mathbf{Z},\ H_T(\tau_b) + \boldsymbol{\alpha}_0^T\mathbf{Z}\,]$ and $\{G_T^{-1}$ on $[\,H_T(\tau_a) + \boldsymbol{\alpha}_0^T\mathbf{Z},\ H_T(\tau_b) + \boldsymbol{\alpha}_0^T\mathbf{Z}\,]$, $F_{XY\mathbf{Z},\Delta^\circ=1}\} \mapsto F_{WY\mathbf{Z},\Delta^\circ(\boldsymbol{\alpha}_0)=1}$ follows the results in van der Vaart and Wellner (1996, Section 3.9) directly. In terms of the map $\{F_{X\mathbf{Z}}, F_{X\mathbf{Z},\Delta=1}\} \mapsto \boldsymbol{\Psi}_1(\boldsymbol{\alpha}_0)$, we can see from equality (2.5.2) that this map is further decomposed into the inner and outer integrations. The inner one is linear and continuous, and hence Hadamard-differentiable, whereas the Hadamard-differentiability of the outer one follows the Lemma 5.1 of Gill et al. (1995). The remaining piece of the decomposed maps, $F_{WY\mathbf{Z},\Delta^\circ(\boldsymbol{\alpha}_0)=1} \mapsto \boldsymbol{\Psi}_2\left(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \theta_0\right)$, can be shown to be Hadamard-differentiable, by using similar argument as that of the map $\{F_{X\mathbf{Z}}, F_{X\mathbf{Z},\Delta=1}\} \mapsto \boldsymbol{\Psi}_1(\boldsymbol{\alpha}_0)$. The chain rule of the Hadamard-differentiability (e.g. van der Vaart and Wellner 1996, Lemma 3.9.3) now results in the Hadamard-differentiability of the mapping from $F_{XY\mathbf{Z}\Delta}$ to $\boldsymbol{\Psi}\left(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \theta_0\right)$. By the functional delta

method then, the asymptotic normality of the empirical distribution $\widehat{F}_{XY\mathbf{Z}\Delta}$ yields that of $\boldsymbol{\Psi}_2\left(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \theta_0\right)$. □

Theorem below establish the asymptotic linearity of $\boldsymbol{\Psi}_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta)$ around a small neighborhood of $(\boldsymbol{\alpha}_0^T, \boldsymbol{\beta}_0^T, \theta_0)^T$.

Theorem 2.5. *Define* $\boldsymbol{\eta} \equiv (\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T, \theta)^T$*, under conditions* C1-C5*, for any positive sequence* $\delta_n \to 0$ *in probability,*

$$\sup_{\|\boldsymbol{\eta}-\boldsymbol{\eta}_0\| \leq \delta_n} \|\boldsymbol{\Psi}_n(\boldsymbol{\eta}) - \boldsymbol{\Psi}_n(\boldsymbol{\eta}_0) - \boldsymbol{\Psi}(\boldsymbol{\eta})\| \;=\; o_p(n^{-1/2} \vee \|\boldsymbol{\eta} - \boldsymbol{\eta}_0\|). \tag{2.13}$$

To prove Theorem 2.5, we prove the following Lemma first.

Lemma 2.3. *Under conditions* C1-C5*, for* $\boldsymbol{\eta}$ *in a small neighborhood around* $\boldsymbol{\eta}_0$*, the expectation of* $I\{\varepsilon_i^y(\boldsymbol{\beta}) \geq \varepsilon_j^y(\boldsymbol{\beta})\}$ *conditioning on* $X_i, X_j, \mathbf{Z}_i, \mathbf{Z}_j, \Delta_i = 1, \Delta_j = 1$*,* $i, j = 1, \ldots, n$*, is equal to* $\Phi_2(\theta_0\, W_{ij}(\boldsymbol{\alpha}_0)) + O_p(\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|)$.

Proof. Since $\varepsilon_i^y(\boldsymbol{\beta}) = \varepsilon_i^y(\boldsymbol{\beta}_0) - (\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T \mathbf{Z}_i$, we may have $f_{\varepsilon_i^y(\boldsymbol{\beta})}(s) = f_{\epsilon^u}(s + (\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T \mathbf{Z}_i)$, where $f_{\varepsilon_i^y(\boldsymbol{\beta})}$ is the conditional density function of $\varepsilon_i^y(\boldsymbol{\beta})$ and f_{ϵ^u} is the density function of $\varepsilon_i^y(\boldsymbol{\beta})$ the true error term ε^u on cost scale. Noting that $\varepsilon_i^y(\boldsymbol{\beta})$ and $\varepsilon_j^y(\boldsymbol{\beta})$ are independent conditioning on $X_i, X_j, \mathbf{Z}_i, \mathbf{Z}_j, \Delta_i = 1, \Delta_j = 1$, we could have the following expression of

the conditional density function of $\varepsilon_{ij}^y(\boldsymbol{\beta}) \equiv \varepsilon_i^y(\boldsymbol{\beta}) - \varepsilon_j^y(\boldsymbol{\beta})$,

$$
\begin{aligned}
& f_{\varepsilon_{ij}^y(\boldsymbol{\beta})}(t) \\
&= \int_{-\infty}^{\infty} f_{\varepsilon_j^y(\boldsymbol{\beta})}(s) \cdot f_{\varepsilon_i^y(\boldsymbol{\beta})}(s+t)\, ds \\
&= \int_{-\infty}^{\infty} f_{\varepsilon^u}(s + (\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T\mathbf{Z}_j) \cdot f_{\varepsilon^u}(s+t + (\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T\mathbf{Z}_i)\, ds \\
&= \int_{-\infty}^{\infty} \left\{ \left[f_{\varepsilon^u}(s) + f'_{\varepsilon^u}(s + \lambda_1(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T\mathbf{Z}_j)\,(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T\mathbf{Z}_j \right] \right. \\
&\qquad \left. \cdot \left[f_{\varepsilon^u}(s+t) + f'_{\varepsilon^u}(s+t+\lambda_2(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T\mathbf{Z}_i)\,(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T\mathbf{Z}_i \right] \right\} ds \\
&= \int_{-\infty}^{\infty} \Bigg[f_{\varepsilon^u}(s) f_{\varepsilon^u}(s+t) + f_{\varepsilon^u}(s+t) f'_{\varepsilon^u}(s + \lambda_1(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T\mathbf{Z}_j)\,(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T\mathbf{Z}_j \\
&\qquad + \; f_{\varepsilon^u}(s) f'_{\varepsilon^u}(s+t+\lambda_2(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T\mathbf{Z}_i)\,(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T\mathbf{Z}_i \\
&\qquad + \; f'_{\varepsilon^u}(s + \lambda_1(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T\mathbf{Z}_j)\,(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T\mathbf{Z}_j f'_{\varepsilon^u}(s+t+\lambda_2(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T\mathbf{Z}_i)\,(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T\mathbf{Z}_i \Bigg] ds,
\end{aligned}
$$

where f'_{ε^u} is the first order derivative of f_{ε^u} and $\lambda_1\,, \lambda_2\, \in (0,1)$. Thus, we have

$$
\begin{aligned}
& Pr\left\{ \varepsilon_i^y(\boldsymbol{\beta}_0) \geq \varepsilon_j^y(\boldsymbol{\beta}_0) \,|\, \cdot \right\} \\
&= 1 - \int_{-\infty}^{0} f_{\varepsilon_{ij}^y(\boldsymbol{\beta})}(t)\, dt \\
&= 1 - \int_{-\infty}^{0} \int_{-\infty}^{\infty} f_{\varepsilon^u}(s) f_{\varepsilon^u}(s+t)\, ds\, dt \\
&\qquad - \; (\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T\mathbf{Z}_j \cdot \int_{-\infty}^{0} \int_{-\infty}^{\infty} f_{\varepsilon^u}(s+t) f'_{\varepsilon^u}(s + \lambda_1(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T\mathbf{Z}_j) ds\, dt \\
&\qquad - \; (\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T\mathbf{Z}_i \cdot \int_{-\infty}^{0} \int_{-\infty}^{\infty} f_{\varepsilon^u}(s) f'_{\varepsilon^u}(s+t+\lambda_2(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T\mathbf{Z}_i) ds\, dt \\
&\qquad - \; (\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T\mathbf{Z}_j (\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T\mathbf{Z}_i \int_{-\infty}^{0} \int_{-\infty}^{\infty} f'_{\varepsilon^u}(s + \lambda_1(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T\mathbf{Z}_j) \\
&\qquad \cdot f'_{\varepsilon^u}(s+t+\lambda_2(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T\mathbf{Z}_i)\, ds\, dt
\end{aligned}
$$

From the equality (2.5) in Section 2.2, we have the following expression for the conditional expectation of $I\{\varepsilon_i^y(\boldsymbol{\beta}_0) \geq \varepsilon_j^y(\boldsymbol{\beta}_0)\}$,

$$\begin{aligned} Pr\left\{\varepsilon_i^y(\boldsymbol{\beta}_0) \geq \varepsilon_j^y(\boldsymbol{\beta}_0) \,|\, \cdot\right\} &= 1 - \int_{-\infty}^{0} f_{\varepsilon_{ij}^y(\boldsymbol{\beta}_0)}(t)dt \qquad (2.14) \\ &= 1 - \int_{-\infty}^{0}\int_{-\infty}^{\infty} f_{\varepsilon^u}(s) f_{\varepsilon^u}(s+t)\, ds\, dt = \Phi_2(\theta_0\, W_{ij}(\boldsymbol{\alpha}_0)). \end{aligned}$$

Since both f_{ε^u} and f'_{ε^u} are bounded under Condition *C5*, we have

$$\begin{aligned} &\int_{-\infty}^{0}\int_{-\infty}^{\infty} f_{\varepsilon^u}(s+t) f'_{\varepsilon^u}(s + \lambda_1(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T\mathbf{Z}_j) ds\, dt \qquad (2.15) \\ &= \int_{-\infty}^{\infty} F_{\varepsilon^u}(s)\cdot f'_{\varepsilon^u}(s + \lambda_1(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T\mathbf{Z}_j)\, ds \\ &= \int_{-\infty}^{\infty} F_{\varepsilon^u}(s - \lambda_1(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T\mathbf{Z}_j)\, df_{\varepsilon^u}(s) \\ &= \left[F_{\varepsilon^u}(s - \lambda_1(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T\mathbf{Z}_j)\cdot f_{\varepsilon^u}(s)\right]\Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} f_{\varepsilon^u}(s + \lambda_1(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T\mathbf{Z}_j)\cdot f_{\varepsilon^u}(s)\, ds \\ &= 1\cdot f_{\varepsilon^u}(+\infty) - \int_{-\infty}^{\infty} f_{\varepsilon^u}(s + \lambda_1(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T\mathbf{Z}_j)\cdot f_{\varepsilon^u}(s)\, ds \end{aligned}$$

where F_{ε^u} is the cumulative distribution function of ε^u. Since f_{ε^u} is bounded, if we define $K_1 \equiv \sup_s \{f_{\varepsilon^u}(s)\}$, we would have

$$0 \;\leq\; \int_{-\infty}^{\infty} f_{\varepsilon^u}(s + \lambda_1(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T\mathbf{Z}_j)\cdot f_{\varepsilon^u}(s)\, ds \;\leq\; K_1 \int_{-\infty}^{\infty} f_{\varepsilon^u}(s)\, ds = K_1.$$

Thus, the equality (2.15) is finite. Similarly, we can show that both the integral $\int_{-\infty}^{0}\int_{-\infty}^{\infty} f_{\varepsilon^u}(s) f'_{\varepsilon^u}(s+t) ds\, dt$ and $\int_{-\infty}^{0}\int_{-\infty}^{\infty} f'_{\varepsilon^u}(s) f'_{\varepsilon^u}(s+t)\, ds\, dt$ are finite. Putting these results together with equality (2.14), we would find that the conditional expectation of $I\{\varepsilon_i^y(\boldsymbol{\beta}) \geq \varepsilon_j^y(\boldsymbol{\beta})\}$ is equal to $\Phi_2(\theta_0\, W_{ij}(\boldsymbol{\alpha}_0)) + O_p(\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|)$. □

Now, we prove the Theorem 2.5.

Proof of Theorem 2.5. . Note that $\boldsymbol{\Psi}_n = (\boldsymbol{\Psi}_{1,n}^T, \boldsymbol{\Psi}_{2,n}^T)^T$. The asymptotic linearity of $\boldsymbol{\Psi}_{1,n}(\boldsymbol{\alpha})$ has already established by Ying (1993). All we need to show is the asymptotic linearity of $\boldsymbol{\Psi}_{2,n}(\boldsymbol{\eta})$. Write $\boldsymbol{\Psi}_{2,n}(\boldsymbol{\eta}) \equiv \mathbf{T}_n(\boldsymbol{\alpha}, \boldsymbol{\beta}) + \mathbf{U}_n(\boldsymbol{\alpha}, \theta)$, where $\mathbf{T}_n(\boldsymbol{\alpha}, \boldsymbol{\beta}) \equiv n^{-2} \sum_{i,j=1}^n \Delta_i^\circ(\boldsymbol{\alpha}) \Delta_j^\circ(\boldsymbol{\alpha}) \left(\widehat{W}_{ij}(\boldsymbol{\alpha}), \mathbf{Z}_{ij}^T\right)^T I\{\varepsilon_i^y(\boldsymbol{\beta}) \geq \varepsilon_j^y(\boldsymbol{\beta})\}$ and $\mathbf{U}_n(\boldsymbol{\alpha}, \theta) \equiv n^{-2} \sum_{i,j=1}^n \Delta_i^\circ(\boldsymbol{\alpha}) \Delta_j^\circ(\boldsymbol{\alpha}) \left(W_{ij}(\boldsymbol{\alpha}), \mathbf{Z}_{ij}^T\right)^T \Phi_2(\theta \widehat{W}_{ij}(\boldsymbol{\alpha}))$. In the following, we first show the asymptotic linearity of $\widehat{W}_i(\boldsymbol{\alpha})$ around $\boldsymbol{\alpha}_0$. Then we show the asymptotic linearity of $\mathbf{T}_n(\boldsymbol{\alpha}, \boldsymbol{\beta})$ and $\mathbf{U}_n(\boldsymbol{\alpha}, \theta)$ in order to obtain the asymptotic linearity of $\boldsymbol{\Psi}_{2,n}(\boldsymbol{\eta})$.

From (2.12), we have the asymptotic linearity of $\widehat{F}_T(t, \boldsymbol{\alpha})$ around $\boldsymbol{\alpha}_0$. Since $\widehat{W}_i(\boldsymbol{\alpha}) = \widehat{H}_T(\varepsilon_i^x(\boldsymbol{\alpha})) = \Phi^{-1}(\widehat{F}_T(\varepsilon_i^x(\boldsymbol{\alpha}); \boldsymbol{\alpha}))$, we may consider $\widehat{H}_T(t; \boldsymbol{\alpha}_0) = \Phi^{-1}(\widehat{F}_T(t; \boldsymbol{\alpha}_0))$ on $t \in [\tau_a, \tau_b]$, where the function $\Phi^{-1}(\cdot)$ is defined over the interval $s \subset (0, 1)$ and s is bounded away from both 0 and 1. By using standard Taylor expansion and putting together (2.12) and uniform consistency result in Lemma 2.1, we can establish the asymptotic linearity of $\widehat{W}_i(\boldsymbol{\alpha})$ around $\boldsymbol{\alpha}_0$, i.e.,

$$\sup_{\|\boldsymbol{\alpha}-\boldsymbol{\alpha}_0\| \leq \delta_n} \left\| \widehat{W}_i(\boldsymbol{\alpha}) - \widehat{W}_i(\boldsymbol{\alpha}_0) - W_i(\boldsymbol{\alpha}) + W_i(\boldsymbol{\alpha}_0) \right\| = o_p(n^{-1/2} \vee \|\boldsymbol{\alpha} - \boldsymbol{\alpha}_0\|), \quad (2.16)$$

for any positive sequence $\delta_n \to 0$ in probability.

Define $\mathbf{T}_n^*(\boldsymbol{\alpha}, \boldsymbol{\beta}) \equiv E[\mathbf{T}_n(\boldsymbol{\alpha}, \boldsymbol{\beta}) \mid X_1, ..., X_n, \mathbf{Z}_1, ..., \mathbf{Z}_n]$ as the conditional expectation of $\mathbf{T}_n(\boldsymbol{\alpha}, \boldsymbol{\beta})$ given $X_i, \mathbf{Z}_i, i = 1, ..., n$. With $X_i, \mathbf{Z}_i, i = 1, ..., n$ treated as constants, $\mathbf{T}_n(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is indeed a weighted log-rank type estimating function. Under Conditions *C1–C5*, when fixing $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0$, we can obtain the asymptotic linearity of $\mathbf{T}_n(\boldsymbol{\alpha}_0, \boldsymbol{\beta})$ in a small neighborhood around $\boldsymbol{\beta}_0$ by following Theorems 1 and 2 in Lai and Ying (1988)

or Theorem 2 in Ying (1993), i.e., for any positive sequences $\delta_n \to 0$ in probability,

$$\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|\le\delta_n} \|\mathbf{T}_n(\boldsymbol{\alpha}_0,\boldsymbol{\beta}) - \mathbf{T}_n(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0) - \mathbf{T}_n^*(\boldsymbol{\alpha}_0,\boldsymbol{\beta}) + \mathbf{T}_n^*(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0)\| \;=\; o_p(n^{-1/2} \vee \|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|). \tag{2.17}$$

Using the similar argument in the proof of Lemma 2.2, we can show that $\mathbf{T}_n(\boldsymbol{\alpha}_0,\boldsymbol{\beta})$ converges to $\mathbf{T}_n^*(\boldsymbol{\alpha}_0,\boldsymbol{\beta})$ uniformly in $\boldsymbol{\beta}$. Combining the asymptotic linearity result of $\widehat{W}_i(\boldsymbol{\alpha})$ with the uniform consistency result of $\mathbf{T}_n(\boldsymbol{\alpha}_0,\boldsymbol{\beta})$, and by the Taylor expansion of $\Delta_i^\circ(\boldsymbol{\alpha})W_i(\boldsymbol{\alpha})$ around $\Delta_i^\circ(\boldsymbol{\alpha}_0)W_i(\boldsymbol{\alpha}_0)$, we can obtain that for any positive sequences $\delta_{1,n} \to 0$ and $\delta_{2,n} \to 0$ in probability,

$$\begin{aligned}&\sup_{\|\boldsymbol{\alpha}-\boldsymbol{\alpha}_0\|\le\delta_{1,n},\,\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|\le\delta_{2,n}} \|\mathbf{T}_n(\boldsymbol{\alpha},\boldsymbol{\beta}) - \mathbf{T}_n(\boldsymbol{\alpha}_0,\boldsymbol{\beta}) - \mathbf{T}_n^*(\boldsymbol{\alpha},\boldsymbol{\beta}) + \mathbf{T}_n^*(\boldsymbol{\alpha}_0,\boldsymbol{\beta})\| \\ &\quad = \; o_p(n^{-1/2} \vee \|\boldsymbol{\alpha}-\boldsymbol{\alpha}_0\|).\end{aligned} \tag{2.18}$$

Combing equation (2.17) with equation (2.18) we obtain that under conditions *C1–C5*,

$$\begin{aligned}&\sup_{\|\boldsymbol{\alpha}-\boldsymbol{\alpha}_0\|\le\delta_{1,n},\,\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|\le\delta_{2,n}} \|\mathbf{T}_n(\boldsymbol{\alpha},\boldsymbol{\beta}) - \mathbf{T}_n(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0) - \mathbf{T}_n^*(\boldsymbol{\alpha},\boldsymbol{\beta}) + \mathbf{T}_n^*(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0)\| \\ &\quad o_p(n^{-1/2} \vee \|(\boldsymbol{\alpha}^T,\boldsymbol{\beta}^T)^T - (\boldsymbol{\alpha}_0^T,\boldsymbol{\beta}_0^T)^T\|).\end{aligned} \tag{2.19}$$

Define $\mathbf{T}(\boldsymbol{\alpha},\boldsymbol{\beta}) \equiv E[\mathbf{T}_n(\boldsymbol{\alpha},\boldsymbol{\beta})]$. Write $\mathbf{T}_n^*(\boldsymbol{\alpha},\boldsymbol{\beta}) - \mathbf{T}_n^*(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0) - \mathbf{T}(\boldsymbol{\alpha},\boldsymbol{\beta}) + \mathbf{T}(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0)$ $\equiv (I) + (II)$, where

$$\begin{aligned}(I) \;&=\; \mathbf{T}_n^*(\boldsymbol{\alpha},\boldsymbol{\beta}) - \mathbf{T}_n^*(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0) \\ &-\; n^{-2}\sum_{i,j=1}^n \left\{\Delta_i^\circ(\boldsymbol{\alpha})\Delta_j^\circ(\boldsymbol{\alpha})\,\left(W_{ij}(\boldsymbol{\alpha}),\mathbf{Z}_{ij}^T\right)^T E\left[I\{\varepsilon_i^y(\boldsymbol{\beta}) \ge \varepsilon_j^y(\boldsymbol{\beta})\} \mid X_i, X_j, \mathbf{Z}_i, \mathbf{Z}_j\right]\right\} \\ &+\; n^{-2}\sum_{i,j=1}^n \left\{\Delta_i^\circ(\boldsymbol{\alpha}_0)\Delta_j^\circ(\boldsymbol{\alpha}_0)\,\left(W_{ij}(\boldsymbol{\alpha}_0),\mathbf{Z}_{ij}^T\right)^T E\left[I\{\varepsilon_i^y(\boldsymbol{\beta}_0) \ge \varepsilon_j^y(\boldsymbol{\beta}_0)\} \mid X_i, X_j, \mathbf{Z}_i, \mathbf{Z}_j\right]\right\},\end{aligned}$$

$$
\begin{aligned}
(II) &= n^{-2}\sum_{i,j=1}^{n}\Big\{\Delta_i^{\circ}(\boldsymbol{\alpha})\Delta_j^{\circ}(\boldsymbol{\alpha})\left(W_{ij}(\boldsymbol{\alpha}),\mathbf{Z}_{ij}^T\right)^T E\left[I\{\varepsilon_i^y(\boldsymbol{\beta})\geq\varepsilon_j^y(\boldsymbol{\beta})\}\mid X_i,X_j,\mathbf{Z}_i,\mathbf{Z}_j\right]\Big\} \\
&\quad - n^{-2}\sum_{i,j=1}^{n}\Big\{\Delta_i^{\circ}(\boldsymbol{\alpha}_0)\Delta_j^{\circ}(\boldsymbol{\alpha}_0)\left(W_{ij}(\boldsymbol{\alpha}_0),\mathbf{Z}_{ij}^T\right)^T E\left[I\{\varepsilon_i^y(\boldsymbol{\beta}_0)\geq\varepsilon_j^y(\boldsymbol{\beta}_0)\}\mid X_i,X_j,\right. \\
&\quad \left.\mathbf{Z}_i,\mathbf{Z}_j\right]\Big\} - \mathbf{T}(\boldsymbol{\alpha},\boldsymbol{\beta}) + \mathbf{T}(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0).
\end{aligned}
$$

Under condition *C5*, the conditional density function of $\varepsilon_{ij}^y(\boldsymbol{\beta}) \equiv \varepsilon_i^y(\boldsymbol{\beta}) - \varepsilon_j^y(\boldsymbol{\beta})$ given $X_i, X_j, \mathbf{Z}_i, \mathbf{Z}_j, \Delta_i = 1, \Delta_j = 1$ is bounded. Then we can show by Taylor expansion that $E\left[I\{\varepsilon_i^y(\boldsymbol{\beta})\geq\varepsilon_j^y(\boldsymbol{\beta})\}\mid X_i,X_j,\mathbf{Z}_i,\mathbf{Z}_j\right] = E\left[I\{\varepsilon_i^y(\boldsymbol{\beta}_0)\geq\varepsilon_j^y(\boldsymbol{\beta}_0)\}\mid X_i,X_j,\mathbf{Z}_i,\mathbf{Z}_j\right] + O_p(\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|)$, uniformly in β. Combining this result with the asymptotic linearity result of $\widehat{W}_i(\boldsymbol{\alpha})$, we have

$$
\begin{aligned}
&\sup_{\|\boldsymbol{\alpha}-\boldsymbol{\alpha}_0\|\leq\delta_{1,n},\,\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|\leq\delta_{2,n}}|(I)| \\
&\leq \Big|E\left[I\{\varepsilon_i^y(\boldsymbol{\beta}_0)\geq\varepsilon_j^y(\boldsymbol{\beta}_0)\}\mid X_i,X_j,\mathbf{Z}_i,\mathbf{Z}_j\right]\Big|\sup_{\|\boldsymbol{\alpha}-\boldsymbol{\alpha}_0\|\leq\delta_{1,n}}\Big|\Delta_i^{\circ}(\boldsymbol{\alpha})\Delta_j^{\circ}(\boldsymbol{\alpha})\widehat{W}_{ij}(\boldsymbol{\alpha}) \\
&\quad - \Delta_i^{\circ}(\boldsymbol{\alpha}_0)\Delta_j^{\circ}(\boldsymbol{\alpha}_0)\widehat{W}_{ij}(\boldsymbol{\alpha}_0) - \Delta_i^{\circ}(\boldsymbol{\alpha})\Delta_j^{\circ}(\boldsymbol{\alpha})W_{ij}(\boldsymbol{\alpha}) + \Delta_i^{\circ}(\boldsymbol{\alpha}_0)\Delta_j^{\circ}(\boldsymbol{\alpha}_0)W_{ij}(\boldsymbol{\alpha}_0)\Big| \\
&\quad + \sup_{\|\boldsymbol{\alpha}-\boldsymbol{\alpha}_0\|\leq\delta_{1,n}}\Big|\Delta_i^{\circ}(\boldsymbol{\alpha})\Delta_j^{\circ}(\boldsymbol{\alpha})\widehat{W}_{ij}(\boldsymbol{\alpha}) - \Delta_i^{\circ}(\boldsymbol{\alpha})\Delta_j^{\circ}(\boldsymbol{\alpha})W_{ij}(\boldsymbol{\alpha})\Big|\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|\leq\delta_{2,n}}\Big| \\
&\quad E\left[I\{\varepsilon_i^y(\boldsymbol{\beta})\geq\varepsilon_j^y(\boldsymbol{\beta})\}\mid X_i,X_j,\mathbf{Z}_i,\mathbf{Z}_j\right] - E\left[I\{\varepsilon_i^y(\boldsymbol{\beta}_0)\geq\varepsilon_j^y(\boldsymbol{\beta}_0)\}\mid X_i,X_j,\mathbf{Z}_i,\mathbf{Z}_j\right]\Big| \\
&= o_p(n^{-1/2}\vee\|(\boldsymbol{\alpha}^T,\boldsymbol{\beta}^T)^T-(\boldsymbol{\alpha}_0^T,\boldsymbol{\beta}_0^T)^T\|). \qquad (2.20)
\end{aligned}
$$

The expression (II) involves continuous functions with bounded derivatives, by Taylor expansion and uniform consistency results, it is easy to show that $\sup_{\|\boldsymbol{\alpha}-\boldsymbol{\alpha}_0\|\leq\delta_{1,n},\,\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|\leq\delta_{2,n}} |(II)| = o_p(n^{-1/2}\vee\|(\boldsymbol{\alpha}^T,\boldsymbol{\beta}^T)^T-(\boldsymbol{\alpha}_0^T,\boldsymbol{\beta}_0^T)^T\|)$, which combing with (2.20) leads to the result that under conditions *C1–C5*, for any positive sequences $\delta_{1,n}\to 0$ and $\delta_{2,n}\to 0$ in

probability,

$$\begin{aligned} & \sup_{\|\boldsymbol{\alpha}-\boldsymbol{\alpha}_0\|\le\delta_{1,n},\,\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|\le\delta_{2,n}} \|\mathbf{T}_n^*(\boldsymbol{\alpha},\boldsymbol{\beta}) - \mathbf{T}_n^*(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0) - \mathbf{T}(\boldsymbol{\alpha},\boldsymbol{\beta}) + \mathbf{T}(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0)\| \\ = \quad & o_p(n^{-1/2} \vee \|(\boldsymbol{\alpha}^T,\boldsymbol{\beta}^T)^T - (\boldsymbol{\alpha}_0^T,\boldsymbol{\beta}_0^T)^T\|). \end{aligned} \tag{2.21}$$

From equalities (2.19) and (2.21), we have shown the asymptotic linearity of $\mathbf{T}_n(\boldsymbol{\alpha},\boldsymbol{\beta})$. Note that $\mathbf{U}_n(\boldsymbol{\alpha},\theta)$ is a continuous function of $\widehat{W}_{ij}(\boldsymbol{\alpha})$ with bounded derivatives. By using the asymptotic linearity result of $\widehat{W}_i(\boldsymbol{\alpha})$ and the Taylor expansion, we can show the asymptotic linearity of $\mathbf{U}_n(\boldsymbol{\alpha},\theta)$. Then we can readily establish the asymptotic linearity of $\boldsymbol{\Psi}_{2,n}(\boldsymbol{\eta})$ and hence the asymptotic linearity of $\boldsymbol{\Psi}_n(\boldsymbol{\eta})$, i.e., the equality (2.13). □

Proof of Theorem 2.2. The asymptotically normality of $(\widehat{\boldsymbol{\alpha}}_0^T,\widehat{\boldsymbol{\beta}}_0^T,\widehat{\theta}_0)^T$ can be obtained from Theorems 2.4 and 2.5 by fairly standard arguments. □

Chapter 3

Semiparametric Inference for Successive Durations

A bi-state progressive process is a reasonable model for the course of many chronic diseases. As an example, a national intergroup trial was conducted in 1980's to evaluate the effect of the drugs levamisole and fluorouracil, as adjuvant therapy for resected colorectal carcinoma (Moertel et al. 1990). In this colon cancer trial, 929 eligible patients with Stage C disease were randomized to three study arms, including observation, levamisole alone, and levamisole combined with fluorouracil treatments. During the study, 468 patients experienced cancer recurrence, and 414 of them died by the end of the study; additionally, 38 patients died without cancer recurrence. Such disease progression can be described by a bi-state progression process with cancer-free and recurrence states, where the duration of the recurrence state may be 0. While time to cancer recurrence and time to death can be addressed by standard statistical methods, time between recurrence and death, i.e., the duration of the recurrence state is also of scientific interest but its estimation with incomplete follow-up data is a challenge. In this chapter, we

address the estimation with such a time-between-events outcome.

Let T_k, $k = 1, 2$, be the two durations in order, and C be the censoring time. As we discussed in Chapter 1, under serial censoring, the second duration T_2 is subject to induced dependent censoring, and the marginal distribution of T_2 is nonparametrically nowhere identifiable. Current available nonparametric approaches have limitations in both estimation and prediction.

Apparently, all these issues could be resolved with a fully parametric approach. But the price is model flexibility. Confronted with similar issues in problems with a univariate mark such as lifetime medical cost, Huang and Berry (2006) proposed a semiparametric copula modeling (Sklar 1959) strategy to strike a balance between model flexibility and identifiability. A parametric form is taken for the association structure between time to the event and the mark, while their marginal distributions are left unspecified. Upon viewing the second duration as a mark of the *initiating* event of the second duration, the modeling strategy of Huang and Berry (2006) may be adopted. However, there exist two significant complications. First, the second duration as a mark is subject to further censoring even upon the occurrence of the associated event. Second, the second duration may be 0, corresponding to the situation of reaching death without cancer recurrence in the colon cancer example. Thus, it is necessary to accommodate the situation that T_2 has a probability mass at 0.

In Section 3.1, we introduce this semiparametric model for the bi-state progressive process, along with a more realistic censoring mechanism. An inference procedure is proposed in Section 3.2 and asymptotic properties of the resulting estimators are established in Section 3.3. Section 3.4 gives numerical results for simulation studies and the application to a colon cancer trial. Section 3.5 concludes with discussion. All the proofs are collected in Section 3.6.

3.1 Semiparametric Modeling

With the copula modeling strategy, we parameterize the copula function of the two durations but leave their marginal distributions unspecified. This is particularly desirable when the marginal distributions might be of interest. In the following, we present a normal copula model first and then impose a realistic censoring mechanism.

3.1.1 Normal copula model

There are various copula families; see Shih and Louis (1995), Genest, Ghoudi, and Rivest (1995), and the references therein. In this paper, we specifically study the normal copula (cf. Klaassen and Wellner, 1997; Huang and Berry, 2006). However, continuous marginal distributions are typically assumed with copula modeling. To allow for the generality of the marginals, we specify the normal copula model as follows: there exist increasing transformations $G_k(\cdot)$, $k = 1, 2$, such that

$$\Pr\{T_1 \le t_1,\, T_2 \le t_2\} \;=\; \Phi_2\{G_1(t_1),\, G_2(t_2),\, \rho_0\}, \qquad t_1,\, t_2 \ge 0, \tag{3.1}$$

where $\Phi_2(\cdot, \cdot, \rho_0)$ is the cumulative distribution function of standard bivariate normal distribution with correlation coefficient ρ_0. Write marginal distributions $F_k(t) \equiv \Pr(T_k \le t)$, $k = 1, 2$. It is easy to see that

$$G_k(\cdot) = \Phi^{-1}\{F_k(\cdot)\}, \qquad k = 1, 2, \tag{3.2}$$

satisfy (3.1), where $\Phi(\cdot)$ is the cumulative distribution function of standard univariate normal distribution. Clearly, the marginal distributions of T_k, $k = 1, 2$, are completely unspecified. In the colon cancer trial, a non-negligible portion of patients were observed

dying free of cancer recurrence, which means the second durations T_2 for those patients equal 0. The normal copula model (3.1) is able to accommodate the situation where T_2 has a probability mass at 0. In the following, we focus on the situation that T_1 has a completely continuous distribution, and T_2 has a probability mass at 0 and is continuously distributed beyond 0.

By normal distribution theory, normal copula model (3.1) implies

$$\Pr\{T_2 \le t_2 \,|\, T_1\} \;=\; \Phi\left\{-\theta_0 G_1(T_1) + (1-\rho_0^2)^{-1/2} G_2(t_2)\right\}, \quad t_2 \ge 0, \tag{3.3}$$

with $G_k(\cdot)$, $k = 1, 2$, defined as in (3.2) and $\theta_0 \equiv \rho_0(1-\rho_0^2)^{-1/2}$. Thus, copula model (3.1) may be alternatively viewed as a regression model for T_2 with T_1 as the covariate; see Cook and Lawless (1997) for regression analysis of T_2 given T_1 under different regression models.

Remark 3.1. In the special case where the probability mass of T_2 at 0 is 0, the model (3.3) would reduce to the following linear transformation model (Cheng, Wei, and Ying 1995) with standard normal error ε:

$$(1-\rho_0^2)^{-1/2} G_2(T_2) \,|\, T_1 \;\sim\; \theta_0 G_1(T_1) + \varepsilon, \tag{3.4}$$

That is, upon an unspecified increasing transformation, T_2 is linearly related to $G_1(T_1)$ with a standard normal error.

3.1.2 A realistic censoring mechanism

To accommodate a realistic censoring mechanism, we reformulate the serial censoring. Let C_k be the censoring time of T_k in the sense that

$$X_k \equiv T_k \wedge C_k, \qquad \Delta_k \equiv I(T_k \leq C_k), \qquad\qquad k = 1, 2.$$

Since the second duration is not observable when the first one is censored, the serial censoring has the following constraint on C_2:

$$C_2 = 0 \qquad \text{if } T_1 > C_1.$$

For the purpose of identifiability, we assume that $\Pr\{C_2 = 0 \,|\, T_1 \leq C_1\} = 0$. We take the viewpoint that T_2 is a mark of the initiating event of the second duration, i.e., the transition between the two durations. Consequently, the censoring mechanism imposed consists of two nested components. One is on this marked point process, and the other is on the mark T_2 after its associated event has occurred:

$$\begin{cases} C_1 \;\; \perp \;\; \{T_1, T_2\} \\ C_2 \;\; \perp \;\; T_2 \;\; \Big| \;\; (T_1, T_1 \leq C_1) \end{cases}, \tag{3.5}$$

where $\perp$ represents independence. The dependence of the process history is allowed and unspecified, which is more general and realistic than the censoring mechanism required in Wang and Wells (1998) and Lin, Sun, and Ying (1999).

3.2 Proposed Estimation Procedure

Suppose that the data consist of $\{X_{ki}, \Delta_{ki} : k = 1, 2\}$, $i = 1, \cdots, n$, as n iid replicates of $\{X_k, \Delta_k : k = 1, 2\}$. Among the three parameters with normal copula model (3.1), $F_1(.)$ can be readily estimated. However, the estimation of $F_2(\cdot)$ and ρ_0 is not as obvious and indeed challenging. Throughout this article, our attention is restricted to situations with $\rho_0 \in (-1, 1)$.

The current application of bivariate copula modeling is different from those studied by Klaassen and Wellner (1997), Genest, Ghoudi, and Rivest (1995), and Shih and Louis (1995), where both marginal distributions can be easily estimated and the association parameter is the focus of interest. This problem is similar to that of Huang and Berry (2006), however, with the complication that the mark, i.e., T_2, is still subject to censoring after the occurrence of the transition between the two states.

Given that $C_1 \perp \{T_1, T_2\}$ in censoring mechanism (3.5), it follows that,

$$\Pr\{T_2 \le t_2 \,\big|\, Z, \Delta_1 = 1\} \;=\; \Phi\left\{-\theta_0 Z + H_0(t_2)\right\}, \quad t_2 \ge 0, \tag{3.6}$$

where $H_0(\cdot) \equiv (1-\rho_0^2)^{-1/2} G_2(\cdot)$ and $Z \equiv G_1(X_1)$. Note that $H_0(\cdot)$ and θ_0 are determined by $F_2(\cdot)$ and ρ_0, and vice versa. In the following, we propose an inference procedure for estimation by utilizing the martingale structure associated with survival data.

Using counting process notation for the second duration, we write $N(t) = \Delta_2 I(X_2 \le t, \Delta_1 = 1)$, $Y(t) = I(X_2 \ge t, \Delta_1 = 1)$ and $M_2(t) = N(t) - Y(0)\Phi\{-\theta Z + H(0)\} - \int_{s\in(0,t]} Y(s)\, d\Lambda\{-\theta Z + H(s)\}$, where $\Lambda(\cdot)$ is the cumulative hazard function of standard normal distribution. Since $M_2(t)$ is the martingale process associated with the counting process $N(t)$, we have $E\{M_2(t)|Z\} = 0$. Note that Z in model (3.6) is not directly observed. Nevertheless, it can be estimated. Let $\widehat{F}_1(\cdot)$ be the Kaplan–Meier estimator

of $F_1(\cdot)$ using data $\{X_{1i}, \Delta_{1i}\}$, $i = 1, \cdots, n$. A natural estimator for $Z = G_1(X_1)$ is

$$\widehat{Z} \equiv \Phi^{-1}\{\widehat{F}_1(X_1)\}.$$

Define $\tau_1 = \inf\{t : \Pr(X_1 > t) = 0\}$. To avoid technical difficulties arising from the unboundedness of $\Phi^{-1}(\cdot)$ at 0 and 1 and from the tail instability of $\widehat{F}_1(\cdot)$, we only consider individuals with $X_1 \in [a, b]$ where $0 < a < b < \tau_1$ for constants a and b. Write $\Delta^{ab} = I(a \leq X_1 \leq b)$ and τ_2 as a constant satisfying regularity condition 1 stated in section **??**. Then, we construct the following estimating equations for $\{H_0(\cdot), \theta_0\}$:

$$\sum_{i=1}^{n} \Delta_i^{ab} \widehat{Z}_i \left[N_i(\tau_2) - Y_i(0)\Phi\left\{-\theta\widehat{Z}_i + H(0)\right\} - \int_{t\in(0,\tau_2]} Y_i(t)\, d\Lambda\left\{-\theta\widehat{Z}_i + H(t)\right\}\right] = 0 \tag{3.7}$$

$$\sum_{i=1}^{n} \Delta_i^{ab} \left[N_i(t) - Y_i(0)\Phi\left\{-\theta\widehat{Z}_i + H(0)\right\} - \int_{s\in(0,t]} Y_i(s)\, d\Lambda\left\{-\theta\widehat{Z}_i + H(s)\right\}\right] = 0 \quad (t \geq 0). \tag{3.8}$$

To motivate these equations, with $H(\cdot)$ given, (3.7) may be used to estimate θ. On the other hand, (3.8) is an estimating equation for $H(\cdot)$ with given θ.

Based on Remark 3.1, in the special case where the probability mass of T_2 at 0 is 0 and $G_1(\cdot)$ is given, the model (3.6) is a linear transformation model with independent censoring since the censoring mechanism (3.5) implies $C_2 \perp T_2 \,\big|\, (X_1, \Delta_1 = 1)$. In this case, our proposed estimating equations (3.7) and (3.8) reduce to those for linear transformation model developed by Chen, Jin, and Ying (2002). Compared with the estimation of linear transformation model in Chen, Jin, and Ying (2002), our estimation problem has some unique features. First, T_2 in our set-up has a probability mass at 0, and a probit model is induced for $T_2 = 0$. Specially, the induced model (3.3) implies

that $I(T_2 > 0)$ follows a probit model (McCullagh and Nelder 1989) conditioning on T_1,

$$\Pr\{T_2 = 0 \,\big|\, T_1\} = \Phi\left\{-\theta_0 G_1(T_1) + (1-\rho_0^2)^{-1/2} G_2(0)\right\}. \tag{3.9}$$

Second, Chen et al. (2002) emphasizes on the estimator of the regression parameters (e.g., θ_0 in our setup), while in our case the emphasis is on the estimation of the marginal distribution of T_2.

We denote $\{\widehat{H}(\cdot), \widehat{\theta}\}$ as the solution of (3.7) and (3.8), where $\widehat{H}(\cdot)$ jumps only at observed failure times $0, t_1, ..., t_K$, and propose an iterative algorithm to compute $\{\widehat{H}(\cdot), \widehat{\theta}\}$.

Step 1. Set $m = 0$ and obtain the initial value for $\widehat{\theta}$, denoted by $\widehat{\theta}^{(0)}$, as estimate from the probit model

$$\Pr\{T_2 = 0 \,\big|\, (Z,\, \Delta_1 = 1)\} = \Phi\left\{-\theta_0 Z + H(0)\right\}.$$

Step 2. Find $H^{(m)}(\cdot)$ and a new estimate for θ_0, denoted by $\widehat{\theta}^{(m+1)}$ as follows. First, obtain $H^{(m)}(0)$ by solving

$$\sum_{i=1}^{n} \Delta_i^{ab} \left[N_i(0) - Y_i(0)\Phi\left\{-\theta^{(m)}\widehat{Z}_i + H(0)\right\}\right] = 0.$$

Then, obtain $H^{(m)}(t_k)$, for $k = 1, 2, ..., K$, by sequentially solving the equation

$$\begin{aligned}\sum_{i=1}^{n} \Delta_i^{ab}\, Y_i(t_k)\, \Lambda\left\{-\theta^{(m)}\widehat{Z}_i + H(t_k)\right\} \;=\; & \sum_{i=1}^{n} \Delta_i^{ab} \left[N_i(t_k) - N_i(t_k -)\right. \\ & \left. + Y_i(t_k)\, \Lambda\left\{-\theta^{(m)}\widehat{Z}_i + H(t_k -)\right\}\right].\end{aligned}$$

Next, obtain $\widehat{\theta}^{(m+1)}$ by solving (3.7) with $H = H^{(m)}$.

Step 3. Update m to $m + 1$.

Step 4. Repeat Step 2 and Step 3 until a predetermined convergence criterion is met.

Note that the estimating equations (3.7) and (3.8) alone will not guarantee a unique zero-crossing. By using the the probit model for $T_2 = 0$ as in our algorithm, we have a consistent initial estimate for θ_0, which provides guidance for us to find the zero-crossing close to the truth. After obtaining $\{\widehat{H}(\cdot), \widehat{\theta}\}$ as an estimator of $\{H_0(\cdot), \theta_0\}$, we can get an estimator for $\{F_2(\cdot), \rho_0\}$:

$$\widehat{F}_2(\cdot) = \Phi\left\{\left(1+\widehat{\theta}^{\,2}\right)^{-1/2}\widehat{H}(\cdot)\right\}, \qquad \widehat{\rho} = \widehat{\theta}\left(1+\widehat{\theta}^{\,2}\right)^{-1/2}. \tag{3.10}$$

In terms of the interval estimation for $\{F_2(\cdot), \rho_0\}$, we have two approaches available in general. One is to derive their influence functions as maps from the distribution of $\{X_1, X_2, \Delta_1, \Delta_2\}$. However, the derivation is algebraically complex. An alternative is through resampling. We adopt nonparametric bootstrap here to obtain the estimated standard error. Then, the Wald-type confidence interval will be constructed.

3.3 Asymptotic Study

In this section, we establish the asymptotic properties of the proposed estimators.

We assume the following regularity conditions.

Condition 1. Let τ_2 be a constant such that $Pr(T_2 > \tau_2) > 0$ and $Pr(C_2 > \tau_2) > 0$.

Condition 2. The partial derivative $D_1(t)$ derived in (3.21) is uniformly bounded away from 0 and finite for $t \in [0, \tau_2]$. The partial derivative D_2 derived in (3.22) is bounded away from 0 and finite.

The uniform consistency and weak convergence of Kaplan-Meier estimator $\widehat{F}_1(\cdot)$ are well known. We have the following theorems for the asymptotic properties of the pro-

posed estimators.

Theorem 3.1. *Suppose that the regularity conditions C1-C2 hold. There exists a neighborhood of θ_0 within which $\widehat{\theta}$ is unique, and $\widehat{\theta} \to \theta_0$ and $\sup_{t\in[0,\tau_2]} \|\widehat{H}(t) - H_0(t)\| \to 0$, in probability. Furthermore, $\widehat{\rho} \to \rho_0$ and $\sup_{t\in[0,\tau_2]} \|\widehat{F}_2(t) - F_2(t)\| \to 0$, in probability.*

Theorem 3.2. *Under the regularity conditions C1-C2, $n^{1/2}(\widehat{\theta} - \theta_0)$ is asymptotically normal with mean 0, and $n^{1/2}\{\widehat{H}(t) - H_0(t)\}$ weakly converges to a Gaussian process for $t \in [0, \tau_2]$. Thus, $n^{1/2}(\widehat{\rho} - \rho_0)$ is asymptotically normal with mean 0, and $n^{1/2}\{\widehat{F}_2(t) - F_2(t)\}$ weakly converges to a Gaussian process for $t \in [0, \tau_2]$.*

3.4 Numerical Studies

Simulations were carried out to assess the finite-sample performance of both the estimator of the joint distribution function $F_{12}(t_1, t_2)$ and the estimator of marginal distribution function of T_2.

3.4.1 Simulation Under Normal Copula Model

We consider the general situation where T_2 has a probability mass at 0. To generate the successive durations (T_1, T_2), we first generate (S_1, S_2) from bivariate normal copula with standard exponential margins. Then, let $T_1 = S_1$ and $T_2 = (S_2 - Q_{S_2}(0.05)) \cdot I(S_2 > Q_{S_2}(0.05))$, where $Q_{S_2}(\tau)$ is the τ-th quantile of S_2. Thus, T_2 has a probability mass of 0.05 at $t = 0$. Various association levels with $\rho_0 = 0.8, 0.4, 0, -0.4, -0.8$ were considered, corresponding to strong positive, moderate positive, independent, moderate negative and strong negative associations between the two durations. The follow-up time was subject to right censoring by censoring time C, which is independent of $\{T_1, T_2\}$ and

has an exponential distribution with rate 0.2 but truncated at 2. Thus, any event with $T_1 > 2$, along with associated T_2 is censored, corresponding to the top 13.5% of T_1; thus, the marginal distribution of T_2 is nonparametrically nowhere identifiable. The overall censoring rate for T_1 and T_2 are 24.3% and 46.1% -52.9% respectively. The sample size was 200 and the replication time was 1000.

Table 3.1 presents the findings of the simulation for ρ_0 and the marginal distribution $F_2(t_2)$. The nonparametric bootstrap with size 200 was used to obtain estimated standard deviation and 95% Wald-type confidence interval. As seen, The proposed estimator for ρ_0 is virtually unbiased and the coverage probability of its Wald-type confidence interval is fairly accurate. The proposed estimator for marginal distribution $F_2(t_2)$ is essentially unbiased and the empirical coverage probability of its Wald-type confidence interval is very close to its nominal level in general. When T_1 and T_2 have strong negative association, i.e., $\rho_0 = -0.8$, the coverage probability of Wald-type confidence interval for $\widehat{F}_2(t_2)$ at $t_2 = F_2^{-1}(0.05)$ is much lower than its nominal level, which is due to the round-off error in computation when we have limited observations for $T_2 = 0$ under sample size 200. The coverage probability of Wald-type confidence interval for $\widehat{F}_2(t_2)$ at $t_2 = F_2^{-1}(0.05)$ is close to its nominal level when we increases the sample size to 400. When T_1 and T_2 have strong positive association, i.e., $\rho_0 = 0.8$, moderate bias is observed for $\widehat{F}_2(t_2)$ at $t_2 = F_2^{-1}(0.8)$ and the corresponding coverage probability of Wald-type confidence interval is lower than its nominal level. This is mainly due to the heavy censoring on the tail parts of both T_1 and T_2. In our simulation setting, the upper 13.5% of T_1 is censored, and the associated T_2, which is also censored, occurs mostly in the upper tail when $\rho_0 = 0.8$. For comparison purpose, Table 3.1 also includes the naïve Kaplan-Meier estimator of $F_2(t_2)$. The naïve Kaplan-Meier estimator is generally biased except for the special case of $\rho_0 = 0$, where the naïve Kaplan-Meier estimator is

consistent and efficient.

Table 3.2 summarize the simulation results for the joint distribution $F_{12}(t_1, t_2)$. The performance of the estimator for the joint distribution $F_{12}(t_1, t_2)$ is similar to that of the marginal distribution $F_2(t_2)$.

3.4.2 Simulation With Misspecified Copula Model

We considered Clayton's and Frank's families (Shih and Louis 1995). Copulas in these two families generally differ from normal copulas except when survival time and lifetime medical cost are independent. We adopted the same simulation scenario as described above except for the copula. For comparison, corresponding to each nonzero ρ_0 value considered in the three sets above, we chose a nonnormal copula with the same Kendall's tau. Only positive association was considered for Claytons copula, given its limitation in accommodating negative association. The results are presented in Table 3.3. For Clayton's family, the proposed estimator for the marginal distribution $F_2(t_2)$ has moderate bias at the upper tail, especially when T_1 and T_2 are strong associated. In contrast, the proposed inference procedure seems to be rather robust with Frank's family. These simulation findings suggest that the model checking is important in practice.

3.4.3 Application to a Colon Cancer Study

As a cause of death due to cancer, colon cancer is second only to lung cancer in the United States. In approximately 80% of the patients with colon cancer in the United States, the diagnosis is made at a sufficiently early stage when all apparent diseased tissue can be surgically removed. Those who have regional nodal involvement that is clinically completely resected are referred to as having Duke's Stage C disease (Dukes

Table 3.1: *Simulation Summary Statistics for ρ_0 and the Marginal Distribution $F_2(t_2)$ Under Normal Copulas*

			$\widehat{F}_2$ at the τ-th percentile of F_2				
Method		$\widehat{\rho}$	$\tau =$ 0.05	0.20	0.40	0.60	0.80
			$\rho_0 = 0.8$				
Proposed	(a)	-6, 42	2, 16	0, 30	2, 37	2, 45	-22, 59
	(b)	94.1	93.3	94.0	95.1	94.2	86.6
NKM	(a)		0, 16	43, 36	101, 42	152, 41	147, 42
			$\rho_0 = 0.4$				
Proposed	(a)	-5, 101	1, 16	1, 32	2, 42	0, 49	-7, 54
	(b)	92.1	94.1	94.2	95.1	93.5	92.5
NKM	(a)		-4, 15	23, 34	50, 42	64, 44	62, 43
			$\rho_0 = 0$				
Proposed	(a)	6, 115	1, 18	-2, 35	-4, 43	-7, 46	-10, 48
	(b)	94.1	92.6	93.1	95.2	94.1	92.6
NKM	(a)		-12,14	-12, 33	-9, 41	-8, 44	-5, 45
			$\rho_0 = -0.4$				
Proposed	(a)	9, 103	0, 22	-1, 38	-4, 45	-8, 45	-6, 39
	(b)	94.0	91.9	93.1	93.2	94.0	94.4
NKM	(a)		-25,11	-50, 30	-65, 41	-68, 45	-56, 44
			$\rho_0 = -0.8$				
Proposed	(a)	7, 42	-6, 30	-2, 36	-2, 41	-4, 39	-4, 32
	(b)	95.6	77.2	94.8	94.4	94.8	94.7
NKM	(a)		-41, 5	-103,24	-125,38	-116,45	-85, 43

NOTE: (a): bias ($\times 10^3$), empirical standard deviation ($\times 10^3$); (b): empirical coverage (%) of the 95% Wald-type confidence interval; NKM: naive Kaplan-Meier estimator.

Table 3.2: *Simulation Summary for the Joint Distribution $F_{12}(t_1, t_2)$ With the Proposed Method Under Normal Copulas*

		$\widehat{F}_{12}(t_1, t_2)$									
ρ_0	t_1	$t_2 =$	$F_2^{-1}(0.05)$		$F_2^{-1}(0.2)$		$F_2^{-1}(0.4)$		$F_2^{-1}(0.6)$		$F_2^{-1}(0.8)$	
			(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
0.8	$F_1^{-1}(0.2)$		1, 14	92.9	-1, 22	92.8	0, 26	94.2	1, 28	94.1	1, 28	94.1
	$F_1^{-1}(0.4)$		1, 16	93.2	-1, 27	92.9	0, 31	94.7	1, 33	94.9	0, 35	94.8
	$F_1^{-1}(0.6)$		1, 16	93.2	0, 29	93.7	0, 34	94.9	-1, 36	94.4	-8, 38	93.9
	$F_1^{-1}(0.8)$		2, 16	93.3	0, 30	93.7	1, 36	95.1	0, 40	94.9	-17, 43	90.8
0.4	$F_1^{-1}(0.2)$		0, 9	92.0	0, 17	93.8	0, 22	93.5	0, 25	94.6	0, 27	94.2
	$F_1^{-1}(0.4)$		0, 12	92.9	0, 22	93.6	1, 29	94.3	0, 31	94.5	-2, 34	94.1
	$F_1^{-1}(0.6)$		0, 14	93.2	0, 26	94.5	0, 33	93.9	-2, 35	94.8	-5, 38	94.3
	$F_1^{-1}(0.8)$		1, 16	94.0	1, 29	94.4	1, 37	94.5	-1, 40	95.3	-5, 43	94.2
0	$F_1^{-1}(0.2)$		0, 5	90.9	0, 11	92.9	0, 17	93.1	0, 21	93.2	-1, 25	93.4
	$F_1^{-1}(0.4)$		1, 8	91.9	0, 17	93.1	0, 24	93.7	0, 29	94.5	-1, 34	94.1
	$F_1^{-1}(0.6)$		0, 11	92.3	-1, 23	93.1	-1, 30	93.9	-2, 35	94.5	-3, 40	94.9
	$F_1^{-1}(0.8)$		0, 15	92.5	-2, 28	93.1	-3, 35	94.9	-5, 39	94.1	-7, 43	94.4
-0.4	$F_1^{-1}(0.2)$		0, 1	90.0	0, 6	90.2	0, 11	92.3	0, 16	93.5	-1, 22	92.8
	$F_1^{-1}(0.4)$		0, 4	89.7	0, 11	91.1	-1, 19	92.8	-3, 26	92.8	-3, 33	93.4
	$F_1^{-1}(0.6)$		0, 7	90.5	0, 17	92.2	-2, 27	93.6	-5, 33	93.6	-5, 39	94.4
	$F_1^{-1}(0.8)$		0, 11	90.6	-1, 25	92.6	-3, 35	94.0	-6, 39	94.0	-5, 42	94.6
-0.8	$F_1^{-1}(0.2)$		0, 0	76.8	0, 0	91.3	1, 2	93.1	1, 8	92.9	0, 16	94.0
	$F_1^{-1}(0.4)$		0, 0	76.7	0, 2	89.8	1, 10	92.2	-1, 20	92.5	-3, 30	93.5
	$F_1^{-1}(0.6)$		0, 1	77.2	0, 8	90.6	0, 19	92.3	-3, 30	93.1	-5, 38	94.5
	$F_1^{-1}(0.8)$		0, 5	77.7	0, 18	92.2	-1, 31	93.3	-4, 38	94.6	-5, 40	94.3

NOTE: (a): bias ($\times 10^3$), empirical standard deviation ($\times 10^3$); (b): empirical coverage (%) of the 95% Wald-type confidence interval.

Table 3.3: *Simulation Summary Statistics for the Marginal Distribution $F_2(t_2)$ With the Proposed Method Under Non-normal Copulas*

Corresponding ρ_0	$t_2 =$	$F_2^{-1}(0.05)$		$F_2^{-1}(0.2)$		$F_2^{-1}(0.4)$		$F_2^{-1}(0.6)$		$F_2^{-1}(0.8)$	
		(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
						Clayton's copula					
0.8		-1, 15	91.6	-6, 29	93.8	12, 41	93.4	67, 49	71.2	78, 51	54.6
0.4		-1, 17	92.0	6, 32	95.8	22, 44	93.8	40, 48	84.4	49, 50	74.6
						Frank's copula					
0.8		-1, 15	92.3	-8, 29	92.8	3, 39	94.4	23, 47	90.5	-13, 58	88.7
0.4		-2, 16	92.7	-2, 31	95.1	3, 42	95.4	4, 48	93.3	-7, 54	91.3
-0.4		6, 22	94.2	4, 38	94.3	0, 45	93.6	-1, 44	94.3	-1, 39	94.0
-0.8		24, 32	86.1	10, 38	93.8	2, 40	95.0	3, 36	95.7	1, 31	94.3

NOTE: (a): bias ($\times 10^3$), empirical standard deviation ($\times 10^3$); (b): empirical coverage (%) of the 95% Wald-type confidence interval.

1932). Unfortunately, about one-half of these patients have residual cancer existing in an occult and probably microscopic stage, which leads to recurrence of disease and death within 5 years. A national intergroup trial was conducted in 1980's to evaluate the effect of the drugs levamisole and fluorouracil, as adjuvant therapy for resected colorectal carcinoma (Moertel et al. 1990).

In this trial, 929 eligible patients with Stage C disease were randomized to three study arms. Of those, 315, 310 and 304 patients received observation, levamisole alone, and levamisole combined with fluorouracil treatments, respectively. The patient enrollment was begun in March 1984 and was completed in October 1987. The dataset available on the Mayo Clinic website contains much richer long-term information than that used in the original report by Moertel et al. (1990), with a maximum follow up of more than 8 years. By the end of the study, 177 patients in the observation arm had cancer recurrence, among whom 155 died; 172 patients in the levamisole alone arm had cancer recurrence, among whom 151 died; while in the levamisole plus fluorouracil arm 119 patients had cancer recurrence, among whom 108 died. In addition, 38 of the 929 patients in the trial died without cancer recurrence. Among those, 13, 10 and 15 patients belonged to observation, levamisole alone, and levamisole combined with fluorouracil treatments, respectively.

Figure 3.1 (a) and (b) show the Kaplan-Meier estimates for the cumulative distribution function (CDF) of time to cancer recurrence and time to death since randomization for each of the three treatment groups. As for the comparison between the observation group and levamisole alone group, the p-values of the log-rank statistics are 0.756 and 0.812 for cancer recurrence and death, respectively, suggesting that levamisole alone produced no detectable effect. If we compare the the observation group and levamisole plus fluorouracil treatment group, we can see that the p-values of the log-rank statistics

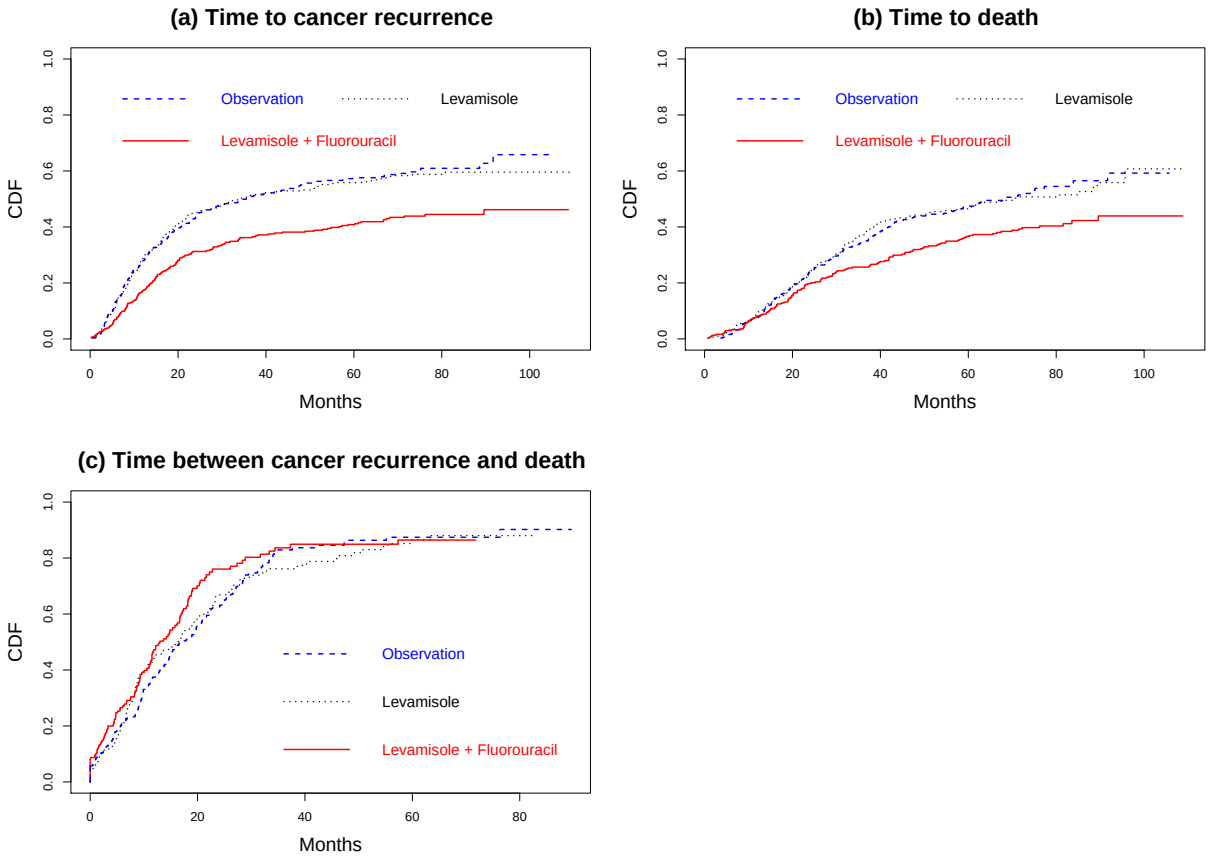

Figure 3.1: Estimated cumulative distribution function (CDF) of T_1, T_2 and $T_1 + T_2$ for each of the three treatment groups in colon cancer study: (a) Kaplan-Meier estimates for the CDF of time to cancer recurrence (T_1), (b) Kaplan-Meier estimates for the CDF of time to death since randomization ($T_1 + T_2$), (c) proposed estimates for the CDF of time between cancer recurrence and death (T_2).

are 2.06×10^{-5} and 1.6×10^{-3} for cancer recurrence and death, respectively, which demonstrated that therapy with levamisole plus fluorouracil produced an unequivocal advantage over observation, and delayed the time to cancer recurrence as well as time to death since randomization.

Under our proposed normal copula model for successive durations, T_1 is the time to cancer recurrence since randomization and T_2 is the time between cancer recurrence and death. For the 38 patients who died without cancer recurrence, T_1 is set to be the time to death since randomization and $T_2 = 0$. Table 3.4 presents the estimates for the marginal distribution $F_2(t_2)$ for $t_2 = 0, 6, 12, ..., 36$ months and ρ_0 in each treatment group. What are also presented are the corresponding 95% Wald-type confidence intervals based on nonparametric bootstrap with size 200. As seen, the estimates for the marginal distribution $F_2(t_2)$ of the levamisole plus fluorouracil treatment group are significantly higher than those of the observation group for $t_2 = 0, 6, 12, ..., 30$ months. In all the three groups, T_1 and T_2 are positively correlated. Figure 3.1 (c) displays the proposed estimates for the CDF of the time between cancer recurrence and death, which indicates that the patients who received levamisole plus fluorouracil treatment died faster after cancer recurrence compared to those in the observation group. This suggests that the treatment with levamisole plus fluorouracil may no longer be beneficial on survival after cancer recurrence.

We also compared the performance of naïve estimates and proposed estimates of cumulative distribution function of time between cancer recurrence and death (T_2) for each of the three treatment groups in colon cancer study. The naïve estimator is obtained by using Kaplan-Meier estimate and ignoring the fact of induced dependent censoring. Figure 3.2 suggests that the naïve estimates are consistently higher than proposed estimates in all three groups.

Table 3.4: *Estimate (95% Wald-type confidence interval) of the marginal distribution of T_2 and ρ_0 in each treatment group.*

t_2 (in months)	Observation	Levamisole	Levamisole + Fluorouracil
0	.057 (.021, .093)	.037 (.011, .064)	.082 (.027, .138)
6	.215 (.149, .281)	.206 (.145, .267)	.266 (.158, .373)
12	.374 (.283, .466)	.446 (.357, .535)	.472 (.328, .616)
18	.510 (.412, .607)	.540 (.445, .636)	.628 (.468, .788)
24	.625 (.525, .725)	.668 (.575, .762)	.760 (.610, .911)
30	.747 (.656, .838)	.732 (.635, .828)	.803 (.657, .949)
36	.829 (.757, .900)	.762 (.667, .865)	.836 (.698, .974)
ρ_0	.330 (.126, .534)	.368 (.166, .569)	.376 (.071, .682)

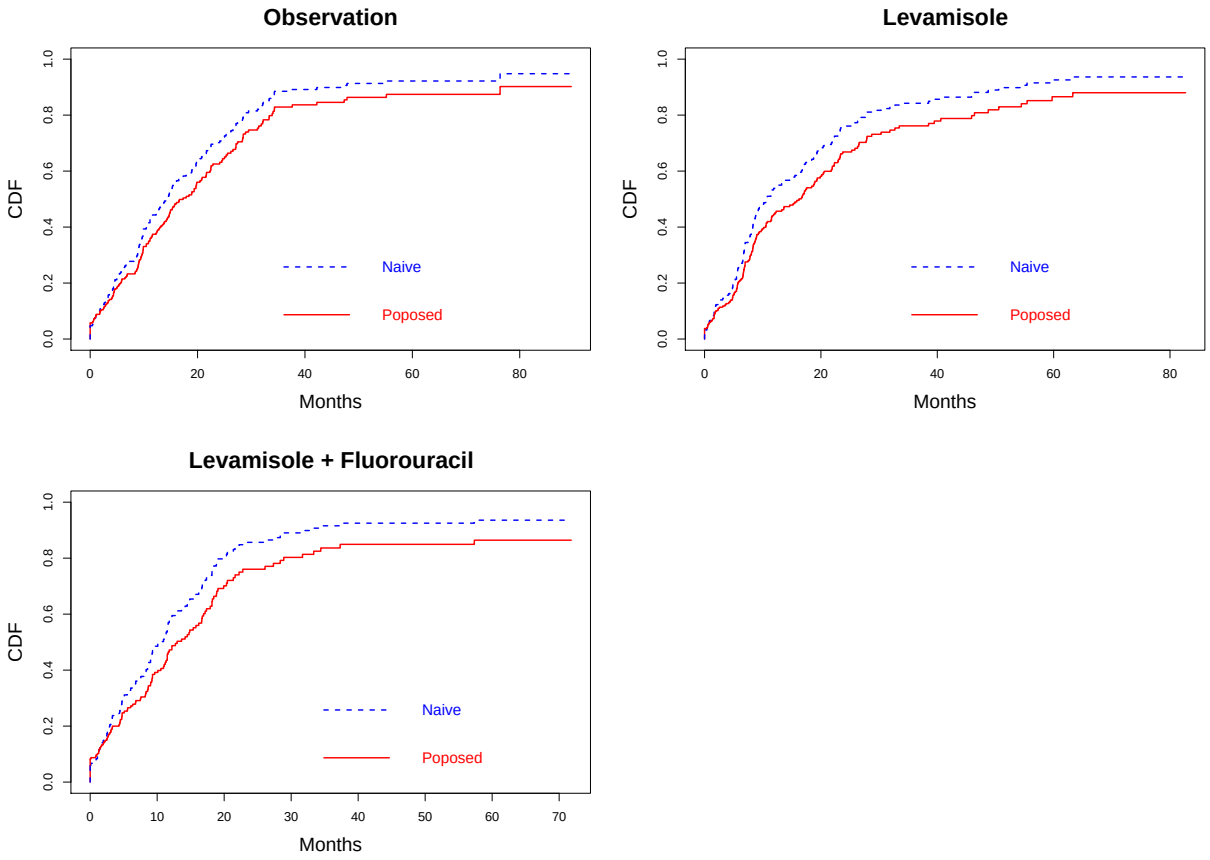

Figure 3.2: Comparison of naïve estimates and proposed estimates of cumulative distribution function of the time between cancer recurrence and death (T_2) for each of the three treatment groups in colon cancer study.

3.5 Discussion

We have proposed a copula-based semiparametric model for successive durations with incomplete follow-up data. This semiparametric model postulates normal copula for the association between the two durations, while leaving the marginals unspecified. Motivated by the colon cancer data example used in our paper, we allow our model to accommodate the situation where the second duration has a probability mass at 0. Under a relatively weak censorship assumption, we propose an inference procedure for the estimation of semiparametric model when the second duration has a probability mass at 0. Along the same line, we can extend the semiparametric model and the inference procedure further to allow multiple probability mass at different time points of the second duration.

The proposed copula-based semiparametric model strikes a balance between model identifiability and robustness. The nonparametric approach such as is more robust, but we do not have the identifiability of the marginal distribution of the second duration. On the other hand, it is generally undesirable to use a method which requires many assumptions. Through parameterizing the association between the two durations but leaving the marginals unspecified, we are able to identify and estimate the marginal distribution of the second duration.

The copula-based semiparametric inference procedure offered is computationally easy. The resulting estimators are shown to be consistent or uniformly consistent, and are shown to be asymptotically normal or weakly converge to a tight Gaussian process. Our simulation study shows that the proposed method performs well in samples of moderate size. As shown in our simulation studies, the inference procedure appears to be quite robust with Frank's family but less robust with Clayton's family. These results suggest the importance of model checking in practice. Graphical model checking procedure

and goodness-of-fit test for the proposed copula-based semiparametric model is under development.

In addition to estimation, of scientific interest is prediction with given history of the bi-state process. For instance, a colon cancer patient might like to know his survival probabilities given occurrence and timing of his cancer relapse. A generalization of the proposed semiparametric model for one-sample problem to a regression analysis set-up is meaningful and desirable.

3.6 Proofs

3.6.1 Consistency and Uniform Consistency

Proof of Theorem 3.1. We first prove the consistency of $\widehat{\theta}$ and the uniform consistency of $\widehat{H}(t)$, by dividing the proof into six steps. In Step A1 to A5, we consider the estimating equations

$$\begin{aligned}\Psi_n(\theta, H) &= n^{-1}\sum_{i=1}^{n}\Delta_i^{ab} Z_i\Big[N_i(\tau_2) - Y_i(0)\Phi\left\{-\theta Z_i + H(0)\right\} \\ &\qquad - \int_{t\in(0,\tau_2]} Y_i(t)\, d\Lambda\left\{-\theta Z_i + H(t)\right\}\Big] = 0 \qquad (3.11)\\ U_n(t; H, \theta) &= n^{-1}\sum_{i=1}^{n}\Delta_i^{ab}\Big[N_i(t) - Y_i(0)\Phi\left\{-\theta Z_i + H(0)\right\} \\ &\qquad - \int_{s\in(0,t]} Y_i(s)\, d\Lambda\left\{-\theta Z_i + H(s)\right\}\Big] = 0\ (t \geq 0) \qquad (3.12)\end{aligned}$$

which have Z_i instead of $\widehat{Z}_i$ as in (3.7) and (3.8). Let $\widehat{\theta}_Z$ and $\widehat{H}_Z(t)$ be the solution of (3.11) and (3.12); the subscript Z is used to differentiate them from $\widehat{\theta}$ and $\widehat{H}(t)$. We show the consistency of $\widehat{\theta}_Z$ and uniform consistency of $\widehat{H}_Z(t)$ through Step A1 to Step A5. In

Step A6, we show that our arguments in Step A1 to Step A5 still can be used when we consider the original estimating equations (3.7) and (3.8), to establish the consistency of $\widehat{\theta}$ and uniform consistency of $\widehat{H}(t)$.

Step A1: In this step, we show the uniform consistency of $\Psi_n(\theta, H)$ and $U_n(t; H, \theta)$, i.e., $\sup_{\theta\in\mathcal{B}(\theta_0), H\in\mathcal{H}}|\Psi_n(\theta, H)-\Psi(\theta, H)| \to 0$ and $\sup_{t\in[0,\tau_2],\, \theta\in\mathcal{B}(\theta_0), H\in\mathcal{H}}|U_n(t; H, \theta)-U(t; H, \theta)| \to 0$ in probability. Here $\Psi(\theta, H) = E[\Delta^{ab} Z\, N(\tau_2)] - E[\Delta^{ab} Y(0)\Phi\{-\theta Z + H(0)\}] - \int_{t\in(0,\tau_2]} E[\Delta^{ab} Z\ Y(t)\, d\Lambda\{-\theta Z + H(t)\}]$, $U(t; H, \theta) = E[\Delta^{ab} N(t)] - E[\Delta^{ab} Y(0)\ \Phi\{-\theta Z + H(0)\}] - \int_{s\in(0,t]}\ E[\Delta^{ab} Y(s)\, d\Lambda\,\{-\theta Z + H(s)\}]$, $\mathcal{B}(\theta_0)$ is a neighborhood of θ_0 and $\mathcal{H}$ is the collection of bounded nondecreasing cadlag functions on $[0, \tau_2]$ with jumps of order $O(n^{-1})$ in size.

Let $\lambda(\cdot)$ be the hazard function of the standard normal distribution and $\dot{\lambda}(\cdot)$ be the derivative of $\lambda(\cdot)$. With standard empirical process arguments, we can show that both the class $\{\Delta_i^{ab} N_i(t), t \in [0, \tau_2]\}$ and $\{\Delta_i^{ab} Y_i(t)\lambda\,\{-\theta Z_i + H(t)\}\,, t \in (0, \tau_2], \theta \in \mathcal{B}(\theta_0), H \in \mathcal{H}\}$ are Glivenko-Cantelli (van der Vaart and Wellner 1996). It is easy to show that $\sup_{\theta\in\mathcal{B}(\theta_0), H(0)\in\mathcal{H}} |U_n(0; H, \theta) - U(0; H, \theta)| \to 0$ in probability. For $t \in (0, \tau_2]$ and $H \in \mathcal{H}$, note that

$$\begin{aligned} n^{-1}\sum_{i=1}^{n} Y_i(t) d\Lambda\{-\theta Z_i + H(t)\} &= n^{-1}\sum_{i=1}^{n} Y_i(t)\lambda\{-\theta Z_i + H^*(t)\} dH(t) \qquad (3.13) \\ &= n^{-1}\sum_{i=1}^{n} Y_i(t)\lambda\{-\theta Z_i + H(t)\} dH(t) + r_n(t), \end{aligned}$$

where $H(t-; \theta) < H^*(t; \theta) < H(t; \theta)$, $dH(t; \theta) = H(t; \theta) - H(t-; \theta)$, $r_n(t) = n^{-1}\sum_{i=1}^{n} Y_i(t)$ $\dot{\lambda}\{-\theta Z_i + H^{**}(t)\}\{H^*(t) - H(t)\} dH(t)$ and $H^*(t) < H^{**}(t) < H(t)$. Since $dH(t; \theta)$ is of the order $O(n^{-1})$ almost surely, the fact that $\dot{\lambda}\{-\theta Z_i + H^{**}(t; \theta)\}$ has a finite upper bound uniformly in $t \in (0, \tau_2]$, $\theta \in \mathcal{B}(\theta_0)$ implies that $r_n(t) = O(n^{-2})$ almost surely.

Thus,

$$\sup_{t\in(0,\tau_2],\theta\in\mathcal{B}(\theta_0),H\in\mathcal{H}}|U_n(t;H,\theta)-U(t;H,\theta)| \tag{3.14}$$

$$= \sup_{t\in(0,\tau_2],\theta\in\mathcal{B}(\theta_0),H\in\mathcal{H}}\Bigg|n^{-1}\sum_{i=1}^{n}\Delta_i^{ab}\Bigg[N_i(t)-Y_i(0)\Phi\left\{-\theta Z_i+H(0)\right\}$$

$$-\int_{s\in(0,t]}\Delta_i^{ab}Y_i(s)\,\lambda\left\{-\theta Z_i+H(s)\right\}dH(s)\Bigg]-E\Bigg(\Delta^{ab}\big[N(t)-Y(0)\Phi\{-\theta Z$$

$$+H(0)\}\big]\Bigg)+\int_{s\in(0,t]}E\big[\Delta^{ab}Y(s)\,\lambda\left\{-\theta Z+H(s)\right\}\big]dH(s)\Bigg|+O_p(n^{-2})$$

$$\leq \sup_{t\in(0,\tau_2]}\left|n^{-1}\sum_{i=1}^{n}\Delta_i^{ab}N_i(t)-E\left[\Delta^{ab}N(t)\right]\right|+\sup_{\theta\in\mathcal{B}(\theta_0),H\in\mathcal{H}}\left|n^{-1}\sum_{i=1}^{n}\Delta_i^{ab}Y_i(0)\Phi\{-\theta Z_i\right.$$

$$\left.+H(0)\}-E\left[\Delta^{ab}Y(0)\Phi\left\{-\theta Z+H(0)\right\}\right]\right|+\int_{s\in(0,t]}\sup\nolimits_{s\in(0,\tau_2],\theta\in\mathcal{B}(\theta_0),H\in\mathcal{H}}\left|n^{-1}\sum_{i=1}^{n}\right.$$

$$\left.\Delta_i^{ab}Y_i(s)\lambda\left\{-\theta Z_i+H(s)\right\}-E\left[\Delta^{ab}Y(s)\lambda\left\{-\theta Z+H(s)\right\}\right]\right|dH(s)+O_p(n^{-2})\to 0.$$

Therefore, $\sup_{t\in[0,\tau_2],\theta\in\mathcal{B}(\theta_0),H\in\mathcal{H}}|U_n(t;H,\theta)-U(t;H,\theta)|\to 0$ in probability. Similarly, we can show that $\sup_{\theta\in\mathcal{B}(\theta_0),H\in\mathcal{H}}|\Psi_n(\theta,H)-\Psi(\theta,H)|\to 0$ in probability.

Step A2: For a given $\theta\in\mathcal{B}(\theta_0)$, we define $\widehat{H}_Z(t;\theta)$ and $\widetilde{H}_Z(t;\theta)$ as the solutions of the equations $U_n(t;H,\theta)=0$ and $U(t;H,\theta)=0$, respectively. In this step, we show that $\sup_{t\in[0,\tau_2],\,\theta\in\mathcal{B}(\theta_0)}|\widehat{H}_Z(t;\theta)-\widetilde{H}_Z(t;\theta)|\to 0$ in probability.

Given any fixed $\theta\in\mathcal{B}(\theta_0)$, the size of $N(t)$ for all $t>0$ is less than or equal to 1 almost surely since T_2 is continuous beyond $t=0$. Thus, for $t\in(0,\tau_2]$, $\widehat{H}_Z(t;\theta)$ satisfies the property

$$\left|n^{-1}\sum_{i=1}^{n}Y_i(t)d\Lambda\{-\theta Z_i+\widehat{H}_Z(t;\theta)\}\right|\leq n^{-1},\quad \text{almost surely.}$$

Following the equality (3.13), we can see that $d\widehat{H}_Z(t;\theta) = \widehat{H}_Z(t;\theta) - \widehat{H}_Z(t-;\theta)$ is uniformly bounded in $t \in (0, \tau_2]$ and is of the order $O(n^{-1})$ almost surely. Thus, $\widehat{H}_Z(t;\theta)$ belongs to $\mathcal{H}$.

In Step A1, we have established that $\sup_{t\in[0,\tau_2],\theta\in\mathcal{B}(\theta_0),H\in\mathcal{H}}|U_n(t;H,\theta) - U(t;H,\theta)| \to 0$ in probability. Given the fact that $U_n(0;\widehat{H}_Z,\theta) = 0$ and $U(0;\widehat{H}_Z,\theta) = 0$ for $\theta \in \mathcal{B}(\theta_0)$, it is straightforward to establish $\sup_{\theta\in\mathcal{B}(\theta_0)}|\widehat{H}_Z(0;\theta) - \widetilde{H}_Z(0;\theta)| \to 0$ in probability. Since $\widehat{H}_Z(t;\theta) \in \mathcal{H}$, $U(t;\widehat{H}_Z,\theta) = o_p(1)$, uniformly in $t \in (0,\tau_2]$ and $\theta \in \mathcal{B}(\theta_0)$. Then $U(t;\widehat{H}_Z,\theta) = 0$ suggests that

$$
\begin{aligned}
& U(t;\widehat{H}_Z,\theta) - U(t;\widehat{H}_Z,\theta) \\
&= \int_{s\in(0,t]} E\left(\Delta^{ab}Y(s)\, d\left[\Lambda\left\{-\theta Z + \widehat{H}_Z(s;\theta)\right\} - \Lambda\left\{-\theta Z + \widetilde{H}_Z(s,\theta)\right\}\right]\right) \\
&\quad + E\left(\Delta^{ab}Y(0)\left[\Phi\{-\theta Z + \widehat{H}(0;\theta)\} - \Phi\{-\theta Z + \widetilde{H}(0;\theta)\}\right]\right) \\
&= \int_{s\in(0,t]} E\left(\Delta^{ab}Y(s)\, d\left[\Lambda\left\{-\theta Z + \widehat{H}_Z(s;\theta)\right\} - \Lambda\left\{-\theta Z + \widetilde{H}_Z(s,\theta)\right\}\right]\right) + o_p(1) \\
&= o_p(1), \qquad (3.15)
\end{aligned}
$$

where the expectations $E[\Delta^{ab}Y(0)\Phi\{-\theta Z+\widehat{H}(0;\theta)\}$ and $E[\Delta^{ab}Y(t)d\Lambda\{-\theta Z+\widehat{H}_Z(t;\theta)\}]$ in (3.15) are defined as $E[\Delta^{ab}Y(0)\Phi\{-\theta Z + H(0)\}$ and $E[\Delta^{ab}Y(t)d\Lambda\{-\theta Z + H(t)\}]$ evaluated at $H(t) = \widehat{H}_Z(t;\theta)$, $t \geq 0$. Application of the mean value theorem to (3.15) leads to

$$
\begin{aligned}
& \int_{s\in(0,t]} \left\{\widehat{H}_Z(s-;\theta) - \widetilde{H}_Z(s;\theta)\right\} E\left[\Delta^{ab}Y(s)\, d\lambda\left\{-\theta Z + H_Z(s;\theta)\right\}\right] \qquad (3.16) \\
&\quad + \int_{s\in(0,t]} E\left[\Delta^{ab}Y(s)\lambda\left\{-\theta Z + H_Z(s;\theta)\right\}\right] d\left\{\widehat{H}_Z(s;\theta) - \widetilde{H}_Z(s;\theta)\right\} = o_p(1),
\end{aligned}
$$

where $H_Z(t;\theta)$ lies between $\widehat{H}_Z(t;\theta)$ and $\widetilde{H}_Z(t;\theta)$.

Let $\mathcal{F} = \{f : (0, \tau_2] \mapsto \mathcal{R}, f \text{ is cadlag on } (0, \tau_2)\}$, and consider a map from $\mathcal{F}$ to $\mathcal{F}$ for $f \in \mathcal{F}$, i.e., $\phi\{f(t)\} = \int_{s\in(0,t]} f(s-)A_1(\theta, ds) + \int_{s\in(0,t]} A_2(\theta, s)\, d\, f(s)$, where $A_1(\theta, t) = \int_{s\in(0,t]} E[\Delta^{ab} Y(s) d\lambda\{-\theta Z + H_Z(s;\theta)\}]$ and $A_2(\theta, t) = E[\Delta^{ab} Y(s)\lambda\{-\theta Z + H_Z(s;\theta)\}]$. Then it is implied in (3.16) that

$$\phi\left\{\widehat{H}_Z - \widetilde{H}_Z\right\}(t;\theta) = o_p(1). \tag{3.17}$$

With the product integration theory (Andersen et al., 1993, Section II.6) and similar arguments as in Peng and Huang (2007), there exists an inverse of ϕ, say, ϕ^{-1}, which has a close form. For $g(t) \in \mathcal{F}$,

$$\phi^{-1}\{g(t)\} = \int_{s\in(0,t]} \mathcal{I}_{(s,t]} A_2^{-1}(\theta, s)\, d\, g(s) + o_p(1), \tag{3.18}$$

where $\mathcal{I}_{(s,t]} = \prod_{u\in(s,t]} \left\{1 - A_2^{-1}(\theta, u)A_1(\theta, du)\right\}$, and $\prod$ is the product integral notation. From (3.17), $g(t) = 0$ for $t \in (0, \tau_2]$. therefore, (3.18) implies $\sup_{t\in(0,\tau_2], \theta\in\mathcal{B}(\theta_0)} |\widehat{H}_Z(t;\theta) - \widetilde{H}_Z(t;\theta)| \to 0$ in probability, which combined with $\sup_{\theta\in\mathcal{B}(\theta_0)} |\widehat{H}_Z(0;\theta) - \widetilde{H}_Z(0;\theta)| \to 0$ in probability completes the proof in this step.

Step A3: Here we show that $(\partial/\partial\theta)\widehat{H}_Z(t;\theta)$ and $(\partial/\partial\theta)\Psi_n\{\theta, \widehat{H}_Z(\cdot;\theta)\}$ at $\theta = \theta_0$ converge to $(\partial/\partial\theta)\widetilde{H}_Z(t;\theta)$ and $(\partial/\partial\theta)\Psi\{\theta, \widetilde{H}_Z(\cdot;\theta)\}$ at $\theta = \theta_0$ respectively, both of which are bounded away from 0 and are finite for $t \in [0, \tau_2]$. The deduction here is inspired by Step A3 of Chen et al. (2002).

Let $\dot{\Phi}(\cdot)$ be the derivative of $\Phi(\cdot)$, the cumulative distribution function of the standard normal distribution. When $t = 0$, $(\partial/\partial\theta)\widehat{H}_Z(0;\theta) = E[\Delta^{ab} Y(0) Z\dot{\Phi}\{-\theta_0 Z + H_0(0)\}]/$ $E[\Delta^{ab} Y(0)\dot{\Phi}\{-\theta_0 Z + H_0(0)\}]$, which is bounded away from 0 and finite under the probit model.

For $t \in (0, \tau_2]$ and $\theta \in \mathcal{B}(\theta_0)$, differentiating $U(t; \widehat{H}_Z, \theta) = 0$ with respect to θ, we

have that

$$\int_{s\in(0,t]} E\left(\Delta^{ab}Y(s)\,d\left[\lambda\left\{-\theta Z+\widetilde{H}_Z(s;\theta)\right\}\left\{-Z+\frac{\partial\widetilde{H}_Z(s;\theta)}{\partial\theta}\right\}\right]\right)=0. \tag{3.19}$$

Define

$$B_1(t)=E\big[\Delta^{ab}Y(t)\lambda\{-\theta_0 Z+H_0(t)\}\big], \qquad B_2(t)=E\big[\Delta^{ab}Y(t)\dot{\lambda}\{-\theta_0 Z+H_0(t)\}\big],$$
$$B_3(t)=E\big[\Delta^{ab}ZY(t)\dot{\lambda}\{-\theta_0 Z+H_0(t)\}\big], \qquad B_4(t)=E\big[\Delta^{ab}ZY(t)\lambda\{-\theta_0 Z+H_0(t)\}\big]$$
$$W(t)=\int_{s\in(0,t]} B_1(s)^{-1}B_3(s)dH_0(s) \qquad B(s,t)=\exp\left\{\int_{u\in[s,t]}\frac{B_2(u)}{B_1(u)}\,dH_0(u)\right\}.$$

By some algebra,

$$\left\{\left.\frac{\partial\widetilde{H}_Z(t;\theta)}{\partial\theta}\right|_{\theta=\theta_0}\right\}+\int_{s\in(0,t]} B_1(s)^{-1}B_2(s)\left\{\left.\frac{\partial\widetilde{H}_Z(s;\theta)}{\partial\theta}\right|_{\theta=\theta_0}\right\}dH_0(s)$$
$$=\left\{\left.\frac{\partial\widetilde{H}_Z(0;\theta)}{\partial\theta}\right|_{\theta=\theta_0}\right\}-W(t), \tag{3.20}$$

which is a Volterra integral equation, and the unique solution of which (Andersen et al. 1993) is

$$\left.\frac{\partial\widetilde{H}_Z(t;\theta)}{\partial\theta}\right|_{\theta=\theta_0}=-\int_{s\in(0,t]} B(s,t)^{-1}dW(s)+B(0,t)^{-1}\left.\frac{\partial\widetilde{H}_Z(0;\theta)}{\partial\theta}\right|_{\theta=\theta_0}=D_1(t). \tag{3.21}$$

Under regularity condition 2, $(\partial/\partial\theta)\widetilde{H}_Z(t;\theta)$ at $\theta=\theta_0$ is bounded away from 0 and is finite for $t\in[0,\tau_2]$. To derive the expression of $(\partial/\partial\theta)\widehat{H}_Z(t;\theta)$ at $\theta=\theta_0$, we replace the expectation E and $\widetilde{H}_Z(t;\theta)$ in (3.19) by the summation notation $n^{-1}\sum_{i=1}^n$ and $\widehat{H}_Z(t;\theta)$ respectively. By the uniform consistency of $\widehat{H}_Z(t,\theta_0)$ in Step A2 and the Glivenko-Cantelli theorem, we have a stochastic Volterra integral equation for $\{\partial\widehat{H}_Z(t,\theta)/\partial\theta\}|_{\theta=\theta_0}$,

which takes a similar form as equation (3.20). With the production integration theory, its unique solution (Andersen et al. 1993) takes the form

$$\left.\frac{\partial \widehat{H}_Z(t;\theta)}{\partial\theta}\right|_{\theta=\theta_0} = \left.\frac{\partial \widetilde{H}_Z(t;\theta)}{\partial\theta}\right|_{\theta=\theta_0} + o_p(1) = D_1(t) + o_p(1).$$

By Glivenko-Cantelli theorem and similar as Step A3 of Chen et al. (2002), we can write

$$\begin{aligned}
&\left.\frac{\partial \Psi_n\{\theta, \widehat{H}_Z(\cdot;\theta)\}}{\partial\theta}\right|_{\theta=\theta_0} \\
&= -n^{-1}\sum_{i=1}^{n}\Delta_i^{ab}Z_i\left[\lambda\left\{-\theta_0 Z_i + \widehat{H}_Z(X_{2i};\theta_0)\right\}\left\{-Z_i + \left.\frac{\partial \widehat{H}_Z(X_{2i};\theta)}{\partial\theta}\right|_{\theta=\theta_0}\right\} + Y_i(0)\right. \\
&\quad \left.\cdot\left(\dot{\Phi}\left\{-\theta_0 Z_i + \widehat{H}_Z(0;\theta_0)\right\} - \lambda\left\{-\theta_0 Z_i + \widehat{H}_Z(0;\theta_0)\right\}\right)\left\{-Z_i + \left.\frac{\partial \widehat{H}_Z(0;\theta)}{\partial\theta}\right|_{\theta=\theta_0}\right\}\right] \\
&= \int_{t\in(0,\tau_2]} E\left[\left\{\Delta^{ab}Z + B_4(X_2)B(t,X_2)^{-1}B_1^{-1}(t)\right\}Z\dot{\lambda}\{-\theta_0 Z + H_0(t)\}Y(t)\right]dH_0(t) \\
&\quad + \left(E\left[\frac{\Delta^{ab}Y(0)Z\dot{\Phi}\{-\theta_0 Z + H_0(0)\}\Phi\{-\theta_0 Z + H_0(0)\}}{1-\Phi\{-\theta_0 Z + H_0(0)\}}\right] - B_4(X_2)B(0,X_2)^{-1}\right) \\
&\quad \cdot \frac{E\left[\Delta^{ab}Y(0)Z\dot{\Phi}\{-\theta_0 Z + H_0(0)\}\right]}{E\left[\Delta^{ab}Y(0)\dot{\Phi}\{-\theta_0 Z + H_0(0)\}\right]} - E\left[\frac{\Delta^{ab}Y(0)Z^2\dot{\Phi}\{-\theta_0 Z + H_0(0)\}}{1-\Phi\{-\theta_0 Z + H_0(0)\}}\right. \\
&\quad \left.\cdot\,\Phi\{-\theta_0 Z + H_0(0)\}\right] + o_p(1) \\
&= \left.\frac{\partial \Psi\{\theta, \widetilde{H}_Z(\cdot;\theta)\}}{\partial\theta}\right|_{\theta=\theta_0} + o_p(1) \\
&= D_2 + o_p(1), \qquad (3.22)
\end{aligned}$$

which is bounded away from 0 and is finite by regularity condition 2.

Step A4: Here we show the consistency of $\widehat{\theta}_Z$, which is the solution to $\Psi_n\{\theta, \widehat{H}_Z(\cdot;\theta)\} = 0$. In Step A1, we have shown that $\sup_{\theta\in\mathcal{B}(\theta_0), H\in\mathcal{H}}|\Psi_n(\theta, H)-\Psi(\theta, H)| \to 0$ in probability. Since $\widehat{H}_Z(\cdot;\theta) \in \mathcal{H}$, we have

$$\sup_{\theta\in\mathcal{B}(\theta_0)} \left|\Psi_n\{\theta, \widehat{H}_Z(\cdot;\theta)\} - \Psi\{\theta, \widehat{H}_Z(\cdot;\theta)\}\right| \to 0 \quad \text{in probability.}$$

In Step A2, we have shown that $\sup_{t\in[0,\tau_2],\, \theta\in\mathcal{B}(\theta_0)}|\widehat{H}_Z(t;\theta) - \widetilde{H}_Z(t;\theta)| \to 0$ in probability. Note that

$$\Psi\{\theta, \widehat{H}_Z(\cdot;\theta)\} - \Psi\{\theta, \widetilde{H}_Z(\cdot;\theta)\} = \int_{t\in(0,\tau_2]} E\left(\Delta^{ab} ZY(t) d\left[\Lambda\{-\theta Z + \widehat{H}_Z(t;\theta)\} - \Lambda\{-\theta Z + \widetilde{H}_Z(t;\theta)\}\right]\right) + E\left(\Delta^{ab} ZY(0)\left[\Phi\{-\theta Z + \widehat{H}_Z(0;\theta)\} - \Phi\{-\theta Z + \widetilde{H}_Z(0;\theta)\}\right]\right).$$

By applying the mean value theorem, we can easily show that

$$\sup_{\theta\in\mathcal{B}(\theta_0)} \left|\Psi\{\theta, \widehat{H}_Z(\cdot;\theta)\} - \Psi\{\theta, \widetilde{H}_Z(\cdot;\theta)\}\right| \to 0 \quad \text{in probability.}$$

Also, we have shown in Step A3 that $(\partial/\partial\theta)\Psi\{\theta, \widetilde{H}_Z(\cdot;\theta)\}$ at $\theta = \theta_0$ is bounded away from 0. Thus, $\Psi\{\theta, \widetilde{H}_Z(\cdot;\theta)\}$ has a unique zero crossing in $\mathcal{B}(\theta_0)$, and $\widehat{\theta}_Z \to \theta_0$ in probability.

Step A5: Here we show the uniform consistency of $\widehat{H}_Z(t;\widehat{\theta}_Z)$ for $t \in [0, \tau_2]$.

When sample size n is large enough, $\widehat{\theta}_Z$ must be in the neighborhood $\mathcal{B}(\theta_0)$, therefore,

$$\left|\widehat{H}_Z(t, \widehat{\theta}_Z) - \widetilde{H}(t, \widehat{\theta}_Z)\right| \le \sup_{t\in[0,\tau_2],\, \theta\in\mathcal{B}(\theta_0)} \left|\widehat{H}_Z(t, \theta) - \widetilde{H}_Z(t, \theta)\right|, \tag{3.23}$$

where the right hand side converges to 0 in probability based on Step A2. Since the partial derivative of $\widetilde{H}_Z(t;\theta)$ with respect to θ at $\theta = \theta_0$ is bounded away from 0 and

finite under regularity condition 2, by Taylor expansion as well as the consistency of $\widehat{\theta}_Z$ as established in Step A4, we can show that $\sup_{t\in[0,\tau_2]}|\widetilde{H}_Z(t;\widehat{\theta}_Z) - \widetilde{H}_Z(t;\theta_0)| \to 0$ in probability. Then the uniform consistency of $\widehat{H}_Z(t) = \widehat{H}_Z(t,\widehat{\theta}_Z)$ follows from the inequality (3.23) and the triangular inequality

$$\begin{aligned}\left|\widehat{H}_Z(t,\widehat{\theta}_Z) - H_0(t)\right| &= \left|\widehat{H}_Z(t,\widehat{\theta}_Z) - \widetilde{H}_Z(t,\theta_0)\right| \\ &\leq \left|\widehat{H}_Z(t,\widehat{\theta}_Z) - \widetilde{H}_Z(t,\widehat{\theta}_Z)\right| + \left|\widetilde{H}_Z(t,\widehat{\theta}_Z) - \widetilde{H}_Z(t,\theta_0)\right|.\end{aligned}$$

Step A6: In Step A1 to Step A5, we have established the consistency of $\widehat{\theta}_Z$ and the uniform consistency $\widehat{H}_Z(t;\widehat{\theta}_Z)$ for $t \in [0,\tau_2]$ with Z_i, which is the true value of the estimated covariate $\widehat{Z}_i$. In this step, we show that our arguments in Step A1 to Step A5 still hold when we have the estimated covariate $\widehat{Z}_i$ in the estimating equations.

First, we can show that $\sup_{x\in[a,b]}|\Phi^{-1}\{\widehat{F}_1(x)\} - \Phi^{-1}\{F_1(x)\}| \to 0$ in probability, which follows the well established uniform consistency result of the Kaplan-Meier estimator $\widehat{F}_1(\cdot)$.

When Z_i is replaced by $\widehat{Z}_i$, instead of having the equations $\Psi_n(\theta,H) = 0$ and $U_n(t;H,\theta) = 0$ as in (3.11) and (3.12) for $\theta \in \mathcal{B}(\theta_0)$, we will have $\Psi_n(\theta,H) = o_p(1)$ and $U_n(t;H,\theta) = o_p(1)$, which is based on $\sup_{x\in[a,b]}|\Phi^{-1}\{\widehat{F}_1(x)\} - \Phi^{-1}\{F_1(x)\}| \to 0$ in probability. Then the arguments in Step A1 to A5 still hold, which lead to the consistency of $\widehat{\theta}$ and the uniform consistency $\widehat{H}(t;\widehat{\theta})$ for $t \in [0,\tau_2]$.

With the results in Step A6, it is straightforward to show that $\widehat{\rho} \to \rho_0$ and $\sup_{t\in[0,\tau_2]}|\widehat{F}_2(t) - F_2(t)| \to 0$, in probability, since the mappings $\theta \mapsto \rho$ and $\{\theta, H(\cdot)\} \mapsto F_2(\cdot)$ are both continuous and (uniformly) bounded based on the equalities in (3.10). This completes the proof of Theorem 3.1. □

3.6.2 Asymptotic Normality and Weak Convergence

Proof of Theorem 3.2. Let $\Psi_n^*(\theta, H)$ and $U_n^*(t; H, \theta)$ be n^{-1} times the left hand side of (3.7) and (3.8) respectively. Note that the difference between $(\partial/\partial\theta)\Psi_n^*\{\theta, \widehat{H}(\cdot;\theta)\}$ at $\theta = \theta_0$ and $(\partial/\partial\theta)\Psi_n\{\theta, \widehat{H}_Z(\cdot;\theta)\}$ at $\theta = \theta_0$ is $o_p(1)$. Thus, the expression of the former partial derivative follows the latter one, which has been derived in Step A3. We divide the proof of asymptotic normality of $\widehat{\theta}$ and weak convergence of $\widehat{H}(t)$ into four steps. The asymptotic normality of $\widehat{\rho}$ and weak convergence of $\widehat{F}_2(t)$ then follow straightforward.

Step B1: We show that $U_n^*(t; H_0, \theta_0)$ can be expressed asymptotically as the summation of n independent and identically distributed influence functions and is Donsker.

$$
\begin{aligned}
& U_n^*(t; H_0, \theta_0) \\
= \;& n^{-1}\sum_{i=1}^{n}\int_{s\in(0,t]}\Delta_i^{ab}dM_{2i}(s) + n^{-1}\sum_{i=1}^{n}\Delta_i^{ab}\left[N_i(0) - Y_i(0)\Phi\left\{-\theta_0 Z_i + H_0(0)\right\}\right] \\
& + \; n^{-1}\sum_{i=1}^{n}\Delta_i^{ab}Y_i(0)\left[\Phi\left\{-\theta_0 Z_i + H_0(0)\right\} - \Phi\left\{-\theta_0\widehat{Z}_i + H_0(0)\right\}\right] \\
& + \; n^{-1}\sum_{i=1}^{n}\int_{s\in(0,t]}\Delta_i^{ab}Y_i(s)d\left[\Lambda\left\{-\theta_0 Z_i + H_0(s)\right\} - \Lambda\left\{-\theta_0\widehat{Z}_i + H_0(s)\right\}\right].
\end{aligned}
$$

where $dM_2(t) = dN(t) - Y(t)\lambda\{-\theta_0 Z + H_0(t)\}dH_0(t)$. Since $\{\Delta_i^{ab}N_i(t), t \in [0, \tau_2]\}$ and $\{\Delta_i^{ab}\lambda\{-\theta_0 Z_i + H_0(t)\}, t \in [0, \tau_2]\}$ are Donsker classes, it follows that $\{\int_{s\in(0,t]}\Delta_i^{ab}dM_{2i}(s), t \in [0, \tau_2]\}$ is a Donsker class. It is obvious that $\{\Delta_i^{ab}[N_i(0) - Y_i(0)\Phi\{-\theta_0 Z_i + H_0(0)\}]\}$ is Donsker.

Define $M_1(t) = I(X_1 \le t, \Delta_1 = 1) - \int_{u\in(0,t]} I(X_1 \ge u)d\Lambda_{T_1}(u)$, where $\Lambda_{T_1}(t)$ is the cumulative hazard function of T_1. It follows from the martingale integral representation

for $\widehat{F}_1 - F_1$ (Gill 1980, page 36-37) that

$$\begin{aligned}
&\Phi^{-1}\{\widehat{F}_1(x)\} - \Phi^{-1}\{F_1(x)\} \qquad (3.24)\\
&= \dot{\Phi}^{-1}\{F_1(x)\}\{\widehat{F}_1(x) - F_1(x)) + o_p(n^{-1/2})\\
&= \dot{\Phi}^{-1}\{F_1(x)\}\{1 - F_1(x)\} \sum_{i=1}^{n} \int_{s\in(0,x)} \frac{\{1 - \widehat{F}_1(s-)\}}{\{1 - F_1(s)\}\sum_{j=1}^{n} I(X_{1j} \geq s)} dM_{1i}(s)\\
&\quad + o_p(n^{-1/2})\\
&= \dot{\Phi}^{-1}\{F_1(x)\}\{1 - F_1(x)\}\, n^{-1} \sum_{i=1}^{n} \int_{s\in(0,x)} \frac{1}{\pi(s)} dM_{1i}(s) + o_p(n^{-1/2}),
\end{aligned}$$

for any $x \in [a, b]$, where $\dot{\Phi}^{-1}(\cdot)$ is the derivative of $\Phi^{-1}(\cdot)$ and $\pi(s) = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} I(X_{1i} \geq s)$.

It then follows that

$$\begin{aligned}
&n^{-1} \sum_{i=1}^{n} \Delta_i^{ab} Y_i(0) \left[\Phi\left\{-\theta_0 Z_i + H_0(0)\right\} - \Phi\left\{-\theta_0 \widehat{Z}_i + H_0(0)\right\}\right]\\
&= \quad n^{-1} \sum_{i=1}^{n} \Delta_i^{ab} Y_i(0)\theta_0(\widehat{Z}_i - Z_i)\dot{\Phi}\left\{-\theta_0 Z_i^* + H_0(s)\right\}\\
&= \quad n^{-1} \sum_{k=1}^{n} \int_{s\in(0,\infty)} \frac{q(s)}{\pi(s)}\, dM_{1k}(s) + o_p(n^{-1/2}), \qquad (3.25)
\end{aligned}$$

where Z_i^* lies between Z_i and $\widehat{Z}_i$, and

$$q(s) = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \Delta_i^{ab} \theta_0 Y_i(0) \dot{\Phi}\{-\theta_0 Z_i + H_0(0)\}\, \dot{\Phi}^{-1}\{F_1(X_{1i})\}\{1 - F_1(X_{1i})\} I(X_{1i} \geq s) + o_p(1).$$

Similarly,

$$n^{-1}\sum_{i=1}^{n}\int_{s\in(0,t]}\Delta_i^{ab}Y_i(s)d\left[\Lambda\left\{-\theta_0 Z_i+H_0(s)\right\}-\Lambda\left\{-\theta_0\widehat{Z}_i+H_0(s)\right\}\right]$$
$$=n^{-1}\sum_{k=1}^{n}\int_{s\in(0,\infty)}\frac{r(s,t)}{\pi(s)}\,dM_{1k}(s)+o_p(n^{-1/2}), \tag{3.26}$$

where

$$\begin{aligned} r(s,t) &= \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\Delta_i^{ab}\theta_0\dot{\Phi}^{-1}\{F_1(X_{1i})\}\{1-F_1(X_{1i})\}I(X_{1i}\geq s) \\ &\quad\cdot\int_{u\in(0,t]}Y_i(u)d\Lambda\{-\theta_0 Z_i+H_0(u)\}]+o_p(1), \end{aligned}$$

Since $\{I(X_{1i}\leq s,\Delta_{1i}=1), s\in[0,\tau_1]\}$ and $\{I(X_{1i}\geq s), s\in[0,\tau_1]\}$ are Donsker classes, it follows that $\int_{s\in(0,\infty)}\{q(s)/\pi(s)\}\,dM_{1i}(s)$ and $\int_{s\in(0,\infty)}\{r(s,t)/\pi(s)\}\,dM_{1i}(s)$ are Donsker. Thus, $U_n^*(t;H_0,\theta_0)$ is Donsker.

Step B2: We show the weak convergence of $n^{1/2}\{\widehat{H}(t,\theta_0)-H_0(t)\}$. First, we can show that $n^{1/2}\{\widehat{H}(0,\theta_0)-H_0(0)\}=(E[Y(0)\dot{\Phi}\{-\theta_0 Z+H_0(0)\}])^{-1}\,n^{1/2}U_n^*(0;H_0,\theta_0)+o_p(n^{-1/2})$, which converges to mean 0 normal distribution based on Step B1. With the uniform

convergence of $\widehat{H}(t,\theta_0)$ and the fact that $U_n^*(t;\widehat{H},\theta_0\} = 0$, for $t \in (0,\tau_2]$, we can write

$$
\begin{aligned}
& U_n^*(t;\widehat{H},\theta_0) - U_n^*(t;H_0,\theta_0) - \left\{U_n^*(0;\widehat{H},\theta_0) - U_n^*(0;H_0,\theta_0)\right\} \\
&= n^{-1}\sum_{i=1}^{n}\int_{s\in(0,t]} \Delta_i^{ab} Y_i(s)\, d\left[\Lambda\left\{-\theta_0\widehat{Z}_i + \widehat{H}(s;\theta_0)\right\} - \Lambda\left\{-\theta_0\widehat{Z}_i + H_0(s)\right\}\right] \\
&= n^{-1}\sum_{i=1}^{n}\int_{s\in(0,t]} \Delta_i^{ab} Y_i(s)\, d\left[\lambda\left\{-\theta_0 Z_i + H_0(s)\right\}\left\{\widehat{H}(s;\theta_0) - H_0(s)\right\}\right] + o_p(n^{-1/2}) \\
&= n^{-1}\sum_{i=1}^{n}\int_{s\in(0,t]} \Delta_i^{ab} Y_i(s)\lambda\left\{-\theta_0 Z_i + H_0(s)\right\}\, d\left\{\widehat{H}(s;\theta_0) - H_0(s)\right\} \\
&\quad + n^{-1}\sum_{i=1}^{n}\int_{s\in(0,t]} \Delta_i^{ab} Y_i(s)\dot{\lambda}\left\{-\theta_0 Z_i + H_0(s)\right\}\left\{\widehat{H}(s-;\theta_0) - H_0(s)\right\} dH_0(s) \\
&\quad + o_p(n^{-1/2}).
\end{aligned}
$$

Applying the Glivenko-Cantelli theorem to each item on the right-hand side of the above equation, we obtain that

$$
\begin{aligned}
& -n^{1/2}\left\{U_n^*(t;H_0,\theta_0) - U_n^*(0;H_0,\theta_0)\right\} \qquad (3.27) \\
&= n^{1/2}\int_{s\in(0,t]} \left[B_1(s) d\left\{\widehat{H}(s;\theta_0) - H_0(s)\right\} + B_2(s)\left\{\widehat{H}(s-;\theta_0) - H_0(s)\right\} dH_0(s)\right] \\
&\quad + o_p(1).
\end{aligned}
$$

Let $\mathcal{F} = \{f : [0,\tau_2] \mapsto \mathcal{R}, f \text{ is cadlag on } (0,\tau_2)\}$, and consider a map from $\mathcal{F}$ to $\mathcal{F}$ for $f \in \mathcal{F}$, i.e., $\varphi\{f(t)\} = \int_{s\in(0,t]} B_1(s)df(s) + \int_{s\in(0,t]} B_2(s)f(s-)dH_0(s)$. Then it is implied in (3.27) that

$$
\varphi\left[n^{1/2}\left\{\widehat{H}(t;\theta_0) - H_0(t)\right\}\right] = -n^{1/2}\left\{U_n^*(t;H_0,\theta_0) - U_n^*(0;H_0,\theta_0)\right\}. \qquad (3.28)
$$

By the product integration theory (Andersen et al., 1993, Section II.6) and similar ar-

guments as in Peng and Huang (2007), there exists an inverse of φ, say, φ^{-1}, which has a close form. For $g(t) \in \mathcal{F}$,

$$\varphi^{-1}\{g(t)\} = \int_{s\in(0,t]} \mathcal{J}_{(s,t]} B_1^{-1}(s)\, d\, g(s) + g(0)\, \mathcal{J}_{[0,t]}, \tag{3.29}$$

where $\mathcal{J}_{(s,t]} = \prod_{u\in(s,t]} \left\{1 - B_1^{-1}(u)B_2(u)dH_0(u)\right\} = B(s,t)^{-1}$, and $\prod$ is the product integral notation. In Step B1, we have shown that $U_n^*(t; H_0, \theta_0)$ is Donsker, and $-n^{1/2}\{U_n^*(t; H_0, \theta_0) - U_n^*(0; H_0, \theta_0)\}$ converges weakly to a tight Gaussian process $G(t)$ with covariance $\Sigma(s,t) = E[\iota(s)\iota(t)]$, where $\iota(t) = \int_{s\in(0,t]} \Delta^{ab}\, dM_2(s) + \int_{s\in(0,\infty)} \{r(s,t)/\pi(s)\}$ $dM_1(s)$. Combined with (3.28), the continuous mapping theorem then suggests that $n^{1/2}\{\widehat{H}(t;\theta_0) - H_0(t)\}$ converges weakly to $\varphi^{-1}\{G(t)\}$, which is also a Gaussian process in $\mathcal{F}$ since φ^{-1} is a linear map.

Step B3: We show that $\Psi_n^*\{\theta_0, \widehat{H}(\cdot, \theta_0)\}$ can be expressed asymptotically as the summation of n independent and identically distributed influence functions.

$$\Psi_n^*\{\theta_0, \widehat{H}(\cdot, \theta_0)\} = \mathbf{I} + \mathbf{II} + \mathbf{III}, \tag{3.30}$$

where

$$\begin{aligned} \mathbf{I} \;=\; & n^{-1} \sum_{i=1}^{n} \Delta_i^{ab} (\widehat{Z}_i - Z_i) \left[N_i(\tau_2) - Y_i(0)\Phi\left\{-\theta_0 \widehat{Z}_i + \widehat{H}(0;\theta_0)\right\} \right. \\ & \left. - \int_{t\in(0,\tau_2]} Y_i(t)\, d\Lambda\left\{-\theta_0 \widehat{Z}_i + \widehat{H}(t;\theta_0)\right\} \right], \end{aligned}$$

$$\begin{aligned}\mathbf{II} &= n^{-1}\sum_{i=1}^{n}\Delta_i^{ab}Z_i\left[N_i(\tau_2)-Y_i(0)\Phi\left\{-\theta_0 Z_i+\widehat{H}(0;\theta_0)\right\}\right.\\ &\qquad\left.-\int_{t\in(0,\tau_2]}Y_i(t)\,d\Lambda\left\{-\theta_0 Z_i+\widehat{H}(t;\theta_0)\right\}\right],\end{aligned}$$

and

$$\begin{aligned}\mathbf{III} &= n^{-1}\sum_{i=1}^{n}\Delta_i^{ab}Z_iY_i(0)\left[\Phi\left\{-\theta_0 Z_i+\widehat{H}(0;\theta_0)\right\}-\Phi\left\{-\theta_0\widehat{Z}_i+\widehat{H}(0;\theta_0)\right\}\right]\\ &\quad+\ n^{-1}\sum_{i=1}^{n}\int_{t\in(0,\tau_2]}\Delta_i^{ab}Z_iY_i(t)d\left[\Lambda\left\{-\theta_0 Z_i+\widehat{H}(t;\theta_0)\right\}-\Lambda\left\{-\theta_0\widehat{Z}_i+\widehat{H}(t;\theta_0)\right\}\right].\end{aligned}$$

It follows from the martingale integral representation for $\Phi^{-1}\{\widehat{F}_1(x)\}-\Phi^{-1}\{F_1(x)\}$ in (3.24) that $\mathbf{I}$ in (3.30) can be written as

$$\mathbf{I}=n^{-1}\sum_{k=1}^{n}\int_0^{\infty}\frac{v(t)}{\pi(t)}\,dM_{1k}(t)+o_p(n^{-1/2}),$$

where

$$\begin{aligned}v(t) &= \lim_{n\to\infty}\ n^{-1}\sum_{i=1}^{n}\Delta_i^{ab}\left[N_i(0)-Y_i(0)\Phi\{-\theta_0 Z_i+H_0(0)\}+\int_{t\in(0,\tau_2]}dM_{2i}(t)\right]\\ &\qquad\cdot\ \dot{\Phi}^{-1}\{F_1(X_{1i})\}\{1-F_1(X_{1i})\}I(X_{1i}\geq t)+o_p(1).\end{aligned}$$

The martingale properties associated with X_2 implies that $v(t)\to 0$ in probability, and thus $\mathbf{I}=o_p(n^{-1/2})$.

Based on the results in Step B2, we can write

$$
\begin{aligned}
\mathbf{II} &= n^{-1}\sum_{i=1}^{n}\int_{t\in(0,\tau_2]}\Delta_i^{ab}Z_i dM_{2i}(t) - n^{-1}\sum_{i=1}^{n}\int_{t\in(0,\tau_2]}\Delta_i^{ab}Z_iY_i(t)\,d\Big[\Lambda\left\{-\theta_0 Z_i+\widehat{H}(t;\theta_0)\right\}\\
&\quad -\Lambda\left\{-\theta_0 Z_i+H_0(t)\right\}\Big] + n^{-1}\sum_{i=1}^{n}\Delta_i^{ab}Z_i\left[N_i(0)-Y_i(0)\Phi\left\{-\theta_0 Z_i+\widehat{H}_0(t;\theta_0)\right\}\right]\\
&= n^{-1}\sum_{i=1}^{n}\int_{t\in(0,\tau_2]}\Delta_i^{ab}Z_i dM_{2i}(t) - n^{-1}\sum_{i=1}^{n}\Delta_i^{ab}Z_i\left[\Lambda\left\{-\theta_0 Z_i+\widehat{H}(X_{2i};\theta_0)\right\}\right.\\
&\quad \left.-\Lambda\left\{-\theta_0 Z_i+H_0(X_{2i})\right\}\right] + n^{-1}\sum_{i=1}^{n}\Delta_i^{ab}Z_iY_i(0)\left[\Lambda\left\{-\theta_0 Z_i+\widehat{H}(0;\theta_0)\right\}\right.\\
&\quad \left.-\Lambda\left\{-\theta_0 Z_i+H_0(0)\right\}\right] + n^{-1}\sum_{i=1}^{n}\Delta_i^{ab}Z_i\left[N_i(0)-Y_i(0)\Phi\left\{-\theta_0 Z_i+\widehat{H}_0(0;\theta_0)\right\}\right]\\
&= n^{-1}\sum_{i=1}^{n}\left(\int_{t\in(0,\tau_2]}\Delta_i^{ab}Z_i dM_{2i}(t) - \varphi^{-1}\left\{\int_{t\in(0,\tau_2]}\Delta_i^{ab}w(t)dM_{2i}(t)\right\}\right.\\
&\quad -\varphi^{-1}\left\{\int_{s\in(0,\infty)}\frac{r(s,\tau_2)B_4(X_2)}{\pi(s)}\,dM_{1i}(s)\right\} + \Delta_i^{ab}Z_i[N_i(0)-Y_i(0)\Phi\{-\theta_0 Z_i\\
&\quad + H_0(0)\}] + \frac{E[\Delta^{ab}ZY(0)(\lambda\{-\theta_0 Z+H_0(0)\}-\dot{\Phi}\{-\theta_0 Z+H_0(0)\})]}{E[\Delta^{ab}Y(0)\dot{\Phi}\{-\theta_0 Z+H_0(0)\}]}\,\Delta_i^{ab}\Big[N_i(0)\\
&\quad \left.-Y_i(0)\Phi\{-\theta_0 Z_i+H_0(0)\}+\int_{s\in(0,\infty)}\frac{q(s)}{\pi(s)}\,dM_{1i}(s)\Big]\right) + o_p(n^{-1/2}),
\end{aligned}
$$

where $w(t)=\lim_{n\to\infty}\, n^{-1}\sum_{i=1}^{n}\Delta_i^{ab}Z_i\lambda\{-\theta_0 Z_i+H_0(X_{2i})\}I(X_{2i}\geq t)$. Since $\{\Delta_i^{ab}Z_iN_i(t), t\in[0,\tau_2]\}$ and $\{\Delta_i^{ab}Z_i\lambda\{-\theta_0 Z_i+H_0(t)\}, t\in[0,\tau_2]\}$ are Donsker classes, it follows that $\{\int_{s\in(0,t]}\Delta_i^{ab}Z_i dM_{2i}(s),\, t\in[0,\tau_2]\}$ and $\{\int_{t\in(0,\tau_2]}\Delta_i^{ab}w(t)dM_{2i}(t)\}$ are Donsker. Following the arguments in Step B1 and B2, $\int_{s\in(0,\infty)}\{r(s,\tau_2)B_4(X_2)/\pi(s)\}\,dM_{1i}(s)$ and $\int_{s\in(0,\infty)}\{q(s)/\pi(s)\}\,dM_{1i}(s)$ are also Donsker. Since φ^{-1} is a linear map, **II** is Donsker.

Similarly to (3.26), we can write

$$
\mathbf{III} \;=\; n^{-1}\sum_{k=1}^{n}\int_0^{\infty}\frac{\zeta(t)}{\pi(t)}\,dM_{1k}(t)+o_p(n^{-1/2}),
$$

where

$$\begin{aligned}\zeta(t) &= \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\Delta_i^{ab}\theta_0 Z_i\left[Y_i(0)\dot{\Phi}\{-\theta_0 Z_i + H_0(0)\} + \int_{s\in(0,t]} Y_i(s)d\Lambda\{-\theta_0 Z_i + H_0(s)\}\right] \\ &\quad \cdot \dot{\Phi}^{-1}\{F_1(X_{1i})\}\{1-F_1(X_{1i})\}I(X_{1i}\geq t) + o_p(1).\end{aligned}$$

Using the arguments in Step B1, we can show that $\int_{t\in(0,\infty)}\{\zeta(t)/\pi(t)\}\,dM_{1i}(t)$ is a Donsker class. Therefore, by Donsker theorem, $n^{1/2}\Psi_n^*\{\theta_0, \widehat{H}(\cdot;\theta_0)\} = n^{1/2}(\mathbf{I}+\mathbf{II}+\mathbf{III})$ converges to a normal distribution with mean 0.

Step B4: Finally, we show the asymptotic normality of $\widehat{\theta}$ and the weak convergence of $\widehat{H}(t,\widehat{\theta})$. Because of the consistency of $\widehat{\theta}$ and $\Psi_n^*\{\widehat{\theta}, \widehat{H}(\cdot;\widehat{\theta})\} = 0$. we can write

$$n^{1/2}\left(\widehat{\theta}-\theta_0\right) = -n^{1/2}\left[\left.\frac{\partial\Psi_n^*\{\theta,\widehat{H}(\cdot;\theta)\}}{\partial\theta}\right|_{\theta=\theta_0}\right]\Psi_n^*\{\theta_0,\widehat{H}(\cdot;\theta_0)\} + o_p(1), \tag{3.31}$$

which converge to a normal distribution with mean 0, based on the results in Step B3.

Then, by combining the result of $(\partial/\partial\theta)\widehat{H}(t;\theta)$ at $\theta=\theta_0$ in Step A3 with equations (3.31), we can write

$$\begin{aligned}&n^{1/2}\{\widehat{H}(t,\widehat{\theta}) - H_0(t)\} \\ &= n^{1/2}\{\widehat{H}(t,\widehat{\theta}) - \widehat{H}(t,\theta_0)\} + n^{1/2}\{\widehat{H}(t,\theta_0) - H_0(t)\} \\ &= \left.\frac{\partial\widehat{H}(t,\theta)}{\partial\theta}\right|_{\theta=\theta_0} n^{1/2}\left(\widehat{\theta}-\theta_0\right) + n^{1/2}\{\widehat{H}(t,\theta_0) - H_0(t)\} + o_p(1) \\ &= -D_1\, n^{1/2}\left[\left.\frac{\partial\Psi_n^*\{\theta,\widehat{H}(\cdot;\theta)\}}{\partial\theta}\right|_{\theta=\theta_0}\right]\Psi_n^*\{\theta_0,\widehat{H}(\cdot;\theta_0)\} + n^{1/2}\{\widehat{H}(t,\theta_0) - H_0(t)\} + o_p(1) \\ &= -D_1\, D_2 n^{1/2}\Psi_n^*\{\theta_0,\widehat{H}(\cdot;\theta_0)\} + \varphi^{-1}\left[-n^{1/2}\{U_n^*(t;H_0,\theta_0) - U_n^*(0;H_0,\theta_0)\}\right] + o_p(1).\end{aligned}$$

Based on the results in Steps B1 to B3 and Donsker theorem, $n^{1/2}\{\widehat{H}(t,\widehat{\theta}) - H_0(t)\}$ converges weakly to a tight Gaussian process.

With the results in Step B4, it is straightforward to show that $n^{1/2}(\widehat{\rho} - \rho_0)$ is asymptotically normal with mean 0, and $n^{1/2}\{\widehat{F}_2(t) - F_2(t)\}$ weakly converges to a Gaussian process for $t \in [0, \tau_2]$, since the mappings $\theta \mapsto \rho$ and $\{\theta, H(\cdot)\} \mapsto F_2(\cdot)$ are both continuous and (uniformly) bounded based on the equalities in (3.10). This completes the proof of Theorem 3.2. □

Chapter 4

Summary and Future Research

4.1 Summary

This dissertation focuses on semiparametric methods to analyze biomedical outcomes subject to induced dependent censoring. Typical examples of such outcomes include the lifetime medical cost and the sojourn times in successive disease progression, which are both important for health care evaluation, health policy and management. However, analysis of these outcomes is complicated by induced dependent censoring and the issue of identifiability, which arise from the incomplete follow-up data in clinical trials. The statistical methods developed in this dissertation deal with these issues in appropriate ways, and provide us with practically useful tools in the study of these outcomes of interest.

First, we develop a copula-based semiparametric model for lifetime medical cost with incomplete follow-up data. This conceptually simple regression model is semiparametric in the sense that the marginal error distribution of both lifetime medical cost and survival time are completely unspecified. This model is appealing because its parame-

ters are identifiable from incomplete follow-up data and have meaningful interpretation in terms of lifetime medical cost as well. Under this model, we can also quantify the covariate effect in terms of difference in dollar amount as desired. The proposed copula-based semiparametric regression model strikes a balance between model identifiability and robustness. We develop an inference procedure which is computationally easy and we prove that the resulting estimators are consistent and asymptotically normal. Simulation studies show that the proposed method performs well in samples of moderate size. The proposed method is applied to a SWOG lung cancer clinical trial. Our proposed model only requires and exploits uncensored lifetime medical cost in addition to the standard survival data, and thus can be applied to a wide range of cost data collection schemes. This copula-based semiparametric regression model is practically most useful in medical studies where only a small portion of patients survive beyond the study duration.

Next, we propose a semiparametric inference procedure for the successive durations in this bi-state progressive disease process. We suggest a semiparametric model that postulates normal copula for the association between the two durations, while leaving the marginals unspecified. Motivated by the colon cancer data example where some patients reach death without cancer recurrence, we allow our model to accommodate the situation that the second duration has a probability mass at 0. With normal distribution theory, the proposed normal copula model can be written as a regression model for the second duration with the first duration as the covariate. By using the martingale features associated with the data, we proposed an inference procedure for estimation. In the special case when the probability mass of the second duration at zero is zero, our inference procedure reduces to that for the linear transformation model developed in Chen et al. (2002). The proposed estimation procedure is computationally easy. We provide rigorous proof of the asymptotic properties of the resulting estimators, which

are shown to be consistent or uniformly consistent and to be asymptotically normal or weakly converge to a tight Gaussian process. Simulation studies show that the proposed method performs well in samples of moderate size. We apply the proposed method to the motivating example of a national intergroup colon cancer clinical trial.

4.2 Future Research

In this section, we present several topics which are worth investigating in future research on censored medical cost and successive durations.

In Chapter 2, we propose a copula-based semiparametric model for lifetime medical cost, which is useful for a wide range of cost data collection schemes. However, this approach for lifetime medical cost has its limitation in certain practical situations. First, we achieve the marginal identifiability of lifetime medical cost distribution through modeling. This modeling strategy may not be desirable when a large portion of patients survive beyond the study duration. Second, additional data might be available in practice. For example, the cost accumulation process was observed at scheduled visits in the study we discussed in Section 2.3.3, or one may observe accumulated cost at censoring time for each censored individual. In those cases, the ignorance of the accumulated medical cost up to the time of censoring, as our method for lifetime medical cost, seems undesirable, especially when the censoring rate is heavy.

Some of the available approaches for censored medical cost such as Bang and Tsiatis (2000), could take advantage of the information of the accumulated medical cost up to the time of censoring for censored individuals. However, the target of the analysis is *time-restricted* medical cost. As we have discussed in Chapter 1, in the analysis of *time-restricted* medical cost, an artificial time limit is imposed to ensure the potential

censoring time has positive support beyond this time limit. Since the time limit is artificial and the covariates may impact survival time as well, attempt to interpret time-restricted results in terms of lifetime medical cost, as desired, is inappropriate.

To make appropriate cost and cost-effectiveness analysis, it is desirable to consider new measurements of medical cost. One future research direction is to develop the quantile inference for censored medical cost, which balances the pros and cons of time-restricted medical cost and lifetime medical cost approaches. In the medical cost context, it is hard to alleviate the problem of bias in estimating the mean lifetime medical cost, because the cost data distribution is usually highly skewed to the right and the medical studies have a limited duration. Therefore, it is helpful to consider the *quantile-restricted* medical cost $U\{Q_T(\tau)\}$ $(0 < \tau < 1)$, which is the accumulative medical cost up to a certain quantile of survival time. This quantity is of scientific interest in both one-sample inference and two-sample comparison. By targeting at the quantile-restricted medical cost, we may be able to use the accumulated cost at censoring time for each censored individual, and avoid the difficulty of interpretability of time-restricted medical cost.

In Chapter 3, we propose a copula-based semiparametric model for successive durations with incomplete follow-up data. The model is useful to estimate the distribution of the sojourn time between successive events (e.g., cancer recurrence and death) in practice. As evident from the colon cancer example, it is often of scientific interest to compare the sojourn time distributions among different treatment groups. Our proposed estimators and their variance estimates enable such comparison at a set of fixed time points, however, they are not able to compare the entire distributions of different groups. One future research direction is to develop log-rank-type tests along the line of the proposed semiparametric model for the bi-state progressive disease process.

In addition to estimation of the the sojourn time distribution, of scientific interest is

prediction with given history of the bi-state process. For example, it might be desirable to know a colon cancer patient's survival probabilities given occurrence and timing of his cancer relapse. On a continuous time scale the existing nonparametric approaches do not provide a reliable estimator of the second duration distribution given the first duration. Future research is worth investigating to develop the inference procedure for prediction under our proposed semiparametric model for the bi-state progressive disease process.

Bibliography

Andersen, P. K., Borgan, O., Borgan, O., Borgan, O., Gill, R. D., and Keiding, N. (1993), *Statistical Models Based on Counting Processes*, Springer-Verlag Inc.

Bang, H. and Tsiatis, A. A. (2000), "Estimating Medical Costs with Censored Data," *Biometrika*, 87, 329–343.

— (2002), "Median Regression with Censored Cost Data," *Biometrics*, 58, 643–649.

Brown, M., Lipscomb, J., and Snyder, C. (2001), "The Burden of Illness of Cancer: Economic Cost and Quality of Life," *Annual Review of Public Health*, 22, 91–113.

Chen, K., Jin, Z., and Ying, Z. (2002), "Semiparametric Analysis of Transformation Models with Censored Data," *Biometrika*, 89, 659–668.

Cheng, S. C., Wei, L. J., and Ying, Z. (1995), "Analysis of Transformation Models with Censored Data," *Biometrika*, 82, 835–845.

Cook, R. J. and Lawless, J. F. (1997), "Marginal Analysis of Recurrent Events and a Terminating Event," *Statistics in Medicine*, 16, 911–924.

Cox, D. R. (1972), "Regression Models and Life-tables (with Discussion)," *Journal of the Royal Statistical Society, Series B: Methodological*, 34, 187–220.

Cox, D. R., Fitzpatrick, R., Fletcher, A. E., Gore, S. M., Spiegelhalter, D. J., and Jones, D. R. (1992), "Quality-of-life Assessment: Can We Keep It Simple? (Disc: P375-393)," *Journal of the Royal Statistical Society, Series A: Statistics in Society*, 155, 353–375.

Dabrowska, D. M. (1988), "Kaplan-Meier Estimate on the Plane," *The Annals of Statistics*, 16, 1475–1489.

Diehr, P., Yanez, D., Ash, A., Hornbrook, M., and Lin, D. Y. (1999), "Methods for Analyzing Health Care Utilization and Costs," *Annual Review of Public Health*, 20, 125–144.

Dudley, R. A., Harrell, F. E., Smith, L. E., Mark, D. B., and Califf, R. M. (1993), "Comparison of analytic models for estimating the effect of clinical factors on the cost of coronary artery bypass graft surgery," *Journal of Clinical Epidemiology*, 46, 261–271.

Dukes, C. (1932), "The Classification of Cancer of the Rectum," *Journal of Pathological Bacteriology*, 35, 323.

Fenn, P., McGuire, A., Phillips, V., Backhouse, M., and Jones, D. (1995), "The Analysis of Censored Treatment Cost Data in Economic Evaluation," *Medical Care*, 33, 851–863.

Fygenson, M. and Ritov, Y. (1994), "Monotone estimating equations for censored data," *The Annals of Statistics*, 22, 732–746.

Gehan, E. A. (1965), "A generalized Wilcoxon test for comparing arbitrarily singly-censored samples," *Biometrika*, 52, 203–223.

Gelber, R. D., Gelman, R. S., and Goldhirsch, A. (1989), "A Quality-of-life-oriented Endpoint for Comparing Therapies," *Biometrics*, 45, 781–795.

Genest, C., Ghoudi, K., and Rivest, L.-P. (1995), "A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions," *Biometrika*, 82, 543–552.

Gill, R. D. (1980), *Censoring and Stochastic Integrals, Mathematical Centre Tract No. 124*, Amsterdam: Mathematisch Centrum.

— (1989), "Non- and Semi-parametric Maximum Likelihood Estimators and the Von Mises Method (Part 1) (C/R: P124-128)," *Scandinavian Journal of Statistics*, 16, 97–124.

— (1994), "Lectures on Survival Analysis," in *Lectures on Probability Theory,* ed. P. Bernard, Berlin Heidelerg: Springer-Verlag, 115-241.

Gill, R. D. and Johansen, S. (1990), "A Survey of Product-integration with a View toward Application in Survival Analysis," *The Annals of Statistics*, 18, 1501–1555.

Gill, R. D., van der Laan, M. J., and Wellner, J. A. (1995), "Inefficient Estimators of the Bivariate Survival Function for Three Models," *Annales de l'Institut Henri Poincaré: Probabilités et Statistiques*, 31, 545–597.

Glasziou, P. P., Simes, R. J., and Gelber, R. D. (1990), "Quality Adjusted Survival Analysis," *Statistics in Medicine*, 9, 1259–1276.

Glidden, D. V. (2000), "A Two-stage Estimator of the Dependence Parameter for the Clayton-Oakes Model," *Lifetime Data Analysis*, 6, 141–156.

Hallstrom, A. P. and Sullivan, S. D. (1998), "On Estimating Costs for Economic Evaluation in Failure Time Studies," *Medical Care*, 36(3), 433–436.

Hiatt, R. A., Quesenberry, C. P., Selby, J. V., Fireman, B. H., and Knight, A. (1990), "The Cost of Acquired Immunodeficiency Syndrome in Northern California: the Experience of a Large Prepaid Health Plan," *Archives of Internal Medicine*, 150, 833–838.

Honoré, B. E. and Powell, J. L. (1994), "Pairwise difference estimators of censored and truncated regression models," *Journal of Econometrics*, 64, 241–278.

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

Huang, Y. (2002), "Calibration Regression of Censored Lifetime Medical Cost," *Journal of the American Statistical Association*, 97, 318–327.

Huang, Y. and Berry, K. (2006), "Semiparametric Estimation of Marginal Mark Distribution," *Biometrika*, 93, 895–910.

Huang, Y. and Louis, T. A. (1998), "Nonparametric Estimation of the Joint Distribution of Survival Time and Mark Variables," *Biometrika*, 85, 785–798.

Jain, A. K. and Strawderman, R. L. (2002), "Flexible Hazard Regression Modeling for Medical Cost Data," *Biostatistics (Oxford)*, 3, 101–118.

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

Kaplan, E. L. and Meier, P. (1958), "Nonparametric Estimation from Incomplete Observations," *Journal of the American Statistical Association*, 53, 457–481.

Kelly, K., Crowley, J., Bunn, P., Presant, C., Grevstad, P., Moinpour, C., Ramsey, S., Wozniak, A., Weiss, G., Moore, D., Israel, V., Livingston, R., and Gandara, D.

(2001), "Randomized phase III trial of paclitaxel plus carboplatin versus vinorelbine plus cisplatin in the treatment of patients with advanced non–small-cell lung cancer: a Southwest Oncology Group trial." *Journal of Clinical Oncology*, 19, 3210–3218.

Klaassen, C. A. J. and Wellner, J. A. (1997), "Efficient Estimation in the Bivariate Normal Copula Model: Normal Margins Are Least Favourable," *Bernoulli*, 3, 55–77.

Kosorok, M. R. (2008), *Introduction to Empirical Processes and Semiparametric Inference*, New York: Springer.

Koul, H., Susarla, V., and Van Ryzin, J. (1981), "Regression Analysis with Randomly Right-censored Data," *The Annals of Statistics*, 9, 1276–1288.

Lai, T. L. and Ying, Z. (1988), "Stochastic Integrals of Empirical-type Processes with Applications to Censored Regression," *Journal of Multivariate Analysis*, 27, 334–358.

Lin, D. (2003), "Regression Analysis of Incomplete Medical Cost Data," *Statistics in Medicine*, 22, 1181–1200.

Lin, D. Y. (2000a), "Linear Regression Analysis of Censored Medical Costs," *Biostatistics*, 1, 35–47.

— (2000b), "Proportional Means Regression for Censored Medical Costs," *Biometrics*, 56, 775–778.

Lin, D. Y., Feuer, E. J., Etzioni, R., and Wax, Y. (1997), "Estimating Medical Costs from Incomplete Follow-up Data," *Biometrics*, 53, 419–434.

Lin, D. Y., Sun, W., and Ying, Z. (1999), "Nonparametric Estimation of the Gap Time Distributions for Serial Events with Censored Data," *Biometrika*, 86, 59–70.

Lin, D. Y., Wei, L. J., and Ying, Z. (1998), "Accelerated failure time models for counting processes," *Biometrika*, 85, 605–618.

Lipscomb, J., Ancukiewicz, M., Parmigiani, G., Hasselblad, V., Samsa, G., and Matchar, D. B. (1998), "Predicting the cost of illness: a comparison of alternative models applied to stroke," *Medical Decision Making*, 18(2), S39–S56.

Ma, S. and Huang, J. (2005), "Regularized ROC method for disease classification and biomarker selection with microarray data," *Bioinformatics*, 21, 4356–4362.

Mantel, N. (1966), "Evaluation of Survival Data and Two New Rank Order Statistics Arising in its Consideration," *Cancer Chemotherapy Report*, 50, 163–170.

McCullagh, P. and Nelder, J. A. (1989), *Generalized Linear Models (Second Edition)*, Chapman & Hall Ltd.

Moertel, C., Fleming, T., and Macdonald, J. e. a. (1990), "Levamisole and fluorouracil for adjuvant therapy of resected colon carcinoma," *New England Journal of Medicine*, 322, 352 –358.

Neuhaus, G. (1971), "On Weak Convergence of Stochastic Process with Multidimensional Time Parameter," *The Annals of Mathematical Statistics*, 42, 1285–1295.

Peng, L. and Huang, Y. (2007), "Survival Analysis with Temporal Covariate Effects," *Biometrika*, 94, 719–733.

Prentice, R. L. and Cai, J. (1992), "Covariance and Survivor Function Estimation Using Censored Multivariate Failure Time Data (Corr: 93V80 P711-712)," *Biometrika*, 79, 495–512.

Quesenberry, C. P., Fireman, B. H., Hiatt, R. A., and Selby, J. V. (1989), "A Survival Analysis of Hospitalization Among Patients with Acquired Immunodeficiency Syndrome," *American Journal of Public Health*, 79, 1643–47.

Ries, L., Melbert, D., Krapcho, M., Stinchcomb, D., Howlader, N., Horner, M., Mariotto, A., Miller, B., Feuer, E., Altekruse, S., Lewis, D., Clegg, L., Eisner, M., Reichman, M., and Edwards, B. (2007), *SEER Cancer Statistics Review, 1975-2005, National Cancer Institute. Bethesda, MD, based on November 2007 SEER data submission, posted to the SEER web site, 2008.*

Robins, J. M. and Rotnitzky, A. (1992), "Revocery of Information and Adjustment for Dependent Censoring Using Surrogate Markers," In *AIDS Epidemiology: Methodological Issues,*, Eds N.P.Jewell, K.dietz and V.T. Farewell, pp.297–331. Boston, MA: Birkhäuser.

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994), "Estimation of Regression Coefficients When Some Regressors Are Not Always Observed," *Journal of the American Statistical Association*, 89, 846–866.

Rudin, W. (1976), *Principles of Mathematical Analysis*, New York: McGraw-Hill.

Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statistics*, New York: John Wiley & Sons.

Shih, J. H. and Louis, T. A. (1995), "Inferences on the Association Parameter in Copula Models for Bivariate Survival Data," *Biometrics*, 51, 1384–1399.

Sklar, A. (1959), "Fonctions de répartition à n dimensions et leurs marges," *Publications de I'Institut de Statistique de I'Université de Paris*, 8, 229–231.

Strawderman, R. L. (2000), "Estimating the Mean of an Increasing Stochastic Process at a Censored Stopping Time," *Journal of the American Statistical Association*, 95, 1192–1208.

Tsiatis, A. A. (1990), "Estimating Regression Parameters using Linear Rank Tests for Censored Data," *The Annals of Statistics*, 18, 354–372.

van der Laan, M. J. (1997), "Nonparametric Estimators of the Bivariate Survival Function under Random Censoring," *Statistica Neerlandica*, 51, 178–200.

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

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

Visser, M. (1996), "Nonparametric Estimation of the Bivariate Survival Function with an Application to Vertically Transmitted AIDS," *Biometrika*, 83, 507–518.

Wang, W. and Wells, M. T. (1998), "Nonparametric Estimation of Successive Duration Times under Dependent Censoring," *Biometrika*, 85, 561–572.

Wei, L. J., Ying, Z., and Lin, D. Y. (1990), "Linear Regression Analysis of Censored Survival Data Based on Rank Tests," *Biometrika*, 77, 845–851.

Yang, S. (1997), "A Generalization of the Product-limit Estimator with an Application to Censored Regression," *The Annals of Statistics*, 25, 1088–1108.

Ying, Z. (1993), "A Large Sample Study of Rank Estimation for Censored Regression Data," *The Annals of Statistics*, 21, 76–99.

Zhao, H. and Tsiatis, A. A. (1997), "A Consistent Estimator for the Distribution of Quality Adjusted Survival Time," *Biometrika*, 84, 339–348.

— (2001), "Testing Equality of Survival Functions of Quality-adjusted Lifetime," *Biometrics*, 57, 861–867.